D7.1 SPECTRUM Strategic Research Innovation and Deployment Agenda

Status: UNDER EC REVIEW

Dissemination Level: Public

Funded by
the European Union

Abstract

Key Words: Data-intensive science, High Performance Computing, HPC, High Energy Physics, HEP, Radio Astronomy, RA, compute and data continuum, European Union, EU, supercomputing, AI, ML, quantum computing, SRIDA, Strategy, Strategic priorities

The SPECTRUM Strategic Research, Innovation and Deployment Agenda (SRIDA) presents a coordinated European strategy for the compute and data continuum serving data-intensive science. Grounded in the requirements of High Energy Physics (including the HL-LHC programme and the WLCG) and Radio Astronomy (including the SKA Observatory and LOFAR), the agenda addresses challenges shared across data-intensive disciplines: exabyte-scale data, heterogeneous computing, federated access, AI adoption, environmental sustainability, and workforce development. Thirteen strategic priorities are organised within four pillars: Policy, Trust and Governance; Architecture and Interoperability; Software and Science Enablers; and Human Capital and Responsibility. The SRIDA identifies investment areas spanning computing and data infrastructures, software, governance, and human capital, and defines a phased implementation roadmap from near-term governance foundations through medium-term technical integration to long-term operational maturity. The agenda builds on evidence from the following project deliverables: Technical Blueprint (D6.1), use case analysis (D5.1), landscape survey (D5.3), access policy analysis (D5.2), and community of practice (D3.1). The agenda is informed by structured community consultation and is intended to guide coordinated European action across scientific communities, service providers, and policy makers.

<table>
<tr><th colspan="4">Document Description</th></tr>
<tr><td colspan="4">D7.1 SPECTRUM Strategy Research Innovation and Deployment Agenda</td></tr>
<tr><td colspan="4">Work Package Number 7</td></tr>
<tr><td>Document Type</td><td colspan="3">Report</td></tr>
<tr><td>Document status</td><td>Under EC Review</td><td>Version</td><td>2.0</td></tr>
<tr><td>Dissemination Level</td><td colspan="3">Public</td></tr>
<tr><td>Copyright status</td><td colspan="3">CC BY
This material by Contributing Parties of the SPECTRUM Consortium is licensed under <u>Creative Commons Attribution 4.0 International License</u>.</td></tr>
<tr><td>Lead Partner</td><td colspan="3">EGI Foundation</td></tr>
<tr><td>Document link</td><td colspan="3"><u>https://documents.egi.eu/document/4075</u></td></tr>
<tr><td>Digital Object Identifier</td><td colspan="3"><u>https://zenodo.org/records/20182878</u></td></tr>
<tr><td>Author(s)</td><td colspan="3">Lead Editors:
● Sergio Andreozzi (EGI Foundation)
● Xavier Salazar (EGI Foundation)

Contributors:
● Concezio Bozzi (INFN)
● Florian Berberich (FZJ)
● Luis Cifuentes (FZJ)
● Stefano Bagnasco (INFN)
● Luca dell'Agnello (INFN)
● Tommaso Boccali (INFN)
● John Swinbank (ASTRON)
● Chiara Ferrari (CNRS/OCA)
● Jeff Wagg (CNRS/OCA)
● Shan Mignot (CNRS/OCA)
● Raymond Oonk (SURF)
● Maria Girone (CERN)
● Eric Wulff (CERN)</td></tr>
<tr><td>Reviewers</td><td colspan="3">● Luciano Gaido (INFN)
● Mylene Pischella (NEOVIA Innovation)</td></tr>
<tr><td>Moderated by</td><td colspan="3">Patricia Ruiz (EGI Foundation)</td></tr>
<tr><td>Approved by</td><td colspan="3">AMB</td></tr>
</table>

Revision History			
Version	**Date**	**Description**	**Contributors**
V0.1	14/01/2025	First version of table of contents with brief description of focus for each section reviewed by WP7	Sergio Andreozzi, Xavier Salazar
V0.2	11/04/2025	Second version of table of contents	Sergio Andreozzi, Xavier Salazar
v0.3	31/07/2025	Template for priority descriptions agreed	Sergio Andreozzi, Xavier Salazar
v0.4	30/09/2025	Added section on methodology	Sergio Andreozzi
V0.5	03/11/2025	Added priorities into annex	Xavier Salazar with contribution from partners: P01 S. Andreozzi, EGI; CoP WG6 - P02 C. Bozzi, INFN; CoP WG2, WG3, WG6 - P03 F. Berberich; FZJ. CoP WG6, WG1 - P04 S. Andreozzi, EGI - P05 S. Bagnasco, INFN; CoP WG1 - P06 C. Bozzi, INFN; CoP WG2, WG3, WG6 - P07 L. dell'Agnello, INFN; CoP: WG4, WG3, WG6 - P08 J. Swinbank, ASTRON; CoP WG4 - P09 X. Salazar - P10 X. Salazar - P11 C. Ferrari CNRS, Jeff Wagg CNRS; CoP WG5 - P12 S. Mignot; CoP WG5 - P13 T. Boccali INFN; CoP WG5
V0.6	15/11/2025	Added priority grouping into pillars and descriptions	Sergio Andreozzi (based on feedback from WP7 members)
V0.7	21/11/2025	Added analysis of gathered evidence	Xavier Salazar
V0.8	8/12/2025	Document complete for internal review	Sergio Andreozzi, Xavier Salazar
V1.0	11/12/2025	Addressed comments and final edits for public draft	Sergio Andreozzi, Xavier Salazar
V1.1	15/04/2026	Completed all priorities in annex A and addressed comments from Experts meeting	Sergio Andreozzi, Stefano Bagnasco, Tommaso Boccali, Florian Berberich, Concezio Bozzi, Luca dell'Agnello, Chiara Ferrari, Shan Mignot, Xavier Salazar, John Swinbank
V1.2	24/04/2026	Developed sections 5 and 6	Sergio Andreozzi
V1.3	03/05/2026	Review completed	Luciano Gaido, Mylène Pischella Raymond Oonk (as WP internal reviewer)
V1.4	26/05/2026	Addressed comments from reviews	Sergio Andreozzi
V1.5	25/06/2026	Final version	Sergio Andreozzi
V1.6	30/06/2026	Approved by AMB	
V2.0	01/07/2026	**Final**	

Terminology / Acronyms	
Terminology / Acronym	**Definition**
AI	Artificial Intelligence
AARC BPA	AARC Blueprint Architecture
CADA	Cloud and AI Development Act
CoP	Community of Practice
CPU	Central Processing Unit
CUE	Carbon Usage Effectiveness
DB	Database
DOI	Digital Object Identifier
DPU	Data Processing Unit
EAB	External Advisory Board
EOSC	European Open Science Cloud
ERA	European Research Area
ERDF	European Regional Development Fund
ESA	European Space Agency
ESCAPE	European Science Cluster of Astronomy and Particle Physics ESFRI research infrastructures
ESFRI	European Strategy Forum on Research Infrastructures
FAIR	Findable, Accessible, Interoperable, Reusable
FIAM	Federated Identity Access Management
FPGA	Field-Programmable Gate Array
GDPR	General Data Protection Regulation
GPU	Graphics Processing Unit
HEP	High Energy Physics
JENA	Joint ECFA-NuPECC-APPEC
LHC	Large Hadron Collider
HL-LHC	High Luminosity Large Hadron Collider
HTC	High-Throughput Computing
HPC	High Performance Computing

KER	Key Exploitable Result
KPI	Key Performance Indicator
LOFAR	LOw Frequency ARray
MFA	Multi-factor authentication
ML	Machine Learning
MoU	Memorandum of understanding
NGI	National federation of shared computing, storage and data resources part of EGI
NREN	National Research and Education Network
PUE	Power Usage Effectiveness
RA	Radio Astronomy
RI	Research Infrastructure
SKA	Radio Telescopes managed by SKAO
SKAO	SKA Observatory: Intergovernmental organisation
SLA	Service Level Agreement
SRCNet	SKA Regional Centre Network
SRIDA	Strategic Research, Innovation and Deployment Agenda
SSO	Single Sign On
TI	Technology Infrastructure
VA/TNA	Virtual Access / Transnational Access
WLCG	Worldwide LHC Computing Grid

Table of Contents

List of Figures

List of Tables

Executive summary

European data-intensive science is approaching a step change in scale. The instruments coming online over the next decade, including the High Luminosity Large Hadron Collider (HL-LHC), the SKA Observatory and LOFAR 2.0, will generate data volumes and processing demands an order of magnitude beyond current infrastructure capacity. The HL-LHC alone is projected to require tens of exabyte storage and tens of millions of CPU-equivalent cores from 2030 onwards; SKA Phase 1 will produce hundreds of petabytes per year before scaling further. These facilities operate on multi-decade timelines, so the architectural and governance choices made in the next five years will determine whether Europe can extract the scientific return on infrastructure investments already committed. Meeting this challenge requires coordinated action across scientific communities, computing and data service providers, and policy makers. This Strategic Research, Innovation and Deployment Agenda (SRIDA) presents thirteen priorities, five investment areas, and a phased multi-annual roadmap that turn that coordination into action.

Europe's research infrastructure landscape combines EU-level coordination, through ESFRI for long-term research infrastructure planning, EuroHPC for supercomputing capability, and EOSC for federated data and open science, with national investments that still provide the majority of operational storage and computing capacity. This agenda builds on operational experience from federated infrastructures already coordinating nationally-funded resources at European scale: EGI for national and research institution computing facilities, WLCG for LHC-related computing/storage facilities, the SKA Regional Centre Network (SRCNet) and GEANT for the networking. These existing federations show what is feasible and where the next generation of coordination must reach further.

Seven drivers shape the strategic environment. **Infrastructure Scale**: flagship instruments will move from petabyte to exabyte operations, requiring architectural evolution across compute, storage, and networking, with heterogeneous accelerators (GPU, FPGA, DPU, and emerging quantum devices) becoming the default rather than the exception. **AI Adoption**: artificial intelligence is being adopted rapidly across the research data lifecycle, with most applications still moving from proof-of-concept to production and demand for AI-capable infrastructure growing across European research communities. **Environmental Sustainability**: energy availability, water use, and embodied carbon in hardware are becoming primary constraints on infrastructure growth, making efficiency a design parameter rather than an afterthought. **Security and Trust**: post-quantum cryptographic transitions, evolving authentication requirements, and stricter data governance frameworks add complexity to federated operations. **Long-term Preservation**: instrument timelines of 20 to 50 years require software, data, and AI models to remain usable across hardware generations, with FAIR principles extending to workflows and provenance. **Workforce Capacity**: skills profiles are shifting toward research software engineering, data engineering, and applied AI; career structures, recognition, and competition with industry for technical talent remain unresolved. **Digital Sovereignty**: European policy increasingly treats digital capacity as a strategic asset, requiring open technologies, data processing within European jurisdictions, and resilience against geopolitical disruption.

Responding to these drivers requires progress against **five strategic goals**: **coherent governance** connecting thematic research infrastructures with horizontal e-Infrastructures; **seamless resource access** through multi-year allocation and federated authentication; **technical interoperability** across heterogeneous computing and cross-facility workflows; **sustainability across software portability and environmental impact**; and **human capacity** built on expertise combining domain science with computing skills.

The agenda organises **thirteen priorities within four pillars to deliver these goals**. **Pillar 1: Policy, Trust and Governance** establishes cross-infrastructure governance, multi-year resource allocation, and federated identity as foundations for everything that follows. **Pillar 2: Architecture and Interoperability** enables heterogeneous computing, federated data management, and cross-facility workflow orchestration. **Pillar 3: Software and Science Enablers** advances AI/ML in production, code portability and performance, scientific reproducibility, and the long-term preservation of data, software, and workflows. **Pillar 4: Human Capital and Responsibility** addresses community collaboration and co-design, environmental sustainability, and workforce development. The pillars are interdependent: technical interoperability without governance fails to scale, governance without skilled people cannot be implemented, and none of it is sustainable without explicit attention to environmental footprint.

Investment spans five areas: computing infrastructures, data infrastructures, software and tools, governance and coordination, and human capital. Open science and environmental sustainability cut across all five. Three categories are routinely underestimated and are essential for realising value from capital investments: coordination and federation costs, sustained software development beyond initial deployment, and long-term preservation of data, software, and workflows under FAIR principles. Effective investment depends on coherent action across EuroHPC, EOSC, Horizon Europe, the Digital Europe Programme, national research and infrastructure programmes, and operational funding for research infrastructures themselves.

Implementation follows three phases. The short-term foundation phase (1 to 3 years) establishes governance mechanisms, federated identity, and standardised interfaces, and begins co-design between thematic research infrastructures and e-Infrastructures with the aim of aligning services and interfaces, not merging organisations. The medium-term integration phase (3 to 5 years) deploys heterogeneous computing capabilities at scale, federated data management, and workflow orchestration spanning HPC, HTC, cloud, and edge resources. The long-term maturation phase (5 years and beyond) achieves AI/ML at production scale, end-to-end reproducibility, and a stable career framework for research computing. Governance and identity foundations must precede technical integration; workforce development and environmental management run across all phases to keep the effort sustainable.

The agenda places concrete tasks on each stakeholder group. **Scientific communities** should publish multi-year resource plans, take part in co-design with e-Infrastructures, identify high-impact use cases for moving AI/ML to production, and contribute to shared software and data stewardship. **Service providers** should establish formal liaison mechanisms with thematic research infrastructures, deploy federated identity and standardised interfaces, expose heterogeneous resources through common APIs, and provide training that allows researchers to use them efficiently. **Policy makers** should ensure that data-intensive science requirements inform EuroHPC and EOSC priorities, enable multi-year allocation across heterogeneous resources, fund/reward sustained software development and data stewardship as first-class activities, support career pathways for research software and data engineers, and set environmental budgets that infrastructure operators are required to meet.

Broader impact reaches past the instruments that drove the agenda, and Section 7 sets it out through the RI-PATHS framework for research infrastructures. The workforce comes first: building the continuum needs research software and data engineers, specialists in heterogeneous architectures and AI/ML deployment, and governance professionals who staff cross-infrastructure forums and operate multi-year allocation frameworks, and the skills they develop transfer to industry, finance, and public administration. Procurement across computing, storage, networking, and software strengthens European industrial capacity and extends the value chain that EuroHPC and the RISC-V DARE programme have begun; open-source scientific software and technology transfer, from ROOT and Geant4 to the radio-astronomy patents behind Wi-Fi, add value across sectors. The methods built for exabyte-scale processing and federated data management apply equally to climate modelling, public health, and disaster response, and they reach those domains through horizontal federations such as EGI, which pools national digital infrastructures for research across Europe. Open science keeps publicly funded outputs available for reuse and scrutiny, and the geographic and gender composition of the workforce is tracked and reported. On policy, the governance, allocation, and identity models developed here inform EuroHPC, EOSC, and national infrastructure decisions, provide templates for cross-border resource sharing, and align the agenda with the AI Continent Action Plan, the Cloud Sovereignty Framework already operational in EU procurement, the CADA, and the forthcoming ERA Act.

The Technical Blueprint (D6.1) provides the architectural foundation; the SRIDA provides the strategic framework for investment, governance, and implementation. Together they chart the path toward a compute and data continuum where researchers move workloads across HPC, HTC, cloud, and edge resources with one identity, predictable allocation, and portable software, and where environmental footprint is measured and managed. Whether Europe reaches it depends on choices made in the next five years; this agenda sets out those choices and the stakeholders who must make them.

1. Introduction

This section introduces the SPECTRUM Strategic Research, Innovation and Deployment Agenda (SRIDA), its relationship to the Technical Blueprint, and the audiences it serves. It establishes the scope, vision, and structure that frame the strategic priorities presented in subsequent sections.

1.1. Purpose and Scope

European data-intensive science faces a structural challenge. The instruments coming online over the next decade, the High Luminosity Large Hadron Collider, the SKA Observatory, LOFAR 2.0, and facilities across multiple domains, will generate data at scales that exceed current infrastructure capacity by an order of magnitude. Meeting this challenge requires more than incremental growth. It requires coordinated European action on governance, architecture, software, sustainability, and workforce.

This document presents the SPECTRUM SRIDA. It defines at an EU-level strategic priorities, investment areas, and a multi-annual roadmap for European research computing and data infrastructure. The agenda complements and builds upon existing European strategies: EuroHPC for supercomputing capability, EOSC for federated data and open science, and ESFRI for long-term research infrastructure planning. Where those strategies address broad European digital capacity, the SRIDA focuses specifically on the compute and data continuum required for data-intensive science, grounded in documented requirements from scientific communities. The network underpinning the continuum is another important factor, however it is not treated in detail in this work and would require a dedicated focus in follow up work.

SPECTRUM (Computing Strategy for Data-intensive Science Infrastructures in Europe) brings together research infrastructures, digital infrastructure providers, and policy stakeholders under Horizon Europe. The project has gathered evidence through systematic use case analysis (D5.1), landscape survey (D5.3), access policy analysis (D5.2), community consultation via the Community of Practice (D3.1, D4.1), and expert review. This evidence base informs both the SRIDA and the companion Technical Blueprint (D6.1).

Relationship with the Technical Blueprint

The SRIDA works in tandem with the Technical Blueprint. The Blueprint defines the architectural vision and technical capabilities for the compute and data continuum, identifying sixteen consolidated gaps. The SRIDA builds on these technical requirements, expands into non-technical areas and formulate strategic actions: it explains why these capabilities matter for European research competitiveness, identifies who should lead implementation, proposes when investments should be made, and recommends where resources should be directed.

Scope

The SRIDA focuses on infrastructures supporting data-intensive scientific research. Two flagship domains anchor the analysis:

- **High Energy Physics (HEP)**, including the HL-LHC programme and the Worldwide LHC Computing Grid (WLCG), generating tens of exabytes of data from 2030 onwards and requiring distributed processing across heterogeneous compute resources at hundreds of sites globally
- **Radio Astronomy (RA)**, including the SKA Observatory, LOFAR 1.0 and LOFAR 2.0, with data rates scaling to exabyte levels during full operations

Scope and Transferability

SPECTRUM's priorities are grounded primarily in the documented needs of HEP and RA, complemented by additional science cases from life sciences, neuroscience, and meteorology collected in D5.1. Many of these needs are shared across data-intensive science, but the SRIDA does not presume to speak for communities that were not directly consulted. To enable other communities to assess which findings apply to their context, the following characteristics define the infrastructure assumptions underlying this agenda:

- **Centrally reconstructed data**: experimental data is correlated and/or reconstructed at central facilities (CERN Tier-0, SKA correlator) before it is distributed for further processing and analysis
- **Research data, not personal data**: primary datasets are physical measurements with no GDPR constraints; infrastructure operations (login, accounting) handle personal data separately
- **Open science by default**: data is intended for broad sharing with time-limited embargoes, there are no permanent access restrictions
- **Throughput over latency**: processing is optimised for high volume throughput (batch processing and real-time filtering at source) and not for low-latency delivery of results
- **Global federated governance**: experiments span multiple countries with formal governance structures
- **Multi-decade experiment lifecycles**: instruments operate for decades with forecastable data growth (HL-LHC schedule, SKA phases, LOFAR phases)

Many priorities in this agenda apply broadly across data-intensive science. Workforce capacity, energy efficiency, software sustainability, federated authentication, and long-term data preservation are challenges shared with life sciences, environmental monitoring, earth observation, and other domains. Where contexts differ, adaptation is needed: life sciences operates under GDPR for primary data; IoT and environmental monitoring work from independent streams without central correlation; real-time earth observation prioritises latency over throughput. In the following, the agenda distinguishes between domain-specific findings grounded mainly in HEP and RA evidence (see section 4, sub-section "Challenge" and Annex A) and cross-cutting recommendations applicable to data-intensive science more broadly (see section 4, sub-section "Recommendations").

1.2. Vision and Strategic Goals

Long-Term Vision

"By 2035, European researchers will access a seamlessly integrated compute and data continuum spanning High Performance Computing, High Throughput Computing, cloud resources, quantum computing, and edge infrastructure. Scientists will execute complex workflows across heterogeneous resources without concern for infrastructure boundaries, authentication barriers, or data locality constraints. Europe will maintain digital sovereignty over its research data whilst participating in global scientific collaborations."

Realising this vision requires progress across several dimensions. Researchers need **unified access** to European computing and data resources through federated governance, harmonised allocation mechanisms, and standardised interfaces. **FAIR data management** must enable exabyte-scale workflows with automated placement, transport, and long-term preservation. **Portable and reproducible workflows** require distributable software stacks and architecture-optimized code that exploits heterogeneous hardware. **Embedded AI/ML capabilities** must support simulation, reconstruction, and analysis throughout scientific workflows. Operations must balance scientific ambition with **environmental responsibility** through green computing practices. **Federated identity** must ensure cybersecurity and **European digital sovereignty** across all resources.

Strategic Goals

The SRIDA pursues the following five overarching goals, which structure the priorities presented in Section 4:

1. **Coherent governance** connects vertical, thematic research infrastructures with horizontal e-Infrastructures through formal coordination mechanisms;
2. **Seamless resource access** enables researchers to work across infrastructure boundaries through multi-year allocation for RIs, aligned with evolution of their facilities, and federated authentication;
3. **Technical interoperability** standardises interfaces for heterogeneous computing and cross-facility workflow orchestration;
4. **Sustainability** keeps software portable across architectures and reproducible over decades-long programmes, and holds environmental impacts under control through appropriate tools and policies;
5. **Human capacity** develops expertise combining domain science with computing skills, supported by stable career pathways.

1.3. Target Audience

The SRIDA addresses three primary stakeholder groups whose coordinated action is essential for realising the compute and data continuum.

Scientific communities and research infrastructures include domain scientists, software developers, workflow engineers, and operations teams conducting data-intensive research. For this audience, the SRIDA provides a strategic framework and investment priorities (see Section 5) for digital infrastructure evolution aligned with published scientific roadmaps (e.g. the HL-LHC programme plan, the SKA Construction and Science Plan, and the LOFAR 2.0 plan), helping communities articulate requirements and engage with infrastructure planning.

Service providers and e-Infrastructures include national and community computing and data centres, EuroHPC sites, network operators (e.g. GÉANT), the EGI federation and the EOSC Federation. For this audience, the SRIDA offers strategic direction for service development and investment priorities, identifying where interoperability and coordination can enhance service delivery.

Policy makers and funding bodies include ESFRI delegates, the EuroHPC Governing Board, the European Commission, and national research councils. For this audience, the SRIDA provides evidence-based recommendations for strategic investment, grounded in documented requirements from scientific communities.

The SRIDA may also be of interest to secondary audiences including independent researchers outside large collaborations, who benefit from accessible entry points to shared infrastructure; universities, industry, and SMEs developing technologies or collaborating with research infrastructures; and, where relevant, citizens engaged with research infrastructures through public-engagement channels operated by the RIs themselves.

1.4. Document Structure

The SRIDA is organised as follows:

- Section 2, **Integrated Strategic Foresight Framework and Methodology**: Strategic foresight framework and evidence gathering approach
- Section 3, **Strategic Context and Trends**: European landscape and five transformations shaping infrastructure evolution
- Section 4, **Strategic Priorities**: Thirteen priorities organised within the Four Pillars framework, with challenge analysis and recommendations for each pillar
- Section 5, **Investment Areas**: Resource requirements mapped to priorities and investment types
- Section 6, **Multi-Annual Roadmap**: Implementation phases across short-term, medium-term, and long-term horizons
- Section 7, **Broader Impact**: Brief analysis of socio-economic impact
- Section 8, **Conclusion**: Brief summary of the document
- Annexes, Detailed priority specifications, priority template and summary of supporting evidence

2. Integrated Strategic Foresight Framework and Methodology

2.1. Purpose and Rationale

The development of SPECTRUM's Strategic Research, Innovation and Deployment Agenda (SRIDA) and Technical Blueprint employs an integrated strategic foresight framework that builds upon the foundational work completed in WP3/WP4 (Community of Practice establishment) and WP5 (use cases, landscape analysis, and gap identification). This framework transforms the comprehensive analytical base into forward-looking strategies for data intensive European research infrastructure, informed by HEP and RA, over the next years.

Rather than developing multiple scenarios or alternative architectures, SPECTRUM's approach focuses on creating an agreed vision that can adapt to different rates of technological change and varying implementation conditions. The framework employs foresight methods not to create competing futures, but to stress-test and refine a coherent strategy that addresses identified priorities whilst maintaining flexibility for evolution.

Figure 2.1.1: Integrated Strategic Foresight Framework Diagram.

2.2. Relationship to Technical Blueprint

The SRIDA and Technical Blueprint are complementary elements of a unified strategy. The Blueprint provides the technical architecture that enables achievement of the strategic priorities, whilst the SRIDA ensures that investments, governance, and implementation sequencing support the Blueprint's realisation.

Both documents were built upon substantial preparatory work. The Community of Practice (launched under WP3 and further developed through WP4) has established collaborations with over 50 participating entities, creating working groups across domains and technical areas that provide the distributed expertise essential for strategy development. It has also enabled the project to carry out a series of consultations and interviews with key stakeholders. The landscape analysis (WP5) has delivered comprehensive use case analysis, infrastructure mapping, and gap identification that grounds forward-looking strategies in current realities and validated requirements, all with the support of a Community of Practice (WP3).

The relationship between documents serves different audiences. Technical implementers read the Blueprint for architecture then consult SRIDA priorities for stakeholder coordination and investment sequencing. Policymakers read the SRIDA for strategic rationale then reference the Blueprint for technical validation.

Research communities use both documents together to understand what digital infrastructure can provide and how to influence its development.

2.3. Strategic Development Approach

The SRIDA development follows three integrated phases that move from analysis through synthesis to implementation planning.

The first phase, **Anticipate**, synthesises WP5 findings with forward-looking analysis to establish strategic context. This involves examining how identified gaps might evolve, projecting technology maturation timelines, and assessing how external factors (e.g., policy changes, funding landscapes, technological breakthroughs, environmental impacts) could influence implementation. Through Community of Practice (CoP) engagement, SPECTRUM validates and expands the vision for data intensive European research infrastructure, establishing the parameters within which both the Technical Blueprint and strategic priorities must operate.

The second phase, **Analyse**, develops and evaluates strategic components in parallel tracks. The Technical Blueprint track translates requirements into a coherent, high-level architecture design that addresses identified gaps whilst maintaining flexibility for technology evolution. Simultaneously, the SRIDA track analyses and prioritises strategic initiatives across governance, architecture, software, and workforce dimensions. Cross-impact analysis ensures technical feasibility aligns with strategic priorities, whilst the Community of Practice validates both architectural approach and priority rankings through structured consultation.

The third phase, **Prepare**, integrates the Technical Blueprint architecture with SRIDA priorities into a coherent implementation pathway (i.e., a roadmap). Using backcasting from the target state, SPECTRUM identifies critical milestones, resource requirements, and decision points. The Community of Practice transitions to co-designers and advocates, helping define implementation sequences and becoming champions for adoption within their organisations.

Throughout this process, evidence is drawn from internal sources (SPECTRUM project deliverables, Technical Blueprint requirements, CoP working group outputs) and external sources (European and international roadmaps from external sources, categorised in Annex C.1, including policy and funding bodies, strategic coordination groups, e-Infrastructure providers, and thematic research facilities stakeholder interviews, and interactive sessions at conferences and workshops). Engagement with the three primary stakeholder groups (scientific communities and research infrastructures, computing and data service providers, and policy makers and funding bodies) occurs through technical workshops, CoP sessions, consultations, and the SPECTRUM Knowledge Hub.

Each priority undergoes systematic assessment across strategic importance, implementation complexity, stakeholder engagement, and measurability. The standardised priority template (Annex B) ensures consistent treatment across all thirteen priorities, with explicit traceability from identified gaps through strategic interventions to technical implementation. This methodology enables systematic comparison across priorities and provides clear accountability frameworks for implementation.

3. Strategic Context and Trends

European data-intensive science is entering a period of structural change. The instruments coming online over the next decade will generate data at scales that exceed current infrastructure capacity by an order of magnitude and more. Addressing this challenge requires coordinated European action across governance, technology, sustainability, and workforce dimensions. This section identifies the key drivers shaping this environment and their implications for the European research infrastructure strategy and the data intensive sciences with a focus on HEP and RA.

3.1. European Context

SPECTRUM operates within a European research and digital infrastructure ecosystem in which policy frameworks, coordination bodies, federations, sites and the scientific instruments they serve interact across multiple levels. Understanding this ecosystem is essential for positioning the SRIDA's priorities and identifying where its recommendations complement, extend, or depend upon existing initiatives. Figure 3.1.1 provides an overview, structured across five organisational levels (L0 scientific instruments, L1 sites, L2 federations, L3 coordination, L4 policy and regulation) and two tracks: a *thematic* track that is domain-specific and follows a single scientific community, and a *horizontal* (or *digital*) track that delivers cross-domain compute, data and network services. The remainder of this section describes each layer from bottom (L0) to top (L4) and closes with a note on the evolving infrastructure categories that intersect this picture.

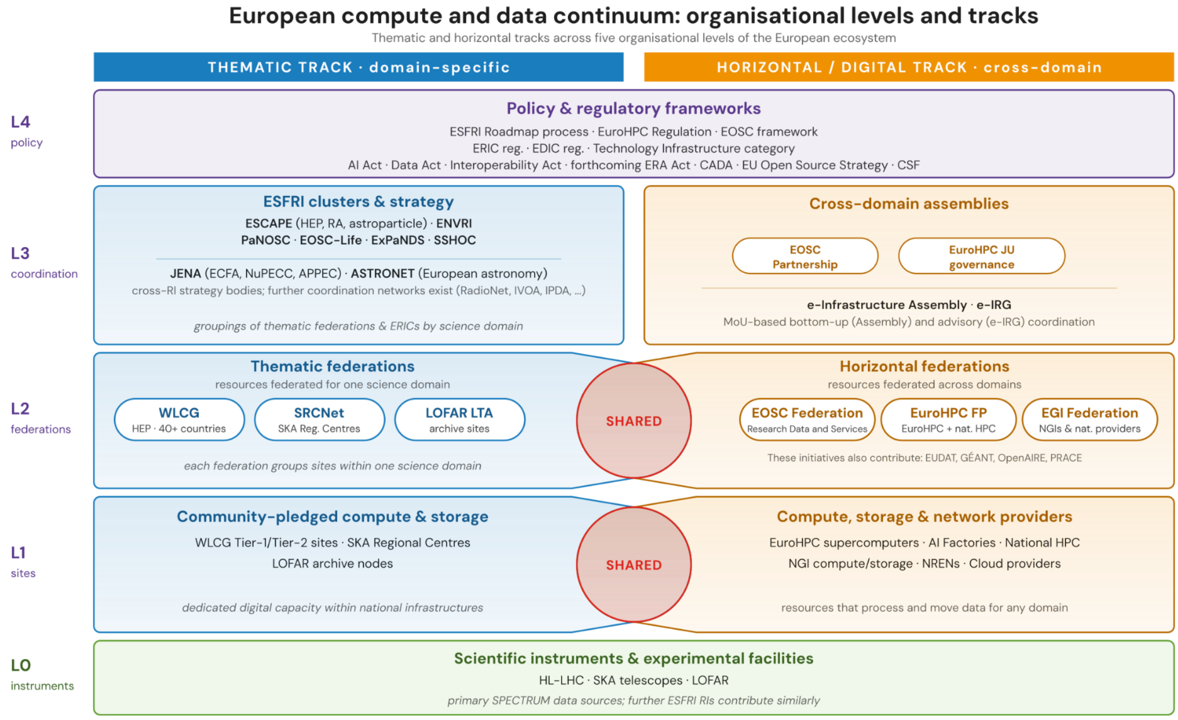


Figure 3.1.1: European compute and data continuum (organisational levels and tracks).

(L0) Scientific instruments

The instrument layer is what the rest of the ecosystem ultimately exists to serve. For SPECTRUM, the primary instruments are the High-Luminosity LHC at CERN, the SKA telescopes operated by SKAO, and the LOFAR telescope operated by ASTRON. Further ESFRI research infrastructures generate comparable scientific data and contribute to the broader continuum but are not the SRIDA's primary focus.

(L1) Sites

The site layer is where compute, storage and network capacity is physically delivered.

Thematic sites carry community-pledged capacity dedicated to specific science domains. For HEP these are the WLCG Tier-0 (CERN), Tier-1 (currently 14 national centres) and Tier-2 (around 150 institutional centres) sites. For SKA these are the SKA Regional Centres in their national or regional implementations. For LOFAR these are archive nodes that host the federated long-term archive.

Horizontal sites carry capacity that processes and moves data for any domain. EuroHPC-procured supercomputers (such as LUMI, LEONARDO and MareNostrum 5), hosted by national consortia under EuroHPC governance, deliver European-level high-performance capability. National HPC centres, funded predominantly through national programmes, provide the bulk of additional HPC capacity. NGIs (national e-Infrastructures) and other compute providers deliver the resources federated by EGI. NRENs deliver the network capacity that GÉANT connects at European level. Commercial cloud providers supply elastic capacity where research workflows allow. The EuroHPC AI Factories, becoming operational from 2025-2026, provide AI-optimised compute alongside national HPC.

(L2) Federations

The federation layer pools resources from many sites into a single operational service that researchers can use without needing to interact with each site individually.

Thematic federations pool sites within one scientific community. The Worldwide LHC Computing Grid (WLCG) coordinates federated computing and data resources across more than 40 countries for the LHC experiments, operating under a multi-party memorandum of understanding (MoU) governance model. The SKA Regional Centres Network (SRCNet) is being designed from inception as a federated network of national and regional centres serving SKA data to end users and supporting their analysis needs. The LOFAR Long-Term Archive (LTA) coordinates distributed data processing across European LOFAR stations and a federated data archive.

Horizontal federations pool sites across many scientific domains. The EGI Federation aggregates compute and storage delivered by National Grid Initiatives (NGIs) and other providers. GÉANT federates the pan-European networking backbone delivered by the National Research and Education Networks (NRENs). The EuroHPC operations layer, consisting of EuroHPC-procured supercomputers and the national HPC centres that operate them, functions as the horizontal compute federation parallel to EGI and GÉANT. PRACE (Partnership for Advanced Computing in Europe) operates a continuing HPC partnership association alongside EuroHPC, with its own access calls and community programmes. The EUDAT Collaborative Data Infrastructure (CDI) federates more than 20 European research organisations and data centres delivering the B2-suite data services and is on a path to join the EOSC Federation as a Node, parallel to EGI's onboarding.

(L3) Coordination

Two complementary coordination structures organise European research infrastructure collaboration. The *thematic* track brings together research infrastructures within scientific domains. The *horizontal* track federates the cross-cutting compute, data and networking services on which all scientific domains depend.

Thematic coordination. The ESFRI Cluster projects, launched from 2019 under H2020 INFRAEOSC-05 and named as such in the ESFRI Roadmap 2021, connect research infrastructures listed on the ESFRI Roadmap around shared challenges in open science, data management and computing, with explicit links to the European Open Science Cloud. Five clusters cover astronomy, particle physics and astroparticle physics (ESCAPE), environmental science (ENVRI-FAIR), photon and neutron science (PaNOSC), life sciences (EOSC-Life) and social sciences and humanities (SSHOC); a sixth cluster for digital research infrastructures is in preparation. Cross-RI strategy is coordinated by JENA (Joint ECFA-NuPECC-APPEC) for particle, nuclear and astroparticle physics, and by ASTRONET for European astronomy. Further coordination networks operate alongside, including astronomy-specific networks (RadioNet, Europlanet, Opticon, Solarnet) and international standards alliances (IVOA, IHDEA and IPDA for virtual observatory, heliophysics and planetary data coordination respectively).

Horizontal coordination. On the digital infrastructure side, two tiers operate side by side. The formal EC-driven tier comprises the EOSC Partnership, the Horizon Europe co-programmed partnership that operationalises the EOSC framework, and the EuroHPC JU governance; both coordinate top-down with EC policy and funding. A bottom-up advisory tier comprises the e-Infrastructure Assembly, an MoU-based forum where the major European e-Infrastructures (EGI, GÉANT, PRACE, EUDAT and others) coordinate operationally, and the e-Infrastructure Reflection Group (e-IRG), a Member-State-level advisory forum that aligns digital research infrastructure policy with national programmes. OpenAIRE coordinates open-access scholarly infrastructure, EVERSE addresses research software quality and sustainability, and European Data Spaces establish frameworks for cross-sectoral data sharing.

(L4) Policy and regulatory frameworks

European policy frameworks provide strategic direction, governance and funding for research infrastructure at a European level. This matters because research, which is inherently cross-institutional and trans-national, requires coordinated decisions that single Member States cannot take alone. The EuroHPC Joint Undertaking, established under the EuroHPC Regulation, coordinates European supercomputing on a 50/50 co-funding model between the European Commission and participating states, with exascale systems becoming operational in 2025-2026. The European Open Science Cloud (EOSC) establishes the framework for federated data infrastructure, FAIR practices, and open science services; its Federation and EU Node are under active development. ESFRI, through its Roadmap and Landscape Analysis, has long recognised the strategic importance of the research infrastructures that anchor this agenda: the LHC programme (with HL-LHC operations extending to the late 2030s), the SKA Observatory (with full operations continuing beyond 2040), and LOFAR (operating since 2010, with LOFAR 2.0 extending operations into the next decade). RAISE (Resource for AI Science in Europe), launched in November 2025, coordinates AI capabilities for science across the EU.

The regulatory and policy environment is evolving. The AI Act, Data Act, Interoperability Act, the forthcoming ERA (European Research Area) Act, and ongoing Omnibus developments establish new requirements for research infrastructure operations covering AI governance, data sharing, interoperability standards, and cross-border resource circulation. On the sovereign-supply dimension, the European Technological Sovereignty Package (COM(2026) 503, 3 June 2026) comprises four interlinked initiatives: a Chips Act 2.0, the Cloud and AI Development Act (CADA), an EU Open Source Strategy, and a Strategic Roadmap for Digitalisation and AI in the Energy Sector. CADA aims to develop European cloud and AI infrastructure, complemented by the Cloud Sovereignty Framework (CSF), used by the Commission since April 2026 to grade providers across sovereignty criteria for public-sector procurement. The EU Open Source Strategy, included in the package, promotes secure, transparent and auditable open source ecosystems for critical systems and digital infrastructures, the model on which HEP and radio astronomy computing already operate.The European Commission's 2025 Strategy on Research and Technology Infrastructures identifies research infrastructures (both the scientific facilities and the digital infrastructures that operate them) as underpinning European technological and data sovereignty. The 2028-2034 Multiannual Financial Framework allocates sustained investment through the European Competitiveness Fund, creating a funding pathway from research to deployment of digital research infrastructures.

Evolving infrastructure categories

The European infrastructure landscape is expanding beyond the traditional Research Infrastructure model. The European Commission is introducing new infrastructure categories and pushing towards a more integrated landscape in which research, technology, and digital infrastructures complement one another. Three legal-instrument categories now coexist:

- **Research Infrastructures** (ESFRI Roadmap, ERIC regulation): the established pathway for publicly funded scientific facilities. CERN, SKAO and LOFAR ERIC operate under this model. The ESFRI Roadmap process signals strategic importance and facilitates engagement with national ministries.
- **Technology Infrastructures (TIs)**: a category promoted by the EC, focused on facilities for developing, testing, and validating technology to foster industrial competitiveness. TIs are oriented towards the commercial sector, with the expectation that industry pays for services. The RI/TI distinction is one of primary mission and beneficiary (research community vs industrial sector) rather than of operational maturity: many RIs, including the LHC programme and SKA Observatory, operate at very high TRL.
- **Digital Infrastructures**: distinct from RIs in legal form (the EDIC regulation, under the Digital Decade, provides the cross-border legal instrument, parallel to the ERIC regulation for RIs) and in primary orientation (Digital Decade goals for public and commercial digital transformation).

These categories mark a reorganisation in which "**e-Infrastructures**", a named funding priority under the Horizon 2020 regulation (Regulation (EU) No 1291/2013 and its Specific Programme, Council Decision 2013/743/EU), is being retired as a term. EuroHPC focuses on supercomputing for both industry and science, EOSC is being shaped as the Common European Data Space for Research and Innovation, and ESFRI focuses on research facilities. GÉANT has traditionally been supported by a Framework Partnership Agreement. EGI, by contrast, plays a key role in the European landscape by federating national and community computing facilities that underpin many research infrastructures, yet has no comparable structural EC support; the same applies to EUDAT and OpenAIRE. The e-infrastructures hence play an important role in connecting these facilities for transnational use and they have led the development of federation processes and tooling over the last 2 decades. How this federation layer will be supported once the e-Infrastructures category is retired is an open gap, which Section 5 addresses as an investment priority.

The EC is actively promoting integration across these categories, seeking to avoid silos and build a continuum of "knowledge infrastructures" spanning research, technology, digital, and data dimensions. For SPECTRUM, this matters in two ways. First, the research-IT compute and data continuum serves publicly funded research communities. Historically, digital research infrastructures have been funded through Horizon work programmes rather than the ESFRI Roadmap, but ESFRI's recognition of EOSC and the rise of the DIGIT working group are moving them into closer alignment with the RI tradition. Second, the technologies and services this continuum develops (federated computing, data management, AI platforms) have broader applicability that the EC's integrated infrastructure vision recognises. Engagement with these emerging categories, particularly where the EC encourages collaboration between RIs and TIs, is an active question for positioning the agenda's priorities.

Foundation for this agenda

SPECTRUM builds on two decades of operational experience from the federations and sites described above. National funding provides most of computing capacity across all of them: EuroHPC operates on a 50/50 co-funding model, while the resources underpinning WLCG, SRCNet, LOFAR and EGI are predominantly nationally funded. Within EOSC, an in-kind co-funding model is being trialled as a potential mechanism for sustaining pan-European services, complementing direct national and EU funding streams. The documented requirements and identified gaps from these communities, combined with analysis of European strategy documents, inform the strategic priorities presented in Section 4. For detailed descriptions of European initiatives, see D6.1 Technical Blueprint Annex B.

3.2. The Evolving Landscape

Seven external drivers are reshaping the strategic environment for European data-intensive science. These drivers emerge from the landscape analysis within WP5, of scientific use cases (Deliverable 5.1), infrastructure capabilities (Deliverable 5.3), access policies (Deliverable 5.2), community consultation, and alignment with European strategy documents. Each driver represents a force that current infrastructure arrangements cannot adequately address through incremental change alone. Together, they define the transformation that the strategic priorities in Section 4 are designed to deliver.

Figure 3.2.1: Seven strategic drivers shaping the European compute and data continuum.

3.2.1. Digital Infrastructure Scale

European flagship instruments will generate data volumes that exceed current processing, storage, and networking capacity by an order of magnitude from 2030 onwards. Both flagship programmes are projected to generate hundreds of petabytes per year at full operation (HL-LHC during Run 4-6, SKA1 from AA4 onwards), with multi-exabyte aggregates over their lifetimes; SKA's challenge is concentrated in the real-time signal path, HL-LHC's in long-term storage and reprocessing. Meeting this scale through linear growth alone (e.g., adding nodes, disks, and bandwidth) would breach power, cost and procurement envelopes by the early 2030s. Architectural change is therefore required across compute, storage, and networking (heterogeneous computing, near-data processing, federated storage), as set out in detail in the Technical Blueprint (D6.1 Section 3).

Meeting this scale requires architectural change across compute, storage, and networking (heterogeneous computing, near-data processing, federated storage), not linear scaling. Two factors compound the pressure. On the supply side, efficiency gains from new hardware tend to be absorbed by increased demand (the rebound effect documented in D6.1 §3.2), and in the current European energy mix the fabrication-phase environmental footprint of replacing hardware can outweigh usage-phase savings. On the demand side, AI workloads (§3.2.2) increasingly share the same infrastructure as data-intensive science, pulling resources in directions that traditional simulation does not. The architectural alternatives, scheduling and life-cycle practices needed are set out in the Technical Blueprint (D6.1 §3.2, §4 capabilities and §5.3 software portability).

European exascale systems offer new and transformational computational capabilities, but realising their potential for science requires addressing a fundamental mismatch: strategically important data intensive research infrastructures (e.g., ESFRI roadmap) operate on multi-decade timelines, yet access to HPC resources typically follows annual competitive cycles. Major experiments need planning certainty over 5-10 year horizons; current competitive, rather than strategic, allocation models provide 1-2 years at most. Bridging this temporal gap is essential for effective (digital and research) infrastructure utilisation, planning and alignment.

Architecture diversity adds complexity. Computing platforms now span multiple processor families and accelerator types, each requiring distinct software optimisation. Scientific workflows must operate across this heterogeneous landscape. Investment in portable software infrastructure (e.g., EESSI - the European Environment for Scientific Software Installations - and containers) can reduce fragmentation, but requires coordinated community effort, in-depth hardware architecture skills and sustained funding. In addition, the European systems need to consider high-volume data processing throughput as an essential criterion in the design of these systems. This means optimizing the full chain from data repository access (connecting to distributed storage systems), networking (connecting datacentres) and the system configuration within the datacentre (e.g., network, staging/processing storage, compute server layout).

Strategic implication: Europe needs allocation mechanisms that bridge the mismatch between infrastructure timelines and funding cycles, alongside investment in software portability that enables efficient use of diverse architectures. High-volume data processing throughput needs to be an essential criterion in the design of European systems, and these systems need to be connected through federated identity, interoperable interfaces and unified data movement so that researchers can compose pipelines across HPC, HTC and cloud resources without per-stack glue. Strategic planning, allocations and co-design with the data-intensive research communities is essential for achieving this.

3.2.2. AI Adoption

Artificial intelligence is being adopted rapidly across the research data lifecycle as a tool that supports simulation, reconstruction, analysis, and operations. Adoption is uneven and the scientific value of AI depends on the circumspection with which it is used, but demand for AI-capable infrastructures is growing fast across European research communities. Examples already in production span across scientific domains: AlphaFold for protein structure prediction, ECMWF machine-learning weather models that rival physics-based simulation for medium-range forecasting, and CNN-based gravitational-wave signal detection in LIGO/Virgo. The pattern is consistent: AI accelerates specific stages of the research workflow rather than replacing the scientific method that surrounds them.

A maturity gap constrains progress. Most AI applications in physics remain at the proof-of-concept stage rather than production deployment. Scaling from demonstration to operational use is challenging, often dependent on small teams rather than sustainable capacity. The gap between AI potential and AI reality represents a strategic bottleneck.

Hardware evolution creates both opportunity and tension. The computing industry is optimising for AI workloads that differ from traditional scientific simulation, creating divergence between commercial hardware roadmaps and scientific computing needs. Some workloads (e.g., AI inference, signal processing, parts of generative simulation) can exploit reduced-precision AI-optimised hardware directly; others (e.g., Lattice QCD, climate, high-precision lattice models) remain FP64-bound and risk diminishing alignment with mainstream technology development unless mixed-precision strategies or dedicated capability systems are sustained alongside AI-optimised platforms.

Scientific communities must either adapt to AI-optimised hardware or accept diminishing alignment with mainstream technology development.

Alternative approaches are emerging. Smaller, more efficient AI models can run on distributed infrastructure closer to data sources, reducing the need for massive, centralised resources. Domain-specific scientific AI models, developed by and for research communities, provide capabilities tailored to scientific needs rather than commercial applications, but their training, validation, and continued availability depend on curated scientific datasets, dedicated training compute, and governance arrangements that only sustained research infrastructure can underwrite.

European policy is also shaping AI ambition. The AI Continent Action Plan commits substantial public investment to AI infrastructure (over €10 billion for AI Factories, €20 billion for AI Gigafactories, and a target of tripling European data-centre capacity), and the AI Act sets the regulatory baseline for trustworthy deployment. For research specifically, RAISE (Resource for AI Science in Europe), launched in November 2025, coordinates AI resources along two dimensions, "Science for AI" and "AI in Science", through a Secretariat, Digital Hub, and Academy; its first project, SCIANCE, is mapping the European AI-in-science landscape. The direction of AI in research is no longer set by demand and technology alone; it is increasingly shaped by these European policy choices.

Strategic implication: Europe needs community-driven AI infrastructures that serve scientific requirements. Governance arrangements, including coordination instruments such as RAISE, must let research communities shape AI deployment rather than reduce them to consumers of commercial offerings.

3.2.3. Environmental Sustainability

Energy availability and environmental impact are becoming primary constraints on infrastructure growth. Technically speaking, some locations face grid and cooling capacity limits. Politically speaking, in a context of multiple environmental crises, societal expectations for environmental responsibility are rising. Meanwhile, the use and the energy intensity of AI and large-scale computing are increasing rapidly, so that data centre power demand and environmental impact are projected to grow substantially over the coming decade, overwhelmingly driven by commercial AI workloads. Research-IT growth, while non-trivial, remains a small share of that total, and European public research infrastructures lead on efficiency (see Green500 standing of EuroHPC systems).

Environmental accountability is becoming a condition of public funding and social licence to operate. To resolve this tension, hardware and software tools have emerged to measure and report resource usage and methodologies based on life-cycle assessments and even ISO standards defining systems and metrics are being developed to quantify environmental impact. However, this is done piecemeal and a posteriori, instead of globally and proactively to contain impacts and avoid transfers.

Europe has demonstrated leadership in energy-efficient high-performance computing. European systems achieve world-leading efficiency metrics, as shown in EuroHPC supercomputers global standing in Green500 [85]. Emerging computing paradigms (photonics, neuromorphic architectures) offer potential for further efficiency gains. But realising this potential requires sustained investment and coordinated adoption. Sustainability also stands at the core of European policy initiatives, the Digital Decade and the European

Green Deal frame the "twin transitions" (digital and green) introduced under the 2019-2024 Commission, and is a key driver of future European competitiveness. At EU level, sustainability is becoming a regulated property of infrastructure rather than a voluntary aspiration: the European Technological Sovereignty Package (June 2026) pairs the CADA requirement that new data centre capacity be planned in line with sustainability goals with a Delegated Regulation establishing an EU-wide scheme to rate the sustainability of data centres in Europe. Federated scientific infrastructures will increasingly need to report against such schemes, which reinforces this agenda's emphasis on energy-aware operations and transparent accounting.

Strategic implication: Europe needs a strategic, coordinated approach to sustainability in research computing, bringing digital infrastructures, research infrastructures, and the Commission together, supported by an EU-level action that catalyses the collaboration, common sustainability standards, accountability mechanisms covering the full facility life cycle, and investment in energy-efficient technologies and operational practices.

3.2.4. Security and Trust

The security landscape is evolving in ways that require proactive adaptation. Quantum computing developments create urgency for cryptographic transition. Human-based authentication requirements increasingly, due to single person failure, conflict with the automation that scientific workflows require as well as for collaborative asset-maintenance and the incorporation of AI like AI agents. Federated operations across organisational and national boundaries depend on several trust frameworks that in terms of cross-framework alignment require additional and coordinated effort at an EU-level.

Data governance is evolving. European Data Spaces initiatives are establishing standards for secure data sharing that research infrastructures must align with to ensure interoperability with the broader European regulatory frameworks (GDPR, Data Act, Data Governance Act, and AI Act). The scientific data and software produced by European research communities has important societal and commercial impacts and hence are unique assets requiring dedicated stewardship (at an EU-level) that cannot be outsourced. Therefore, the constraints arisen from digital sovereignty drivers also have an impact on security and trust

Strategic implication: Europe needs proactive security evolution, federated trust frameworks enabling cross-boundary collaboration grounded in EU law and regulations whilst remaining fit for purpose, and alignment with emerging European data governance frameworks.

3.2.5. Long-Term Preservation

Multi-decade instrument timelines create preservation challenges that extend beyond immediate operational concerns. Scientific software and data must remain usable across hardware generations and organisational changes spanning at least 10-50 years. The expanding scope of the FAIR principles (now encompassing software, workflows, and AI models alongside data) reflects a growing recognition that reproducibility depends on preserving the complete digital context, not just datasets.

Active preservation requires sustained investment, and an explicit ownership model for the long-term cost of keeping data assets accessible, covering software functionality as underlying platforms evolve, data-format migration as standards change, and documentation sufficient for future reproduction. Software reuse should be encouraged via libraries usable across projects. Current funding models rarely support such long-term stewardship.

Strategic implication: Europe needs preservation strategies ensuring long-term (decades) reproducibility, sustained funding for software and data stewardship, and FAIR practices extended across the complete computational context. Re-use of data and reproducibility of processes are key elements for not only the scientific method but also for societal and commercial impact.

3.2.6. Workforce Capacity

The research computing workforce is undergoing fundamental change. In scientific domains, there used to be little distinction between researchers working on hardware, software and analysis tasks; this is shifting due to the larger specialisation in experimental physics, and with the advent of new technologies which need a more solid background in computing: skills requirements are shifting as heterogeneous architectures, machine learning methods, and security demands reshape what expertise is needed. Career structures designed for traditional academic paths do not accommodate the IT roles that modern research infrastructure requires. Competition with industry for technical talent has intensified as the same skills command premium salaries in the commercial sector. At the same time the research sector provides a unique and innovative frontier for digital challenges. Hence, the capability of upskilling workforce spilling over from academia to industry is also a driver for economic and social social impact

Workforce capacity affects capacity across all other areas. Scale requires specialists in heterogeneous computing, software portability, and performance engineering. AI depends on specialised engineers who can bridge the gap between proof-of-concept and production deployment. Environmental sustainability goals require expertise in code optimisation, energy-aware scheduling, and efficiency monitoring. Trust and continuity demand security specialists and preservation experts who understand both technical and policy dimensions.

Europe faces a significant skills gap in precisely these areas. Community surveys indicate that most of research groups identify technical capacity as a concern, yet lack the resources to address it. Scientific career structures often do not recognise or reward contributions to infrastructure, software development, or operational efficiency. Research software engineers, who bridge domain science and computing expertise, typically lack defined career paths. This is further explained, for example, in the Briefing Book from the European Strategy Particle Physics update [143]. Incentivizing the creation and sustainability of IT skills together with the research community can be another important opportunity in the co-design process.

Strategic implication: Europe needs coordinated career frameworks that recognise and reward research and digital infrastructure contributions, training programmes aligned with emerging skill requirements, and retention strategies that can compete with industry for technical talent.

3.2.7. Digital Sovereignty

European research infrastructure increasingly operates within a policy environment that treats digital capacity as a strategic asset. The European Commission's 2025 Strategy on Research and Technology Infrastructures identifies research infrastructures, both the scientific facilities and the digital infrastructures that operate them, as "underpinning European technological and data sovereignty". Frameworks alone are insufficient: sovereignty also depends on sustained European data infrastructure (storage capacity and the institutional expertise to curate it). The Cloud Sovereignty Framework (CSF v1.2.1, October 2025, EC DG DIGIT) defines eight Sovereignty Objectives (notably SOV-4 Operational Sovereignty, on the EU's practical ability to run and evolve technology independently) and a five-level SEAL assurance scale used in public-sector procurement. The 2028-2034 MFF allocates sustained investment through the European Competitiveness Fund, creating a funding pathway from research to European strategic autonomy. The strategic backdrop is quantified: a Commission study estimates a European compute capacity gap of around 20 GW by 2036, and three hyperscalers currently hold approximately 70% of the European cloud market.

For data-intensive science, sovereignty concerns manifest in three dimensions. First, dependence on non-European technology providers for critical components (processors, interconnects, cloud platforms) creates supply chain vulnerabilities that, for example, the EuroHPC Joint Undertaking addresses through open technologies such as RISC-V and the European Open Stack. Second, scientific data and software produced by European research communities are unique assets requiring processing within European jurisdictions (GDPR, Data Act, Data Governance Act, AI Act) under European governance, particularly as AI workflows (see §3.2.2 and P07) increasingly involve sensitive or strategically valuable datasets. Third, reliance on commercial AI models for scientific applications raises questions of reproducibility, auditability, and long-term availability that open-source, community-developed alternatives can address.

These concerns are not hypothetical. Recent international developments, including funding cuts, database closures, and restrictions on research collaboration, demonstrate that open science requires resilience frameworks capable of withstanding geopolitical disruption. European federated infrastructures, built on open-source technology stacks with multi-jurisdictional governance, provide this resilience by design.

Europe's procurement response operates within tight legal constraints. Equal-treatment obligations (2019 EC guidance on third-country bidders) prevent excluding bidders on origin alone; available levers are the

International Procurement Instrument (Regulation (EU) 2022/1031) for reciprocal restrictions, IPCEI for European industrial-capacity state aid, and CSF tender-grading within those rules. Stronger preferences likely require new legislation: the CADA, proposed on 3 June 2026 within the European Technological Sovereignty Package, targets faster permitting and deployment of energy-efficient data centres, at least a tripling of EU data centre capacity within five to seven years, and a common EU-level procurement framework that lets public administrations pool purchasing power and choose European sovereign solutions.The Berlin Declaration (November 2025, 17 Member States at ministerial level) anchors the political framing, including protection against extraterritorial application of non-EU laws.

Strategic implication: Europe needs sovereignty-aware infrastructure strategies that ensure European control over critical research computing assets through open technologies, federated governance, and sustained investment in European digital capacity, while maintaining the openness and international collaboration essential to scientific progress and at the core of European digital norms and values (e.g., the European Declaration on Digital Rights and Principles). Among these levers, open technologies are increasingly recognised as a sovereignty asset in their own right: the Commission's EU Open Source Strategy (COM(2026) 503, June 2026) names CERN's open tools (ROOT, the Open Hardware Licence, Open Research Europe) and EOSC as European exemplars, reinforcing the open-technology priorities (P08) and the case for sustaining these tools as a sovereignty investment.

3.3. From Drivers to Priorities

Responding to these drivers requires coordinated action across four dimensions: governance, technical architecture, workforce, and the business/funding models that sustain operation over the lifetime of the supporting infrastructures. No single driver maps to a single response: Infrastructure Scale demands both architectural evolution (heterogeneous computing, data management) and institutional change (multi-year allocation mechanisms). Digital Sovereignty requires open technology choices, federated governance models, and workforce investment in European capabilities. Environmental Sustainability spans infrastructure design, operational practices, and policy frameworks. This cross-cutting nature shapes the structure of this agenda.

This agenda organises thirteen strategic priorities within four pillars, each addressing a distinct dimension of intervention. **Pillar 1 (Policy, Trust and Governance)** establishes the institutional frameworks enabling coordinated action, including cross-infrastructure governance, resource allocation mechanisms, and federated identity. **Pillar 2 (Architecture and Interoperability)** addresses the technical infrastructure for heterogeneous computing, data management, and workflow orchestration. **Pillar 3 (Software and Science Enablers)** develops the tools and practices that translate infrastructure into scientific capability, spanning AI integration, portability, reproducibility, and preservation. **Pillar 4 (Human Capital and Responsibility)** ensures the workforce capacity, community collaboration, and environmental sustainability foundations for long-term success.

The pillars follow a sequencing logic reflected in the roadmap (Section 6). Governance and trust foundations (Pillar 1) must be established before technical integration can proceed at scale (Pillars 2 and 3), while workforce development and sustainability (Pillar 4) span all phases as enabling conditions. Priorities within each pillar are interdependent and reinforce each other; dependencies across pillars are addressed in Section 4 and the implementation roadmap.

Section 4 presents these priorities with their challenges, evidence base, and recommendations. Annex A provides detailed specifications for each priority, including implementation pathways and success indicators.

4. Strategic Priorities

Building on the drivers and pillar framework introduced in Section 3.3, this section sets out the thirteen strategic priorities and their evidence base. Each priority addresses gaps identified through systematic analysis of scientific use cases, infrastructure capabilities, and community requirements, aligned with European strategy documents from EOSC, EuroHPC, ESFRI, and domain-specific roadmaps. Recommendations in section 4.x.3 follow the cross-cutting transferability convention introduced in section 1.1; HEP- and RA-specific evidence sits in the section 4.x.1 Challenge sub-sections.

4.1. The Four Pillars Framework

The thirteen priorities are organised into a **Four Pillars framework**. Each pillar represents a distinct dimension of intervention required to achieve the continuum vision for data intensive research, and together they form a coherent strategy where progress in one dimension enables and depends upon progress in others.

Pillar 1: Policy, Trust and Governance establishes the institutional foundations for coordinated infrastructure. This pillar addresses funding coordination mechanisms, multi-year resource allocation frameworks, and federated identity systems enabling authenticated access across organisational and national boundaries.

Pillar 2: Architecture and Interoperability creates the technical machinery for a coherent continuum. This pillar enables heterogeneous computing integration across heterogeneous computing (e.g., HPC, HTC, cloud, quantum and accelerated platforms); data federation and transport at exabyte scale; and workflow orchestration across distributed resources.

Pillar 3: Software and Science Enablers bridges infrastructure capabilities with scientific applications. This pillar advances AI (including ML/DL methods and emerging LLM/agentic approaches) and accelerated computing, software portability across architectures, reproducibility over multi-year timescales, and code preservation across the decades-long lifetimes of major research facilities (e.g., HL-LHC, SKA, LOFAR 2.0).

Pillar 4: Human Capital and Responsibility addresses organisational and environmental dimensions. This pillar builds collaborative communities through co-design processes, establishes environmental sustainability through impact-aware facility management, and develops career pathways combining domain science with computing expertise.

The pillar structure reflects a fundamental principle: all four dimensions must advance together. No single pillar succeeds in isolation. Research infrastructures cannot effectively access diverse computing resources (Pillar 1) without data management capabilities for exabyte-scale workflows (Pillar 2). Technical capabilities prove unsustainable without workforce development (Pillar 4) and lose scientific value without portability and reproducibility mechanisms (Pillar 3). All four pillars rest on a federated digital infrastructure that supplies compute, storage and network capacity through thematic and national centres, EuroHPC and federation initiatives such as EGI; sustaining this physical foundation is a precondition for any of the pillar interventions to deliver value.

Mapping to the Strategic Goals

Each pillar realises one or more of the five overarching goals introduced in Section 1.2 (see Table 4.1 for the mapping).

Table 4.1.1: Mapping SRIDA Goals to SRIDA Pillars.

Pillar	Goals from Section 1.2
Pillar 1: Policy, Trust & Governance	Coherent governance; Seamless resource access
Pillar 2: Architecture & Interoperability	Technical interoperability
Pillar 3: Software & Science Enablers	Sustainability (software portability, reproducibility, preservation)
Pillar 4: Human Capital & Responsibility	Sustainability (environmental); Human capacity

The diagram below illustrates the thirteen priorities organised within their respective pillars.

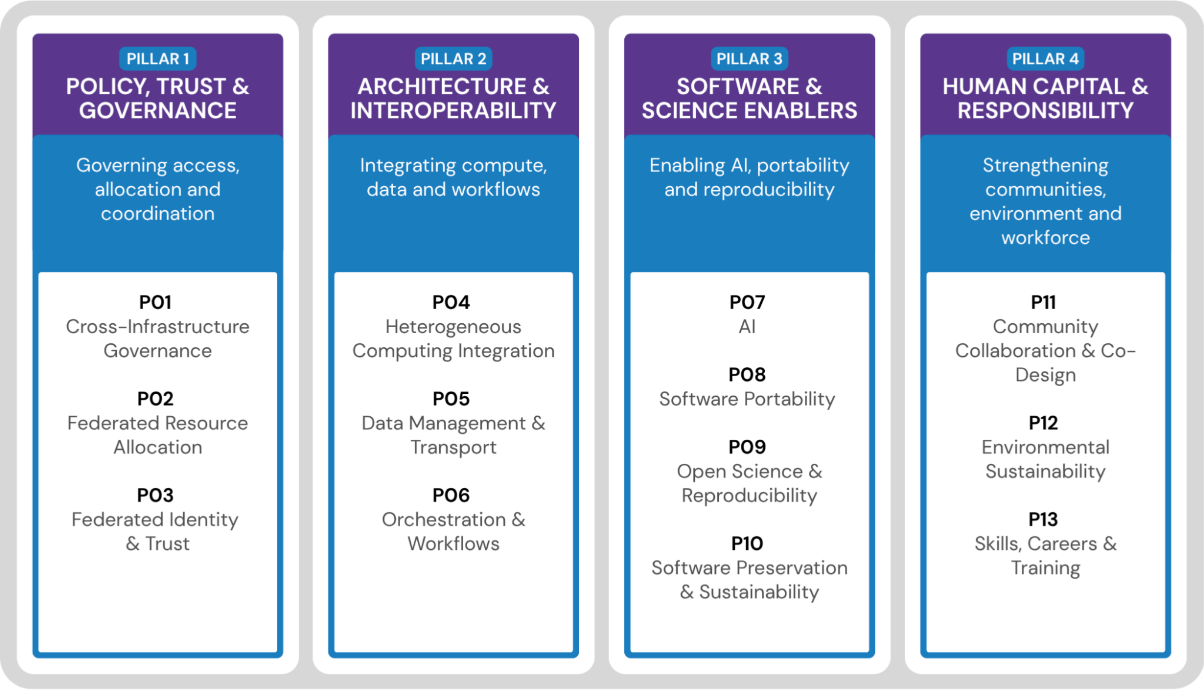


Figure 4.1.1: Strategic Pillars of the European Compute & Data Continuum for Research.

Sustainability spans Pillars 3 and 4 because its two sides, keeping codes portable and reproducible across decades-long programmes and keeping infrastructure environmentally responsible, are operationally distinct interventions delivered by different communities. Governance and access are grouped within Pillar 1 because both are institutional foundations, addressing the question of who coordinates resources and how researchers gain access to them, before the technical machinery of the continuum is introduced in Pillar 2.

4.2. Pillar 1: Policy, Trust & Governance

Governing access, allocation and coordination

4.2.1. Challenge

Europe invests in research computing and data infrastructure through several routes: EuroHPC JU is the principal EC-level channel for compute capacity; EOSC coordinates the federated data-services layer; the bulk of operational HTC, cloud and storage capacity is funded through national programmes and the thematic research infrastructures themselves; Horizon Europe (and successor FP10) work-programme lines support development and integration projects. Yet these investments face a structural challenge: computing and storage requirements are projected to grow by an order of magnitude by the 2030s, while EC-level capacity growth (channelled chiefly through EuroHPC and the AI Factories) is concentrated on exascale and AI-optimised systems, and national operational baselines that fund the bulk of HTC, storage and thematic-RI capacity are not projected to track the demand curve.

Moreover, researchers cannot easily use existing resources across organisational boundaries: the limiting factor is typically national or organisational funding eligibility rather than technical interoperability. EuroHPC and EOSC partially resolve this at EC level, but their access modes are not yet adapted to the long-running, data-heavy workflows characteristic of HEP and RA. A workflow requiring EuroHPC cycles, EGI federation services (which deliver capacity from national and thematic providers), and EOSC data repositories must navigate three separate governance structures, allocation processes, and authentication systems. This fragmentation reduces infrastructure utilisation and creates barriers that favour well-resourced teams over smaller research groups.

Three interconnected challenges prevent the compute and data continuum from functioning as coherent European capability:

- **Governance fragmentation**: Thematic research digital infrastructures (e.g. WLCG for HL-LHC, SRCNet for SKAO) and pan-European multi-disciplinary digital infrastructures (e.g. EuroHPC, EOSC, EGI) operate under different governance models with limited systematic coordination. Success stories exist: the long-standing cooperation between WLCG and EGI demonstrates that effective co-design is achievable. However, structured engagement between data-intensive science communities and EuroHPC remains underdeveloped. Wider coordination mechanisms would benefit all parties.

- **Allocation cycle mismatch**: Research infrastructures operate on multi-decade timelines (the LHC has been operational since 2008 and via the HL-LHC upgrade will run through 2041; SKA is foreseen to be operational for 50+ years from the start of operations), yet must navigate incompatible allocation models across the infrastructure landscape. Federated infrastructures like WLCG operate on pledged resources committed by national funding agencies, whilst EuroHPC centres use competitive peer-reviewed proposals typically granted for one year. Experiments need guaranteed access for 3-5 year periods compatible with their planning cycles, but no unified framework spans these different approaches. This creates planning uncertainty, repetitive submission burden, and discourages investment in workflow optimisation.

- **Authentication fragmentation**: federated digital infrastructures and HPC centres operate partially incompatible identity management systems. Lack of federated identity management and multi-factor authentication (MFA) barriers for unattended workflows prevent seamless authenticated access. Without this, technical interoperability alone provides no benefit to researchers.

4.2.2. Priorities Overview

Pillar 1 establishes governance, allocation, and trust foundations across both compute and data dimensions of the continuum. Three priorities form a layered intervention: governance coordination (P01) creates the institutional basis for allocation agreements (P02), which in turn require authenticated identity (P03) to track usage and enforce entitlements.

P01: Cross-Infrastructure Governance. Establish a coordination forum enabling thematic research digital infrastructures (e.g. WLCG, SRCNet, LOFAR ERIC) to align positions and seek formal representation within pan-European multi-disciplinary digital infrastructures (EuroHPC, EOSC, EGI), with policy-maker representation from national funding agencies and relevant EC units. The forum addresses interoperability across legal, organisational, semantic, and technical layers, coordinates an innovation agenda for data-intensive science distinct from classical HPC optimisation, and develops policy recommendations bridging operational versus project-based funding. National programmes provide the majority of HTC, cloud, and thematic-RI storage capacity, while the EC channel via EuroHPC concentrates on exascale and AI-optimised compute; this division of labour is deliberate, not legacy, and the strategic question is alignment across the two funding streams (consistent with the forthcoming ERA Act framework), not migration of one into the other. (Detail: Annex A, P01)

P02: Federated Resource Allocation. Establish multi-year allocation mechanisms aligned with timelines on which research facilities evolve, supported by standardised, transparent accounting frameworks for heterogeneous resources (CPU, GPU, storage, network) across governance boundaries. Existing models demonstrate feasibility: WLCG enables 5-10 year resource planning through national commitments, and EuroHPC's Strategic Access designation provides a pathway for enabling guaranteed multi-year allocations at the EU-level. Funding bodies can quantify return on investment through standardised accounting metrics tied to multi-year allocation checkpoints, addressing the operational-versus-project funding mismatch that currently undermines long-term scientific programmes and their societal impact. (Detail: Annex A, P02)

P03: Federated Identity and Trust. Deploy federated authentication enabling researchers to access infrastructure across organisational boundaries, with dynamic authorisation, automated delegation, and service-account support for long-running and unattended scientific workflows. The major community and federation AAI proxies, including Indigo-IAM (WLCG, SRCNet), EGI Check-in (EGI), MyAccessID (EuroHPC, EOSC, GEANT) are already AARC-BPA-compliant and accept eduGAIN identities; the integration challenge is interoperation across these proxies (attribute mapping, group and role assertion handling, IdP discovery), together with bridging to site-local SSO systems (e.g. CERN-SSO) for resources under separate authorisation domains. Adopt the AARC BPA as the architectural baseline, contribute to AARC-TREE work on cross-proxy attribute exchange, and make collaborative and service-account operation a first-class capability with delegation patterns that displace the single-person MFA workaround for unattended workflows (Detail: Annex A, P03).

Interdependencies: the three priorities are mutually load-bearing. Allocation (P02) cannot operate without governance agreements (P01) to legitimise cross-boundary use, and depends on federated identity (P03) to attribute usage; trust frameworks (P03) require P01 to gain currency outside their home organisation; and P01's coordination work becomes operationally meaningful only when P02 and P03 are in place.

4.2.3. Recommendations

For Scientific Communities and Research Infrastructures:

- Seek formal representation in horizontal e-Infrastructure advisory structures to ensure scientific requirements shape infrastructure decisions (P01)
- Publish multi-year resource plans aligned with experimental timelines to inform infrastructure planning (P02)
- Adopt federated identity frameworks enabling seamless access across infrastructure boundaries (P03)

For e-Infrastructures and Service Providers:

- Establish liaison mechanisms with thematic research infrastructure governance bodies to ensure scientific requirements shape infrastructure decisions (P01)

- Technically implement multi-year allocation frameworks providing planning certainty for data intensive research facilities beyond annual cycles (PO2)
- Deploy AARC Blueprint Architecture for federated authentication supporting unattended workflows (PO3)

For Policy Makers and Funding Bodies:

- Recognise data-intensive research infrastructures which their long-term needs for digital infrastructure as an important stakeholder category in EuroHPC and EOSC governance (PO1)
- Fund coordination mechanisms bridging thematic and horizontal infrastructure governance (PO1)
- Enable strategic (multi-year) access designations aligning allocation cycles with the planning horizons for data-intensive research facilities (PO2)
- Support trust framework harmonisation enabling federated authentication across European digital Infrastructures and national centres (PO3)

4.3. Pillar 2: Architecture & Interoperability

Integrating compute, data and workflows

4.3.1. Challenge

European research infrastructures operate across heterogeneous computing environments that were not designed to work together. A scientific workflow may require GPU clusters for machine learning training, HPC systems for large-scale simulation, and distributed storage for multi-petabyte datasets. Currently, each resource type often requires separate interfaces, data formats, and operational procedures.

Three interconnected challenges prevent the compute and data continuum from functioning as an integrated technical capability for research:

- **Computing integration fragmentation**: HTC and Cloud communities have established federation within their domains (WLCG federates 170 sites; EGI provides brokering across European e-Infrastructures)The HPC federation is still developing, with the EuroHPC Federation Platform under development. Each HPC centre uses its own portals, schedulers, and software stacks, requiring experiments to custom-adapt to each site. Integration across the different resource types remains limited and standardised integration policies are needed to enable unified orchestration across HPC, HTC, Cloud, and emerging quantum resources (D6.1, Gaps #3, #5).
- **Data management fragmentation**: Exabyte-scale workflows require moving data between long-term archives and transient compute resources, yet transfer mechanisms are fragmented and often rely on ad-hoc tools. No unified architectural approach exists for FAIR storage across domains, including standardised metadata and provenance tracking. Between large Tier-0/1 sites (CERN, the main WLCG and SRCNet hubs) the network backbone is generally adequate, with GÉANT and dedicated overlays (LHCONE, LHCOPN) providing high-capacity transport (D6.1 Section 4.2). Reaching beyond those core sites is the binding constraint: D6.1 Section 5.4 identifies network connectivity as a persistent technical constraint and notes that "at minimum, HPC sites need strong connectivity to WLCG centres and SRCNet nodes", with realistic demand validation required at non-core centres.
- **Workflow orchestration limitations**: Scientific workflows increasingly span multiple facilities, yet orchestration systems remain fragmented across communities. Cross-facility workflows lack resilience and provenance capture. Resource-aware scheduling that considers data locality, compute availability, energy and cost optimisation remains immature.

4.3.2. Priorities Overview

Pillar 2 establishes the technical machinery enabling heterogeneous computing integration, data federation and transport, and workflow orchestration across distributed resources. Three priorities address complementary layers of the technical stack: computing abstraction (P04), data management (P05), and workflow coordination (P06).

P04: Heterogeneous Computing Integration. Provide unified access to heterogeneous computing resources (HPC, HTC, cloud, edge, and emerging quantum) so that researchers can compose workflows across providers without per-stack glue. This requires standardised integration policies adopted across the major European federations, EuroHPC Federation Platform, EGI Federation, WLCG, SRCNet, alongside national pre-exascale and exascale centres. Move beyond reliance on CLI-based batch interfaces toward API-based access mechanisms compatible with workflow engines including the LEXIS Platform (based on Apache Airflow), Common Workflow Language, Dask and HTCondor; introduce shared ontologies (with reference to Numpex) as a politically tractable first-stage requirement. Integration must connect with data infrastructures and data providers, including the European Space Agency (ESA) initiative serving Earth observation and astronomy communities, since heterogeneous compute is rarely useful without proximate data. (Detail: Annex A, P04)

P05: Data Management and Transport. Establish cross-domain standards for federated FAIR data management, automated large-scale data movement with advance-planning and event-driven capabilities, and flexible architectures for sensitive data handling (a transversal need across data-intensive science beyond HEP/RA). For datasets too large to move repeatedly, remote computational access (where supported by the target site) complements bulk transfer rather than substituting for it, since many HPC sites block outbound connections from compute nodes. Convergence on common transfer protocols supporting both POSIX-like and object-storage access patterns, deployment of proven movement services at HPC centres, and operational observability of throughput, latency, error rates, and end-to-end staging time complete the operational picture. Network capacity underlies all of this: transport between core sites is provided by GÉANT, the NRENs and dedicated overlays (e.g., LHCONE, LHCOPN), but assured end-to-end capacity beyond the Tier-0/1 core remains a binding constraint (D6.1 Section 5.4), and a dedicated assessment of exabyte-era network requirements is flagged as a follow-up action (Section 8). (Detail: Annex A, P05)

P06: Workflow Orchestration and Management. Evolve workflow orchestration into a cross-facility capability with built-in resilience, provenance capture, and support for interactive compute use cases, mixed workloads combining embarrassingly parallel tasks with tightly coupled AI processes, and unattended execution of long-running workflows. Build on D6.1's Orchestration and Workflows capability (Workflow Management, Resource Orchestration, Task Scheduling, Provenance) and on the European workflow-management projects catalogued in the Technical Blueprint already advancing this space. Cross-priority dependencies on P01-P03 (governance, allocation, identity) carry the orchestration prerequisites identified in the D6.1 section 4.4 cross-references. (Detail: Annex A, P06)

Interdependencies: The three priorities form a layered structure. Workflow orchestration (P06) depends on heterogeneous computing integration (P04) to execute across diverse resources and on data management (P05) to stage data appropriately. Effective data transport requires understanding of compute and storage resource availability; efficient computing requires data locality awareness. Existing projects demonstrate this integration: interTwin [111] delivers workflow orchestration spanning HPC and HTC resources, whilst Rucio (the WLCG-originated distributed data-management framework, now adopted across multiple experiments) provides data management across the WLCG and SRCNet federation.

4.3.3. Recommendations

For Scientific Communities and Research Infrastructures:

- Down-select to a small set of supported interfaces, data-management tools, and workflow engines so that e-Infrastructures can build against a stable target (P04, P05, P06)
- Ensure the capability to execute on diverse compute platforms via code porting and/or the use of high level programming frameworks (P04)
- Contribute to use cases that drive data management and transport requirements (P05)

- Engage in co-design of workflow orchestration capabilities through community working groups and with service providers (P06)

For e-Infrastructures and Service Providers:

- Implement standardised interfaces for heterogeneous resource description, discovery, and access (P04)
- Deploy interoperable data management services with FAIR-compliant metadata (P05)
- Support cross-facility workflow execution through common orchestration frameworks (P06)

For Policy Makers and Funding Bodies:

- Fund architecture-ecosystem development across European digital infrastructures (EuroHPC, EOSC, EGI) (P01, P04)
- Fund uptake of federated identity and access management for workflows and the compute, data, and network components those workflows interface with (P03, P04)
- Fund development of integration and deployment testbeds enabling validation across HPC/HTC/Cloud resources (P04)
- Support (smart) network capacity investments enabling exabyte-scale data movement (P05)
- Require interoperability standards in infrastructure procurement (P04, P05, P06)

4.4. Pillar 3: Software & Science Enablers

Enabling AI, portability and reproducibility

4.4.1. Challenge

Scientific software represents decades of accumulated knowledge and investment, yet faces growing tensions between longevity requirements and rapid technological change. Major experiments like HL-LHC and SKA will operate for 10-50 years, requiring software that remains functional across multiple generations of hardware and computing paradigms. Simultaneously, AI is transforming scientific workflows, encompassing ML/DL methods used for surrogate modelling and event classification, and increasingly LLM-based and agentic systems that can support code generation, infrastructure operations, and documentation, but most applications remain at proof-of-concept stage rather than production deployment.

Four interconnected challenges constrain the software layer bridging infrastructure and science:

- **AI maturity gap**: Machine learning methods are increasingly central to simulation, reconstruction, and analysis, yet most infrastructures treat them as add-ons rather than integral capabilities. Scaling from demonstration to operational deployment is challenging, often dependent on small teams rather than sustainable capacity. The gap between AI potential and AI reality represents a strategic bottleneck.
- **Portability limitations**: Scientific codes optimised for one architecture often cannot exploit another without substantial rewriting. The computing landscape now spans CPUs, GPUs, FPGAs and emerging accelerators, each with different programming models and each with typical evolution cycles of 2-5 years. Taking advantage of heterogeneous resources requires software that is portable across centres and automatically exploits available hardware capabilities; this can be realized by careful porting to all the architectures of interest (e.g., EESSI), or by using high-level frameworks designed for portability (e.g. Kokkos, SYCL/oneAPI, OpenMP target offload, or ALPACA).
- **Reproducibility requirements**: Scientific insights must be reproducible over a long period requiring preservation of complete computational context including software, workflows, documentation and execution environments. Current practices often fail to capture sufficient information for later reproduction.
- **Long-term preservation**: Instrument lifetimes of 10-50 years require software (together with the complete execution environment) that survives across hardware generations and organisational changes. Code developed today must remain executable and maintainable through multiple technology transitions.

4.4.2. Priorities Overview

Pillar 3 enables AI and accelerated computing methods, software portability across heterogeneous architectures, reproducibility of scientific insights, and code preservation across instrument lifetimes. Four priorities address the software lifecycle from development through long-term preservation.

P07: AI. Establish AI as first-class infrastructure across the European compute and data continuum. AI surrogate models can deliver two to three orders of magnitude computational cost reduction for high-impact HEP and RA use cases [10, 24]. Realising this depends on an integrated AI ecosystem that spans HPC, HTC, cloud, and instrument-edge resources, supports federated learning and FAIR model sharing, and is not concentrated in a single class of facility. The strategic lever is active engagement with European AI policy (RAISE, the AI Office, the AI Act), supported by resource-accounting reform that recognises GPU and AI-Factory allocations as pledged for the major experiments (Detail: Annex A, P07).

P08: Software Portability. Develop scientific codes using portable programming models that execute efficiently across diverse architectures (CPUs, GPUs, FPGAs, and emerging accelerators). Adopt hardware abstraction layers such as Kokkos, SYCL, and Alpaka for architecture-agnostic code; package applications in containers (e.g. Apptainer) or modules (e.g., Lmod); and follow FAIR principles for performance portability and reusable software (e.g., FAIR4RS). Build on European tooling already in production, including EESSI through the EuroHPC Centre-of-Excellence MultiXscale and the EuroHPC Federation Platform's Federated Software Catalogue, and on the EOSC Cloud Container Platform for containerised application deployment. (Detail: Annex A, P08)

P09: Open Science and Reproducibility. Ensure scientific workflows and results remain reproducible over 5-10 year timescales through preservation of (a sufficiently) complete computational context, including standardised metadata, provenance tracking, and FAIR-compliant practices for software and workflows alongside data. Connect with e.g., EVERSE for research software best-practice consolidation and with the DORA/sfdora research assessment principles for career recognition based on FAIR and Open Science contributions. Open data alliances (IVOA, IHDEA, IPDA) provide established community standards that the priority builds on. (Detail: Annex A, P09)

P10: Software Preservation and Sustainability. Ensure scientific codes remain functional across the 20-50 year lifetimes of major instruments (HL-LHC programme to its successors, SKA1 to SKA2, LOFAR through LOFAR 2.0) through systematic preservation and modernisation, with documentation as an integral part of preservation policies and Software Management Plans. HEP's long collaborative legacy from LHC informs emerging RA infrastructures such as SRCNet (still in build-out, so the comparison must be drawn carefully). The priority can be supported by e.g., EVERSE, the SPECTRUM Knowledge Hub, the ESCAPE ESFRI Cluster, and networks such as Radionet, with advocacy from DORA and the ADORE declaration of the Research Software Alliance. Promote open-source and open-standards policies to prevent sovereignty issues and avoid proprietary vendor lock-in. (Detail: Annex A, P10)

Interdependencies: The four priorities form a coherent approach to sustainable scientific software. Portability (P08) enables reproducibility (P09) by decoupling software from specific environments. AI/ML integration (P07) depends on portable software methods (P08) to deploy models across heterogeneous resources. Long-term preservation (P10) requires the documentation and standardisation practices developed for reproducibility (P09).

4.4.3. Recommendations

For Scientific Communities and Research Infrastructures:

- Invest in transitioning AI applications from proof-of-concept to production deployment (P07), including evaluation of LLM-assisted ("vibe-coding") workflows as a possible accelerator for porting scientific codes to new architectures and for sustaining them across hardware generations (P10)
- Engage with the RAISE AI for Science pillar to secure HEP and RA representation and contribute use cases for AI Factory pilot workloads (P07)
- Adopt portable programming models and (micro-)architecture optimized distribution for new software developments (P08)

- Implement FAIR Principles for research softwares (e.g., FAIR4RS) guidelines for software documentation and preservation (P09, P10)
- Engage with the i.a., the EVERSE software-quality framework and the EOSC science-cluster software repositories such as the ESCAPE Open-source Scientific Software and Service Repository (OSSR) (P08, P09, P10)

For e-Infrastructures and Service Providers:

- Deploy AI platforms as integral infrastructure capabilities, not add-ons (P07)
- Provide federated-learning infrastructure for privacy- and sovereignty-constrained training, and operate AI Factory pilots that integrate HEP and RA workloads with EuroHPC AI infrastructure (P07)
- Support common software repositories and distribution systems (P08)
- Align service offerings with the community-driven and maintained software-excellence frameworks and science-cluster software repositories (e.g., EVERSE, ESCAPE OSSR (P08, P09))
- Provide execution environments and support workflow tooling that enable reproducibility and provenance capture (P09)

For Policy Makers and Funding Bodies:

- Fund transition of AI applications from proof-of-concept to production deployment (P07)
- Fund inference infrastructure integration across distributed computing resources (P07)
- Require reproducibility standards in funded research (P09)
- Fund software preservation as an essential investment component of research and digital infrastructures (P10)
- Reform resource-accounting frameworks so that GPU and AI-Factory allocations are accepted as pledged resources within WLCG and SRCNet accounting (P07)
- Engage the AI Office on AI Act application to scientific AI, including model documentation, dataset transparency, and risk classification for safety-relevant operational AI such as detector control and observatory operations (P07)
- Sustain pan-European software-excellence frameworks such as EVERSE, science-cluster software repositories such as the ESCAPE Open-source Scientific Software and Service Repository (OSSR), and optimized software compilation, packaging and distribution such as EESSI as cross-priority enablers for software quality, training, and FAIR-for-software practices (P08, P09, P10)

4.5. Pillar 4: Human Capital & Responsibility

Strengthening communities, improving environmental sustainability and workforce

4.5.1. Challenge

The European compute and data continuum depends on people as much as technology. Technical capabilities require skilled professionals to implement, operate, and evolve them. Environmental sustainability requires not just efficient hardware, human expertise to optimise software and operations but also the management of environmental impacts over the whole life cycle of infrastructures or scientific projects, as well as the means to ensure actors at all levels (end users, data centre operators, policy makers and funders) adopt environmentally responsible behaviours. Cross-infrastructure coordination requires communities that collaborate effectively across organisational boundaries.

Three interconnected challenges constrain human capital and organisational capacity:

- **Workforce skills gap**: The drivers described in Section 3 require expertise that is in short supply. Heterogeneous computing demands specialists in portability and performance engineering. AI/ML adoption depends on machine learning engineers who can bridge proof-of-concept and production. Sustainability goals require transverse hardware-software expertise to minimise impacts through continuous code optimisation, impact-aware scheduling and infrastructure operation. Sustained progress requires a triangular collaboration model: researchers driving concept and validation, code developers responsible for implementation and optimisation, and digital-infrastructure specialists owning operations and performance. This collaboration needs to be sustained by science funders as a multi-year commitment rather than assembled from ad-hoc project additions, with return on investment evaluated where appropriate. Community surveys indicate that most research groups identify technical capacity as a concern, yet lack resources and expertise to address it.

- **Career structure misalignment**: Scientific career structures do not recognise or reward contributions to infrastructure, software development, or operational efficiency. Research software engineers, who bridge domain science and computing expertise, typically lack defined career paths. This creates retention challenges as the same skills command premium salaries in the commercial sector. The Amsterdam Declaration on Funding Research Software Sustainability identifies recognition of software as a first-class research output, but operational uptake has been uneven: most large-scale digital infrastructures provide only limited code-optimisation support, and deep domain-level RSE work is rarely embedded in their core remit. Partial implementations exist via EuroHPC Centres of Excellence, RI Train, and HPC Train, but no European-level funding line yet sustains the triangular researcher–RSE–DI collaboration the priority calls for. The Declaration treats recognition as a research contribution as essential to workforce sustainability.
- **Fragmented community coordination**: Research infrastructures and the federated digital infrastructures serving them are coordinated unevenly: WLCG and the Grid stack have a mature multi-party governance, but the coupling between data-intensive science needs and EuroHPC remains under-articulated, and structured engagement between scientific RIs and the broader European digital-infrastructure providers is still partial. Scientific requirements are often invisible when infrastructure decisions are made. Effective co-design requires sustained collaboration that current funding models do not reliably support.

4.5.2. Priorities Overview

Pillar 4 builds collaborative communities through co-design processes, establishes environmental sustainability through impact-aware management of facilities, and develops career profiles combining domain science with computing expertise. Three priorities address the human and organisational dimensions of the continuum. The pillar combines the two strands that its title names: human capital, P11 and P13, which concern the people and communities that operate the continuum; and responsibility, P12, which is different in kind. P12 is a technical and operational priority addressing the environmental footprint of computing rather than the workforce. It sits here because environmental sustainability, like community and skills, is a continuous responsibility and an enabling condition rather than a one-off deliverable, and its technical detail is developed in the Technical Blueprint (D6.1) and the Environmental Sustainability driver (Section 3.2.3).

P11: Community Collaboration and Co-Design. Establish structured mechanisms for scientific communities to feed requirements, roadmaps and white papers into the cross-infrastructure coordination forum (P01), building on the SPECTRUM CoP and the existing networks (e.g., APPEC, ASTRONET, RadioNet, JENA). These networks contribute strategy and requirements but hold no operational mandate, so commitment and uptake depend on the research facilities and digital infrastructure providers convened in that forum. The European Commission should give the forum formal standing and recognise data-intensive research infrastructures as stakeholders in EuroHPC and EOSC governance. (Detail: Annex A, P11)

P12: Environmental Sustainability. Embed environmental accountability for multiple impacts into research computing operations, using lifecycle assessment methodologies (ISO 14040:2006 and ISO 14044:2006 on Environmental Management, Life Cycle Assessment; the EU Product Environmental Footprint (PEF) method) to track impacts across procurement, fabrication, operation, and end-of-life. For this accounting to drive behaviour rather than displacement, it must be paired with policy on research conduct (procurement criteria, allocation rules), otherwise raised EU-tracked costs risk pushing workloads to non-European providers and net-out the sustainability gain.

Allocate environmental impacts as credits flowing from policy makers to data centres and end users, replacing core hours and abstract metrics with visible impact accounting. The credit metaphor, rather than direct euros, allows top-down political targets (e.g. EU Green Deal limits) to flow through the allocation chain without entangling with market-cost dynamics; the two metrics are complementary, not substitutable, that supports responsible behaviour.

Multi-objective optimisation balancing performance with carbon emissions, energy consumption, and resource utilisation is a candidate output, with best-practices for European data centres (best practices and shared reference patterns rather than rigid blueprints, given the diversity of facility designs) an option for community discussion. The priority addresses growing requirements for environmental responsibility as a condition of public funding and social licence to operate. (Detail: Annex A, P12).

Relaxing the environmental tension requires addressing the whole value chain leading to the generation of science data products in order to reduce the environmental impact: to jointly optimise the research problem (recognising that responsibility for resource-appropriateness lies with the researcher's problem framing as well as with software and hardware) in combination with software and hardware, to make maximal use of minimal resources, to tailor execution to environmental impact in the sense of choosing what resources to use where and at what moment in time, to reduce digital waste through the reuse of data and software, framing failures during execution and resilient execution by adapting to changing operating conditions. Optimising scheduling and hence avoiding idle cycles is equally important. This relies on combining tools, policies and a structured collaborative effort between research and digital infrastructures.

Key to their successful application is a change of paradigm which consists in allocating environmental impacts at all levels, instead of core hours for end users and only funds to data centres. This change of metric is intended, on one hand, to ensure these impacts are visible in order to foster responsible behaviours by the actors and, on the other hand, to establish a framework allowing a high-level allocation made by policy makers in line with European or national objectives to then be used by data centres, typically through hardware procurement and operations, with the latter itself resulting from impact allocation to end users. These allocations serve as credit which each actor is then free to manage (within a framework that is collaborative set in e.g., a Commons approach by science funders, research infrastructures and digital infrastructures) and consumed by arbitrating between their priorities. At the level of the compute and data continuum, this is expected to require coordination between institutional actors for the continuum to not just provide extended resources to users but also a lever to reduce the global and local environmental impacts of their activities.

P13: Skills, Careers and Training. Develop the hybrid expertise combining domain science with computing skills that the continuum requires, with training programmes aligned with emerging drivers (AI/ML integration, heterogeneous computing, sustainability practices), delivered through a mix of integrated curricula and AI-assisted training tools (which can extend reach where curriculum capacity is the bottleneck), and career frameworks following the Research Software Engineer model. The current arrangement carries an avoidable systemic cost: trained experts leave at later career stages, taking with them critical, slow-to-rebuild institutional knowledge, and forcing complete system replacements that would not otherwise be needed. The intent is not to compete with industry on salary but to retain the specific, hard-to-substitute knowledge that the long-lived experiments (HL-LHC, SKA) depend on. A modest number of stable, well-placed career positions would deliver a better return on the substantial investments already committed to scientific computing. Short-term actions may leverage current EU instruments (e.g., RI Train, HPC Train, EuroHPC Centres of Excellence); medium-term actions integrate training into European university curricula (Detail: Annex A, P13).

Interdependencies: Community collaboration (P11) provides the foundation for this pillar and connects across all pillars through co-design processes that shape governance decisions (Pillar 1), technical architecture (Pillar 2), and software priorities (Pillar 3). Within Pillar 4, P11 coordinates sustainability practices (P12) and enables knowledge transfer for skills development (P13). These priorities create the organisational foundation on which technical capabilities depend.

4.5.3. Recommendations

For Scientific Communities and Research Infrastructures:

- Increase and incentivise participation in co-design processes and CoPe Working Groups (P11)
- Integrate and reward environmental impact awareness into software development practices (P12)
- Recognise and value research software engineering contributions within projects and collaborations (P13)

For e-Infrastructures and Service Providers:

- Maintain community engagement mechanisms through dedicated liaison roles and coordination structures (P11)
- Implement a sustainability governance through impact reporting and impact-driven resource allocation (P12), supported where possible by EU-level regulation that institutions can act on (the post-CSRD/Omnibus reporting landscape is still settling, so the regulatory hook for research computing remains to be defined)
- Provide automation (e.g, AI-driven), training and documentation enabling efficient use of heterogeneous resources (P13)

For Policy Makers and Funding Bodies:

- Fund sustained community coordination mechanisms beyond project timelines (P11)
- Define admissible impact budgets for infrastructures and evaluate investments and operation in this frame (e.g., by implementing reporting requirements at the EU-level) (P12)
- Establish career frameworks for research software engineers (across emerging skill areas including AI/ML, performance engineering, and scientific code preservation) and for data stewards (responsible data management, hierarchical-storage management, and long-term curation) (P13)

5. Investment Areas

Realising the compute and data continuum requires coordinated investment across five areas. This section maps the thirteen strategic priorities to those areas and to the pillar framework introduced in Section 4, identifies the European funding instruments available, and highlights investment patterns that traditional hardware-and-floor-space budgeting tends to underestimate (coordination labour, federation-software engineering, RSE effort, and long-horizon data curation).

5.1. Investment Areas, Pillars, and Priorities

Table 5.1.1 maps the five investment areas to pillars, priorities, dominant expenditure types, and the specific investments involved.

Table 5.1.1: Investment areas mapped to pillars, priorities, and expenditure types.

Investment Area	Associated Pillar	Primary Priorities	Dominant Expenditure Type	Key Investments
Computing Infrastructure	Pillar 2 (Architecture and Interoperability)	P04 (primary), P07 (secondary)	Capital + operational	Integration of thematic and national interfaces (especially community-specific interfaces for data-intensive science) with the EU-level federation ecosystem (EuroHPC Federation Platform, EOSC); enabling technologies that bridge thematic edge/continuum resources with EU-level capacity; cross-facility workflow orchestration capacity at execution sites; software distribution infrastructure; large-scale AI training infrastructure and distributed inference resources
Data Infrastructure	Pillar 2 (Architecture and Interoperability)	P05 (primary), P09 (secondary)	Capital + operational	Tiered storage systems with standard access protocols; high-bandwidth transfer networks with advance reservation; FAIR repositories with standardised metadata and provenance tracking. European Data Spaces target a single market for data, but the distributed research-data infrastructure that this agenda depends on is not market-driven; a more active EC coordination role is needed to address this market failure for research data.

Investment Area	Associated Pillar	Primary Priorities	Dominant Expenditure Type	Key Investments
Software and Tools	Pillar 3 (Software and Science Enablers)	P06, P08, P10 (primary), P07, P09 (secondary)	Sustained development + operational	Hardware abstraction layers (e.g. Kokkos, SYCL), modules (e.g., Lmod) and containers (e.g. Apptainer); CI/CD, compilation and software distribution infrastructure; workflow management systems and orchestration tooling; domain repositories and modernisation processes for software preservation
Governance and Coordination	Pillar 1 (Policy, Trust and Governance); Pillar 4 (P11, P12)	P01, P02, P03, P11, P12	Coordination	Liaison mechanisms and advisory representation; multi-year allocation frameworks; federated identity, authentication, and authorisation infrastructure (based on the AARC Blueprint Architecture, full AAI stack including attribute release and token translation, with eduGAIN as the underlying identity federation); sustained co-design processes and European/international coordination; sustainability governance and evaluation
Human Capital	Pillar 4 (Human Capital and Responsibility)	P13	Operational (permanent positions)	Training curricula for emerging skills; career frameworks for research computing and data professionals (including research software engineers, data specialists, and domain scientists dedicated to computing); retention strategies competing with industry; adoption and incorporation of AI (e.g., agentic AI), requiring sustained development effort in addition to training and recruitment

Cross-cutting: P09 (Open Science and Reproducibility) spans computing, data, software, and governance investments. P12 (Environmental Sustainability) applies across all infrastructure and operations. Operational expenditure is present in every priority; the table records its dominance where no other type is primary.

5.2. Investment Profiles Across Implementation Phases

The investment profile changes across the roadmap phases described in Section 6. Aligning funding instruments with actual expenditure patterns requires recognising this shift.

In the **short-term foundation phase** (years 1-3), investment is dominated by coordination costs and design work. Establishing governance forums (P01), bootstrapping trust frameworks (P03), conducting gap analyses (P04, P05, P06), and launching community engagement (P11) require sustained staff effort rather than capital procurement. The primary resource needed is people: liaison officers, policy specialists, AAI operations staff, and community coordinators. Pilot projects and testbeds require modest computing and data resources alongside engineering capacity.

In the **medium-term integration phase** (years 3-5), the profile shifts toward technical deployment. Prototyping federated data management across research infrastructures (P05), deploying workflow orchestration (P06), validating heterogeneous computing integration (P04), and consolidating portable software stacks (P08, P10) require both infrastructure investment and sustained engineering effort. Multi-year allocation frameworks (P02) move from design to operational deployment, requiring structural funding for accounting systems and MoU-based coordination.

In the **long-term maturation phase** (years 5+), the dominant need is sustained operational funding. Production-grade federated AI/ML environments (P07), software preservation repositories (P10), reproducibility services (P09), and sustainability reporting frameworks (P12) must operate reliably over multi-decade scientific programme timescales. Career frameworks for research computing and data professionals (P13) require permanent positions, not project funding. To ensure that operational funding is applied in an optimal and effective manner over time, sustained development and innovation funding are also needed but these do not dominate the overall cost.

This progression means that short-term investment can largely be addressed through project-based funding (e.g., individual Horizon Europe and FP10 calls, national pilot programmes), but medium- and long-term success depends on structural funding commitments that outlast individual project cycles and requires a longer-term programmatic approach (e.g., via EuroHPC, EOSC).

This phasing assumes that the infrastructure the community is asked to engage with remains available over the relevant timescales. EuroHPC in particular is an evolving programme, and the community's medium and long-term investment in adopting EuroHPC and the AI Factories is contingent on EuroHPC continuing to provide sustainable, science-oriented IT resources on the multi-decade horizons of the HL-LHC and SKA. This dependency should be explicit: it requires a Commission commitment that EuroHPC will provide sustained resources for data-intensive science over the evolution timescales relevant to these communities.

5.3. Mapping of priorities across European Funding

Investment in the compute and data continuum draws on multiple European and national funding streams. No single instrument covers all investment areas or phases; effective implementation requires coordinated use of complementary sources. Section 3.1 describes the broader European ecosystem within which these instruments operate; this section focuses on their funding mechanics, committed amounts, and alignment with investment areas.

National funding provides the most computing capacity for European research infrastructure. The resources underpinning WLCG, SRCNet, LOFAR, EGI, and national HPC centres are predominantly nationally funded. Without sustained national commitments, long-term operational continuity is not feasible; EU Framework Programmes typically provide approximately 10% of research infrastructure budgets.

EuroHPC Joint Undertaking has committed approximately EUR 3.08 billion across the 2021-2027 Multiannual Financial Framework (combining EU contributions with matched contributions from participating states and private partners) for digital infrastructure at an EU-level. This covers exascale supercomputing systems, AI Factories, quantum computing platforms, and the EuroHPC Federation Platform (under development). EuroHPC is directly relevant to priorities P04 (heterogeneous computing), P07 (AI/ML), and P03 (federated identity through the Federation Platform). Access for data-intensive science communities remains governed by EuroHPC's allocation mechanisms, which Priority P02 seeks to align with multi-year scientific planning horizons for research infrastructures.

The **AI Continent Action Plan** (2025) and the CADA proposed on 3 June 2026 within the European Technological Sovereignty Package provide the principal funding pathway for P07 (AI), with over EUR 10 billion committed to AI Factories and over EUR 20 billion for AI infrastructure overall. The RAISE programme coordinates AI capabilities for science across the EU; HEP and RA representation in the RAISE AI for Science pillar is the actionable engagement route. IPCEI-CIS provides a co-funding pathway for sovereign federated cloud relevant to scientific AI infrastructure, and the Berlin Declaration (November 2025, signed by 17 Member States) accelerates implementation of IPCEI-CIS and aligned AI infrastructure investments.

European Open Science Cloud (EOSC) covers priorities P05 (data management), P09 (open science), and P03 (federated identity). EOSC operates as a co-programmed European Partnership: EOSC-A Members have pledged approximately EUR 500 million in in-kind contributions to co-finance the partnership over 2021-2030, matching public funding allocated through the EOSC partnership envelope. The EOSC EU Node and Federation are under active development.

Horizon Europe (2021-2027) funds research infrastructure through dedicated call topics including INFRA-DEV (design and development), INFRA-TECH (technology), INFRA-SERV (services), and INFRA-EOSC (EOSC integration). These instruments support short-term design, prototyping, and pilot work. ESFRI infrastructures and ERICs benefit from explicit eligibility criteria in many calls, facilitating consortia formation. Current projects relevant to SPECTRUM priorities include ENSURE (INFRA-TECH, relevant to P02 federated allocation design), interTwin (INFRA-TECH, HPC federation, relevant to P04), ODISSEE [140] (INFRA-TECH, relevant to P04, P05, P06, P07, P08), and EVERSE (INFRA-EOSC, research software quality, relevant to P08, P09, P10).

The Horizon Europe 2026-2027 Work Programme for Research Infrastructures includes structural developments relevant to this agenda. The INFRA-DEV programme now supports thematic cluster consolidation across six ESFRI domains, with "Data, Computing and Digital Research Infrastructures" introduced as a domain. This creates a formal home within the ESFRI cluster framework for the kind of federated computing and data infrastructure SPECTRUM addresses. Alongside INFRA-DEV, the INFRA-TECH programme is the principal vehicle for sustained technical development of distributed research-IT, and the INFRA-EOSC line plus the EuroHPC JU work programmes are the two main EU-level streams whose direction the data-intensive science communities need to influence. The INFRA-SERV programme groups ESFRI domains into integrated access areas. We note the absence of INFRA-SERV projects led by the data and computing domain: the combined area is newly introduced and, in earlier SERV calls, access provision was taken up by physical-science, engineering and environmental RIs rather than digital ones."

A project that clusters the shared digital-infrastructure needs of HEP and RA research infrastructures, federated identity, data movement, workflow orchestration, sustainability accounting, and channels them into the EuroHPC, EOSC, and EGI roadmaps would fit this funding structure directly, without implying any merger or deep integration at the RI level. These programme directions provide a funding pathway for the short-term foundation activities described in Section 6.

Framework Programme 10 (FP10, 2028-2034) is under negotiation. The proposed structure includes four pillars, with research and technology infrastructures positioned in a new Pillar IV (European Research Area) with an estimated allocation of approximately EUR 16 billion, of which research and technology infrastructures may receive approximately EUR 8 billion. The proposed FP10 "20% construction" allocation refers to new critical Research infrastructures, EuroHPC continues to operate at its existing 50% EC / 50% Member State co-funding model. The European Competitiveness Fund will be linked to FP10 to ensure that European industry benefits from research results. The European Parliament and the Council of the EU (in the Warsaw Declaration) have expressed that R&I must be funded by a specific instrument. This leaves the e-Infrastructure federations that pool digital resources for research without a clear successor to the dedicated e-Infrastructures line they held under Horizon 2020. A structural funding route for this federation layer connecting national and research community facilities (i.e., those outside of EuroHPC) for transnational use and that is aligned with other major policy initiatives is needed if the continuum is to be sustained.

ERA Act and Policy Agenda. The forthcoming European Research Area Act will establish a binding regulatory framework for cross-border research cooperation, realising the "fifth freedom" of free movement of researchers, knowledge, and technology. The ERA Policy Agenda (2025-2027) includes active actions that map to SPECTRUM priorities: research infrastructures (25 Member States participating), research careers (27 MS), AI in science (23 MS), enabling open science (24 MS), and research security (26 MS). The ERA action on Research Management as a distinct career path is particularly relevant to Priority P13.

ESFRI provides strategic direction and legitimacy through its Roadmap process. While ESFRI has no direct funding authority, ESFRI status facilitates discussions with national ministries and provides eligibility advantages in EU calls. The ESFRI Report on Funding of Research Infrastructures (2025) identifies persistent challenges including long-term operational sustainability, coordination between funding sources, and human resources. Its recommendations align closely with SPECTRUM priorities P01 (governance), P02 (allocation), and P13 (workforce).

European Regional Development Fund (ERDF) can contribute to research infrastructure construction, with co-funding rates up to 85% in some regions. Operations funding through ERDF is more constrained.

5.4. Underestimated Investment Needs

Three categories of investment are consistently underestimated in research and digital infrastructure budgets. In particular these categories require structural funds and without them, capital investments fail to deliver returns.

Coordination costs. Every pillar of this agenda requires coordination investment: liaison capacity bridging governance structures (P01), policy specialists aligning legal and privacy constraints (P03), community coordinators sustaining co-design processes (P11), and sustainability governance staff (P12). This coordination can be delivered through in-kind contributions from participating organisations or through existing institutional frameworks, but it must be sustained beyond individual project cycles (typically 3-year). The ESFRI Funding Report (ESFRI LTS-WG, 2025) identifies discontinuity in coordination funding as a persistent challenge.

Software maintenance and support. Software makes infrastructure usable. The Amsterdam Declaration on Funding Research Software Sustainability calls for recognition of software maintenance as ongoing infrastructure cost rather than a one-time project deliverable. Priorities P06 (workflow orchestration), P08 (portability), P09 (reproducibility), and P10 (preservation) all depend on sustained software development and maintenance that current project-based funding models do not reliably support. The shift from "software as a project output" to "software as a maintained service" is a funding model change, not a technical one.

Human capital retention. Research infrastructures cannot compete with industry on salary for the technical skills this agenda requires: heterogeneous computing specialists, machine learning engineers, security architects, sustainability analysts. Priority P13 addresses career frameworks and training, but the underlying investment need is for permanent positions with competitive conditions. Training programmes create belonging and community, but retention requires institutional commitment. The ERA Policy Agenda's emerging action on Research Management as a career path signals policy-level recognition; similar efforts are needed for research software engineers and infrastructure technical staff.

Capital expenditure on hardware yields returns only when coordination, software, and people investments are sustained alongside it.

The European Commission's role. Across these investment areas the Commission could: recognise data-intensive research infrastructures as stakeholders in EuroHPC and EOSC governance; give formal standing to a cross-infrastructure coordination forum (P01, P11); enable multi-year, cross-infrastructure allocation (P02); guarantee sustained, science-oriented EuroHPC and AI Factories resources on HL-LHC and SKA timescales; provide a successor funding route for horizontal e-Infrastructure federations as the Horizon 2020 e-Infrastructures category is retired; fund the underestimated categories, coordination, software development, data stewardship, preservation and career frameworks; and align research computing with the sovereignty and open-source agenda (Section 3.2).

6. Multi-Annual Roadmap

This section sequences the thirteen priorities across time horizons, reflecting dependencies between priorities, realistic implementation timescales, and alignment with European funding cycles. The roadmap spans 2026-2036 and beyond, covering the preparation and operational phases of HL-LHC and SKA Observatory.

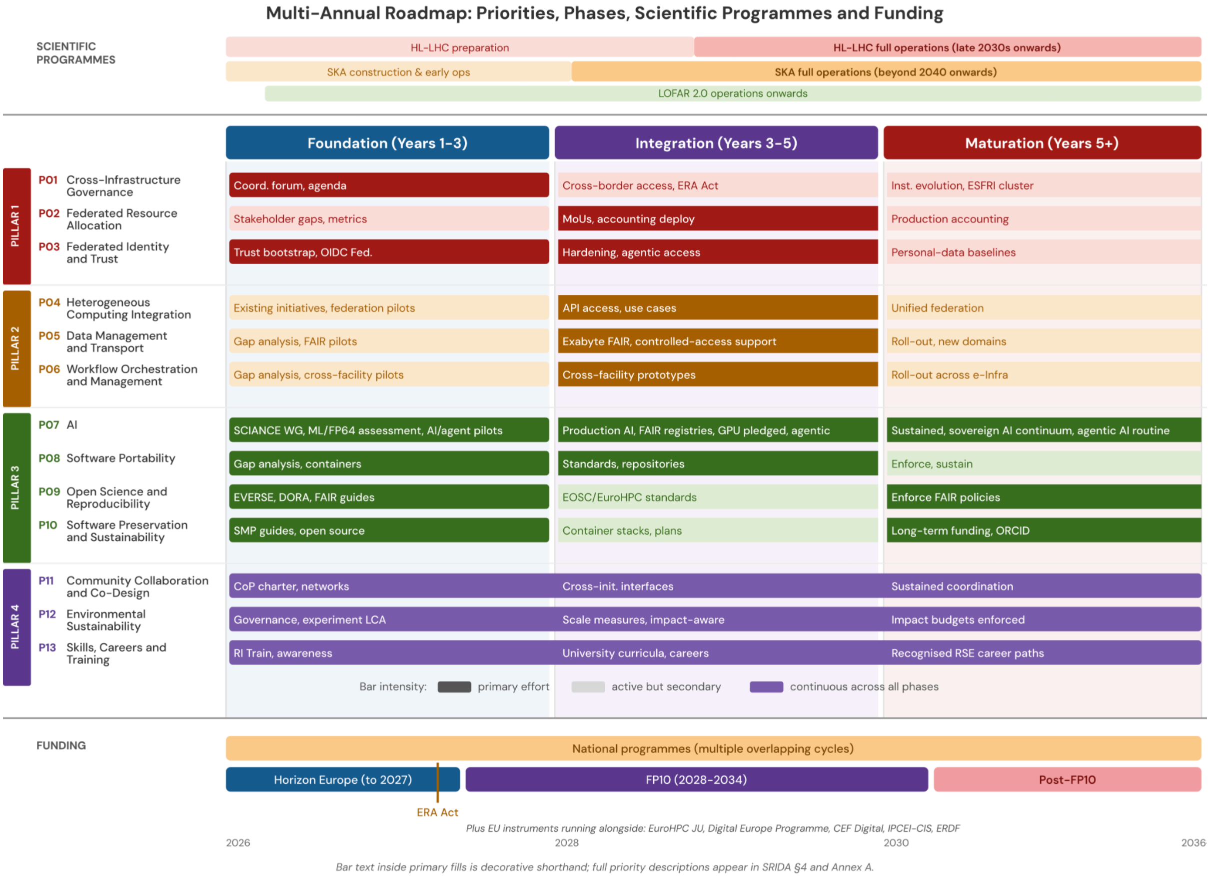
6.1. Phasing Rationale

Priority sequencing follows four principles:

1. **Governance and trust ready in time for integration at scale:** Cross-infrastructure coordination (P01), federated identity (P03), and allocation frameworks (P02) advance in parallel with the technical design and piloting of computing, data and workflow integration (P04-P06), which also begins in the foundation phase. The governance foundations need to be ready before integration proceeds at scale, not fully in place before technical work starts; sequencing the two tracks to converge avoids losing time. These foundations underpin interoperability at the legal, organisational, semantic, and technical layers (European Interoperability Framework [49]).
2. **Infrastructure and application in parallel:** Computing and data capabilities (Pillar 2) and the software that exploits them at scale (Pillar 3) develop in parallel through a phased, agile, validated scale-up; infrastructure capacity needs to be at least at pilot maturity before scaled-out application deployment becomes useful.
3. **Sustained investment:** Human capital (P13), community engagement (P11), and environmental sustainability (P12) require continuous attention across all phases. These are enabling conditions, not sequential steps.
4. **Alignment with European funding cycles:** The transition from Horizon Europe (ending 2027) to FP10 (2028-2034) creates a natural boundary. Short-term actions should use remaining Horizon Europe instruments for design and piloting. Medium-term actions should target FP10 calls and the forthcoming European Competitiveness Fund for deployment and integration. Long-term actions require structural commitments from national programmes and multi-year (e.g., MoU-based) frameworks that outlast individual funding cycles.

6.2. Roadmap Overview

Figure 6.2.1 shows all thirteen priorities across the full timeline, organised by pillar, alongside European funding cycles and scientific programme timescales. Filled cells indicate where each priority's primary effort falls; outlined cells show phases where the priority is active but secondary. Pillar 4 priorities span all phases continuously.



Pillar	Priority	Foundation (Years 1-3)	Integration (Years 3-5)	Maturation (Years 5+)
PILLAR 1	**P01** Cross-Infrastructure Governance	Coord. forum, agenda	Cross-border access, ERA Act	Inst. evolution, ESFRI cluster
	P02 Federated Resource Allocation	Stakeholder gaps, metrics	MoUs, accounting deploy	Production accounting
	P03 Federated Identity and Trust	Trust bootstrap, OIDC Fed.	Hardening, agentic access	Personal-data baselines
PILLAR 2	**P04** Heterogeneous Computing Integration	Existing initiatives, federation pilots	API access, use cases	Unified federation
	P05 Data Management and Transport	Gap analysis, FAIR pilots	Exabyte FAIR, controlled-access support	Roll-out, new domains
	P06 Workflow Orchestration and Management	Gap analysis, cross-facility pilots	Cross-facility prototypes	Roll-out across e-Infra
PILLAR 3	**P07** AI	SCIANCE WG, ML/FP64 assessment, AI/agent pilots	Production AI, FAIR registries, GPU pledged, agentic	Sustained, sovereign AI continuum, agentic AI routine
	P08 Software Portability	Gap analysis, containers	Standards, repositories	Enforce, sustain
	P09 Open Science and Reproducibility	EVERSE, DORA, FAIR guides	EOSC/EuroHPC standards	Enforce FAIR policies
	P10 Software Preservation and Sustainability	SMP guides, open source	Container stacks, plans	Long-term funding, ORCID
PILLAR 4	**P11** Community Collaboration and Co-Design	CoP charter, networks	Cross-init. interfaces	Sustained coordination
	P12 Environmental Sustainability	Governance, experiment LCA	Scale measures, impact-aware	Impact budgets enforced
	P13 Skills, Careers and Training	RI Train, awareness	University curricula, careers	Recognised RSE career paths

Plus EU instruments running alongside: EuroHPC JU, Digital Europe Programme, CEF Digital, IPCEI-CIS, ERDF

Bar text inside primary fills is decorative shorthand; full priority descriptions appear in SRIDA §4 and Annex A.

Figure 6.2.1: Multi-annual roadmap

6.3. Detailed Phase Descriptions

6.3.1 Phase 1: Foundation (Years 1-3)

The foundation phase establishes the institutional and trust prerequisites without which technical integration cannot proceed at scale. Investment in this phase is predominantly coordination, development and people, not capital. The Horizon Europe 2026-2027 Work Programme provides a concrete funding pathway for foundation-phase actions, including thematic-cluster coordination calls and large-scale integrated-access calls covering the new "Data, Computing and Digital Research Infrastructures" domain, with the caveat that integrated-access call cost-eligibility currently excludes capital expenditure, so the pathway is workable for coordination, software and trans-national access but not for hardware investment. Computing and Digital Research Infrastructures" ESFRI domain together with physics-cluster access. Call identifiers are listed in Section 5.3 and the accompanying policy brief.

Governance and coordination (P01). Establish a coordination forum bringing together thematic and multidisciplinary digital infrastructures with policy maker representation. Define its substantive agenda: procurement alignment, innovation coordination for data-intensive science, and bridging the operational-versus-project funding mismatch. Map existing coordination mechanisms and identify gaps the forum addresses. Develop shared metrics for infrastructure utilisation and cross-infrastructure workflow performance.

Federated identity and trust (P03). Bootstrap trust governance across participating infrastructures. Agree operational KPIs for federated authentication. Review site-local policy and risk-acceptance barriers to agentic access (service accounts), which currently prevent automated workflow execution across HPC centres. Adopt the AARC Blueprint Architecture as the architectural baseline and evaluate OIDC Federation as the scalable technical mechanism for OIDC and OAuth deployments.

Resource allocation design (P02). Continue stakeholder conversations to close the two identified gaps: long-term assured access and planning of resource allocations, and e-Infrastructure-wide accounting valid across the whole infrastructure landscape. Establish the interoperability framework for resource accounting: agree metrics and checkpoints, identify tasks and responsibilities, deploy first testbeds and feasibility studies. Identify appropriate Horizon Europe calls for design and prototyping of federated accounting systems.

Computing integration (P04). Build on existing initiatives (e.g., WLCG federation, EGI HPC integration, EuroHPC Federation Platform). Agree on a common descriptive language, shared ontologies mapping across domain-specific vocabularies and HPC/HTC systems, as a first-stage requirement for resource sharing. Begin policy discussions on standardised interfaces across HPC, HTC, cloud, and emerging quantum providers, and on extending coverage to public scientific data archives of broader European interest, since heterogeneous compute is rarely useful without proximate data. Move beyond reliance on a single CLI-type batch interface and towards API-based access mechanisms compatible with workflow engines.

Data and workflow pilots (P05, P06). Conduct gap analyses for selected use cases. Identify the interfaces between research infrastructures, e-Infrastructures, and community APIs that constrain workflow execution. Pilot representative use cases for FAIR-compliant cross-RI data federation, automated data staging from archives into HPC environments, GDPR-sensitive workflows using secure enclaves or logically separated data zones, and cross-facility workflow orchestration. Begin developing operational observability across federated digital services - throughput, latency, error rates, and end-to-end staging time.

AI pilot (P07). Pilot the AI-specific layer for HEP and RA, interactive analytics platforms, AI/ML training and inference, and access to foundation models, on the shared capabilities it depends on (federated storage P05, authentication P03, module-based and containerised software P08, workflow orchestration P06), across a subset of EuroHPC AI Factories, e-infrastructure sites (EGI federated cloud and HTC) and community-specific sites (e.g. WLCG, SCRNet). Stand up operational HEP and RA pilot workloads on selected AI Factories, and begin the resource-accounting reform so that GPU and AI-Factory allocations can count as "pledged" resources for RA/HEP. Engage with the RAISE programme through the SCIANCE Astronomy and Fundamental Physics working group.

Software foundations (P08, P10). Complete a gap analysis of existing portability tools and systems. Establish and document common software stacks for major research infrastructures. Increase adoption of containerisation for software distribution. Develop guidelines, best practices, and policies for research software preservation as an integral part of Software Management Plans, drawing on the Practical Guide to SMPs as a template. Promote open-source and open-standards policies to prevent sovereignty issues and avoid proprietary vendor lock-in.

Open science and reproducibility (P09). Consolidate guidelines and best practices for FAIR reproducible workflows under Open Science, in connection with EVERSE and aligned with the DORA/sfdora research assessment principles. Implement cross-domain collaboration tools (execution environments, harmonised metadata standards, workflow systems) that streamline findability, analysis, reproducibility, and provenance across European digital infrastructures. Establish Open Science-based career-recognition best practices and policies.

Community and workforce (P11, P13). Strengthen the SPECTRUM CoP by developing its charter, ensuring continuity, and aligning with existing networks (APPEC, ASTRONET, RadioNet, JENA, ESCAPE). Develop shared metrics, tools, and methodologies for usage, performance, and efficiency assessment, and create a shared knowledge base. Start delivering actions through current and future EU instruments . Raise awareness across academic and research institutions of the need for stable career paths for research software engineers and computing experts, framed not as competing with industry salaries but as recovering the value lost when expertise leaks at later career stages.

Environmental sustainability (P12). Institute sustainability governance at data-centre and governing-body levels through dedicated committees. Put available solutions into practice to reduce IT environmental impacts and provide some relief to the environmental crises in the short term, with limited disruption. In parallel, plan, develop, and experiment with the solutions required for the change of paradigm: lifecycle assessment tooling (cradle-to-grave), forecasting impacts, and impact-aware scheduling that exposes environmental costs to users at job-submission time alongside core hours.

End of Phase 1	**Governance forum operational; AARC Blueprint Architecture adopted as baseline and OIDC Federation evaluation completed for federated authentication, with first cross-infrastructure authentication pilots running between participating sites; FAIR workflow and software preservation guidelines established.**

6.3.2. Phase 2: Integration (Years 3-5)

With governance and trust frameworks in place, investment shifts towards scaling up the infrastructure deployment and its use, standards consolidation, and operational testing.

Multi-year allocation (P02). Move from testbed to production: pilot e-Infrastructure-wide resource allocations and quotas. Deploy the standardised accounting system progressively across resource providers, infrastructures, and research communities. Establish Memoranda of Understanding for multi-year access aligned with the multi-decade timescales of HL-LHC, SKA Observatory, and other ESFRI scientific programmes.

Heterogeneous computing (P04). Deliver first API-based access implementations across initiatives, addressing the different requirements of HTC batch processing, AI training, inference, and interactive analysis. These APIs are designed to consume the federated identity and authorisation tokens delivered by P03 in this phase (see Phase 2 P03 actions), so that API access and federated AAI converge rather than diverge. Iterate based on lessons learned from foundation-phase pilots. Begin coordinating with public scientific data archives of broader European interest so that heterogeneous compute integrates with proximate data services rather than being delivered in isolation.

Data management (P05). Prototype and demonstrate integrated FAIR data management across multiple research infrastructures and e-Infrastructure providers. Validate architectures for exabyte-scale data volumes. Expand to high-frequency data ingestion (radio astronomy streaming, gravitational-wave low-latency data, WLCG bulk LHC data ingestion to Tier-1s), multi-protocol storage access, and, where applicable for the small share of workflows touching GDPR-sensitive data, secure enclaves or logically separated data zones. Develop standards for FAIR metadata schemas, access-control and identity integration, and operational interfaces for cross-domain data movement. Consolidate the Phase 1 draft AARC Blueprint Architecture (AARC BPA — see also D6.1 §3 for the technical blueprint context) for European data management and transfer services into a community-reviewed version.

Workflow orchestration (P06). Prototype and demonstrate cross-facility orchestration with selected e-Infrastructure providers. Validate fitness-for-purpose across multiple workflow classes (HEP, SRCNet, LOFAR, and others). Explore standardisation of orchestration interfaces and converge on a limited set of interoperable workflow systems with support for unattended execution, provenance tracking, and cross-infrastructure scheduling. Consolidate the Phase 1 draft Blueprint Architecture for workflow systems into a community-reviewed version, scoped to where convergence is feasible across HEP, RA, and adjacent communities.

Software portability and preservation (P08, P10). Ensure adequate tooling is available to support hardware abstractions at all levels. Document commonly-accepted standards for portability across the (SPECTRUM) Compute and Data Continuum. Publish software stacks through common repositories. Consolidate standardised, community-based module stacks and container images across next-generation heterogeneous hardware. Develop plans for software sustainability as an integral part of scientific programme preparation, particularly for major experiments. Integrate research-software training into European university curricula and evaluate AI support.

Federated identity hardening (P03). Harden the trust framework established in Phase 1: onboarding and offboarding procedures, audit expectations across infrastructures, and operational Service Level Agreement (SLAs) for credential issuance and revocation. Enable agentic access (service accounts for unattended workflows) and community-coordinated authorisation, including any compensatory controls or trust-framework arrangements required by participating sites. Federated authentication, established in Phase 1 pilots, moves to widespread production deployment in this phase, so that the API-based access in P04, the data-federation services in P05, the workflow orchestrators in P06, and the AI/ML platforms in P07 all consume the same identity layer. Establish a trust and security coordination framework that includes federated community participants, with continuous readiness evaluation.

Open science (P09). Consolidate aligned interoperability standards between major EU initiatives (EOSC, EuroHPC JU, ESFRI). Consolidate FAIR-by-design services available across borders and across scientific disciplines. Develop preservation plans for data, software, and results as an integral part of scientific research preparation. Promote Open Science as part of educational curricula in university programmes.

AI integration (P07). Move the piloted AI layer into production across a wider set of HPC, HTC and cloud resources, integrating distributed training, inference deployment and foundation-model access with the shared federated layer (P03, P05, P06) so workflows span the EuroHPC AI Factories, the EGI federation and community sites (e.g. WLCG, SRCNet) without re-authentication or manual staging. Secure the mid-term milestone: GPU and AI-Factory allocations accepted as pledged resources within the WLCG and SRCNet accounting frameworks. Stand up shared model registries following FAIR principles, and sustain HEP and RA input to the SCIANCE working group.

Governance consolidation (P01). Consolidate coordination to deliver cross-border access to data and compute services, piloting integrated access across thematic and digital infrastructure clusters. Develop policy recommendations on the operational-versus-project funding mismatch, aligned with the ERA Act framework. Assess options for institutional evolution, including formal RI representation within EuroHPC and EOSC governance and coordination through the emerging ESFRI digital cluster. Establish joint working groups on cross-cutting technical challenges connecting the forum to Pillar 2 implementation.

Community coordination (P11). Foster interaction, coordination, and potential convergence with related domain initiatives. Formalise cross-initiative interfaces: maintained inventory of European and international programmes, infrastructures, and policy processes; named liaison functions; periodic review points.

Environmental sustainability (P12). Consolidate the experimental measures from Phase 1 and put them into wide-scale practice across European data centres and research infrastructures. Scale impact-aware scheduling so it is exposed to users at job-submission time alongside core hours. Mature the lifecycle assessment tooling so that procurement, maintenance, and usage impacts are tracked consistently and can be compared across sites. Begin experimenting with the more complex measures: environmental impact budgets allocated by policy makers, and credit-based allocation flowing to data centres and end users. Consolidate sustainability metrics (Power Usage Effectiveness -PUE-, Carbon Usage Effectiveness -CUE-, embedded carbon, water usage, repairability, HEPscore/watt) to enable their usage as (standard) procurement criteria.

Workforce (P13). Continue awareness campaigns. Work with academic institutions and research organisations to establish dedicated computing career paths through future project proposals and integration into educational university programmes and curricula. Establish shared training catalogues and formal recognition of training across communities. Facilitate contacts between research computing professionals and industry through secondment opportunities. Evaluate the use of AI to support the existing workforce and help train/support the future workforce.

End of Phase 2	**Scientific workflows span HPC, HTC, and cloud through unified interfaces with automated data staging; MoUs are operational for supporting multi-year allocations.**

6.3.3. Phase 3: Maturation (Years 5+)

By year five (approximately 2032), the integrated continuum should reach production-scale operation, just before HL-LHC enters full luminosity from the late 2030s and as SKAO advances through its phased construction toward full operations beyond 2040. Phase 3 is therefore a phase of maturation, consolidation, and adaptation: hardware, software, and AI capabilities will continue to evolve substantially over the next decade, and the priorities below define the directions to maintain rather than the endpoints reached.

AI at scale (P07). By the maturation phase, AI is available across a wider AI-ready ecosystem that couples compute and data, not only the EuroHPC AI Factories but also EuroHPC supercomputers, national HPC, the e-infrastructure federated cloud and HTC (EGI), and the community and thematic providers (WLCG, SRCNet), integrated with the federated data layer (P05) so that AI workloads reach both accelerator capacity and the datasets they consume, wherever data locality, cost and availability dictate. Operate AI as first-class, production-grade infrastructure on this base: sustained federated AI and AI-driven services, a FAIR AI lifecycle with model registries and long-term preservation, and AI agents that plan and execute workflow steps under human oversight (analysis triage, anomaly investigation, calibration response), drawing on P03 service-account access, P06 orchestration and P09 provenance. Embed AI as a strategic capability supporting Europe's exascale science leadership, with ongoing HEP and RA representation in EU AI policy through RAISE and its SCIANCE working group.

Reproducibility and open science (P09). Enforce FAIR standard policies for publicly funded research. Provide long-term funding for data and software preservation. Establish Open Science-based career recognition as standard practice across European research institutions, supported by DORA-aligned research assessment frameworks.

Software preservation (P10). Ensure long-term funding for software preservation across scientific disciplines. Define career paths for research software engineers, with contributions made visible and assessable through persistent identifiers and citation practices (ORCID for people, DOIs for software releases via Zenodo or similar, CodeMeta and CITATION.cff for software metadata) and through DORA-aligned research assessment frameworks adopted by funders and institutions. Software stacks remain executable across hardware generations through sustained maintenance, with documentation embedded as a first-class element of preservation policies and Software Management Plans. Invoke AI support for the software process from design to operations, preservation and re-use.

Federated allocation in production (P02). Federated, flexible resource allocation processes are widely adopted. Standardised resource accounting is routinely used in production across European e-Infrastructures, and funding bodies receive coordinated evidence on the scientific return from their investments.

Computing federation (P04). Unified federation across HPC, HTC, cloud, and emerging quantum platforms, based on lessons learned from earlier phases. Standardised interfaces and shared ontologies enable scientific workflows to exploit the full compute continuum with minimal per-site adaptation, confined to performance tuning rather than re-engineering.

Data management and transfer (P05). Roll out federated data management and automated transport across European e-Infrastructures. Onboard additional scientific communities beyond HEP and RA, including new scientific domains. Ensure sustainability through collective governance, coordinated investment, and community-driven standards (e.g., using a digital commons approach).

Workflow orchestration (P06). Roll out cross-facility workflow orchestration across European e-Infrastructures, expanding the community base and supporting interactive use cases at scale.

Software portability (P08). Require that all software intended for reuse beyond its original authoring group, that is, libraries, frameworks, workflow components, and tools that support scientific publications, is published through established standards and repositories. Ensure sustained long-term support for enabling technologies and portability frameworks (e.g., Kokkos, SYCL, Alpaka, EESSI, CernVM-FS, Apptainer).

Governance maturation (P01). Assess the impact of governance coordination. Determine whether informal coordination or a formalised institutional structure provides the right balance for multi-decade planning, drawing on options including formal RI representation within EuroHPC and EOSC governance and coordination through the ESFRI digital cluster. Coordinate European representation in international collaborations spanning multiple continents, so that the European compute and data continuum is presented coherently to partners in WLCG, SRCNet, and adjacent global initiatives, rather than as a federation of separate national positions.

Identity operations (P03). Personal-data baselines (REFEDS DPCoCo and equivalents) are operational platform capabilities rather than project-by-project integrations, with auditing and policy enforcement built in. Federated identity operates as transparent infrastructure across HPC, HTC, and cloud federations.

Community and workforce (P11, P13). Maintain continuous coordination and strategic alignment across the broader network of scientific communities. Permanent computing career paths established across European research institutions through reshaped funding and personnel policies. Long-term plans guarantee sustained support for research software engineering, infrastructure operations, and community coordination roles.

Environmental sustainability (P12). All proposed measures (or adaptations) are consolidated and put into practice. Integrated sustainability policies, accounting, and management operate from project level to infrastructure level, with coordinated supervision by European HEP and RA institutions and EuroHPC/EOSC. Environmental impact budgets are defined by policy makers and enforced. Continuous monitoring of measure effectiveness adapts to the evolution of hardware, software, and the environmental situation.

End of Phase 3	**Compute and data continuum operates as integrated production capability for HL-LHC and SKA; career paths for research computing and data professionals established; environmental impact reporting operational.**

6.4. Cross-Priority Dependencies

The priorities are interdependent. Table 6.4.1 shows the primary dependencies that constrain sequencing. A dependency means that progress on the target priority is limited without progress on the source priority.

Table 6.4.1: Key cross-priority dependencies.

Target Priority	Depends On	Nature of Dependency
P02 (Allocation)	P01 (Governance), P03 (Identity), P09 (Open Science)	Multi-year allocation requires governance agreements, authenticated access, and FAIR data practices
P04 (Compute)	P01 (Governance), P02 (Allocation), P03 (Identity)	Heterogeneous computing integration requires governance agreements, allocation frameworks, and identity foundations
P05 (Data)	P01 (Governance), P02 (Allocation), P03 (Identity), P04 (Compute)	Federated data management requires governance agreements, allocation frameworks, authentication, and compute integration
P06 (Workflows)	P01 (Governance), P02 (Allocation), P03 (Identity), P04 (Compute), P05 (Data), P08 (Portability)	Cross-facility workflows depend on governance, allocation, identity, compute, data, and portable software
P07 (AI/ML)	P04 (Compute), P05 (Data), P06 (Workflows)	Production AI/ML depends on integrated compute, data, and orchestration
P08 (Portability)	P04 (Compute), P09 (Open Science), P10 (Preservation), P11 (Community), P13 (Skills)	Portability standards require computing environments, open science practices, preservation, community co-design, and trained workforce
P09 (Reproducibility)	P04 (Compute), P05 (Data), P08 (Portability), P10 (Preservation), P13 (Skills)	Reproducibility requires compute access, FAIR data, portable software, preserved artefacts, and trained data stewards
P10 (Preservation)	P04 (Compute), P08 (Portability), P09 (Reproducibility), P13 (Skills)	Preservation requires compute infrastructure, portability standards, FAIR practices, and trained staff

Priorities P08, P09, and P10 are co-dependent rather than sequential; they must evolve together, with progress in each reinforcing the others.

Pillar 4 priorities (P11, P12, P13) are enabling conditions that span all phases and support progress across all other priorities. P11 (Community Collaboration) provides the co-design mechanisms through which governance decisions (Pillar 1), technical architecture (Pillar 2), and software priorities (Pillar 3) are shaped by scientific community requirements.

6.5. Alignment with European Timescales

The roadmap aligns with three external timescales:

Scientific programmes. HL-LHC operations begin in the period 2028-2032 (Run 4 onwards) and ramp through the late 2030s to data volumes requiring the full compute and data continuum. SKA Observatory operations continue beyond 2040. The foundation and integration phases must deliver operational capabilities before these timescales, not after. Delivery in FP10 is therefore critical.

European funding cycles. Horizon Europe (to 2027) supports design and piloting in the foundation phase. FP10 (2028-2034), with research infrastructures in Pillar IV, aligns with the integration phase. The 2028-2034 MFF European Competitiveness Fund may provide a pathway from research to infrastructure deployment. Long-term maturation requires structural commitments that outlast individual framework programmes.

Policy evolution. The forthcoming ERA Act will establish a binding framework for cross-border research cooperation, and the ERA Policy Agenda (2025-2027) creates short-term opportunities for career development and open science. EOSC Federation development, EuroHPC Federation Platform deployment, and ESFRI Roadmap cycles provide institutional milestones that the roadmap should use as coordination points.

The roadmap will adapt to funding outcomes. The dependencies in Section 6.4 are structural; the specific timelines within each phase will be adjusted based on progress.

7. Broader Impact

For data-intensive science, a working compute and data continuum changes what large-scale European research does day to day and promises an enormous efficiency gain. A researcher accesses reconstructed data from federated sites across several member states, runs each computational step on the resource best suited to it (e.g. HTC for batch processing, HPC for large-scale simulation, cloud for interactive analysis), and publishes the resulting datasets through shared open archives, all through one identity, one allocation, and one portable software stack. Simultaneously AI will deliver new ways to effectively use and interact with this continuum in a seamless manner.

The investments proposed in this agenda serve this scientific shift first: enabling HL-LHC, SKA, LOFAR, and adjacent programmes to extract their scientific return, and producing methods, software, and infrastructure designs that other data-intensive sciences with similar characteristics identified in Section 1.1. adopt. They also generate impacts on workforce development, the economy, society, and European policy that strengthen the case for sustained public investment.

This section outlines the principal impact pathways, drawing on the RI-PATHS framework (Griniece et al., H2020) for socio-economic impact assessment of research infrastructures. RI-PATHS traces each pathway from Resources and Activity through Outputs and Outcomes to Impacts, and groups thirteen impact pathways under four areas, addressed below in turn: Human Resources (P5, P6), Economy and Innovation (P2, P4, P7), Society (P8, P10, P13), and Policy (P12). Scientific and technological impact is addressed throughout Sections 3 to 6; this section concentrates on the socio-economic pathways that complement them.

7.1. Impact on Human Resources

The continuum invests in workforce capacity through P01, P11 and P13 (resources and activity), producing research software engineers, federated-operations engineers, AI/ML specialists, sustainability analysts, and governance and coordination professionals who run cross-infrastructure forums, operate multi-year allocation frameworks, and bridge thematic and horizontal infrastructure governance (outputs). These professionals enable European scientific computing to operate at HL-LHC and SKA scale and sustain the multi-country governance that makes federation work (outcome); the skills they develop transfer to industry, finance, public administration, and adjacent sciences (impact).

Workforce impact is a societal and economic impact in its own right, and is also a precondition for the impacts described in Sections 7.2 and 7.3. Implementing the compute and data continuum requires professionals who combine domain science with computing expertise: research software engineers, specialists in heterogeneous architectures and AI/ML deployment, federated operations engineers, sustainability analysts, and governance and coordination professionals who staff cross-infrastructure forums, operate multi-year allocation frameworks, and bridge thematic and horizontal infrastructure governance. These roles do not yet have well-established career paths in most European research institutions. Priority 1 generates the demand for governance and coordination capacity; P13 develops career frameworks for technical and coordination workforce together; P11 connects these professionals to scientific communities through co-design. The workforce implications extend across the entire agenda. Simultaneously, AI is incorporated in a human-centric manner, compliant with EU regulations and values, to support this workforce and bolster its impact.

The skills developed are transferable. HEP and RA communities have historically trained professionals who move into industry, finance, and technology, carrying expertise in large-scale data processing, statistical analysis, and distributed system management. The composition of this workforce, by geographic origin, by gender, and by career stage, is a strategic indicator in its own right. **Section 7.3** returns to this point with explicit framing, and Section 4.5 (Pillar 4) places reporting commitments under P11 and P13.

Embedding sustainability practices into infrastructure operations (P12) develops a further set of competences, in energy-aware computing, lifecycle assessment, and impact reporting, that are increasingly sought across sectors.

This section addresses workforce capacity and the priorities that build it. Section 7.3 takes up workforce composition (geographic, gender, recognition) within the broader societal-distribution argument.

7.2. Impact on Economy and Innovation

Research infrastructures, digital infrastructures, national computing centres, and institutional providers procure hardware, software, and services across these layers of the continuum (resources and activity), generating European technology contracts in processors, interconnects, system integration, storage, and scientific software (outputs). These contracts strengthen industrial capacity in high-performance computing, data services, and scientific software supply (outcome) and feed into European digital competitiveness (impact).

Research infrastructure procurement creates economic multiplier effects. EuroHPC has committed EUR 3.08 billion to supercomputing, stimulating European technology development in processors (RISC-V through the DARE programme at EUR 240 million), interconnects, and system integration. The compute and data continuum extends this value chain to storage, networking, and software services.

Technology transfer flows both ways. Scientific computing has produced widely adopted tools, from ROOT and Geant4 to containerised workflows and federated identity protocols; in adjacent science domains, technology transfer from radio astronomy contributed key patents underpinning Wi-Fi (CSIRO indoor multipath). Open-source software, a defining feature of HEP and RA computing, generates economic value through reduced duplication and shared development costs.

The sovereignty argument (Section 3.2.7) has direct economic implications. Investing in European open technologies retains procurement value within Europe and builds industrial capacity for digital competitiveness. This connection is already operational at the procurement layer: in April 2026 the Commission used its Cloud Sovereignty Framework, which grades providers across eight sovereignty criteria on SEAL (Sovereignty Effectiveness Assurance Levels) from SEAL-0 to SEAL-4, to award up to EUR 180 million in sovereign cloud contracts over six years to four European providers (Post Telecom with OVHCloud and CleverCloud; STACKIT; Scaleway; Proximus with S3NS, Clarence, and Mistral). The European Competitiveness Fund in the 2028-2034 MFF positions the funding instrument that supports this trajectory at programme level, and the CADA proposed on 3 June 2026 as part of the Commission's European Technological Sovereignty Package, operationalises it for cloud and AI infrastructure capacity in line with the AI Continent Action Plan cited in Section 3.2.2.

The SRIDA's priorities contribute to this trajectory, particularly P04 (heterogeneous computing), P07 (AI/ML), and P08 (software portability), which build the scientific-computing layer of the European cloud and AI base that CSF assesses and CADA is intended to strengthen.

7.3. Impact on Society

The continuum operates as a federated multi-country research infrastructure (resources and activity), producing infrastructure, software, and analytical methods for exabyte-scale processing, federated data management, and AI-assisted analysis (outputs). These artefacts apply wherever large-scale data serves public interest, including climate modelling, public health surveillance, environmental monitoring, and disaster response (outcome), reaching these adjacent domains through the same federated channels that already serve the flagship communities (impact).

Open science practices (Priority 9) keep publicly funded outputs (data, software, methods) available for reuse and for public scrutiny. Accountability for public investment in research depends on this availability.

Environmental sustainability (Priority 12) shapes how research computing operates as a public actor. Energy-aware scheduling, lifecycle impact assessment, and sustainability reporting within research infrastructure can set standards that the broader data-centre industry adopts. Operating within defined impact budgets is the practice that makes accountability concrete.

Structurally, the continuum is multi-country by design at both of its layers. Thematic organisations (HEP through HL-LHC and the WLCG; Radio Astronomy through the SKA Observatory and LOFAR) have operated as multinational federations from their inception. Horizontal digital infrastructures (EuroHPC, EOSC, the EGI Federation, and GÉANT) are designed as European endeavours that pool investment, governance, and capacity across member states. The agenda's recommendations on distribution, recognition, and inclusion follow from this dual structure, with P11 and P13 setting out the reporting expectations.

Geographic balance shapes the investment pattern the continuum assumes. On the thematic side, WLCG distributes computing across more than 40 partner countries through national Tier-1 and Tier-2 sites; SKAO and LOFAR-ERIC each operate as IGO- or ERIC-style federations distributing regional centres and station infrastructure across their member states. On the horizontal side, EuroHPC sites are hosted by consortia spanning multiple member states; the EOSC partnership pools in-kind contributions from organisations across the EU; the EGI Federation coordinates compute and data services through national providers; and GÉANT delivers connectivity through the network of National Research and Education Networks. The continuum extends this distributed pattern, including by recognising research software engineering as a career destination accessible from any participating country.

Gender balance in research computing roles remains a challenge across the field. The thematic organisations (CERN, SKAO, LOFAR-ERIC) and the horizontal e-Infrastructures (EuroHPC, EOSC, EGI, GÉANT) each publish or aggregate staff-composition data, and EU-level R&I monitoring tracks gender balance across the research workforce, but no consolidated figure currently exists for the share of women in research software engineering and operations roles across the continuum. P11 calls for consolidated annual reporting on geographic and gender distribution of community engagement, drawing on these existing sources; P13 calls for the same on research software engineering positions established under this agenda. Closing the visibility gap is the precondition for closing the participation gap.

7.4. Impact on Policy

The continuum's governance, allocation, and trust mechanisms operate across institutional and national boundaries (resources and activity), producing concrete models for cross-infrastructure governance, multi-year resource allocation, and federated identity (outputs). These models inform EuroHPC, EOSC, and national infrastructure policy (outcome) and provide templates for cross-border resource sharing beyond HEP and RA, including science diplomacy and European positioning in international standards (impact).

This agenda provides evidence-based inputs to European policy. The priorities and roadmap inform investment decisions within EuroHPC, EOSC, and national programmes with the specificity that resource allocation requires.

The governance models in Pillar 1 (multi-year allocation mechanisms, federated identity frameworks, cross-infrastructure coordination) address structural barriers in European research policy. Successfully implemented, they serve as templates for cross-border resource sharing beyond HEP and RA, and align with the European policy environment that shapes research computing: the AI Continent Action Plan (Section 3.2.2), the CADA, the Commission's Cloud Sovereignty Framework already operational in EU procurement, and the forthcoming ERA Act.

Science diplomacy is a further dimension. The HL-LHC and SKA programmes span over 40 countries. European leadership in the computing infrastructure that supports these programmes strengthens Europe's position in international scientific cooperation and its capacity to shape global standards for data management, software sustainability, and federated governance.

8. Conclusion

The SRIDA presents thirteen priorities, five investment areas, and a phased multi-annual roadmap for realising the European compute and data continuum. The priorities address gaps identified through systematic analysis of scientific use cases, infrastructure capabilities, and community requirements, aligned with EOSC, EuroHPC, ESFRI, and domain-specific roadmaps.

The Four Pillars framework (Section 4) organises these priorities across governance, architecture, software, and human responsibility. The investment framework (Section 5) maps priorities to funding sources and investment types and names the categories, coordination, sustained software development, and long-term preservation, that traditional infrastructure budgets routinely underestimate. The multi-annual roadmap (Section 6) sequences actions across foundation, integration, and maturation phases. Progress requires advancement across all four pillars together: governance foundations enable technical integration, software development unlocks infrastructure value, workforce capacity sustains all other investments, and managing environmental impact makes them durable.

Realising this requires coordinated action. Scientific communities and research infrastructures as thematic verticals need to publish multi-year resource plans, seek representation in horizontal digital infrastructure governance, and invest in transitioning AI to production. Digital infrastructures and service providers need to establish liaison mechanisms with thematic research infrastructures, deploy federated identity and standardised interfaces, maintain community engagement, and support environmentally sustainable development and operation. Policy makers and funding bodies need to recognise data-intensive research infrastructures as stakeholders in EuroHPC and EOSC, enable multi-year allocation mechanisms, set and steer environmental sustainability commitments, and fund software preservation, data stewardship, and career frameworks alongside hardware.

Policy makers and funding bodies need to recognise data-intensive research infrastructures as stakeholders in EuroHPC and EOSC, give standing to a cross-infrastructure coordination forum, enable multi-year allocation, sustain funding for the e-Infrastructure federations and for EuroHPC's science use over the relevant timescales, and fund software preservation, data stewardship, and career frameworks alongside hardware.

The choices made in the next funding cycle will determine whether Europe converts the instruments it has already committed to into the science they were built to produce. The digital continuum at the EU-level is a critical and necessary component to deliver Europe’s strategic ambitions and return of investment. The thirteen priorities, the five investment areas, and the three-phase roadmap formulated in this document are the instrument for that conversion. The next step is for all of us to act upon them.

9. References

This section lists the cited sources, research papers, policy documents, and technical materials used to develop the SRIDA. The bibliography also includes documents collected through the SPECTRUM CoP that informed the agenda's evidence base, even where individual priorities do not cite them inline. The complete document repository, with summaries and provenance, is maintained at the SPECTRUMCoP Document Repository: https://confluence.egi.eu/display/SPECTRUMCoP/Document+Repository.

[1] AARC BPA 2025. AARC-G080: AARC Blueprint Architecture 2025. Available: https://aarc-community.org/guidelines/aarc-g080/
[2] AARC Policy Development Kit. AARC Policy Development Kit (AARC-PDK). Available: https://aarc-community.org/policy/policy-development-kit
[3] AARC Tree Project. Available: https://aarc-community.org/aarc-tree-project/
[4] Aartsen et al. 2014 (IceProd). Aartsen, M. G. et al. (2014). The IceProd Framework: Distributed Data Processing for the IceCube Neutrino Observatory. Available: https://confluence.egi.eu/display/SPECTRUMCoP/The+IceProd+Framework%3A+Distributed+Data+Processing+for+the+IceCube+Neutrino+Observatory
[5] Abdurachmanov et al. 2014 (ARM/Xeon). Abdurachmanov, D. et al. (2014). Explorations of the viability of ARM and Xeon Phi for physics processing. Available: https://confluence.egi.eu/display/SPECTRUMCoP/Explorations+of+the+viability+of+ARM+and+Xeon+Phi+for+physics+processing
[6] Acosta et al. 2023 (CMS-BSC). Acosta, C. et al. (2023). Integration of the Barcelona Supercomputing Center for CMS computing: Towards large-scale production [Presentation]. Available: https://confluence.egi.eu/display/SPECTRUMCoP/Integration+of+the+Barcelona+Supercomputing+Center+for+CMS+computing%3A+Towards+large+scale+production
[7] Adelman-McCarthy et al. 2023. Adelman-McCarthy, J. et al. (2023). Extending the distributed computing infrastructure of the CMS experiment with HPC resources [Conference Paper]. Available: https://confluence.egi.eu/display/SPECTRUMCoP/Extending+the+distributed+computing+infrastructure+of+the+CMS+experiment+with+HPC+resources
[8] ADORE Software Declaration. Research Software Alliance, Amsterdam Declaration on Funding Research Software Sustainability (1.1), 2024. Zenodo. Available: https://doi.org/10.5281/zenodo.13735888
[9] AI Continent Action Plan. European Commission, "The AI Continent Action Plan", COM(2025) 165. Available: https://digital-strategy.ec.europa.eu/en/library/ai-continent-action-plan
[10] AI4ES (US DOE). U.S. Department of Energy, "AI for Energy and Science Report", 2023.
[11] ALICE O2 evolution 2019. ALICE Collaboration (2019). Evolution of the O2 system [Technical Report]. Available: https://confluence.egi.eu/display/SPECTRUMCoP/The+ALICE+Collaboration%3A+Evolution+of+the+O2+system
[12] Alice Recoque. Available: https://www.eurohpc-ju.europa.eu/contract-signed-alice-recoque-europes-new-exascale-supercomputer-2025-11-18_en
[13] Andreozzi e-IRG 2025. S. Andreozzi, "AI Needs in Compute/Data Intensive Research from HEP and RA — The SPECTRUM Project Perspective", e-IRG Workshop, June 2025.
[14] Arrhenius. Available: https://www.naiss.se/news/2024/11/request-for-proposals-for-new-supercomputer-arrhenius
[15] Atif et al. 2023. Atif, M. et al. (2023). Evaluating Portable Parallelization Strategies for Heterogeneous Architectures in High Energy Physics. Available: https://confluence.egi.eu/display/SPECTRUMCoP/Evaluating+Portable+Parallelization+Strategies+for+Heterogeneous+Architectures+in+High+Energy+Physics
[16] ATLAS 2024 (Google Cloud TCO). ATLAS Collaboration (2024). Total cost of ownership and evaluation of Google cloud resources for the ATLAS experiment at the LHC. Available: https://confluence.egi.eu/display/SPECTRUMCoP/Total+cost+of+ownership+and+evaluation+of+Google+cloud+resources+for+the+ATLAS+experiment+at+the+LHC
[17] ATLAS HL-LHC Roadmap. ATLAS Collaboration (2022). ATLAS Software and Computing HL-LHC Roadmap. Available: https://confluence.egi.eu/display/SPECTRUMCoP/ATLAS+Software+and+Computing+HL-LHC+Roadmap
[18] Apptainer. https://apptainer.org/

[19] Bagnasco et al. 2023 (ET Computing). Bagnasco, S. et al. (2023). Computing Challenges for the Einstein Telescope Project [Conference Paper]. Available: https://confluence.egi.eu/display/SPECTRUMCoP/Computing+Challenges+for+the+Einstein+Telescope+Project

[20] Bagnasco 2024 (LVK Computing). Bagnasco, S. for the Virgo Collaboration and the LIGO Scientific Collaboration (2024). The LIGO-Virgo-KAGRA Computing Infrastructure for Gravitational-wave Research. Available: https://confluence.egi.eu/display/SPECTRUMCoP/The+Ligo-Virgo-KAGRA+Computing+Infrastructure+for+Gravitational-wave+Research

[21] Banerjee et al. 2023 (HECAP+). Banerjee, S. et al. (2023). Environmental sustainability in basic research: a perspective from HECAP+ [Technical Report]. Available: https://confluence.egi.eu/pages/viewpage.action?pageId=288784649

[22] BDVA DataLabs. BDVA DataLabs Initiative. Available: https://bdva.eu/task-forces/bdva-data-labs-initiative/

[23] Beche et al. 2018. Beche, A. et al. (2018). Supercomputers, Clouds and Grids powered by BigPanDA for Brain studies [Conference Paper]. Available: https://confluence.egi.eu/display/SPECTRUMCoP/Supercomputers%2C+Clouds+and+Grids+powered+by+BigPanDA+for+Brain+studies

[24] Bertone et al. 2024 (NP Tools). Bertone, V. et al. (2024). Nuclear Physics Tools — Machine Learning, Artificial Intelligence, and Quantum Computing [Report]. Available: https://confluence.egi.eu/display/SPECTRUMCoP/Nuclear+Physics+Tools+%E2%80%93+Machine+Learning%2C+Artificial+Intelligence%2C+and+Quantum+Computing

[25] Beyer et al. 2020 (DESY Photon). Beyer, C. et al. (2020). Beyond HEP: Photon and accelerator science computing infrastructure at DESY [Conference Paper]. Available: https://confluence.egi.eu/display/SPECTRUMCoP/Beyond+HEP%3A+Photon+and+accelerator+science+computing+infrastructure+at+DESY

[26] Boccali 2019 (Computing Models HEP). Boccali, T. (2019). Computing models in high energy physics. Available: https://confluence.egi.eu/display/SPECTRUMCoP/Computing+models+in+high+energy+physics

[27] Boccali et al. 2019 (Cross-experiment HPC). Boccali, T. et al. (2019). Summary of the cross-experiment HPC workshop [Workshop Summary]. Available: https://confluence.egi.eu/display/SPECTRUMCoP/Summary+of+the+cross-experiment+HPC+workshop

[28] Boccali et al. 2021 (INFN-CINECA). Boccali, T. et al. (2021). Enabling HPC Systems for HEP: The INFN-CINECA Experience. Available: https://confluence.egi.eu/display/SPECTRUMCoP/Enabling+HPC+Systems+for+HEP%3A+The+INFN-CINECA+Experience

[29] Böhler et al. 2021 (Opportunistic WLCG). Böhler, M. et al. (2021). Transparent Integration of Opportunistic Resources into the WLCG Compute Infrastructure [Conference Paper]. Available: https://confluence.egi.eu/display/SPECTRUMCoP/Transparent+Integration+of+Opportunistic+Resources+into+the+WLCG+Compute+Infrastructure

[30] Boveia et al. 2024. Boveia, A. et al. (2024). Snowmass 2021 Cross Frontier Report: Dark Matter Complementarity (Extended Version). Available: https://confluence.egi.eu/pages/viewpage.action?pageId=288785184

[31] Breen et al. 2021 (SKAO SDP). Breen, S. et al. (2021). SKAO Science Data Products: A Summary [Data Product Summary]. Available: https://confluence.egi.eu/display/SPECTRUMCoP/SKAO+Science+Data+Products%3A+A+Summary

[32] Britton 2024 (Carbon Cost). Britton, D. (2024). Simulating the Carbon Cost of Grid Sites [Presentation]. Available: https://confluence.egi.eu/display/SPECTRUMCoP/Simulating+the+Carbon+Cost+of+Grid+Sites

[33] Britton, Campana, Panzer-Stradel 2024 (WLCG Energy). Britton, D., Campana, S. & Panzer-Stradel, B. (2024). A holistic study of the WLCG energy needs for the LHC scientific program [Conference Paper]. Available: https://confluence.egi.eu/display/SPECTRUMCoP/A+holistic+study+of+the+WLCG+energy+needs+for+the+LHC+scientific+program

[34] Bruers et al. 2023. Bruers, B. et al. (2023). Resource-aware Research on Universe and Matter: Call-to-Action in Digital Transformation [Meeting Summary]. Available: https://confluence.egi.eu/display/SPECTRUMCoP/Resource-aware+Research+on+Universe+and+Matter%3A+Call-to-Action+in+Digital+Transformation

[35] Casper et al. 2022 (Small HEP). Casper, D. et al. (2022). Software and Computing for Small HEP Experiments [Workshop Summary]. Available: https://confluence.egi.eu/display/SPECTRUMCoP/Software+and+Computing+for+Small+HEP+Experiments

[36] CernVM-FS. Available: https://cernvm.cern.ch/fs/

[37] Chue Hong et al. 2011. Chue Hong, N. P., Aragon, S., Hettrick, S., & Jay, C. (2011). The future of research software is the future of research. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC12416084/

[38] Cloud Container Platform. Available: https://open-science-cloud.ec.europa.eu/services/cloud-container-platform

[39] CMS HPC Integration 2020. CMS Offline Software and Computing (2020). HPC Resources Integration at CMS. Available: https://confluence.egi.eu/display/SPECTRUMCoP/HPC+Resources+Integration+at+CMS

[40] CMS Phase-2 Computing Model. CMS Offline Software and Computing (2022). CMS Phase-2 Computing Model: Update. Available: https://confluence.egi.eu/display/SPECTRUMCoP/CMS+Phase-2+Computing+Model%3A+Update?src=contextnavpagetreemode

[41] Copernicus Earth-observation data space. Available: https://dataspace.copernicus.eu/

[42] Couvares 2021 (3G GW). Couvares, P. (2021). Gravitational-Wave Data Analysis: Computing Challenges in the 3G Era [Technical Report]. Available: https://confluence.egi.eu/display/SPECTRUMCoP/Gravitational-Wave+Data+Analysis%3A+Computing+Challenges+in+the+3G+Era

[43] CSCS. Available: https://www.cscs.ch/services/overview

[44] D3.1. SPECTRUM Project, "D3.1 Community of Practice — First Period Report", 2024. Available: https://zenodo.org/records/15100586

[45] D4.1. Community of Practice and sustainability plan - Final report. To be published in June 2026

[46] D5.1. SPECTRUM Project, "D5.1 Representative Use Cases: Analysis and Alignment", 2024. Available: https://zenodo.org/records/15310223

[47] D5.2. SPECTRUM Project, "D5.2 Interoperable Access Policies: Analysis and Recommendations", 2024. Available: https://zenodo.org/records/15647608

[48] D5.3. SPECTRUM Project, "D5.3 Landscape of Research Infrastructures: Technologies, Services, Gaps", 2024. Available: https://zenodo.org/records/15647755

[49] D6.1. SPECTRUM Project, "D6.1 Technical Blueprint for Compute and Data Continuum", 2026. https://zenodo.org/records/20138626

[50] Daedalus. Available: https://www.eurohpc-ju.europa.eu/daedalus-call-acquire-new-european-world-class-supercomputer-greece-2024-06-25_en

[51] Deucalion. Available”: https://rnca.fccn.pt/en/deucalion/

[52] Dewdney et al. 2022 (SKA1). Dewdney, P. et al. (2022). SKA1 Design Baseline Description [Design Baseline Description]. Available: https://confluence.egi.eu/display/SPECTRUMCoP/SKA1+Design+Baseline+Description

[53] Diefenthaler et al. 2022. Diefenthaler, M. et al. (2022). Future Trends in Nuclear Physics Computing [Editorial]. Available: https://confluence.egi.eu/display/SPECTRUMCoP/Future+Trends+in+Nuclear+Physics+Computing

[54] Digital Decade 2030. European Commission, "Digital Decade Policy Programme 2030", 2021. Decision (EU) 2022/2481 of the European Parliament and of the Council of 14 December 2022 establishing the Digital Decade Policy Programme 2030. ELI: http://data.europa.eu/eli/dec/2022/2481/oj. Available: https://digital-strategy.ec.europa.eu/en/library/digital-decade-policy-programme-2030

[55] Discoverer. Available:https://sofiatech.bg/en/petascale-supercomputer/

[56] DORA. Allen, L., Barbour, V., Cobey, K., Faulkes, Z., Hazlett, H., Lawrence, R., Lima, G., Massah, F., & Schmidt, R. (2025). A Practical Guide to Implementing Responsible Research Assessment at Research Performing Organizations. Declaration on Research Assessment (DORA). Available: https://sfdora.org/

[57] DUNE Computing CDR. The DUNE Collaboration (2022). DUNE Offline Computing Conceptual Design Report [Technical Report]. Available: https://confluence.egi.eu/display/SPECTRUMCoP/DUNE+Offline+Computing+Conceptual+Design+Report

[58] EBRAINS neuroscience platform. Available: https://www.ebrains.eu/

[59] EGI 2024 Annual Report. EGI Federation, "EGI Federation Annual Report 2024", 2024. Available: https://www.egi.eu/

[60] EGI EOSC Federation Discussion Paper. EGI Federation, "EGI Contribution to the EOSC Federation — Discussion Paper v1.2", May 2024. Available: https://www.egi.eu/
[61] EGI FedCloud. Available: https://www.egi.eu/egi-infrastructure/#layers-federated-cloud
[62] EGI HTC-oriented infrastructure. Available: https://www.egi.eu/egi-infrastructure/#layers-federated-htc
[63] eIDAS 2.0. European Commission, "eIDAS 2.0. European Identity Framework". Regulation (EU) No 910/2014 of the European Parliament and of the Council of 23 July 2014 on electronic identification and trust services for electronic transactions in the internal market and repealing Directive 1999/93/EC. ELI: http://data.europa.eu/eli/reg/2014/910/oj. Available: https://digital-strategy.ec.europa.eu/en/policies/eidas-regulation
[64] Elvira et al. 2022. Elvira, V. D. et al. (2022). The Future of High Energy Physics Software and Computing: Report of the 2021 US Community Study on the Future of Particle Physics. Available: https://confluence.egi.eu/display/SPECTRUMCoP/The+Future+of+High+Energy+Physics+Software+and+Computing%3A+Report+of+the+2021+US+Community+Study+on+the+Future+of+Particle+Physics
[65] EOSC Federation. Available: https://eosc.eu/building-the-eosc-federation/
[66] EOSC SRIA. European Open Science Cloud, "EOSC Strategic Research and Innovation Agenda v1.3", November 2024. Available: https://eosc.eu/sria
[67] EPCC (UK). Available: https://www.epcc.ed.ac.uk/
[68] EP Resolution on ERA Act. European Parliament, "Resolution of 10 March 2026 on the upcoming European Research Area (ERA) Act", P10_TA(2026)0068, 2026. Available: https://www.europarl.europa.eu/doceo/document/TA-10-2026-0068_EN.html
[69] ErUM4 Data Hub. Available: https://erumdatahub.de/en/
[70] ESCAPE Project's Open-source Scientific software and Service Repository. Available: https://projectescape.eu/services/open-source-scientific-software-and-service-repository-ossr
[71] ESFRI Landscape Analysis 2024. European Strategy Forum on Research Infrastructures, "ESFRI Landscape Analysis 2024", 2024. Available: https://www.esfri.eu/
[72] ESFRI Position Paper on ERA Act. ESFRI, "Position Paper on the European Research Area Act (ERA Act)", FI26-WP-20260115, January 2026.
[73] ESFRI Roadmap 2021. European Strategy Forum on Research Infrastructures, "ESFRI Roadmap 2021", 2021. Available: https://www.esfri.eu/
[74] ETP4HPC SRA6. European Technology Platform for High Performance Computing, "ETP4HPC Strategic Research Agenda 6", 2024. Available: https://www.etp4hpc.eu/
[75] European Interoperability Framework. European Commission, "European Interoperability Framework — Implementation Strategy", COM(2017) 134 final, 2017. Updated under Regulation (EU) 2024/903 (Interoperable Europe Act), 2024. Available: https://joinup.ec.europa.eu/collection/nifo-national-interoperability-framework-observatory/european-interoperability-framework-detail
[76] European Particle Physics Strategy 2020. European Particle Physics Strategy, "2020 Update of the European Strategy for Particle Physics", 2020.
[77] EU Data Strategy 2020. European Commission, "European Data Strategy", 2020. Available: https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy
[78] EU Data Union Strategy. European Commission, "European Data Union Strategy". Available: https://digital-strategy.ec.europa.eu/en/policies/data-union
[79] EuroHPC Federation Platform. EuroHPC Federation Platform. Available: https://www.eurohpc-ju.europa.eu/supercomputers/eurohpc-federation-platform_en
[80] EuroHPC Joint Undertaking (JU). Available:https://www.eurohpc-ju.europa.eu/index_en
[81] EuroHPC MAWP 2024. EuroHPC Joint Undertaking, "EuroHPC Multi-Annual Work Programme", 2024.
[82] EuroHPC Strategic Agenda. EuroHPC Joint Undertaking, "EuroHPC Strategic Agenda 2024", 2024. Available: https://eurohpc-ju.europa.eu/
[83] EuroHPC Summit 2024. EuroHPC (2024). EuroHPC Summit 2024: Interconnecting EuroHPC Supercomputers for Scientific and Industrial Advancement [Presentation]. Available: https://confluence.egi.eu/display/SPECTRUMCoP/EuroHPC+Summit+2024%3A+Interconnecting+EuroHPC+Supercomputers+for+Scientific+and+Industrial+Advancement
[84] EVERSE Software. Available: https://everse.software
[85] EuroHPC Supercomputers. Available: https://www.eurohpc-ju.europa.eu/supercomputers/our-supercomputers_en

[86] EVERSE. EVERSE Project. Psomopoulos, F., Capella-Gutierrez, S., Portell-Silva, L., & Pechlivanis, N. (2024, January 18). EOSC-EVERSE: Paving the way towards a European Virtual Institute for Research Software Excellence. Zenodo. Available: https://doi.org/10.5281/zenodo.10526785

[87] EXTRACT. EXTRACT Project. Available: https://extract-project.eu/

[88] EXTRACT-TASKA (Cecconi). Cecconi, B., Le Sidaner, P. & Aicardi, S. (2026). EXTRACT-TASKA: A Distributed Workflow Orchestrator for Radio Astronomy Data Processing. EOSC Winter School 2026, Nice. Zenodo. Available: https://doi.org/10.5281/zenodo.19093528

[89] FAIR4RS. FAIR for Research Software (FAIR4RS): A summary, August 2022. Available: https://www.researchsoft.org/blog/2022-08/

[90] Federated Software Catalogue. Available: https://www.eessi.io/docs/blog/2026/03/09/EESSI-in-EFP-update/#eessi-as-a-base-for-the-efp-federated-software-catalog

[91] Filipcic, Girone & Boccali 2024 (HEP-HPC Europe). Filipcic, A., Girone, M. & Boccali, T. (2024). HEP/HPC Strategy Meeting — Europe [Meeting Summary]. Available: https://confluence.egi.eu/pages/viewpage.action?pageId=298322551

[92] Fischer et al. 2020 (Heterogeneous). Fischer, M. et al. (2020). Effective Dynamic Integration and Utilization of Heterogeneous Compute Resources [Conference Paper]. Available: https://confluence.egi.eu/display/SPECTRUMCoP/Effective+Dynamic+Integration+and+Utilization+of+Heterogenous+Compute+Resources

[93] Fischer et al. 2020 (Opportunistic). Fischer, M. et al. (2020). Lightweight dynamic integration of opportunistic resources [Conference Paper]. Available: https://confluence.egi.eu/display/SPECTRUMCoP/Lightweight+dynamic+integration+of+opportunistic+resources

[94] Fisk et al. 2024 (HEP-HPC USA). Fisk, I. et al. (2024). HEP/HPC Strategy Meeting — USA [Meeting Summary]. Available: https://confluence.egi.eu/pages/viewpage.action?pageId=298322541

[95] Gal et al. 2019 (KM3NeT). Gal, T. et al. (2019). KM3NeT Report on Documentation Strategy, Environment, and Software [Deliverable]. Available: https://confluence.egi.eu/display/SPECTRUMCoP/KM3NeT+Report+on+Documentation+Strategy%2C+Environment%2C+and+Software

[96] GANIL Data Policy 2020. GANIL (2020). GANIL Data Policy [Technical Report]. Available: https://confluence.egi.eu/display/SPECTRUMCoP/GANIL+Data+Policy

[97] GCS. Available: https://www.gauss-centre.eu/for-users/hpc-access

[98] GENCI. Available: http://www.idris.fr/eng/info/gestion/demandes-heures-eng.html

[99] Girone & Boccali 2025 (HEP-HPC All Regions). Girone, M. & Boccali, T. (2025). HEP/HPC Strategy Meeting — All Regions [Meeting Summary]. Available: https://confluence.egi.eu/pages/viewpage.action?pageId=298322557

[100] Girone 2019 (HPC for LHC). Girone, M. (2019). Common challenges for HPC integration into LHC computing [Technical Report]. Available: https://confluence.egi.eu/display/SPECTRUMCoP/Common+challenges+for+HPC+integration+into+LHC+computing

[101] GSI/FAIR OS Licence Guidelines. GSI/FAIR Collaboration (2021). Open source software licences at GSI/FAIR — Guidelines. Available: https://confluence.egi.eu/pages/viewpage.action?pageId=288785215

[102] GSI/FAIR Upload Instructions. GSI/FAIR Collaboration (2023). Instructions for uploading and linking research data and software at GSI Helmholtzzentrum für Schwerionenforschung GmbH. Available: https://confluence.egi.eu/pages/viewpage.action?pageId=288785219

[103] HEPSF HL-LHC Review 2020. HEP Software Foundation (2020). HL-LHC Computing Review: Common Tools and Community Software [Technical Report]. Available: https://confluence.egi.eu/display/SPECTRUMCoP/HL-LHC+Computing+Review%3A+Common+Tools+and+Community+Software

[104] HEP-SF Strategic Plan 2026. High Energy Physics Software Foundation, "HEP Software Foundation Strategic Plan 2026", 2023.

[105] HEP-SF Roadmap. High Energy Physics Software Foundation, "HEP Software Foundation Roadmap", 2022.

[106] HLRS. Available: https://www.hlrs.de/

[107] Hollowell et al. 2023. Hollowell, C. et al. (2023). Financial Case Study: Use of Cloud Resources in HEP Computing [Presentation, CHEP 2023]. Available: https://confluence.egi.eu/display/SPECTRUMCoP/Financial+Case+Study%3A+Use+of+Cloud+Resources+in+HEP+Computing

[108] HyperCON Final Report. EuroHPC Joint Undertaking, "HyperCON Final Report", 2023.
[109] ICSC. Available: https://www.supercomputing-icsc.it/
[110] IHDEA (international heliophysics data environment alliance. Available: https://ihdea.net
[111] InterTwin. interTwin Project. Available: https://www.intertwin.eu/
[112] IVOA (international virtual observatory alliance. Available: https://ivoa.net
[113] IPDA, International Planetary Data Alliance. Available: https://planetarydata.org
[114] JSC. Available: https://www.fz-juelich.de/en/jsc
[115] JENA Computing White Paper. Joint ECFA-NuPECC-APPEC Activities (JENA), "Computing White Paper", 2023.
[116] JENA WG1 (HPC). JENA Working Group 1, "HPC for Fundamental Science", 2023.
[117] JENA WG2 (Software). JENA Working Group 2, "Software and Applications", 2023.
[118] JENA WG3 (Data). JENA Working Group 3, "Data Management and Access", 2023.
[119] JENA WG4 (Infrastructure). JENA Working Group 4, "Infrastructure and Facilities", 2023.
[120] JENA WG5 (Federation). JENA Working Group 5, "Federation and Interoperability", 2023.
[121] JUPITER. Available: https://www.fz-juelich.de/en/ias/jsc/jupiter
[122] Kahn et al. 2022 (Dark Matter Snowmass). Kahn, Y. et al. (2022). Snowmass2021 Cosmic Frontier: Modeling, statistics, simulations, and computing needs for direct dark matter detection. Available: https://confluence.egi.eu/display/SPECTRUMCoP/Snowmass2021+Cosmic+Frontier%3A+Modeling%2C+statistics%2C+simulations%2C+and+computing+needs+for+direct+dark+matter+detection
[123] Karolina. Available: https://www.it4i.cz/en/infrastructure/karolina
[124] Kokkos. Available: https://github.com/kokkos/kokkos
[125] Laycock et al. 2024 (Einstein Telescope). Laycock, P. et al. (2024). Preparatory Phase for the Einstein Telescope Gravitational Wave Observatory: Computing and Data Requirements [Deliverable]. Available: https://confluence.egi.eu/display/SPECTRUMCoP/Preparatory+Phase+for+the+Einstein+Telescope+Gravitational+Wave+Observatory%3A+Computing+and+Data+Requirements
[126] Leonardo. Available: https://leonardo-supercomputer.cineca.eu/
[127] LOFAR ERIC Access Policy. Interim LOFAR ERIC Council (2021). LOFAR ERIC Access Policy to Scientific User Services [Policy Document]. Available: https://confluence.egi.eu/display/SPECTRUMCoP/LOFAR+ERIC+Access+Policy+to+Scientific+User+Services
[128] LOFAR ERIC Science Data Policy. LOFAR ERIC Council (2023). Science Data Policy of LOFAR ERIC [Policy Document]. Available: https://confluence.egi.eu/display/SPECTRUMCoP/Science+Data+Policy+of+LOFAR+ERIC
[129] LOFAR Long-Term Archive (LTA). Available: https://lta.lofar.eu/
[130] LOFAR2.0 Summary 2023. LOFAR (2023). LOFAR2.0 compared to LOFAR: a short summary [Technical Summary]. Available: https://confluence.egi.eu/display/SPECTRUMCoP/LOFAR2.0+compared+to+LOFAR%3A+a+short+summary
[131] LRZ. Available: https://www.lrz.de/en/
[132] LUMI. Available: https://www.lumi-supercomputer.eu/about-lumi/
[133] Maeno et al. 2024 (PanDA in ATLAS). Maeno, T. et al. (2024). Utilizing Distributed Heterogeneous Computing with PanDA in ATLAS [Conference Paper]. Available: https://confluence.egi.eu/display/SPECTRUMCoP/Utilizing+Distributed+Heterogeneous+Computing+with+PanDA+in+ATLAS
[134] MareNostrum 5. Available: https://www.bsc.es/marenostrum/marenostrum-5
[135] Megino et al. 2022 (US ATLAS/CMS HPC). Barreiro Megino, F. et al. (2022). US ATLAS and US CMS HPC and Cloud Blueprint [Technical Report]. Available: https://confluence.egi.eu/display/SPECTRUMCoP/US+ATLAS+and+US+CMS+HPC+and+Cloud+Blueprint
[136] Meluxina. Available: https://www.luxprovide.lu/meluxina/
[137] MultiXscale . Available: https://www.multixscale.eu/
[138] NHR Alliance. Available: https://www.nhr-verein.de/en/computing-time
[139] NuPECC TWG9. NuPECC TWG9 (2024). Open Science and Data. Available: https://confluence.egi.eu/display/SPECTRUMCoP/NuPECC+TWG9+%3A+Open+Science+and+Data
[140] ODISSEE project. Available: https://odissee-project.eu/
[141] OSCARS. OSCARS Project. Nicoletta, C., Carrillo, R., Fiser, D., Wolff-Boenisch, B., Schmidt-Tremmel, F., Lamanna, G., Petzold, A., & Hienola, A. (2023). OSCARS project funded to foster the uptake of Open Science in Europe. Zenodo. Available: https://doi.org/10.5281/zenodo.10379119

[142] Physics Briefing Book 2020. European Strategy for Particle Physics (2020). Physics Briefing Book: Input for the European Strategy for Particle Physics Update 2020 [Strategy Report]. Available: https://confluence.egi.eu/display/SPECTRUMCoP/Physics+Briefing+Book+%3A+Input+for+the+European+Strategy+for+Particle+Physics+Update+2020

[143] Physics Briefing Book 2026. Physics Briefing Book: Input for the 2026 update of the European Strategy for Particle Physics, 2025. Available: https://arxiv.org/abs/2511.03883, https://cds.cern.ch/record/2944678/files/Physics%20Briefing%20Book.pdf

[144] Pérez-Calero Yzquierdo 2020. Pérez-Calero Yzquierdo, A. (2020). CMS strategy for HPC resource exploitation [Conference Paper]. Available: https://confluence.egi.eu/display/SPECTRUMCoP/CMS+strategy+for+HPC+resource+exploitation

[145] Pizzo et al. 2023 (LOFAR2.0 DM). Pizzo, R. et al. (2023). LOFAR2.0 Data Management Capabilities [Note]. Available: https://confluence.egi.eu/display/SPECTRUMCoP/LOFAR2.0+Data+Management+Capabilities

[146] PUNCH4NFDI. Available: https://www.punch4nfdi.de/

[147] Reppin et al. 2021 (DESY Jupyter). Reppin, J. et al. (2021). Interactive Analysis Notebooks on DESY Batch Resources: Bringing Jupyter to HTCondor and Maxwell at DESY. Available: https://confluence.egi.eu/display/SPECTRUMCoP/Interactive+Analysis+Notebooks+on+DESY+Batch+Resources%3A+Bringing+Jupyter+to+HTCondor+and+Maxwell+at+DESY

[148] RES. Available: https://www.res.es/en/access-to-res

[149] Roberts et al. 2019 (DANCE). Roberts, A. et al. (2019). Dark-matter And Neutrino Computation Explored (DANCE) Community Input to Snowmass [Meeting Summary]. Available: https://confluence.egi.eu/display/SPECTRUMCoP/Dark-matter+And+Neutrino+Computation+Explored+%28DANCE%29+Community+Input+to+Snowmass

[150] Rohr & Eulisse 2023 (ALICE O2 GPU). Rohr, D. & Eulisse, G. (2023). The O2 software framework and GPU usage in ALICE online and offline reconstruction in Run 3 [Presentation]. Available: https://confluence.egi.eu/display/SPECTRUMCoP/The+O2+software+framework+and+GPU+usage+in+ALICE+online+and+offline+reconstruction+in+Run+3

[151] Salgado et al. 2023 (SRCNet). Salgado, J. et al. (2023). SKA Regional Centres Network (SRCNet) Software Architecture Document [Design Description Document]. Available: https://confluence.egi.eu/display/SPECTRUMCoP/SKA+Regional+Centres+Network+%28SRCNet%29+Software+Architecture+Document

[152] SKA SRCNet. Available: https://www.skao.int/en/science-users/119/ska-regional-centres

[153] SCIENCE Clusters. SCIENCE Clusters. Available: https://science-clusters.eu/

[154] Sfiligoi et al. 2014. Sfiligoi, I. et al. (2014). CMS experience of running glideinWMS in High Availability mode [Conference Paper]. Available: https://confluence.egi.eu/display/SPECTRUMCoP/CMS+experience+of+running+glideinWMS+in+High+Availability+mode

[155] Simili et al. 2024 (ARM vs x86). Simili, E. et al. (2024). ARMing HEP for the future: Energy Efficiency of WLCG sites (ARM vs. x86) [Conference Paper]. Available: https://confluence.egi.eu/pages/viewpage.action?pageId=302876052

[156] Simpl Data Federation Platform. Available: https://digital-strategy.ec.europa.eu/en/policies/simpl

[157] SMP Template (Martinez-Ortiz). Martinez-Ortiz, C., Martinez Lavanchy, P., Sesink, L., Olivier, B. G., Meakin, J., de Jong, M., & Cruz, M. (2023). Practical guide to Software Management Plans (1.1). Zenodo. Available: https://doi.org/10.5281/zenodo.7589725

[158] Spider. Available: https://servicedesk.surf.nl/wiki/spaces/WIKI/pages/9797664/Data+Processing+-+Spider

[159] SRCSC WG6 2021. SKA Regional Centre Steering Committee (SRCSC) WG6, Task Package 1 (2021). SKA Science Regional Centres Community Input Questionnaire [Survey]. Available: https://confluence.egi.eu/display/SPECTRUMCoP/SKA+Science+Regional+Centres+Community+Input+Questionnaire

[160] Stagni 2023 (DiracGrid). Stagni, F. (2023). Status of DiracGrid projects [Presentation]. Available: https://confluence.egi.eu/display/SPECTRUMCoP/Status+of+DiracGrid+projects

[161] SURF's Grid. Available: https://servicedesk.surf.nl/wiki/spaces/WIKI/pages/1474704/Data+Processing+-+GRID

[162] Vega. Available: https://izum.si/en/vega-en/

[163] Wagg et al. 2021. Wagg, J. et al. (2021). SKA1 Scientific Use Cases [Technical Document]. Available: https://confluence.egi.eu/display/SPECTRUMCoP/SKA1+Scientific+Use+Cases

[164] WLCG Strategy. Worldwide LHC Computing Grid (WLCG), "WLCG Strategy Document 2027", 2023. Available: https://wlcg.web.cern.ch/

Annexes

Annex A: Priorities

Pillar 1: Policy, Trust & Governance

P01 Cross-Infrastructure Governance

Strategic Rationale and Evidence Base

Strategic Importance: The compute and data continuum vision requires research workflows to span resources governed by different entities: thematic research digital infrastructures (e.g. WLCG for HL-LHC, SRCNet for SKAO), pan-European multi-disciplinary digital infrastructures (e.g. EuroHPC, EOSC, EGI), nationally funded facilities, and community-governed federations. Each governance model reflects legitimate objectives. Without coordination mechanisms bridging these structures, researchers experience governance boundaries as friction that limits infrastructure utility.

Governance coordination underpins both interoperability and efficient resource use. Following the European Interoperability Framework [49], interoperability operates at four layers: legal (compatible access policies and liability), organisational (aligned governance processes and roles), semantic (shared accounting metrics and vocabularies), and technical (standardised interfaces and protocols). Beyond interoperability, governance coordination enables complementary investment across national and European programmes, planning certainty for multi-decade scientific programmes, and alignment of infrastructure evolution with scientific requirements.

Current State: European compute and data infrastructure operates through multiple governance structures:

- **Thematic digital infrastructures** serve specific communities through community-driven governance. WLCG has sustained federated operations for over twenty years across 160 sites in more than forty countries, coordinated through 65 Memoranda of Understanding with national funding agencies. SKA Observatory is building comparable governance through SRCNet. LOFAR ERIC, established in 2024, is building its computing governance model as LOFAR 2.0 operations approach, with growing data volumes that will require coordinated multi-year commitments from member states.
- **Pan-European digital infrastructures** provide cross-domain services through different models: EuroHPC coordinates thirty-five states through direct co-funded infrastructure investment, with governance oriented toward industrial competitiveness and technological sovereignty alongside research; EOSC federates existing infrastructure under common FAIR principles; EGI Federation provides federated HTC and cloud services governed by its member institutions; and the eInfrastructure Assembly coordinates horizontal initiatives including EUDAT, OpenAIRE, PRACE and GÉANT.

National programmes provide the majority of HTC and cloud capacity, whilst the EC contributes to HPC via EuroHPC and supports transnational and virtual access to research infrastructure services through Horizon Europe INFRA-SERV calls. Alignment across these funding streams is necessary for a coherent compute continuum.

The WLCG-EGI cooperation demonstrates that coordination across governance models is achievable when mechanisms exist. Structured engagement between data-intensive science communities and EuroHPC, however, remains underdeveloped. National and European funding routes serve different purposes: national programmes provide persistent operational capacity, whilst European programmes support coordination, innovation, and cross-border access. Bridging these complementary roles, rather than treating them as competing alternatives, is a governance challenge this priority addresses.

Community Evidence: D5.2 identifies Gap #2 (lack of coordinated governance frameworks) and Gap #16 (sustainability models not aligned across infrastructures). The JENA Computing White Paper recommends that "requirements of communities must be considered in design and procurement of future HPC systems." The WLCG Strategy notes "there is no WLCG common funding" for federating services, demonstrating that federated governance requires explicit coordination investment.

Community consultation on this priority confirmed that governance scope should focus on alignment and co-design rather than centralised control. Participants identified three topics requiring coordination beyond existing mechanisms: aligning procurement with data-intensive science requirements; coordinating innovation for full-chain workflow optimisation (as distinct from classical HPC algorithmic optimisation); and bridging the structural mismatch between operational computing needs and project-based funding.

Technical Foundation: D6.1 Section 5.2 (Co-Design of Computing, Data, and Networking Infrastructure) provides the rationale: "*HEP and RA must engage with funding agencies during the design and procurement phases of HPC systems... The most effective way to be part of the co-design process in Europe would be to obtain a recognized status within EuroHPC*." D6.1 Section 5.8 and Section 4.6 describe interoperable standards and transparent accounting across federated resources.

European Alignment: Cross-infrastructure governance connects directly to four European strategies:

- **EuroHPC**: governance coordination provides the mechanism for data-intensive science communities to engage with EuroHPC procurement and access processes, ensuring exascale investments serve scientific as well as industrial objectives.
- **EOSC**: the coordination forum aligns thematic RI data governance with EOSC's federated data infrastructure and FAIR principles, bridging domain-specific practices with the EOSC Web of FAIR Data and Services.
- **ESFRI**: research infrastructures on the ESFRI Roadmap operate on 20 to 30 year timelines; governance coordination provides the stability these multi-generational programmes require and aligns with ESFRI's 2025 recommendations on long-term engagement, funding synergies, and community development. The ESFRI thematic clusters (including ESCAPE for astronomy and particle physics) provide a coordination model that this priority extends, particularly as the Horizon Europe 2026-2027 Work Programme introduces "Data, Computing and Digital Research Infrastructures" as a sixth ESFRI domain with its own cluster. ESFRI is also positioning for a more proactive coordination role across the research and technology infrastructure landscape, advocating for consolidation through clustering and a light overarching governance model [72].
- **ERA**: the forthcoming European Research Area Act [68] will establish a binding regulatory framework for cross-border research cooperation, realising the "fifth freedom" of free movement of researchers, knowledge, and technology. The European Parliament resolution of March 2026 calls for strengthened ERA governance, enhanced coordination between national and EU R&I initiatives, and improved accessibility of research infrastructure. The ERA Policy Agenda 2025-2027 includes active actions on research infrastructures (25 Member States participating), research careers, and enabling open science. Governance coordination for the compute and data continuum directly supports these objectives by providing an institutional model for cross-border resource sharing, joint planning, and coordinated investment.

Target Outcomes and Impact Assessment

Strategic Outcomes: This priority delivers governance coordination enabling effective use of European compute and data infrastructure, benefiting thematic research communities, pan-European infrastructure providers, and funding agencies.

Outcome 1: Coordination Forum with Substantive Agenda A coordination forum brings together thematic research digital infrastructures (WLCG, SRCNet, LOFAR ERIC, and others), multidisciplinary digital infrastructures (EuroHPC, EOSC, EGI), and policy maker representatives from national funding agencies and relevant EC units. The forum addresses interoperability across four layers (legal, organisational, semantic, and technical), coordinates the innovation agenda for data-intensive science (full-chain workflow optimisation, real-time processing), and develops policy recommendations bridging national and European funding models. Formal representation replaces ad hoc engagement dependent on personal relationships.

Outcome 2: Coherent European Infrastructure Investment Governance coordination aligns infrastructure development with scientific requirements across all participants. Thematic research infrastructures gain influence over procurement and architecture decisions affecting multi-decade programmes. Pan-European infrastructures gain structured feedback from demanding users, enabling capability development that serves documented scientific needs. Funding agencies gain a coordinated evidence base for investment decisions, reducing duplication and enabling complementary use of national and European funding streams.

Success Indicators

- Coordination forum operational with participation from thematic RI leadership, pan-European infrastructure representatives, and policy maker observers
- Active involvement of scientific communities through topical networks
- Formal engagement channels established with EuroHPC, EOSC, and EGI governance structures
- Joint working groups addressing cross-cutting technical challenges (federated identity, computing interfaces, data management standards, accounting interoperability)
- Coordinated innovation agenda published, distinguishing data-intensive science requirements from classical HPC optimisation
- Policy recommendations on the operational-versus-project funding mismatch delivered to national and European stakeholders
- Documented evidence of coordination influencing infrastructure procurement or access decisions

European Impact

Governance coordination enables Europe to realise scientific return on distributed infrastructure investments and strengthens European digital sovereignty by making European infrastructure more accessible than commercial alternatives. As the ERA Act establishes a binding framework for cross-border research cooperation, governance coordination for the compute and data continuum provides a concrete implementation model for research infrastructure interoperability and coordinated investment.

Implementation Pathway and Stakeholder Coordination

Implementation Phases

Short term (1-3 years):

- Establish a coordination forum bringing together thematic and multidisciplinary digital infrastructures with policy maker representation
- Define the forum's substantive agenda: procurement alignment, innovation coordination, and bridging the operational-versus-project funding mismatch
- Map existing coordination mechanisms and identify gaps the forum addresses
- Develop shared metrics for infrastructure utilisation and cross-infrastructure workflow performance

Mid term (3-5 years):

- Consolidate coordination to deliver cross-border access to data and compute services, piloting integrated access across thematic and digital infrastructure clusters
- Develop policy recommendations on the operational-versus-project funding mismatch, aligned with the ERA Act framework
- Assess options for institutional evolution, including formal RI representation within EuroHPC and EOSC governance and coordination through the emerging ESFRI digital cluster
- Establish joint working groups on cross-cutting technical challenges connecting the forum to Pillar 2 implementation

Long term (5+ years):

- Assess impact of governance coordination and determine whether a formalised institutional structure is warranted for multi-decade infrastructure planning
- Align governance arrangements with the multi-decade timescales of the scientific programmes they serve
- Coordinate European positions in international collaborations spanning multiple continents

Stakeholder Roles

- **Scientific Communities and Research Infrastructures**
  - Responsibilities: Publish multi-year resource plans (5+ year horizons). Designate representatives for the coordination forum. Participate in co-design with e-Infrastructure providers, bringing innovation requirements (e.g. full-chain optimisation, real-time processing) alongside capacity needs. Coordinate positions across thematic infrastructures.
  - Benefits: Voice in decisions affecting multi-decade programmes; access to diverse resources through coordinated allocation; reduced cross-provider friction.
- **Computing and Data Service Providers / e-Infrastructures**
  - Responsibilities: Establish liaison mechanisms with thematic RIs including named contacts and periodic reviews. Participate in the coordination forum and working groups. Accommodate operational access models alongside project-based allocations. Provide technical coordination on interface standardisation, identity, and accounting.
  - Benefits: Demanding users driving capability development; demonstrated scientific impact justifying investment; validated federation models.
- **Policy Makers and Funding Bodies**
  - Responsibilities: Ensure data-intensive science requirements inform EuroHPC and EOSC procurement. Address the operational-versus-project funding mismatch through multi-year allocation mechanisms (P02), strategic access designations, and operational funding streams. Align national and European RI investments, consistent with the ERA Act framework. Participate in the coordination forum.
  - Benefits: Coherent European investments; reduced duplication; clearer accountability for scientific outcomes.

Resource Needs

- Coordination capacity: liaison effort (estimated 2-3 Full-Time-Equivalent -FTE) contributed in-kind by participating organisations or through existing institutional frameworks, rather than requiring a new coordination body
- Policy analysis capacity: expertise on the operational-versus-project funding gap and EC engagement
- Pilot activities: short-term project funding (e.g. Horizon Europe INFRA-DEV calls) for forum establishment and integrated access pilots
- Funding timeline: Horizon Europe INFRA calls (2026-2027) for short-term; FP10 or institutional integration for medium-term; structural funding model for long-term sustainability

Critical Dependencies

Cross-priority dependencies with:

- P02 (Federated Resource Allocation): allocation mechanisms depend on governance agreements; the forum provides the institutional context for negotiation
- P03 (Federated Identity and Trust): trust framework governance depends on cross-institutional agreements the forum facilitates
- P04 (Heterogeneous Computing Integration): technical integration requires governance alignment on interfaces, policies, and access conditions
- P11 (Community Collaboration and Co-Design): the CoP feeds requirements into the coordination forum

PO2 Federated Resource Allocation

Strategic Rationale and Evidence Base

Strategic Importance: Europe invests substantially in research computing and data infrastructure through EuroHPC, national programmes, and federated e-Infrastructures. Yet allocation mechanisms fragment this investment. Scientific computing is increasingly data-intensive, with exploitation of exabyte-scale datasets relying on a diverse and heterogeneous compute continuum spanning HPC, HTC, cloud, and emerging accelerator systems. Research infrastructures increasingly operate on multi-decade timelines, yet face annual application cycles, system-specific allocations that prevent workload-appropriate resource selection, and accounting frameworks that cannot track usage across governance boundaries. The result is less utilised infrastructures, repetitive submission burden, extra overhead e.g. in proposal writing, data collection, steering, and planning uncertainty that undermines long-term scientific programmes.

Federated resource allocation addresses how compute and data resources are allocated, accounted, and funded across the continuum:

- **Compute resources**: CPU cycles, GPU hours, emerging quantum computing allocations across HPC, HTC, and cloud systems, and associated requirements in terms of I/O and memory
- **Data resources**: Storage capacity (HL-LHC requires 4-6 EB tape archival by mid-2030s),
- **Network:** bandwidth for data transfer (SKA and WLCG distributes several hundreds of petabytes per year to regional centres)
- **Data lifecycle management,** including preservation and curation costs.

Balanced and flexible allocation across Europe is essential: governance coordination and federated identity enable access, but without allocation mechanisms that match scientific planning horizons and support diverse user communities, researchers cannot reliably integrate heterogeneous resources into sustainable workflows.

Current Status: Science is outcome driven, whereas allocations are project driven; consequently, predictability in resource allocation is difficult. Whilst a substantial part of the scientific community has (WLCG) or plans for multi-annual, MoU-based access allocations, a significant part of European e-Infrastructures generally provide grant-based access with short-term allocation cycles (months to one year)[1]. This creates several problems:

- **Repetitive submission burden**: Researchers submit repetitive allocation requests, creating administrative overhead for both users and review panels
- **Planning uncertainty**: Multi-decade experiments cannot reliably plan workflows when allocation is uncertain beyond 12 months
- **Fragmentation of resource allocation and accounting:** The allocations are generally valid for specific hosts/sites instead of the whole e-Infrastructure; this limits the overall resource utilisation
- **Utilisation inefficiency**: Short-term allocations discourage investment in workflow optimisation for specific systems

Community Evidence: The analysis of the interoperable access policies (D5.2) and the technical blueprint for compute and data continuum (D6.1) prepared by the SPECTRUM project identifies two gaps:.

#1: Long-term assured access and planning of resource allocations. Long-term (3-5 years) evolution and planning are not substantially informed by the needs of research communities that have been involved in the SPECTRUM project.

#3: Resource allocation and accounting that are valid across the whole e-Infrastructure. In many cases (such as EuroHPC JU), allocations are good for specific hosting sites/systems only, while users have voiced the necessity to use resources across an e-Infrastructure to profit from burst capacity, maximise the efficiency of systems for parts of their workload, and increase resiliency.

---

[1] A notable exception being the EuroHPC Strategic Access framework

Technical Foundation: D6.1 Section 5.8 provides the technical framework: "*Multi-year or renewable annual guaranteed allocations would enable experiments to fully leverage such investments and integrate HPC resources more effectively into their computing models*."

D6.1 4.6 (Resource Federation) specifies requirements for federated access and allocation: "*Usage data, such as CPU/GPU hours, storage consumption, and data transfer volumes, must be collected and reported in a consistent way across all sites. This not only supports fair allocation and equitable participation but also enables funding bodies and infrastructure providers to verify contributions and assess impact*."

Key technical requirements:

- Standardised and transparent accounting framework for heterogeneous resources (CPU, GPU, storage, network) utilisation across governance boundaries
- Cross-infrastructure resource discovery and allocation

European Alignment: Resource allocation management is critical to broader European strategies. For EuroHPC, multi-year allocations for storage and compute would enable better integration of HPC resources into long-term scientific planning, with efficient resource allocation mechanisms minimising the risk of underutilising the resources. A federated resource allocation mechanism benefits interoperability, harmonisation, integration and synergies among the ESFRI research infrastructures, as well as within the EOSC web of fair data and services.

Target Outcomes and Impact Assessment

Strategic Outcomes:

- **Adopt long-term, flexible resource allocation processes:** processes are implemented that allow scientific communities and research infrastructures to obtain long-term access privileges and quotas in a flexible and scalable manner, federated across infrastructures, with transparent accounting and monitoring. [D5.2, R2]
- **Enable e-Infrastructure-wide use of resource allocations and quotas**: Enable e-Infrastructure-wide allocations and harmonized accounting, so projects and users can burst across systems and providers while ensuring fair contributions and usage transparency. [D5.2, R1]
- **Align ESFRI roadmaps with e-Infrastructure roadmaps**
- **Align e-Infrastructures with the European Data Strategy:** interconnect sectoral data spaces by deploying robust e-Infrastructures and advanced interoperability standards

Success Indicators:

- **Metrics and checkpoints** are established for multi-year allocations
- **Standardised accounting framework** deployed across major European facilities
- **Multi-year allocation agreements** operational for data-intensive research infrastructures

European Impact: The European strategy for data aims at creating a single market for data that will ensure Europe's global competitiveness and data sovereignty. This will lead to the creation of Common European Data Spaces. The European leadership in AI requires cooperation and a seamless, integrated ecosystem combining AI Factories, Data Spaces and HPC resources. In this context, Federated resource allocation and standardised resource accounting systems enables research communities to access compute and data infrastructures across organisational boundaries, increasing utilisation of European investments and reducing reliance on non-European providers. Standardised accounting enables to measure the impact on science of EU investments on computing, allowing future better alignment between investments and needs. This strengthens European digital sovereignty by making European infrastructure easier to use than commercial alternatives.

Implementation Pathway and Stakeholder Coordination

Implementation Phases:

Short term (Years 1-3) :

- Continue the conversations between stakeholders, in order to close the identified gaps (Long-term assured access and planning of resource allocations; Resource allocation and accounting that are valid across the whole e-Infrastructure)
- get stakeholders involved to establish an interoperability framework for resource accounting; set up and agree on relevant metrics and checkpoints; identify tasks and responsibilities for the implementation of the federated accounting system; deploy first testbeds and feasibility studies
- Identify appropriate EC calls to get funding for the design and prototyping phases of the federated accounting system (e.g. ENSURE project in INFRA-TECH-01)
- Identify roles of research communities/RIs, resource providers/e-Infrastructures, funding bodies/policy makers in establishing memoranda of understanding

Mid term (Years 3-5):

- Pilot studies of e-Infrastructure-wide use of resource allocations and quotas
- Testbeds for the accounting system are deployed and progressively integrated into resource providers, infrastructures, research communities.
- Memoranda of Understanding are put into place

Long term (Years 5+):

- A federated, flexible resource allocation process is widely adopted
- The standardised resource accounting system is routinely used in production

Stakeholder Roles:

- **Scientific Communities, Research Infrastructures**
    - Responsibilities: establish clear computing resource needs over a timescale of several years, in terms of workflows to execute, storage needed, data access patterns, local- and wide-area network requirements; consume the information aggregated by the accounting system
    - Benefits: researchers obtain long-term access privileges and quotas in a flexible and scalable manner
- **Computing & Data Service Providers / e-Infrastructures**
    - Responsibilities: Enable e-Infrastructure-wide use of resource allocations and quotas; establish MoUs with funding bodies and scientific communities/RIs; agree with Scientific Communities and Research Infrastructures on federated accounting metrics and checkpoints, implement the associated sensors and aggregate information
    - Benefits: underutilisation of resources is minimised
- **Policy Makers / Funding Bodies**
    - Responsibilities**:** establish clear and multi-annual funding plans with the goal of matching eInfrastructure technological evolution with scientific needs; establish MoUs with computing & data service providers and scientific communities/RIs
    - Benefits**:** agencies contributing funding to e-Infrastructures can gauge their return on investment by a more efficient use of the invested funding tailored to scientific community needs

Resource Needs:

- **Funding is needed for**
    - the technical aspects of the accounting system and the multi-annual, federated resource allocation mechanism
    - developing and finalising the policy aspects of federated resource allocation (a pool of people from all stakeholders)
- **Funding timeline**:
    - Short-term, project-related funds for designing/prototyping/testbeds/pilots
    - medium-long term structural funds for commissioning and operations

Critical Dependencies:

Cross-priority dependencies with
- P01-Cross-Infrastructure Governance,
- P03- Federated Identity and Trust,
- P09-Open Science and Reproducibility

P03 Federated Identity and Trust

Strategic Rationale and Evidence Base

Strategic Importance: Federated identity and trust infrastructure enables researchers to access heterogeneous computing and data resources across organisational and national boundaries. Without seamless, secure authentication and authorisation, the compute and data continuum vision remains unattainable. This priority provides the technical identity layer that makes cross-infrastructure workflows executable and data FAIR-compliant. The challenge spans both compute and data access:

- **Compute access**: Authenticating to HPC, HTC, and cloud systems; authorising job submission and resource consumption; enabling automated workflow execution across sites.
- **Data access**: Authenticating to storage systems and data repositories; authorising data read/write/transfer operations; enabling FAIR data access (data cannot be FAIR if it cannot be accessed); protecting sensitive data under GDPR whilst enabling cross-border research.

The scale is substantial: enabling 2 million researchers across 800+ universities [66] to efficiently utilise heterogeneous computing and data infrastructures whilst maintaining security, compliance, and operational efficiency. This priority describes the importance of a sound federation of HPC, HTC and cloud systems and network of trust on the example of the HEP and Radioastronomy community but is equally relevant and important for many other communities as well.

Current State: D6.1 Section 5.6 (Security, Authorization, and Authentication) documents the authentication fragmentation: "*HEP collaborations rely on the WLCG Trust model, which enables sites to grant access to users and workflows based on their membership in federated Virtual Organizations (VOs). This approach simplifies user management but presents challenges when integrating with HPC centres that impose stricter authentication and authorization policies.*"

Current landscape:

- **Federated infrastructures** (EGI, WLCG) have developed technical solutions based on AARC Blueprint Architecture.
- **HPC centres** typically require independent accounts linked to individual individuals or federated identities incompatible with VO-based access control.
- **No federated service accounts**: HPC centres provide accounts linked to identifiable individuals, creating barriers for automated workflows. HPC centres do not support service accounts or robot accounts for automated workflows.

Proven technical foundations demonstrate that federated AAI works at scale. AARC Blueprint Architecture has achieved global adoption across five continents, validating its viability as a European standard. D6.1 Section 4.8 reports current tool usage is "almost equally shared between Indigo-IAM, CERN-SSO and EGI Check-in (with MyAccessID being used by the EFP)" (D6.1). EGI Check-in successfully translates between four protocols, demonstrating technical feasibility of protocol bridging [59]. eduGAIN operates globally with national federations spanning 40+ countries. EuroHPC Federation Platform is developing federated identity management based on MyAccessID service provided by GÉANT [82]. WLCG demonstrates cybersecurity coordination at federation scale through joint OSG/EGI security working groups, providing an operational model for trust framework governance [164].

Community Evidence: D5.2 identifies critical gaps: **Gap #5**: Lack of federated identity management across European e-Infrastructures; **Gap #6**: Unattended execution of long-running workflows, e.g. reauthentication or MFA which creates a "human-in-the-loop" bottleneck preventing unattended execution of long-running workflows or Token-based authentication challenges for command-line and long-running workflows among other blockers.

AARC BPA 2025 (AARC-G080) [1] provides best practice recommendations for federated AAI specifically addressing research infrastructure needs [66]. Also the AARC-Policy-Development-Kit [3] defines policies baseline expectations for all infrastructure members, so that trust can be established. The JENA Computing White Paper identifies standardised policies across the "5-domain gap" (AAI, data transfers, networking, edge services, container support) as essential for HPC integration [115].

Technical Foundation: D6.1 Section 4.8 (Security and Trust) identifies key capabilities: "*The European Compute and Data Continuum will be a federated infrastructure spanning multiple countries, institutions, and administrative domains. In such an environment, trust and security are not add-ons but foundational enablers.*"

Key capabilities from D6.1 4.8:

- **Identity and Access Management (IAM)**: Federated solutions supporting multiple authentication sources, account linking, transparent attribute delivery. Once authenticated, user attributes (collaboration membership, VO roles, project grants) travel with the identity so that access policies can make decisions (D6.1).
- **Trust Framework:** Governance and technical mechanisms for mutual recognition across institutional boundaries. Sites must demonstrate compliance with minimum security baselines, privacy regulations, and facilitate the capabilities necessary for enabling the FAIR principles (D6.1). OIDC Federation could serve as a concrete technical accelerator for the Trust Framework, e.g. by adopting AARC BPA as the architecture baseline, and evaluating OIDC Federation as the scalable technical mechanism to implement the Trust Framework for OIDC/OAuth ecosystems.
- **Policy Enforcement**: Dynamic, context-aware access decisions. D6.1 provides concrete examples: "data from telescope X can only be stored in GDPR-compliant facilities within the EU" or "users must re-authenticate with multi-factor authentication when accessing Tier-0 resources" (D6.1).
- **Privacy & Data Protection**: GDPR compliance throughout the authentication chain, with technical safeguards including encryption at rest and in transit (D6.1).

D6.1 Section 5.6 specifies the requirement: "*The foundation of such a solution would be a federated Authentication and Authorization Infrastructure (AAI) based on industry standards. This would enable users to authenticate once and access multiple sites securely, with consistent authorization policies.*"

European Alignment: Federated identity directly enables European strategic initiatives:

- **EOSC and especially EOSC Web of FAIR Data:** Authentication federation is fundamental, as in general data cannot be FAIR if it cannot be accessed [66], except public open data, which can be FAIR without AuthN.
- **EuroHPC Federation Platform**: Unified access to EuroHPC systems requires federated identity as foundation [82]
- **Digital Decade 2030 [54]**: Secure data spaces and digital public services depend on cross-border authentication
- **eIDAS 2.0 [63]**: European Digital Identity framework provides long-term alignment opportunity

Target Outcomes and Impact Assessment

Strategic Outcomes: Federated identity and trust infrastructure delivers value across three dimensions: seamless researcher access, automated scientific workflows, and FAIR data accessibility.

Outcome 1: Unified Authentication and Workflow Automation

Researchers authenticate once with home institution credentials and access any participating European compute or data infrastructure without additional authentication steps. This eliminates the current fragmentation of passwords, certificates, and tokens whilst maintaining appropriate security standards.

Scientific workflows require extended unattended execution across multiple infrastructures. Token delegation mechanisms and risk-based authentication profiles enable progressive extension of unattended workflow duration, from days to weeks to months, aligned with the continuous processing requirements of HL-LHC and SKA Observatory operations.

Outcome 2: FAIR Data Access and Trust Framework Governance

Data cannot be FAIR if it cannot be accessed [66]. Federated authorisation enables FAIR data access across repositories whilst maintaining GDPR compliance for cross-border transfers. Research collaborations manage access rights in real-time through standardised group and role mechanisms that propagate across connected infrastructures.

Trust framework governance extends the proven WLCG OSG/EGI security coordination model [164], coordinating incident response, establishing mutual recognition across institutional boundaries, and managing liability for federated operations.

Success Indicators

Within control of trust and identity management :

- AARC Blueprint Architecture deployed across EuroHPC Tier-0 systems and major national facilities.
- Specific and realistic protocol translation operational for heterogeneous compute and data systems.
- Unattended workflow execution extended from days to weeks to months.
- Authorisation changes propagating within hours across federated infrastructure
- Trust framework governance coordinating incident response across participating countries out of control of trust and identity management.
- Time-to-first-access reduced from days to hours for new researchers - (AAI aspect).
- Providing interoperability with other major infrastructures.

European Impact: Unified access management transforms European research competitiveness by removing artificial barriers to infrastructure utilisation. Researchers recover time previously lost to authentication procedures. FAIR data principles become achievable at scale, with federated authorisation enabling cross-border access whilst maintaining appropriate protections.

European digital sovereignty is strengthened as authentication remains within European governance. The AARC Blueprint Architecture, originating from European research infrastructure collaboration, has been adopted as a global standard across five continents, demonstrating European leadership in federated identity.

Implementation Pathway and Stakeholder Coordination

Implementation Phases

- **Short term** (1-3 years)
    - Trust Governance bootstrap.
    - Operational KPI agreed.
    - Review of site-local policy and risk acceptance barriers to the use of agentic access (i.e. service accounts).
- **Mid term** (3-5 years)
    - Trust framework hardening (onboarding/offboarding procedures, audit expectations across infrastructures).
    - Enable agentic access and community-coordinated authorisation, including any requisite compensatory controls or trust framework.
    - Establish a trust and security coordination framework that includes federated community participants, including continuous readiness evaluation.
- **Long term** (5+ years)
    - Make GDPR a normal operational capability rather than project-by-project integration (including auditing and policy enforcement).

Stakeholder Roles

- **Scientific communities and Research Infrastructures:** Research Communities and Research Infrastructures including end-users (e.g. community managers, workflow owners, data stewards)
    - Responsibilities: The scientific communities including research infrastructures need to identify the specific requirements for accessing a federated HPC, HTC and cloud infrastructure and agree with other communities on European standards in collaboration with the service providers. In addition, use cases showing the full potential and benefit of an integrated HPC, HTC and cloud infrastructure will be essential to ensure funding for the integrated services and equally important compute time as well.
    - Benefits: Defining European cross domain standards will help to build a sound European Compute Infrastructure. Demonstrating the impact on research and innovation in Europe

will help to secure political support (nationally and European) and required funding for services and compute resources.

- **Service providers and facilities:** Especially HPC HTC and Cloud provider, including Trust Anchor Operators/Federation Authorities, Federated AAI Service Operators (e.g. Proxy/Broker operators), Legal/Data Protection roles (policy owners, DPO), Security Operations & CSIRT.
    - Responsibilities: Service providers need to closely cooperate and exchange with the scientific communities in order to define the new federated access policies and adapt their local AAI policies to be compatible with the enhanced needs of the scientific communities (respecting their security policies).
    - Benefits: By enhancing the access possibilities to European HPC, HTC and cloud services a new European federated compute ecosystem could be created and new communities and users will be able to exploit the established HPC, HTC and cloud infrastructure. A wider and more efficient use of the compute infrastructure would boost European competitiveness and innovation.
- **Policymaker (European-level):**
    - Responsibilities: Policy makers should analyse the potential of a European federated HPC, HTC and cloud Infrastructure and recognise the benefit for research and innovation. Sufficient funding needs to be planned both for developing the federated services across disciplines and the compute resources accessible via the federated infrastructure.
    - Benefits: The most effective use of the available compute resources will help to create synergies in Europe and thus strengthen European competitiveness. The federated compute infrastructure will also help to use synergies and operate a more cost efficient compute infrastructure.

Resource Needs

- Funding:
    - AAI operations capacity (sustained staff for IAM/AAI not only "project-mode" work) (short term).
    - Policy capacity specialist that can align legal/privacy constraints (short term).
    - Pilots and validation across multiple facilities (testbeds, compliance validation campaigns, integration sprints) (medium-term).
    - Training programme (certification-like) (short term).
    - Lifecycle sustainability with funding line for federation core services (not project-only). (long-term).
- Infrastructure
    - Security-by-design setup (short term) including security engineering time to define minimum baseline controls (logging, auditing, key management, token signing, emergency revocation).
    - Authorisation at scale (medium term), including specialist to standardise entitlement semantics and reduce reliance on VO-specific bespoke mappings.
    - Adapted access to services and facilities (short, mid and long term).

Critical Dependencies

Governance agreements. P03 cannot scale without an agreed trust governance approach: who recognises whom, under what baseline, with what incident response, etc. - e.g. resource allocation for community access.

Standardised authorisation semantics. group/role/entitlement definitions and lifecycle management must be interoperable.

Cross-priority Dependencies with:

- P01-Cross-Infrastructure Governance
- P04 Heterogeneous Computing Integration

Pillar 2: Architecture & Interoperability

P04 Heterogeneous Computing Integration

Strategic Rationale and Evidence Base

Strategic Importance: The compute continuum (seamless integration of HPC, HTC, Cloud, edge, and emerging quantum resources) is essential for data-intensive science operating at exabyte scales. HL-LHC will generate 1 EB/year requiring processing across heterogeneous resources (D5.1), whilst SKA Observatory distributes 700 PB/year to regional centres requiring diverse computational capabilities (D5.1). Scientific workflows increasingly require combinations of high-throughput batch processing, high-performance simulation, interactive analysis, and accelerated computing that no single resource type can optimally provide.
European research communities face a fragmented landscape where HPC, HTC, and Cloud resources operate under different allocation models, access policies, and technical interfaces. HEP communities have historically optimised for HTC workloads (e.g. WLCG sustained 7.4B HTC CPU hours in 2024 [59]), whilst emerging requirements for ML training, simulation, and real-time processing demand HPC and GPU capabilities that require new integration approaches.
This priority describes the importance of an integration of heterogeneous computing (continuum computing) including HPC, HTC, cloud, edge and emerging QC systems on the example of the HEP and Astronomy community but is equally relevant and important for many other communities, e.g. numerical weather prediction and foundation model training as well.

Current State: D6.1 Section 5.1 (Standardization of Interfaces) identifies the core integration challenge: "*Today, each supercomputing site uses its own portals, schedulers and software stacks. Therefore, an experiment wishing to run on multiple HPC centres must custom-adapt to each… Without a standardized approach, scaling the number of participating HPC sites remains a labor-intensive challenge.*"

D6.1 Section 2.3.2.5 (Compute Federation Services) documents the compute federation landscape from D5.3 (D5.3): "*It should be mentioned that none of these [federation platforms] support end user access privileges/quotas across all participating systems, and that automatic routing of accesses/compute tasks to the best suited systems is not available.*"

Current landscape shows partial federation within resource types, but limited cross-type integration:

- **HPC federation**: EuroHPC Federation Platform [79] currently under development, will integrate EuroHPC supercomputing facilities, AI factories and Quantum Computers) and will provide AARC-BPA-compliant AAI across EuroHPC systems, with interoperable services for cross-site workflows [82]. Integration with HTC federations is not yet addressed and how Gigafactories will be incorporated remains also unclear.
- **HTC/Cloud federation**: WLCG federates HTC resources for workflow execution via experiment-specific layers, with centres providing access points mediated by batch systems (D6.1). EGI Federation federates HTC and Cloud resources across European e-Infrastructures via common AAI, accounting, and service catalogues [59]. Similarly to WLCG, SRCnet is currently being built, the federated digital infrastructure to process, store and distribute SKA Telescopes scientific data.

The JENA Computing White Paper identifies seven technical gaps between HPC and HTC: edge services, federated access and AAI, workflow management, wide-area networking, data management, software deployment, and programming models [115]. The JENA HPC Working Group specifies six technical areas requiring standardised policies across HPC centres: AAI, data transfers, networking, edge services, container support, and software deployment [116].
Significant progress has been made through European projects. The interTwin project [111] delivered working HPC federation solutions across Cloud, HTC, and HPC testbeds with 10 pilot Digital Twin applications, now serving as "foundation for EGI Federation's HPC integration strategy" [59]. EGI's strategic transition explicitly commits to moving from "compute-focused roots" to "data-and-compute-centric infrastructure" [59], with plans to formalise HPC as a fully-fledged offering. The EXTRACT project [87] has been working on solutions for workflow execution and orchestration.

EuroHPC's Federation Platform addresses unified access to 23+ heterogeneous systems including 8 pre-exascale/exascale HPC systems, 8 quantum computers, and (at the time of writing) 13 AI Factories [82]. However, integration with existing HTC federations (e.g., WLCG with 160 sites across 40+ countries) and emerging Cloud resources remains incomplete.
In the context of BDVA, DataLabs initiative [18] brings together key stakeholders in the digital ecosystems for the implementation of the European Data Union Strategy [78] following the AI Continent Action Plan [9].

Community Evidence: D6.1 Section 2.5 consolidates requirements from WP5 analysis, identifying key gaps relevant to heterogeneous computing integration:

- **Gap #3: Interoperable allocation and accounting**. Allocations and quotas are bound to specific HPC centres, limiting flexibility. Accounting practices differ across infrastructures. Recommendation: Enable e-Infrastructure-wide allocations and harmonised accounting, so projects can burst across systems whilst ensuring fair contributions and usage transparency (D6.1).
- **Gap #4: High-level and portable software environments**. Software stacks differ widely between sites; portability across CPUs, GPUs, and accelerators is inconsistent. Scientists often face brittle interfaces. Recommendation: Provide standardised, efficient software stacks (containers, CVMFS) and high-level domain-specific interfaces, enabling reproducibility and accelerator portability (D6.1).
- **Gap #5: Workflow execution and orchestration**. Workflow engines are fragmented across communities (e.g., HTCondor in HEP vs LEXIS Platform in EuroHPC). Job granularity and scheduling flexibility represent fundamental differences between HTC and HPC that must be addressed. Cross-facility workflows lack resilience and provenance capture. Recommendation: Co-design and converge on common APIs for workflow systems, with support for unattended execution, provenance tracking, and cross-infrastructure orchestration (D6.1).
- **Gap #14: Interactive access to significant-scale computing**. Interactive access to small/medium partitions is possible, but problems include delays in availability and potential underuse of resources. Recommendation: Extend scheduling/orchestration to support interactive compute use cases (D6.1).

Technical Foundation : D6.1 Section 4.1 (Compute Resources) provides the architectural framework: "*The Compute Resources capability is a foundational layer of the European Compute and Data Continuum, providing seamless access to heterogeneous computational resources distributed across HPC centres, community and commercial clouds, edge infrastructures, and emerging quantum platforms.*"

D6.1 Section 5.5 (Workflow Adaptation and Optimization) recommends: "*Integrate existing application specific workflow systems with HPC resource managers to enable native submission, data staging, and monitoring across federated infrastructures to allow for orchestration of complex workflows across grids and supercomputers*" (D6.1).

Multiple approaches exist for addressing these gaps, including meta-schedulers and platform-level orchestration as well as API-based access mechanisms are essential for workflow engine integration.

European Alignment: The ESFRI Landscape Analysis 2024 identifies that "systems of linked RIs are likely to be necessary" for complex scientific challenges, with distributed RIs requiring "central coordination" and European Partnerships expected to "reduce fragmentation of the research and innovation landscape" [71]. This provides policy-level justification for heterogeneous computing integration.

EuroHPC targets federation across its investment in pre-exascale and exascale systems, with the Federation Platform (under development) designed to provide unified access whilst preserving national sovereignty [82]. EGI's 2025+ strategic direction commits to the data-compute continuum transition [59], aligning with EOSC's federated infrastructure model [66]. Both infrastructure providers and user communities have responsibilities in achieving federation goals.
JENA identifies standardised policies across HPC centres for AAI, data transfers, networking, edge services, and container support as prerequisites for effective technical integration [13]. In order to create the compute continuum, policies should focus on integrating existing initiatives rather than forcing individual HPC or HTC centres to change.

In addition it is also critical to introduce a shared language and ontologies in order to enable resource sharing. This is a first-stage requirement that enables many other technical solutions. Ontologies are politically easier to agree on as they do not change existing habits, only require precise definition of needs.

Target Outcomes and Impact Assessment

Strategic Outcomes: Heterogeneous computing integration delivers value across two dimensions: seamless workflow execution and standardised integration infrastructure.

Outcome 1: Standardised Integration Policies Across European HPC, HTC Centres and Cloud Providers
European pre-exascale and exascale centres implement standardised policies for AAI, data transfers, networking, edge services, and container support, enabling plug-and-play integration for data-intensive science workflows. Cross-platform accounting metrics extend established benchmark methodologies (e.g., HEPSCORE) to unified CPU+GPU+quantum measurement, enabling comparable resource tracking across heterogeneous infrastructure. Policies also need to connect with Data Infrastructures and Data Providers

Outcome 2: Seamless Workflow Execution Across Compute Continuum
Researchers submit workflows that automatically decompose across optimal resource types (HTC for throughput, HPC for tightly-coupled simulation, Cloud for burst capacity, GPU for ML training) without manual resource management. The unified submission interface supports automatic resource selection across federated sites, enabling scientific workflows to span organisational boundaries whilst maintaining data locality optimisation and provenance tracking.

Success Indicators:

For outcome 1:

- Integration coverage: Sites meeting standardised integration criteria (AAI, containers, data staging, network access).
- Workflow validation: Representative HEP and RA workflows demonstrated on multiple resource types.
- Orchestration adoption: Production workflows using cross-infrastructure orchestration.
- Accounting interoperability: Common metrics (e.g., HEPSCORE-based) reported across HPC/HTC/Cloud sites.

For outcome 2:

- Workflow portability: Percentage of scientific workflows executable across multiple resource types without modification.
- Resource utilisation efficiency: Utilisation rate of heterogeneous resources under unified allocation.
- Time-to-science: Reduction in time from workflow submission to results for multi-resource workflows.
- Integration coverage: Number of sites implementing standardised integration policies.
- Accounting harmonisation: Common accounting framework operational across federated HPC/HTC/Cloud resources.

European Impact: Heterogeneous computing integration maximises return on Europe's distributed infrastructure investments. Rather than requiring researchers to choose between HPC for simulation, HTC for throughput, or Cloud for flexibility, integrated infrastructure enables optimal resource matching for each workflow component. This increases utilisation efficiency, reduces time-to-science, and enables new computational approaches combining traditional physics simulation with ML-accelerated analysis.

Implementation Pathway and Stakeholder Coordination

Implementation Phases

- **Short term (1-3 years):**
    - Agree on a common (descriptive) language - ontologies - over the domain specific ontologies. Once having them, then they can be mapped to HPC systems.

    - Reliance on and starting with existing initiatives, with e.g. EuroHPC Federation Platform (EFP) establishing a European federated and secure supercomputing, AI and quantum computing service.
    - Start discussion about policies around existing initiatives.
    - Analyse the established policies, develop suitable prototypes and start with the implementation.
- **Mid term (3-5 years):**
    - Show first implementations and API-based access across initiatives EuroHPC, EGI and Cloud providers.
    - Demonstrate different use cases including HTC, batch execution, AI training, and inference with different requirements.
    - Defining next steps based on lessons learned.
- **Long term (5+ years):**
    - Unified federation across different HPC, HTC and cloud providers based on lessons learned from earlier phases.

Stakeholder Roles

- **Scientific Communities, Research Infrastructures**
    - Responsibilities: Community workflows and codes must be adapted for the efficient use of a federated compute continuum infrastructure, especially HPC resources. y
    - Benefits: Provide full ability to use the compute continuum.
- **Computing & Data Service Providers / e-Infrastructures**
    - Responsibilities: Align policies and protocols, ensure federation fits funding models.
    - Benefits: Enable the compute continuum, offer services to a wide range of users.
- **Policy Makers / Funding Bodies**
    - Responsibilities: Provide political agreement, legal framework, and funding support.
    - Benefits: Most efficient use of funding, to help the EU keep being competitive. Realise a federated compute and data infrastructure.

Resource Needs

- A federated European compute continuum involving all relevant services providers requires first support by politics and funding agencies.
- An efficient approach to establish the federated compute continuum would be dedicated funded projects both for service providers and scientific communities.
- Both compute service providers and scientific communities need to agree and engage in the joint projects. To build the legal framework and technical solutions is a basic requirement, however, compute resources (HPC, HTC, Cloud, etc) for scientific communities and research infrastructures are needed as well in order to allow the federated use of compute resources.

Critical Dependencies

First, a common ontology for the different stakeholders (scientific communities, research infrastructures and services providers) needs to be defined. Then, the legal & political framework can be elaborated and defined. Also, a political discussion and decision with a firm funding commitment is needed. The continuum computing priority has of course also cross-dependencies to almost all other priorities in this document.
Cross-priority dependencies with:

- P01: Cross-Infrastructure Governance
- P02: Federated Resource Allocation
- P03: Federated Identity and Trust
- P05: Data Management and Transport
- P06: Workflow Orchestration and Management.
- P08: Software Portability.
- P11: Community Collaboration and Co-Design.

P05 Data Management and Transport

Strategic Rationale and Evidence Base

Strategic Importance SPECTRUM use cases are based on the acquisition and processing of very large amounts of data. The structure of the data (their architecture) and their management are fundamental to European e-Infrastructure competitiveness. The capability to store data and efficiently transport them to HPC centres is essential for HEP and RA communities, whilst native management according to FAIR principles enables cross-domain research and Open Science.

Current State: The management of large amounts of data around HTC computing centres is reasonably addressed by existing tools within each domain. Pan-European research networks (GÉANT) and dedicated overlays (e.g. LHCONE, LHCOPN) provide high-capacity transport, though sustained investment is needed for exabyte-scale data movement. Gaps exist primarily between different types of infrastructures (e.g., data infrastructures and HPC facilities) and across different domains (e.g., standardisation of FAIR storage) (D6.1).

Community Evidence: Three main gaps have been identified in the data domain (D6.1):

- Gap #6 (Data federation and FAIR principles): No unified architectural approach or core implementation for FAIR storage across domains, including standardised metadata and provenance tracking
- Gap #7 (Data transfer and staging): Transfer of large datasets between long-term archives and transient compute resources is highly fragmented and often relies on ad-hoc tools. Remote access in some HPC is not possible as outside connections from nodes is blocked for security reasons.
- Gap #9 (Sensitive data handling): Complexity of integrating GDPR-sensitive and special category data threatens to impact usability and performance for open scientific data use cases

Technical Foundation: D6.1 4.2 (Data Resources) defines four capabilities comprising the Data Continuum (D6.1):

- Federated Storage: Distributed storage presenting unified namespace to users and applications, with data replication, integrity management, and embedded access controls
- Data Transfer: High-performance data movement services (e.g., FTS, Rucio, XRootD) orchestrating bulk transfers across research network backbones (e.g. GÉANT, LHCONE)
- Data Discovery and Cataloguing: Metadata catalogues and registries enabling data findability, with provenance tracking supporting FAIR principles
- Data Lifecycle Management: Automated policies for retention, tiered storage migration, integrity verification, and long-term preservation

These capabilities transform fragmented storage silos into a seamless, policy-compliant ecosystem for exabyte-scale scientific data.

European Alignment: ESFRI identifies data sharing across Research Infrastructures as essential for complex scientific challenges [71]. EOSC provides the policy framework through its federated infrastructure model [66], whilst EuroHPC requires efficient data access to exploit European compute capacity for data-intensive workloads [82].

Target Outcomes and Impact Assessment

Strategic Outcomes

Outcome 1: Federated FAIR Data Management
Cross-domain standards for federated FAIR data management enable Research Infrastructures to share data whilst maintaining custodian control. This includes common metadata catalogues, provenance tracking, and access interfaces built on progressive maturity frameworks. Flexible architectures decouple sensitive and non-sensitive data repositories, enabling GDPR compliance without impacting performance for open scientific use cases.

Outcome 2: Automated Data Movement at Scale
Large-scale data movement infrastructure operates with advance planning capabilities across HTC facilities, HPC centres, and cloud providers. High-performance transfer services (e.g., FTS, Rucio) integrate with HPC storage systems, enabling autonomous data staging for workflows spanning federated European infrastructure. Multi-protocol support ensures compatibility with diverse storage backends.

Success Indicators

- FAIR adoption: Research Infrastructures implementing common FAIR data management frameworks
- Transfer automation: Production workflows with fully automated data pre-staging across infrastructure types
- Integration breadth: HPC centres, cloud providers, and HTC sites integrated via common data movement protocols
- Data accessibility: Representative datasets available for analysis within target timeframes
- Interoperability validation: Cross-domain workflows demonstrated using federated data services

European Impact

Federated FAIR data management strengthens European data sovereignty by enabling Research Infrastructures to share data under common governance whilst retaining custodian control. Automated data movement maximises return on EuroHPC investment by enabling data-intensive workloads to exploit European compute capacity. Shared data infrastructure reduces duplication across Research Infrastructures, improving cost efficiency and sustainability.

Implementation Pathway and Stakeholder Coordination

Implementation Phases

Short term (1-3 years):
- Perform a gap analysis to select pilot use cases related to FAIR storage, high-performance data movement and access, and sensitive data handling, with sufficient coverage for different topics.
- Identify critical interfaces between:
  - Research Infrastructures (RI) generating or stewarding large datasets
  - e-Infrastructures providing compute, network and storage services
  - Community and cross-domain data-access interfaces / APIs (including metadata)
- Identify and evaluate viable architectures for federated FAIR storage, metadata cataloguing, and provenance tracking.
- Pilot one or more representative use cases, such as for example:
  - FAIR-compliant cross-RI data federation
  - Automated data staging from archives into HPC environments
  - Handling of mixed sensitive/non-sensitive datasets

Mid term (3-5 years):
- Prototype and demonstrate integrated FAIR data management across multiple RIs and e-Infrastructure providers.
- Validate that architectures are fit for purpose for exabyte-scale data volumes.
- Expand scope of supported use cases:
  - High-frequency data ingestion (e.g., radio astronomy streaming or GW low-latency data)
  - Multi-protocol access to federated storage
  - GDPR-sensitive workflows using secure enclaves or logically separated data zones
- Develop standards and practices governing:
  - FAIR metadata schemas
  - Access control & identity integration
  - Operational interfaces for cross-domain data movement
  - Compliance requirements for RIs and e-Infrastructure providers
- Draft a Blueprint Architecture for the European Data Management & Transfer ecosystem.

Long term (5 years):
- Roll-out federated data management and automated transport capabilities across European e-Infrastructures.

- Onboard more communities adopting federated FAIR data services, including new scientific domains.
- Ensure sustainability and continual evolution through collective governance, coordinated investment, and community-driven standards.

Stakeholder Roles

- **Scientific Communities, Research Infrastructures**
  Responsibilities:
    - Co-design federated FAIR data architectures, including interoperable metadata standards and provenance models.
    - Provide domain-specific requirements for data retention, curation, sensitivity, and access.
    - Adopt harmonised interfaces and participate in pilot and validation campaigns.

  Benefits:
    - Seamless access to distributed data holdings with uniform FAIR-compliant interfaces.
    - Increased scientific productivity due to reduced time-to-access and improved dataset reuse.
    - Greater visibility, interoperability, and cross-domain collaboration.

- **Computing & Data Service Providers / e-Infrastructures**
  Responsibilities:
    - Deploy scalable federated storage systems with unified namespaces and policy-controlled access.
    - Provide high-performance, automated data-transfer capabilities integrated with workflows.
    - Adopt high-level data-access APIs
    - Provide secure environments for sensitive and special-category data, aligned with GDPR and ethical frameworks.
    - Monitoring services

  Benefits:
    - Maximised utilisation of storage, network, and compute resources through coordinated data movement.
    - Reduced operational overhead thanks to automation and standards-based interfaces.
    - Broader user base as more communities adopt shared data services.

- **Policy Makers / Funding Bodies**
  Responsibilities:
    - Ensure continued investment in high-capacity pan-European research networks (e.g., GÉANT).
    - Support development of FAIR-aligned infrastructure and long-term preservation facilities.
    - Coordinate regulatory and governance frameworks for cross-border data management (GDPR, FAIR metrics, EOSC policies).

  Benefits:
    - Strengthened European data sovereignty and reduced duplication across scientific domains.
    - More efficient use of public infrastructure investments.

Resource Needs

- Compute, network, and storage resources scaling across phases from pilot to exabyte-level deployment.
- Persistent high-speed connectivity through GÉANT and domain overlays (LHCOPN, LHCONE, or equivalents).
- Monitoring services
- Long-term archival and nearline storage; data orchestration tools including intelligent tiering systems; metadata catalogues.
- Engineering resources within RIs and e-Infrastructures to develop, port, test, and integrate FAIR and data-transfer capabilities.
- Governance structures for standards, certification, and service evolution.

Critical Dependencies

- Commitment from infrastructure providers to expose harmonised and interoperable data interfaces.
- Adoption of common metadata standards, provenance models, and access protocols across domains.

Cross-priority dependencies with:

- P01 (Cross-Infrastructure Governance)
- P02 (Federated Resource Allocation), P03 (Access, Security & Trust)
- P04 (Heterogeneous Computing Integration)

P06 Workflow orchestration and management

Strategic Rationale and Evidence Base

Strategic Importance: Scientific workflows are becoming increasingly sophisticated. Advanced workflow management systems capable of orchestrating multi-step processes across distributed resources and domains, ensuring data locality, and optimizing resource utilization, are therefore needed. Unifying workflow environments is crucial for sustaining and expanding innovative approaches such as AI-based workflows. Automated workflow tools are needed to streamline updates and deployments of e.g. AI models in real-time applications. Mixed workloads that combine embarrassingly parallel tasks with tightly coupled AI/ML processes must also be supported. Modern systems must adapt to support this diversity by providing heterogeneous architectures that integrate CPUs, GPUs, and high-speed interconnects. Streamlining workflow software and using automated workflow management for orchestrating, monitoring, accounting, is also important to reduce operational overhead and maximise resource utilisation

Current State Well-developed, automated tools enable workflow management for complex tasks including submission, monitoring, accounting of workloads, and their data management. Integrating these management systems into HPC centres has required careful tailoring to match the computational resources, configuration, and policies. Standard mechanisms for integrating HPC centres into workload management systems are required to maximize HPC adoption. This has been tackled by european projects such as EXTRACT project [88]

Community Evidence The analysis of the interoperable access policies (D5.2) and the technical blueprint for compute and data continuum (D6.1) prepared by the SPECTRUM project identifies following gaps related to workflow orchestration:

- **#5: Workflow execution and orchestration:** Workflow engines are fragmented across communities. Cross-facility workflows lack resilience and robust, standardised provenance capture.
- **#15: High-level end-user interfaces:** Most e-Infrastructures do not offer high-level, abstracted interfaces. Domain scientists should not need to know system details and interact with brittle application interfaces.
- **#16: Unattended execution of long-running workflows** Ensure reliable and unattended execution of long-running workflows and user-operated services.

Technical Foundation Workload Management relies on the following supporting SPECTRUM Compute and Data Continuum capabilities (see D6.1 Section 4.4): Workflow Management, Resource Orchestration, Task Scheduling, Provenance

European Alignment Workflow management is crucial to broader European strategies for sustaining and expanding innovative approaches such as AI-based workflows, and mixed workloads that combine embarrassingly parallel tasks with tightly coupled processes, such as those used for the application of AI/ML. Resource providers will benefit from automated workflow management that makes efficient usage of their resources and their exposure to user communities. The EOSC web of fair data and service benefits from a unified workflow management that is independent of the implementation details and that enables seamless access for European researchers to compute and data resources.

Target Outcomes and Impact Assessment

Strategic Outcomes

- Workflow orchestration evolves into a cross-facility capability with built-in resilience and capturing of provenance.
- Workflow scheduling and orchestration supports interactive compute use cases
- High-level general and domain-specific user interfaces are deployed to lower the threshold to a broader research community.

Success Indicators

- Number of use cases/domains supported by a standardised workflow execution environment (to be

tracked per environment and framework)
- Number of resource providers enabling cross-facility capabilities
- Number of supported use cases and associated number of users

European Impact Standardised workflow management enables researchers to access compute resources and execute their workflows efficiently and independently on the implementation details, thus decreasing the time-to-completion of their research projects and boosting the competitiveness of the European Research Area.

Implementation Pathway and Stakeholder Roles

Implementation Phases

- **Short-term (Years 1-3).**
    - Continuation of the gap analysis that started with the landscape analysis of SPECTRUM for a selection of use cases.
    - Identification of critical interfaces with 1) research infrastructures, 2) e-Infrastructures, 3) (general) user interfaces/API's.
    - Identification and continuation of work for viable orchestration & workflow management frameworks for supporting cross-facility & cross-domain scientific workflows.
    - Pilot one or more use cases (HEP, SRCnet,LOFAR).
- **Medium-term (Years 3-5)**
    - Prototyping & demonstration of selected use cases with selected e-Infrastructure providers.
    - Validation of being fit-for-purpose.
    - Exploration of scope of use cases that can/should be supported (possibly resulting in multiple classes of cases, each requiring a specific solution.
    - Drafting standards and practices that (each) stakeholder must adhere to (e.g. Blueprint Architecture).
- **Long-term (Years 5+)**
    - Roll-out across European e-Infrastructures, grow communities & research infrastructures that build on/benefit.

Stakeholder Roles

- **Scientific Communities, Research Infrastructures, Computing & Data Service Providers / e-Infrastructures**
    - Responsibilities: Co-design and converge on a limited set of interoperable workflow systems, with support for unattended execution, provenance tracking, and cross-infrastructure orchestration.
    - Benefits: (for scientific communities / research infrastructures) scientific output is accelerated; (for Computing & Data Service Providers / e-Infrastructures) resource utilization is optimised.
- **Scientific Communities, Research Infrastructures, Computing & Data Service Providers / e-Infrastructures**
    - Responsibilities: Extend scheduling/orchestration to support interactive compute use cases.
    - Benefits: (for scientific communities): being able to perform interactive data analysis tasks on a scalable infrastructure, with predictable, resilient behaviour and sufficiently low latency/performance; (for Computing & Data Service Providers / e-Infrastructures) efficient use of owned resources.
- **Computing & Data Service Providers / e-Infrastructures**
    - Responsibilities: support deployment of/integration with domain-specific user interfaces by e.g. providing service hosting and integration APIs.
    - Benefits: diverse use of infrastructure; (for users) the high-level part of the scientific workflows can be scaled up to ever-increasing dataset dimensions

Resource Needs

- Compute, network, (temporary) storage resources scaling up through the phases (pilot/demonstration/full deployment).
- Engineering resources for porting/building/demonstrating capabilities (embedded in research infrastructures & e-Infrastructure providers.
- Long-term resources for supporting & maintaining shared capabilities.

Critical Dependencies

- Coordination between research infrastructures & e-Infrastructures towards achieving fit-for-purpose unification of orchestration & workflow frameworks.
- Commitment from e-Infrastructure providers to provide the required unified capabilities through harmonized interfaces.

Cross-priority dependencies with:

- P01 (Cross-Infrastructure Governance),
- P02 (Federated Resource Allocation),
- P03 (Access, Security & Trust),
- P04 (Heterogeneous Computing Integration),
- P05 (Data Management and Transport),
- P08 (Portability - Software Portability)

Pillar 3: Software & Science Enablers

P07 AI

Strategic Rationale and Evidence Base

Strategic Importance: AI methods are reshaping HEP and RA workflows across simulation, reconstruction, calibration, analysis, and instrument operations: pre-trained models at the LHC trigger classify collision events in real time, AI at telescope edges detects astrophysical transients within milliseconds, and digital twins model detector radiation damage, alignment shifts, and telescope-array calibration drift. ML-based fast simulation has demonstrated two-to-three orders of magnitude computational cost reduction for high-impact HEP and RA use cases whilst maintaining scientific accuracy [22]. The European AI moment is open: the AI Continent Action Plan commits over €20 billion to AI infrastructure, AI Factories enter operations in 2025-2026, AI Gigafactories are proposed at €100 million-plus per site, and the AI Act sets the regulatory baseline. With a Commission-estimated European compute capacity gap of around 20 GW by 2036 and three hyperscalers holding roughly 70% of the European cloud market, sovereign-supply development is a load-bearing dimension of the build-out. HEP and RA must be part of this effort or risk being structurally side-lined; GPU resources are not yet accepted as "pledged" allocations for LHC experiments, exposing a structural gap between AI demand and the resource accounting frameworks that fund scientific computing.

AI agents that plan and execute scientific workflow steps under human oversight (analysis triage, anomaly investigation, calibration response) extend AI integration from training and inference into operational decision-making. This requires P03 service-account access, P06 orchestration, and P09 audit and provenance.

Current State: Europe possesses substantial but fragmented AI infrastructure. EuroHPC has procured supercomputers across Europe, including exascale systems, and 19 AI Factories have been selected across three waves (December 2024, March 2025, October 2025) [7]. AI capabilities are not yet integrated as first-class infrastructure across most facilities, which privilege batch-mode access over the elastic processing that production AI workflows require. Access modalities at EuroHPC sites (different AAI stacks, batch-oriented allocation) are not yet aligned with the AI lifecycle. A strategic tension also exists at the hardware layer: industry GPU evolution prioritises low-precision (AI-optimised) computation at the expense of FP64 capability that remains critical for Lattice QCD, parts of HEP detector simulation, and high-precision astrophysical workflows (D6.1 §5.7). Without active engagement at HPC procurement and design stages, future European systems will support either AI workflows or HEP/RA precision needs, not both.

Community Evidence: D6.1 identifies a key gap directly relevant to AI integration [5]: Gap #11: AI/ML integration. AI/ML workloads are increasingly central and are being investigated at every step of data-processing pipelines, from data acquisition to simulation and analysis. Recommendation: enhance services in AI/ML platforms to support distributed training, inference deployment, and access to foundation models, integrated with HPC and domain workflows.

Community surveys reveal that 90% of respondents report software performance improvement as an issue for AI workloads, yet 58% lack the resources to address it [12], quantifying the software investment gap. D6.1 §5.7 sets out the AI-vs-precision tension and three associated actions: identify analysis and simulation tasks suitable for ML acceleration; communicate precision and performance needs to HPC hardware planners; train physicists and astronomers in ML theory, frameworks and tools to fully exploit AI Factories.

Technical Foundation D6.1 Section 4.5 (AI/ML and HPC Applications) defines four capabilities for data-intensive research [5]. P07 builds on each as follows:

- Advanced Analytics and Interactive Exploration. Jupyter, RStudio and Dask-based platforms linked to the data federation that allow seamless scale-out from a researcher's laptop to large external resources. Target: hypothesis-generation workflows that start interactive at PB scale and transition to offline batch without per-site re-engineering.
- AI/ML Platforms. Distributed, GPU-rich training environments spanning HPC, cloud, and edge; scalable inference services; FAIR-compliant model registries that work as searchable "model zoos"

enabling, for example, an astrophysicist to find a transient-classifier and fine-tune it with local data. Federated learning for privacy-preserving training across distributed observatories where bulk data movement is impractical (e.g. RA observatory-level streams; selected HEP collaboration-internal datasets).
- Simulation Frameworks and Digital Twins. Domain simulation (GEANT4 for HEP detectors, N-body and fluid solvers for astrophysics) combined with AI-assisted prediction. Digital twin platforms such as itwinai allow domain experts to use multi-node training and optimisation without dedicated ML engineering. Concrete targets: digital twin of at least one HL-LHC sub-detector for radiation and alignment modelling; digital twin pilot for an SKA precursor station for calibration prediction.
- Large-Scale Data Processing Pipelines. Reconstruction, Monte Carlo production, RA data reduction. AI components integrated into existing pipelines without per-site customisation, running on heterogeneous compute through standardised access protocols (P04).
- Agentic AI. AI agents that plan and execute scientific workflow steps under human oversight, extending AI integration from training and inference into operational decision-making.

European Alignment: Engagement with the European AI policy stack is the strategic backbone of this priority:

- Regulatory. AI Act (Regulation (EU) 2024/1689) sets transparency, documentation, dataset-quality and risk-classification obligations that affect AI used in safety-relevant scientific workflows (detector control, telescope operations), as well as the model-registry and provenance work in this priority.
- Coordination. The AI Office (DG CNECT, 2024) is the central EC coordination point for AI strategy and AI Act enforcement, including scientific-advisory channels.
- Funding. The AI Continent Action Plan (2025) commits over €10 billion for AI Factories and over €20 billion for AI infrastructure overall. The Apply AI strategy and the Cloud and AI Development Act (CADA, proposed 3 June 2026 within the European Technological Sovereignty Package)target a sovereign EU AI and cloud stack.
- Infrastructure. AI Factories (EuroHPC, 19 sites across three waves [7]) carry an explicit obligation under their governing regulation to support the official European Data Spaces. AI Gigafactories proposed at scale capable of training 1 billion-parameter-class foundation models. IPCEI-CIS provides a co-funding pathway for sovereign federated cloud.
- Science. The RAISE programme coordinates AI capabilities for science across the EU via SCIANCE, its AI-in-science landscape-mapping project. The SCIANCE working group for Astronomy and Fundamental Physics provides the route for the RA and HEP communities; substantive engagement and use-case contribution through that working group is the actionable EC route in this priority.
- Data. EOSC enables FAIR-compliant model and dataset sharing across communities [6].
- Strategic frameworks. ESFRI identifies the need for Research Infrastructures to adopt emerging technologies including AI to maintain scientific competitiveness [27].

Target Outcomes and Impact Assessment

Strategic Outcomes:

Outcome 1: AI-Enabled Scientific Workflows. AI capabilities are integrated into scientific workflows as first-class infrastructure rather than add-ons. GPU-rich platforms support distributed training, inference deployment, and access to foundation models across HPC, HTC, and cloud resources. AI surrogate models achieve two to three orders of magnitude computational cost reduction for high-impact HEP and RA use cases, opening research questions previously computationally intractable.

Outcome 2: Sustainable AI Capability. FAIR principles apply not only to training data but also to training processes and models. Federated learning enables model training where data cannot be centralised due to sovereignty, GDPR, or volume constraints. Energy-aware scheduling addresses AI energy consumption whilst building on European Green500 leadership.

Outcome 3: HEP and RA Strategically Positioned in the EU AI Programme. HEP and RA hold formal representation in the RAISE AI for Science pillar. HEP and RA precision requirements are incorporated in AI

Factory and AI Gigafactory procurement specifications, preserving FP64 capability alongside AI-optimised acceleration.

Success Indicators:

- Resource accounting reform. GPU and AI-Factory allocations to HEP and RA experiments accepted as pledged resources within the WLCG and SRCNet accounting frameworks within the mid-term horizon.
- Operational AI workloads on AI Factories. A representative subset of European AI Factories with operational HEP or RA pilot workloads.
- FAIR AI lifecycle. Training data, processes, and models documented and preserved following FAIR principles; shared model registries operational with measurable cross-community reuse.
- Precision-capable AI infrastructure. Lattice QCD (or equivalent FP64 reference workload) sustained at acceptable performance on at least one AI-Factory-class system.
- Programme participation. HEP and RA representation secured in the AI for Science pillar of the RAISE programme; recurring engagement with the AI Office on scientific AI compliance.
- Workflow integration. AI methods integrated into scientific workflows through common orchestration frameworks (P06).
- Energy efficiency. AI workloads scheduled with energy awareness; reporting integrated with P12 sustainability accounting
- Agentic AI workflows. Agent-mediated scientific workflow execution (analysis triage, anomaly investigation, calibration response) operational across HEP and RA collaborations under a documented oversight regime; agent activity audited within WLCG and SRCNet operational reporting.

European Impact: AI integration is a strategic contribution to European digital sovereignty: indigenous AI capability for scientific research reduces dependence on external providers, and the FAIR-compliant model and dataset infrastructure provides a sovereign alternative to commercial model marketplaces. Engagement with the EU AI policy stack positions HEP and RA as evidence-base contributors rather than peripheral users. Workforce uplift through AI training builds capacity that flows back into European industry. Coupled with sustained FP64 capability in procurement specifications, the priority safeguards the long-term feasibility of precision-dependent science within the AI-led infrastructure transition.

Implementation Pathway and Stakeholder Roles

Implementation Phases:

- In the short-term:
    - Formal engagement with the RAISE programme to secure HEP and RA representation in the AI for Science pillar. Identify research-community interlocutors and a governance route.
    - Per-experiment assessment of which analysis and simulation tasks are suitable for ML acceleration, and which require sustained FP64 (per D6.1 §5.7 Action 1).
    - Pilot validation of AI platforms, federated storage and orchestration on a small set of AI Factory sites (P04, P05).
    - Communicate FP64 precision and performance needs to EuroHPC procurement bodies so future procurements specify balanced hardware mixes (per D6.1 §5.7 Action 2).
    - Pilot federated-learning deployment for one cross-observatory RA use case and one HEP collaboration-internal use case.
    - Engage the AI Office on AI Act application to scientific AI (model documentation, dataset transparency, risk classification of operational AI, including the agent-class profile).
    - Define oversight framework for agentic AI in scientific workflows: autonomy scope, approval requirements, audit requirements, interfaces with P03 service-account access and P09 provenance.
    - Pilot one HEP analysis-triage agent and one RA transient-follow-up agent under bounded oversight: P03 service-account access, P06 orchestration, P09 audit and provenance.
- Mid-term:

  - Integrating cloud and HPC infrastructures, with the cloud providing a unified access model, can lower the entry barrier for scientists and accelerate the adoption of larger-scale AI workloads across HEP and research infrastructures. This approach leverages cloud interoperability while providing access to HPC-grade compute for the most demanding workloads.
  - GPU resources accepted as pledged within WLCG and SRCNet accounting frameworks.
  - Agent-mediated workflow execution operational across multiple HEP and RA workflows; oversight framework documented and adopted by participating infrastructures.
- In the long term:
  - Sustained, interoperable European continuum with production-grade federated AI/ML environments, automated provenance tracking, formalized co-design and governance models, and embedded AI/ML as strategic capabilities underpinning Europe's exascale science leadership.
  - Agentic AI as a routine capability for HEP and RA; controls, audit trails, and incident reporting integrated with WLCG and SRCNet operational tooling; AI Act compliance baseline applied to scientific AI agents.

These phases rely on standardization, co-design with scientific communities, advanced resource orchestration, sustainable policies, and a skilled workforce.

Stakeholder Roles

Scientific Communities

- Responsibilities: Identify workloads suitable for AI acceleration and those requiring sustained FP64; produce workload-mix specifications as input to HPC and AI Factory procurement; drive FAIR adoption for models and training processes; engage with RAISE governance via designated SPECTRUM interlocutors.
- Benefits: AI-enabled workflows opening intractable research questions; secured AI Factory and Gigafactory access; recognition of GPU contributions in accounting frameworks.

Research Infrastructures

- Responsibilities: contribute HEP and RA use cases and requirements to the SCIANCE Astronomy and Fundamental Physics working group; coordinate federated-learning pilots across member sites; provide foundational datasets and instrument-edge integration points for AI deployment.
- Benefits: improved utilisation; AI-assisted detector calibration, radiation and alignment modelling, and transient detection at lower marginal cost.

Computing & Data Service Providers / e-Infrastructures

- Responsibilities: integrate AI platform capability across the continuum (EuroHPC supercomputers, AI Factories, national HPC, EGI-federated grid and cloud sites, WLCG, SRCNet); reconcile AAI across community AAIs and AI Factory access; provide elastic allocation modes for AI workloads alongside batch; implement model-registry hosting and federation; coordinate energy-aware scheduling.
- Benefits: standardised interfaces, improved utilisation, reduced operating costs through shared FAIR model infrastructure, and positioning of EGI/WLCG/SRCNet alongside AI Factories as the operational substrate for scientific AI.

Policy Makers / Funding Bodies

- Responsibilities**:** steer coherent EU and national investment across the AI continuum, covering AI Factories, EuroHPC and national facilities, horizontal HTC/Cloud federations like EGI or community federations like WLCG and SRCNet; include HEP and RA precision needs in procurement and provide multi-year compute allocations for science (P02); sustain HEP and RA contribution to SCIANCE; ensure cross-DG coordination (DG RTD, DG CNECT, DG GROW).
- Benefits**:** coherent return on coordinated EU and national AI investment; visible scientific impact across European AI policy and the SCIANCE programme; sovereign capability for high-value science; alignment of AI Act objectives with operational scientific practice.

Resource Needs

- Compute. Annual GPU-hours across the continuum (AI Factories for training, national HPC and EGI-federated grid sites for inference and analysis, WLCG and SRCNet for community workflows), with sustained FP64 capability allocation alongside AI-optimised allocation.
- Storage. Training-data staging capacity and model-registry storage.
- Workforce. ML engineers and ML-physicist hybrid roles within RIs and infrastructure providers, aligned with the P13 framework.

- Software development. Integration of AI capabilities into existing reconstruction, simulation, and analysis pipelines, aligned with P08 and P09.
- Engagement. Sustained coordination capacity for SCIANCE working-group, AI Office and AI Factory engagement.

Critical Dependencies

- SRIDA priorities: P02 (multi-year AI compute allocations), P03 (cross-AAI and service-account access for agentic workflows), P04 (HPC/HTC/cloud integration), P05 (training-data movement and federated-learning access), P06 (orchestration of AI and agentic execution), P08-P10 (FAIR model lifecycle, reproducibility, audit and provenance), P12 (energy-aware scheduling), P13 (ML and hybrid-role capacity).
- Community and infrastructure: WLCG and SRCNet accounting frameworks accepting GPU and AI-Factory allocations as pledged resources; EGI Federation onboarding AI workload patterns (elastic allocation, model-registry hosting); per-experiment HEP and RA workload assessment for ML suitability versus FP64.
- Policy and timeline: AI Act enforcement (AI Office and Member State authorities, including agent-class profile); AI Factory rollout (EuroHPC, 19 sites scheduled); AI Gigafactory selection (proposed, outcome not yet decided); SCIANCE working-group structure and RAISE governance evolution.
- Risk: if the post-2027 EuroHPC procurement cycle drops FP64 in favour of AI-optimised throughput, the precision-dependent fraction of HEP and RA computing loses its European home.

P08 Software Portability

Strategic Rationale and Evidence Base

Strategic Importance SPECTRUM envisions a European Compute and Data Continuum, which will enable research infrastructures and researchers to access and use the most appropriate data centre for any given task in hand. For example, the choice of centre may depend on data locality, connectivity, availability of specialist hardware, or simply on the distribution of load across the continuum. Users will benefit from seamless access to the most effective and efficient e-Infrastructures to address their needs.

Taking advantage of these capabilities requires that the software tools and workflows required by the user are available on the appropriate e-Infrastructure. This requires that they be *portable* across centres: tools developed in one centre should run without modification in other centres, and, where possible, should automatically take advantage of the special capabilities each centre offers, such as accelerator hardware.

In this way, the SPECTRUM Compute and Data Continuum will *maximise the efficiency* of the use of compute resources while simultaneously *eliminating the overhead* imposed on researchers and research infrastructures by a need to modify software depending on locality and *improving reproducibility and sustainability* of the software by decoupling it from specific execution environments.

Current State We identify two essential aspects of the software portability ecosystem: whether the code itself is engineered in such a way as to make it easily portable, and whether there exist software delivery mechanisms that make it possible to deploy, manage and execute the software on multiple platforms. We discuss these aspects independently.

With respect to the first aspect, software engineering, we note that most contemporary HEP and RA software packages are based on large and highly-specialised codebases that have been developed over many years for x86_64 CPU architecture as a primary target. Software is typically optimized at compile-time to the specific characteristics of the system on which it will be executed, and is dependent on specific features of the execution environment, such as system libraries or configuration files, and portability has not been prioritised. In many cases, work is now underway to adapt existing code to accelerator hardware, notably GPUs, to take advantage of their superior energy efficiency and computational throughput. While this work is a step towards portability, it often targets specific hardware environments that are not generically available.

Similarly, although individual compute and data centres generally provide ways, such as modules or containers, for users to configure their execution environment and choose from available software, there is little consistency across the ecosystem. The variety of different approaches and technologies easily leads to brittle environments, poor software portability, and concomitant confusion.

However, we note that this situation is beginning to change. The scientific community is maturing in its approach to software development. In particular, we note the success of projects such as EVERSE [65] that are promulgating best practices in research software engineering across the European scientific landscape, while the FAIR Principles for Research Software (**FAIR4RS** [89]) provide a roadmap for continued improvement.

At the same time, a range of tools that can help are becoming available. These range from low-level abstractions over HPC hardware (e.g. Kokkos [1214]) through containerized systems that encapsulate software together with its dependencies (e.g. Apptainer [18]) to systems like **CernVM-FS** [36] and **EESSI** [90] that can distribute and manage complex software environments.

We therefore identify the coming years as a critical period: the principles and tools that can address the current challenges are available, but coordinated, focused effort across the scientific and technical community is necessary to fully take advantage of them.

Community Evidence The SPECTRUM Technical Blueprint specifically identifies portability as a key requirement (D6.1 Section 2.4 recommendation 4). This requirement is derived from **D5.3** Requirements 5 ("Enable portability of compute tasks across accelerated compute platforms") and 7 ("Establish common SW stacks for compute applications").

Technical Foundation This priority is captured in the following SPECTRUM Compute and Data Continuum capabilities (D6.1 Section 4):

- Software & Execution;
- Compute Continuum.

European Alignment Software portability is critical to the effective use of HPC/HTC resources at European scale. For example, EESSI is supported through the EuroHPC Centre-of-Excellence MultiXscale [137], and is used by the European Federation Platform [79] as the basis for its Federated Software Catalogue [90]. Meanwhile, EOSC specifically supports the deployment of containerized applications through its Cloud Container Platform [38].

The issues surrounding software engineering quality and its impact on portability are also widely acknowledged through major European projects, including, for example, EVERSE [65] and EXTRACT [87], while the ESCAPE Project's **Open-source Scientific software and Service Repository** [70] demonstrated the potential of centralized software distribution systems.

Target Outcomes and Impact Assessment

Strategic Outcomes

- Flexible transfer of software and workflows between data and compute centres with minimal adaptation.
- Minimal implementation effort for software to take advantage of a wide range of compute accelerators.
- Increases overall efficiency of the Compute and Data Continuum and reduces implementation overhead for software authors.

Success Indicators

- Number of standardized software repositories.
- Existence of commonly agreed and applied standards enabling portability of software across multiple data centres.
- Number of codes published making use of hardware abstraction layers like Kokkos.
- Number of software packages made available through tools like CernVM-FS and EESSI.
- Number of data centres on which commonly used HEP and RA processing software can be executed without modification.
- Number of software packages following FAIR4RS guidelines.
- Number of ESFRIs and other research infrastructures including software portability in their coding standards or developer guides.

European Impact Increased homogeneity and portability over different compute systems and architectures increases the overall efficiency of the scientific process, at every level from code implementation to workflow execution. The development of common standards here will spread to other domains, leading to increased reuse and efficiency across the European scientific ecosystem

Implementation Pathway and Stakeholder Roles

Implementation Phases

Short term

- Gap analysis of existing tools and systems supporting portability.
- Establish and document common software stacks for major research infrastructures.
- Increase adoption of well established technologies such as containerization for software distribution.

Medium term

- Based on the gap analysis, ensure that adequate tooling is available to support appropriate

abstractions over hardware at all levels.
- Document commonly-accepted standards for portability over the SPECTRUM Compute and Data Continuum.
- Publish software stacks through common software repositories.
- Provide training to researchers and software engineers on adopting the new technologies.
- Establish a governance model in support of sustainable long-term support for portable software practices by research infrastructures and computing & data service providers.

Long term

- Require and expect that all appropriate software is published through the established standards to the relevant repositories.
- Ensure a process is in place for modernizing and updating standards as well as the supporting services and capabilities with time.
- Provide long-term, sustained support for development and maintenance of key enabling technologies.

Stakeholder Roles

Scientific Communities

- Responsibilities: Define scientific requirements and use cases. Implement software tools following best practices for portability and reuse. Distribute software through standardized repositories.
- Benefits: Reduced overhead in software implementation across multiple e-Infrastructures. Seamless access to compute and accelerator capacity through the Compute and Data Continuum.

Research Infrastructures

- Responsibilities: Incentivize and support the adoption and development and adoption of portable software. Provide training. Invest in appropriate support infrastructures such as software repositories.
- Benefits: Able to make use of the federated compute infrastructure for efficient processing of data from the RI. Enhanced scientific impact.

Computing & Data Service Providers / e-Infrastructures

- Responsibilities: Facilitate the development of common standards and tools for software portability. Ensure their service offering complies with relevant standards and supports standard environments.
- Benefits: A common standard for both interfacing with community software and exchanging software with peer facilities reduces support overhead and increases efficiency, ultimately leading to reduced costs. Make the e-Infrastructure more accessible for end users.

Policy Makers / Funding Bodies

- Responsibilities**:** Define and incentivize strategic goals that recognize the importance of software portability. Ensure the availability of structural funding for key enabling technologies.
- Benefits: Increased impact of resources spent on computing due to improved efficiency. More transparent scientific process due to benefits for reusability and reproducibility.

Resource Needs: Funding for the development of key standards and conventions as well as the necessary supporting and enabling technologies. Infrastructure for establishing software repositories and building and storing software and other artefacts. Development resources where the gap analysis shows existing tools to be inadequate.

Critical Dependencies

- P04 Heterogeneous Computing Integration
- P09 Open Science and Reproducibility
- P10 Software Preservation and Sustainability
- P11 Community Collaboration and Co-Design
- P13 Skills, Careers and Training

P09 Open Science and Reproducibility

Strategic Rationale and Evidence Base

Strategic Importance Open Science and reproducibility are fundamental pillars of modern scientific practice, ensuring that results can be independently verified, reused, and built upon across domains. As European scientific communities move toward exascale data volumes and increasingly complex AI-enhanced workflows, the ability to preserve, share, and reproduce results becomes both more challenging and more essential. In an environment where scientific discovery is data-driven, collaboration-intensive, and reliant on heterogeneous infrastructures, reproducibility emerges as a prerequisite for trust, transparency. Evidence from major SPECTRUM research communities shows that only infrastructures with FAIR-by-design principles, robust provenance tracking, and transparent workflow environments can support long-term scientific value and compliance with open science mandates. For SPECTRUM, Open Science and reproducibility are core enabling conditions for a truly interoperable, cross-domain compute and data continuum. They ensure that scientific knowledge generated in Europe remains accessible, verifiable, and re-usable across communities and generations of infrastructure and crucial way to ensure European sovereignty for research outputs including data, research software and workflows

Current State While many European e-Infrastructures have adopted elements of FAIR data management and workflow packaging, the landscape remains fragmented. Metadata standards vary significantly by community, workflow environments are often tied to specific compute centres, and reproducibility across heterogeneous HPC/HTC/Cloud platforms is not consistently guaranteed. Storage and software preservation practices differ strongly between national and thematic infrastructures and cross-domain interoperability remains limited.

Community Evidence Findings from the SPECTRUM Technical Blueprint, particularly Section 2.5 (software stacks, workflow systems, portability, and reproducibility gaps), reinforce that reproducibility challenges are widespread. Key gaps highlighted by communities include: inconsistent metadata and provenance capture across infrastructures, lack of standardised workflow systems supported by multiple HPC/HTC centres; divergent software stacks and insufficiently portable execution environments; limited long-term preservation of workflows, containers, and software artefacts; absence of domain-agnostic interfaces enabling reproducible execution on heterogeneous architectures.

Technical Foundation This priority is captured in the following SPECTRUM Compute and Data Continuum Capability map (D6.1 Section 4.3 Software Distribution and Execution; Section 4.2 Data Resources)

European Alignment Promotion of Open Science best practices stands at the core of EOSC mission. OpenAIRE provides the research graph connecting publications, datasets, and software with persistent identifiers, enabling discovery and citation tracking across European research outputs. EuroHPC JU is currently developing a federation platform aiming to bring together the diverse computational and data analysis capabilities, advanced software environments, and high-performance storage solutions available on the hosting sites. ESFRI roadmaps highlight long-term data preservation and reproducibility as essential infrastructure requirements for large-scale research facilities. EU Digital Decade prioritises trustworthy digital infrastructures, transparency, and secure data spaces, all of which depend on reproducible, open, verifiable scientific workflows. There are currently several European initiatives such as EVERSE [86], OSCARS [141] directly working towards consolidating Open Science best practices across European Research Infrastructures and Scientific Communities. A number of open data alliances have been introducing open community standards (e.g., the IVOA (international virtual observatory alliance [112]), the IHDEA (international heliophysics data environment alliance [110]), or the IPDA (international planetary data alliance) [113].

Target Outcomes and Impact Assessment

Strategic Outcomes Ensure scientific workflows and results remain reproducible for many years through preservation of complete computational context. This includes standardised metadata, provenance tracking, and FAIR-compliant practices for software and workflows alongside data.

Success Indicators Qualitative measures showing progress toward outcomes
- Increased adoption of FAIR-aligned data services across communities.
- Increased transparency and verifiability of scientific publications.
- Growth in cross-border, cross-domain scientific collaborations enabled by reproducible environments.
- Reduced operational overhead for infrastructure providers.
- Stronger alignment with European Open Science policy frameworks.

European Impact A Europe-wide adoption of reproducibility practices enhances scientific credibility, supports evidence-based policy making, and accelerates joint scientific innovation. It positions Europe as a global leader in trustworthy, transparent, and FAIR science and contributes to strategic objectives under the EOSC, EuroHPC JU, ESFRI, and Digital Decade initiatives as explained in the European alignment.

Implementation Pathway and Stakeholder Roles

Implementation Phases

Short-term (1-3 years)
- Consolidation of guidelines and best practices for implementation of FAIR and reproducible workflows under Open Science
- Implementation specific tools to increase cross-domain collaboration such as the establishment and provision of execution environments streamlining the whole data cycle from findability, analysis, reproducibility and provenance by harmonising meta-data standards, workflow- systems and software stacks across european infrastructures.
- Training & upskilling all stakeholders on Open Science and FAIR paradigms
- Establishment of Open Science-based career recognition best practices and policies

Medium-term (3-5 years)
- Consolidation of aligned interoperability standards between major EU initiatives (EOSC, EuroHPC JU, ESFRI)
- Consolidation of FAIR-by-design services available across borders and across scientific disciplines
- Develop plans for both data & software & results preservation & reproducibility as an integral part of the preparation of scientific research activities & programmes and promoting the reuse of already existing community standards
- Promotion of Open Science part of educational curricula in university programmes

long-term actions (5+ years)
- Enforcement of FAIR standard policies for publicly funded research
- Provision of long-term funding for data and software preservation across scientific disciplines

Stakeholder Roles

Scientific Communities and Research Infrastructures
- Responsibilities: Implement FAIR and reproducible workflow guidelines following Open Science best practices, This includes contributing with domain-specific metadata, standards, and workflow definitions and sharing the best practices across scientific communities, the provision of domain-specific FAIR repositories with provenance tracking that ensure long-term curation of data, software artefacts and metadata.
- Benefits: Increased trust in results, easier collaboration, reduced technical debt, enhanced global visibility. Stronger integration between European infrastructures and enablers for long-term sustainability.

Computing & Data Service Providers / e-Infrastructures
- Responsibilities: Provide execution environments for software and containers that enable reproducibility and provenance capture. This includes the provision of common workflow engines and the integration of reproducibility checks into compute environments.

- Benefits: Reduced operational overhead, simplified user support, increased cross-site interoperability.

Policy Makers / Funding Bodies

- Responsibilities**:** Provide long-term funding for data and software preservation. Require and enforce the use of FAIR standards in publicly funded research Support governance frameworks for Open Science and reproducibility standards. and align national and European priorities and ensure regulatory compliance.
- Benefits: Greater return on investment and increased impact of scientific outputs from publicly funded scientific projects is not hindered due to lack of FAIRness. Increase transparency in the scientific process due to the increased reusability and reproducibility.

Resource Needs

- Updated compute and data infrastructure able to store FAIR scientific data and provide the execution environments
- Manpower including infrastructure operators, software developers to develop and maintain the SW workflows, data stewards / specialists to support the data curation across the whole data-cycle,

Critical Dependencies

Cross-priority dependencies:

- P04- Heterogeneous computing integration,
- P05-Data Management and Transport,
- P08-Software Portability,
- P10-Software Preservation and Sustainability,
- P13 Skills, Careers and Training

P10 Software Preservation and Sustainability

Strategic Rationale and Evidence Base

Strategic Importance: Software is as essential as data in enabling scientific discovery of data-intensive and compute-intensive domains such as HEP and RA. Software environments are becoming increasingly complex, tightly coupled to specific hardware, and subject to rapid technological change. Without long-term preservation strategies, scientific results cannot be reproduced, validated, or reused. SPECTRUM's vision of a federated, cross-domain compute and data continuum depends on stable, portable, and durable software artefacts. Software preservation ensures continuity across successive generations of HPC/HTC/Cloud architectures and enables reproducible science at scale. It reduces technical debt, enhances operational efficiency, and supports scientific autonomy by maintaining Europe's ability to re-run, verify, and extend past analyses

Current State: While some communities (e.g., HEP) have mature practices for preserving code, configurations, and containerised environments, the European landscape remains fragmented. This maturity difference is also reflected when comparing WLCG infrastructure for HEP, which has been already operating for years and SRCnet which is being set up and not yet in operation at the time of writing this document. Many research infrastructures lack systematic approaches for archiving workflows, capturing software dependencies, or ensuring long-term accessibility of container images and modules. Preservation responsibilities are often decentralised, unfunded, or reliant on short-term projects. There is limited standardisation across domains, and current practices do not always support the transition to heterogeneous, accelerated computing environments or new data spaces.

Community Evidence: The SPECTRUM Technical Blueprint, particularly Section 2.5, highlights wide disparities in software stack management across e-Infrastructures; urgent community requirements for common, portable, persistent software environments; gaps in workflow preservation, container longevity, and cross-architecture compatibility; inconsistent support for portable build systems and container runtimes; lack of long-term archival infrastructure for software artefacts; and risks associated with replacing or decommissioning HPC systems without preservation strategies. This has been highlighted by ESFRI related Research Infrastructures and there is an expectation to get those needs supported by compute and data infrastructure providers. On the user side the current legacy know-how spanning from the HEP community in collaborating from many years is providing pathways and learning for the set up of the systems for RA.

Technical Foundation : This priority is captured in the following SPECTRUM Compute and Data Continuum building blocks (D6.1 Section 4): Software & Execution; Compute Continuum.

European Alignment: Software presentation is part of the policy objectives for promoting FAIRness of scientific outputs being pursued under EOSC. EuroHPC needs to ensure portability and continuity across sites and successive upgrades of the supercomputing facilities. ESFRI priorities also focus on long-term sustainability, reproducibility, and cross-RI interoperability. This priority is strongly aligned and supported by the work done under the EVERSE project which can bring solutions and recommendations towards faster achievement of priority outcomes. Advocacy groups such as DORA [56] or ADORE from Research Software Alliance [8] or ESCAPE ESFRI Cluster have also been actively promoting solutions on SW sustainability

Target Outcomes and Impact Assessment

Strategic Outcomes: A sustainable, Europe-wide ecosystem in which scientific software, workflows, execution environments, key enabling SW technologies and containerised stacks are archived, portable, and reusable across generations of HPC/HTC/Cloud systems.

Success Indicators: Qualitative measures showing progress toward outcomes
- Major HPC/HTC/Cloud centres adopt harmonised software preservation policies, including documentation as integral part for of preservation policies / SW management plan
- Community-standard container images and software stacks remain executable across hardware generations.
- Increased reuse of preserved software and workflows across projects and infrastructures.
- Reduced effort required to port legacy codes to new architectures.

European Impact: Software preservation enhances Europe's scientific autonomy, reduces risk during technology refresh cycles, and lowers long-term operational costs.

Implementation Pathway and Stakeholder Roles

Implementation Phases:

Short-term (1-3 years)
- Development and consolidation of guidelines, best practices and policies for research software development that include software preservation to be included as integral part part of SW management plans and related documentation for organizations and initiatives
- Promotion of open source/open standards policies to prevent sovereignty issues and avoid proprietary vendor lock-in
- Establishment and provision of execution environments offering harmonising software stacks across European infrastructures.
- Deploying knowledge hub - in which know-how of different clusters is being placed including harmonized training programs on research software best practices across communities

Medium-term (3-5 years)
- Consolidation of standardized community-based container images and software stacks that can remain executable across next generations of heterogeneous hardware systems.
- Develop plans for software sustainability as an integral part of the preparation for scientific research programmes (including simulations, data analysis) and of the preparation of major experiments in particular
- Training programs on research software best practices integrated into european university curricula

long-term actions (5+ years)
- Ensure the provision of long-term funding is kept for software preservation across scientific disciplines
- Career paths for research software engineers clearly defined, recognised under standard tools like ORCID.

Stakeholder Roles: Stakeholder group responsibilities and benefits

Scientific Communities
Responsibilities: Define domain-specific preservation needs, share best practices, adopt common containers and workflow formats.
Benefits: Long-term stability, reduced porting effort, reproducible science, stronger global collaboration.

Research Infrastructures (ESFRI RIs)
Responsibilities: Provide domain repositories, guarantee metadata quality, integrate software and workflows into preservation systems.
Benefits: Improved sustainability, FAIR compliance, interoperability across Europe.

Computing & Data Service Providers / e-Infrastructures
Responsibilities: Maintain harmonised software stacks, build and preserve container environments, offer long-term archival services.
Benefits: Reduced user support burden, stable environments, easier migration to new HW.

Policy Makers & Funding Bodies
Responsibilities: Provide long-term funding, define sustainability mandates, ensure cross-programme alignment (EOSC, EuroHPC, National).
Benefits: Greater return on investment by fostering reusability of scientific software. Increasing European competitiveness,

Resource Needs: Investment dimensions, intensity levels, and funding alignment
- Heterogeneous compute infrastructure with fully updated software stacks able to provide proper execution environments to keep research software operational over generations.
- Manpower including infrastructure operators, software developers to operate and maintain software stack and to develop and maintain the research software

Critical Dependencies: Key prerequisites and coordination requirements

- Cross-priority dependencies:
    - P04- Heterogeneous computing integration,
    - P08-Software Portability,
    - P09-Open Science and Reproducibility,
    - P13 Skills, Careers and Training

Pillar 4: Human Capital & Responsibility

P11 Community Collaboration and Co-Design

Strategic Rationale and Evidence Base

Strategic Importance

Community management is strategically important because it enables HEP, RA and other data-intensive communities to align around shared priorities, move from fragmented initiatives to a coherent roadmap, and support innovation, interoperability and long-term sustainability in the European compute and data continuum. Building on the SPECTRUM CoPe and existing networking structures, it must go beyond coordination alone by also addressing human-capacity needs, including stable roles, recognised responsibilities, organisational incentives, and retention of key engineering and software expertise. It should also ensure continuity beyond individual projects and maintain structured links with international partners, as major infrastructures operate in a global ecosystem.

Current State

The current landscape offers strong potential for closer convergence across data-intensive domains, but this has not yet been fully translated into coordinated action. Beyond technical integration, the main challenges lie in aligning priorities, collaboration models and governance across diverse communities, while also making European opportunities easier to understand and engage. A more structured coordination mechanism is therefore needed to connect communities, map EU opportunities, and support coherent engagement with international partners while preserving domain-specific needs.

Community Evidence

Key gaps relate not only to technology, but also to coordination and governance. These include: limited long-term resource planning aligned with community needs; insufficiently structured feedback loops between user communities, research infrastructures,service providers and policy stakeholders (such as EOSC and EuroHPC); insufficient involvement of all relevant stakeholders in strategic discussions and decision-making; and the need to better connect shared priorities with domain-specific requirements. Consultation feedback also underlines the importance of maintaining explicit links with European and international initiatives so that collaboration leads to operational convergence rather than parallel or fragmented efforts.

Technical Foundation

The increasing data and computing demands of HEP and RA create a strong basis for deeper collaboration across scientific communities, research infrastructures and e-Infrastructures. Although the two domains differ in maturity and implementation pathways, they increasingly require a common set of capabilities, including sustainability, security, federation, orchestration, distributed software, and support across the compute and data continuum. In this context, the SPECTRUM Technical Blueprint provides not only a technical reference point, but also a shared framework for dialogue, coordination and joint planning across communities working with heterogeneous infrastructures.

European Alignment

The scientific use cases of the SPECTRUM project (WP5) are rooted in major ESFRI projects in HEP and RA, notably SKAO and HL-LHC, as well as ERICs such as LOFAR and JIVE. In this context, community management has an important role in connecting these communities with the broader European landscape, including key initiatives such as EOSC and EuroHPC, as well as related programmes and networks (e.g. APPEC, ASTRONET, RadioNet, Opticon, Europlanet, Solarnet, JENA, ESCAPE). It is also important to address the still not fully articulated alignment between EOSC and EuroHPC, which remains a source of complexity for communities seeking coherent European support across data and compute services. Strengthening these links through structured dialogue, coordination and mutual visibility is essential to help communities better understand, access and shape the European opportunities relevant to their needs (cf. Strategic

Outcomes section). This also supports closer alignment between research infrastructures, e-Infrastructures and scientific communities, while reinforcing Europe's capacity to act coherently in the international data-intensive science landscape.

Target Outcomes and Impact Assessment

Strategic Outcomes Beyond identifying requirements for a sustainable European compute and data continuum to support the scientific exploitation of major European RIs, the effort in community management seeks to build bridges with similar existing and forthcoming multi-disciplinary initiatives, fostering long-term alignment and collaboration. Effective community management depends not only on technical coordination, but on stable human capacity, clear representation, explicit scope, and durable cross-initiative interfaces, since organisational, political, and workforce factors are often as decisive as the underlying technology.

Success Indicators Successful community management can be identified through clear indicators such as: sustained participation of diverse stakeholders across the HEP and RA domains; the development of a collaboration path with other scientific communities; the adoption of a shared community charter ensuring long-term sustainability; effective feedback loops that influence technical and policy outcomes; ease of access; and the creation of structured links with related European and international initiatives. Additional indicators include the establishment of stable community-facing roles and retention pathways for key technical expertise, and the production of clear stakeholder mapping describing how communities are represented.

European Impact Addressing the numerical challenges of HEP and RA science use cases is a critical step towards maximising the impact of major ESFRI and ERIC projects in these domains. The SPECTRUM Community of Practice is a central instrument in this effort, providing a visible and structured framework for coordination across scientific communities, research infrastructures and e-Infrastructures. Through this mechanism, the disruptive data and computing demands of infrastructures such as SKA and the HL-LHC can be articulated collectively, connected more effectively with the EOSC and EuroHPC ecosystems, and used to reinforce the bridges between them. In doing so, the SPECTRUM CoP helps create alignment, foster synergies and consolidate Europe's leadership in data-intensive science.

Implementation Pathway and Stakeholder Roles

Implementation Phases

- **Short-term (Years 1-3):** Strengthen and support the established CoP by developing further the charter developed during the SPECTRUM project, ensuring its continuity and long-term sustainability, and, where relevant, aligning or merging activities with existing networks and consortia. In the computing domain, the CoP should also develop shared metrics, tools and methodologies to assess usage, performance, efficiency and best practices, thereby enabling the definition of common indicators, the collection and analysis of operational data, and the exchange of knowledge across data-driven research communities. This shared knowledge base should be considered a key output in its own right.
- **Medium-term (Years 3-5):** Foster interaction, coordination, and potential convergence with similar initiatives operating in related research domains. Medium-term actions should also include formalisation of cross-initiative interfaces, including a maintained inventory of relevant European and international programmes, infrastructures and policy processes, together with named liaison functions and periodic review points.
- **Long-term (Years 5+):** Maintain continuous coordination and strategic alignment across the broader network of scientific communities.

Stakeholder Roles

- **Scientific Communities** – Responsibilities: dedicate adequate resources to identifying technical needs, collaborate with other scientific communities as a well structured CoP, work with e-Infrastructures on technical developments, and designating clearly identified individuals to coordinate these activities – Benefits: establish a clear pathway to solutions for computing and data requirements, while providing a well-defined interlocutor for technical, political, and financial discussions

- **Research Infrastructures –** Responsibilities: empower community-led initiatives by allocating dedicated expertise and fostering operational collaboration across RIs **–** Benefits: unlock the full potential of existing infrastructures, ensuring efficiency, complementarity, and long-term sustainability through coordinated bottom-up actions
- **Computing & Data Service Providers / e-Infrastructures –** Responsibilities: actively engage in evaluating operational and funding schemes in collaboration with RIs and the scientific community **–** Benefits: establish a clear pathway to fulfilling their mission of advancing open science, enabling high-performance computing and data-intensive research, and supporting scientific progress
- **Policy Makers / Funding Bodies –** Responsibilities**:** facilitate coordination between scientific communities, RIs, and Computing & Data Service Providers, supporting joint development activities through dedicated and long term funding schemes **–** Benefits**:** ensure the optimal use of investments, maximising scientific and technological return while minimising inefficiencies and financial waste**.**

Resource Needs: To fulfill the objectives listed above and make community management efficient, scientific communities require a combination of resources: dedicated human expertise to identify needs and liaise with partners; financial support through flexible and sustained fundings; access to infrastructures for co-design and validation. Resource needs also include recognised career paths and retention measures for research software engineers, infrastructure engineers, and community coordinators, as consultation feedback repeatedly identifies workforce instability and competition with industry as a strategic risk to sustainability. Together, these resources enable scientific communities to address computing and data requirements effectively, coordinate across domains, and act as credible interlocutors in technical, political, and financial discussions

Critical Dependencies Effective community management has critical dependencies on long-term human and technical resource commitments and interoperable tools. A continuous dialogue between RIs, e-Infrastructures, and policy makers must be ensured, supported by formal agreements, cross-community working groups, and structured feedback loops. Together, these elements enable alignment with European initiatives like EOSC and EuroHPC, fostering synergies, avoiding duplication, and ensuring sustainable impact. A representative community structure would help establish a clearer and more transparent channel of communication with European institutions, addressing the lack of visibility and coherence that is often perceived today. Critical dependencies also include clarity of mandate and scope, so that participating communities understand whether the collaboration framework is intended for HEP and RA specifically or for a broader cross-domain landscape, and how this common framework accommodates legitimate domain-specific differences.

Cross-priority dependencies: P01 Cross-Infrastructure Governance

P12 Environmental Sustainability

Strategic Rationale and Evidence Base

Strategic Importance: Energy use has recently become a limiting factor for HPC due to the difficulty of procuring MW, cost and cooling issues. It is, however, just the tip of the iceberg as, in the current context of global warming, collapse of biodiversity and gradually exceeding planetary limits, environmental impacts of IT are significant and need to be reduced at the European scale to fulfil the Paris agreement. Conversely, large science projects, in the HEP and RA domains for instance, call for increased use and growth of HPC infrastructures and will face the challenge of doing more with fewer impacts to be environmentally, politically and socially acceptable. This is not simply a matter of improving the efficiency of the processes, although this will help, but of gaining control over the different impacts of IT in order to allow it to fit in schemes to globally reduce mankind's impacts. A holistic approach is required as science moves into the era of big data and because of the exponential adoption of AI to ensure scientific computing can continue responsibly in spite of the growing environmental pressure.

Current State

The energy challenge and GHG emissions of datacentres have been gaining traction and moves towards the use of renewable energy and more efficient cooling have been taken. While this allows for reducing the environmental impacts of the use phase, it does not allow for addressing them globally. Life-cycle assessment methodologies (ISO 14040, ISO 14044, product environmental footprint) and databases, from the cradle to the grave, have been emerging to quantify the impacts and improve the design, fabrication and overall management of the IT resources, as well as avoid the transfer of impacts between phases, impact types or geographical regions. To date, however, tooling and policies to forecast and manage impacts are still in their infancy and coordination is lacking to gain control on the whole and first stall their growth, then reduce them overall. Impacts can be local (e.g. water use) or global (e.g. climate change) and are generated by diverse communities (end users, datacentre managers, hardware suppliers, waste managers etc.) at the national or international scales, which makes them complex to address consistently. In a similar way, preliminary mitigation of negative social impacts is under development with approaches based on the more recent ISO 14075 norm.

Community Evidence

The optimisation paradigm adopted by RA and HEP is centred on weak scaling: the ability to handle larger problems using a larger amount of resources. In this frame, section 2.5 gap #10 in the Technical Blueprint identifies the need to monitor workflows, performance and energy usage across sites to efficiently exploit a multi-infrastructure compute and storage continuum. This focus on efficiency, as it allows for pushing the limits of what is technically feasible, has and will lead to further increases of storage, network and computing resources in contradiction with the imperative to reduce the environmental impacts. This is something known as the rebound effect or Jevon's paradox.

The D6.1 Technical Blueprint identifies a number of gaps relevant to environmental sustainability. Gap #10 in section 2.5 resulting from WP5 analysis relates to monitoring and observability with the lack of consistent, cross-site monitoring of workflows, performance, and energy usage (D6.1). This forms part of the inefficient resource usage gap in section 3. The same section points out that execution is carried out irrespective of environmental impacts which implies that the latter are only determined a posteriori, ie. too late to be avoided or reduced. A third gap in this section corresponds to significant digital waste due to poor data and software reuse and of failing software. The fragility of operations is finally pointed out, which in the light of the increasing pressure on energy and cooling is susceptible of leading to increased waste or the inability to operate due to shortages of resources.

Technical Foundation

Sections 3.1 and 3.2 of the Technical Blueprint propose a path for reconciling weak-scaling optimisation efforts with environmental sustainability through the development of tools allowing to view them through the prism of environmental impacts. These provide a technical basis for sustainability policies to be enforced at the continuum and datacentre levels to address their impacts with a holistic approach. Central to this approach is using environmental impacts as credits allocated to users at all levels. This will make them visible and promote responsible behaviours (as opposed to abstract quantities like core hours). It will also allow for framing their expenses within a global environmental impact envelope defined by higher-level policy-makers so that the environment is indeed protected. It is expected that datacentres will arbitrate between procurement, maintenance and usage impact expenses on this basis. In doing so, the latter will flow down to end users who will then need to optimise and prioritise their uses accordingly. While the tooling is end-user oriented, these policies are intended to be set by datacentres, governing bodies and higher-level policy-makers in order define a large-scale comprehensive framework for the management and use of European IT infrastructures within determined environmental impact envelopes. Guidelines for defining these policies are provided below but precise definitions based on metrics and setting thresholds is beyond the scope of the project as they are the responsibility of datacentres, governing bodies and high-level policy makers, each at his own level.

In the light of the gaps summarised above, the D6.1 Technical Blueprint recommends the development of software-hardware optimisation to optimally use resources and distribute data and compute, tailoring execution to sustainability in terms of forecasting impacts and providing end users with means to map and prioritise their jobs so they can be scheduled in an impact-aware manner, FAIR data and software publication, in direct relation to Priority P05 (Data Management Continuum), fine software quality assessment prior to the use of production-level resources to reduce the probability and consequences of failures and improvements in the IT's adaptation capabilities, notably through the resilience of software execution to varying or failing hardware resources.

This approach is complemented with policies addressing the other phases of the IT lifecycle.

European Alignment

The European Green Deal aims to make Europe the first climate-neutral continent. This is a significant ambition, a top-level objective addressing climate change globally across the union. Recommendations for technical developments in the Technical Blueprint and policies in this document aim to empower policy-makers in this sense and allow for this objective to flow down in quantitative terms to datacentres and end-users.

EuroHPC JU Strategic Commitment: Europe has demonstrated leadership in sustainable HPC through [82]:

- JUPITER JEDI #1 Green500 ranking
- MareNostrum 5 PUE < 1.08 industry-leading efficiency
- SEANERGYS €16.8M EU investment in energy-efficient workload management

EOSC and Broader European Strategies: Environmental sustainability aligns with broader European strategies. The Digital Decade Policy Programme 2030 emphasises sustainable digitalisation [54], and EOSC principles include environmental responsibility for research infrastructures [66].

JENA Federated Computing Vision: JENA Computing White Paper positions environmental sustainability as one of the central challenges for federated computing, requiring transparency, lifecycle optimisation, and cultural shift towards sustainable computing practices [115].

Target Outcomes and Impact Assessment

Strategic Outcomes

1. **Forecast and Reporting of Environmental Impacts**: Operational capability to simulate, forecast and report environmental impacts of computing workloads, enabling informed resource allocation and scheduling decisions.
2. **Improved Software Quality**: Software development practices that prioritise energy efficiency and resource optimisation alongside performance and scientific accuracy [115], contributions to libraries to increase the lifetime of software, help it mature and standardise practices to facilitate co-design.
3. **Improved Efficiency of Execution**: Multi-objective optimisation of computing workloads balancing performance with environmental impact (carbon emissions, energy consumption, resource utilisation) (D6.1).

4. **Lifecycle Extension**: Procurement and operational policies that extend hardware lifetimes, accounting for fabrication impacts (upstream) and end-of-life disposal (downstream). In the European energy mix, fabrication phase outweighs usage phase, making replacement for efficiency gains rarely beneficial (D6.1).

Success Indicators

1. **Environmental Transparency:** Environmental monitoring (impact budgets, PUE/CUE etc.) operational at major European datacentres with public reporting
2. **User Engagement**: Sustainability metrics visible through dashboards to users during job submission
3. **Procurement Integration**: Sustainability metrics (embedded carbon, water usage, repairability, HEPscore/watt etc.) adopted as standard procurement criteria alongside performance metrics
4. **Lifecycle Extension**: Hardware operational lifetimes extended across SPECTRUM infrastructures
5. **Cultural Shift**: Training programmes on sustainable computing practices operational [115], awareness to environmental impacts and ecologically friendly ways of designing and using IT infrastructures through communities of practice
6. **Policy Alignment**: Integrated sustainability policies, accounting and management from the projects to the infrastructures with coordinated and overarching supervision by European HEP/RA institutions and EuroHPC.

European Impact

Europe already leads in green HPC, with JUPITER JEDI ranked #1 on the Green500 and MareNostrum 5 achieving PUE below 1.08 [82]. Reducing GHG emissions from research infrastructures contributes to European climate neutrality by 2050 and ensures continued public support for large-scale investments such as HL-LHC and SKA. Energy efficiency improvements also reduce operational costs, while lifecycle extension provides greater economic and environmental benefit than frequent hardware replacement (D6.1).

Implementation Pathway and Stakeholder Coordination

The Technical Blueprint recommends a number of tools to support the approach described here. Although part of these tools have been the focus of research activities, achieving the reliability and efficiency required for these tools to be used in production, that is for the effective management of IT resources, requires this effort to be intensified on the short term before the tools can be made mature through consolidation based on their use through an experimental stage described as part of the phases below.

Implementation Phases

Given the European scientific and computational ambitions, the Paris agreement and the corresponding European Union policies call for immediate action with the available means and for their extension. The paragraphs below describe the programmatic aspects while additional detail useful to make this actionable is provided in Annex C.3.

- **Short-term (Years 1-3) :** This first phase has two facets. One consists in putting in practice available solutions to achieve a reduction of IT's environmental impacts and provide some relief to the environmental crises in the short term with limited disruption. The other is preparatory to further this approach. Its purpose is to plan, develop and experiment with the solutions required for the change of paradigm fully integrating the consideration of environmental impacts as a constraining aspect for the development, management and use of the data and compute continuum.
- **Medium-term (Years 3-5) :** In the medium term, as a result of the experimentations carried out in the first phase, the different measures should be consolidated and put into wide-scale practice. It will also be time for experimenting with the more complex measures for which the first phase has laid a basis.
- **Long-term (Years 5+) :** Given the urgency of the environmental crises and the expected growth of IT, all of the different proposed measures (or adaptations thereof) should be consolidated and put into practice in this time frame. Given the likely continued rapid evolution of IT, this will also require continuously monitoring their effectiveness and adapting them to the hardware and software systems, as well as to the environmental situation.

Stakeholder Roles

- **Scientific Communities -** Responsibilities: train users on software development for low environmental impacts and mindful use of resources (priorities, experimentation plan) – Benefits: reduction of digital waste, resilient and adaptive execution, improved software quality
- **Research Infrastructures -** Responsibilities: contribute to sustainability governance, contribute to data and software reuse, incorporate IT environmental impacts in the wider assessment of those of the research infrastructure **-** Benefits: improved coordination between user communities, detailed accounting of environmental impacts resulting from computing and storage
- **Computing & Data Service Providers / e-Infrastructures -** Responsibilities: contribute to sustainability governance, manage the overall environmental impacts of infrastructures, allocate end-user shares such impacts as credits, support users and policy makers in tracking and managing them – Benefits: sustainable operations, detailed accounting of environmental impacts
- **Policy Makers / Funding Bodies -** Responsibilities: define environmental impact budget at the scale of the compute and storage continuum, evaluate sustainability governance - Benefits: meet obligations resulting from the Paris agreement, sustainable operations,s detailed accounting of environmental impacts

Resource Needs

Given the importance of embedded impacts resulting from the fabrication of IT equipment, the transition to a more sustainable management of IT infrastructures throughout their life-cycle relies primarily on the involvement of personnel, both in a dedicated way as part of the sustainability governance staff and the development of the underlying tooling and in terms of specialisation in eco-design of infrastructures and software at the datacentre and end-user levels.

The development of the tooling recommended as part of the Technical Blueprint, although it will build on existing solutions, will require a significant amount of effort to be tailored to environmental impacts. This tooling consists mostly in software.

Critical Dependencies

The management of environmental impacts involves actors at all levels of the IT hierarchy and is by nature transverse as it involves all phases of the life cycle, hardware and software. A holistic, systems engineering view is required to avoid excessive disruption as well as the transfer of impacts. For this reason, coordination is seen as a critical dependency for this effort to be successful.

P13 Skills, Careers and Training

Strategic Rationale and Evidence Base

Strategic Importance: HEP and RA research has become computing intensive, at least since three decades. Large infrastructures and large investments are needed to collect, store and analyse the datasets produced. This requires dedicated teams of experts with specific skills and capabilities to design and operate the required computing systems and software stacks; failing that, the expected impact is a reduced physics exploitation capability for the planned billion Euro investments. In many cases critical pieces of software depend on individuals or small groups of people often in non-permanent positions. Together with workforce ageing, migration to industry and lack of recognition for research software engineering work, this poses a risk to research.

Training, supporting and retaining computing experts is of critical importance for current and next generation initiatives, even more so in a scenario where the same capabilities are in great demand from industry. In several cases researchers are lacking training in research software engineering and will receive the necessary training only when joining a research collaboration. Community established training in this respect is even more vital.

Current State: Computing experts (either by formation or trained inside the initiatives) suffer from inadequate career paths. They are seldom offered permanent positions, and those that do rarely have a strong career path. Academic positions in physics are not open to them, and in the best case they are retained as "technical staff", typically on "soft money" (renewed year on year without long term job security). Computing experts also face an attractiveness deficit relative to industry (which offers higher salaries and more stable positions) and structural under-recognition in scientific outputs (rarely first author on the resulting publications, often not credited as authors at all). The work of research software engineers or researchers who code also often lacks recognition of their contributions within the communities which hampers career progression in addition.

Community Evidence See D6.1 Section 2.5 (Consolidated Gaps and Requirements), Gap #13 (Workforce sustainability and development): the problem of "Workforce, funding, and sustainability" is listed as one of the most critical gaps; the highlighted solution consists of the "… development of a coordinated European funding/governance models, in alignment with national contributions, and in the investment in training and career pathways for research software engineers and data stewards". The Physics Briefing Book "Input for the 2026 update of the European Strategy for Particle Physics" [142], Section 12.6, explains the problem from the angle of High Energy Physics. In the conclusions, the problem of career paths is detailed as "… in the long term, this is more worrying than technological trends, and finding solutions should be a field priority".

Technical Foundation: The problem is not strictly technical, but rooted in the historical way in which computing related tasks have been executed in the target physics domains. Up to 3 decades ago, computing was a post-facto solution, designed after data taking and long after the initial design. This was possible due to its (much) reduced complexity with respect to other technical issues (e.g. building and operating a detector). Today, the size and complexity of scientific computing has increased to a level of criticality comparable with other technical aspects, with the overwhelming majority of experts acknowledging this. Nevertheless, research and academic institutions have failed to keep pace and continue to undervalue scientific computing experts, treating them as replaceable. The result is a fast turn-around of less experienced computing collaborators, who leave after contributing critical software to their communities. Worse still is when this happens at a later career stage, where computing contributions are much more significant. Entire systems need to be replaced, without the knowledge and expertise of the original expert. This issue directly affects the sustainability of large scientific initiatives, and their capability to produce scientific results. It also hampers European strategic efforts to find synergies across scientific computing domains, as the few experts that do exist are too busy mitigating these issues. Finally, while it's desirable to have some early career individuals leave academia for industry, for those who leave later in their careers the transition is typically much less efficient for industry and more stressful on the individual.

European Alignment The career problem for science is now recognized in European initiatives, including the EuroHPC JU and the EOSC, and is leading to promoting positions as technical specialists and data stewards. Still, at the academic level, the career paths offered to computing experts (either by formation or by specialization) are less attractive than, for example, in the construction of equipment and in data analysis.

Sovereignty of computing initiatives and skills is an important aspect which needs to be preserved, in a global scenario where major non-European industries drive the technology innovation. The EOSC EVERSE project, as one example, strives to establish best practices in research software engineering underpinned with services to provide a coherent and connected infrastructure for training materials and events as well as providing recognition for research software engineers and trainers.

Target Outcomes and Impact Assessment

Strategic Outcomes Inform and reiterate among research and academic institutions the need and importance of research personnel fluent in large scale computing; prepare and fund appropriate career paths for such experts and provide means for the recognition of research software engineering work. Feedback to universities and engineering schools on training needs in scientific collaborations

Success Indicators Lower total costs for the computing aspects of European scientific initiatives, driven largely by the ability to create longer-term, more sustainable computing plans, both during the design and operations phases. Improved knowledge sharing in computing across scientific domains thanks to the existence of a small cohort of career computing experts providing a stable and strategic vision for European scientific computing. Of course this cannot be realised by SPECTRUM alone, but must come from a domain-level realization of the issue and its possible solutions. With well-defined career paths, smoother transitions between academia and industry mean individuals destined for industry will start their careers earlier.

European Impact A solid cohort of domain scientists with extended computing skills, able to confront the data deluge expected from next generation scientific initiatives. In perspective, since not all of them will be able to remain in the public research domain, a cohort of scientists is moving to the private / productive sector, and participating in its innovation.

Implementation Pathway and Stakeholder Roles

Implementation Phases

- Short-term actions (deliverable under current initiatives: RI Train, HPC Train, EuroHPC Centres of Excellence) raise awareness of academic and research partners to the problem of providing good career paths to computing experts. Medium-term actions (deliverable through new project proposals or integration into university curricula) target structural change in education and recognition.
- Long term actions require a reshaping of the funding and personnel policies in academic and research institutions.

Stakeholder Roles

Scientific Communities
- Responsibilities: clarify their present / future needs in computing, and the needs in expert personpower; clarify the risks attached to their vacancy. Communicate the challenges to be solved, and the possibility to directly influence scientific results.
- Benefits: secure long term computing positions, for the benefit of decades-long initiatives. Increase the potentially interested community.

Research Infrastructures
- Responsibilities: ensure sufficient staffing and smooth operations (for research and beyond).
- Benefits: long-term planning of resources allows to participate in decades-long initiatives, and to have personnel with a deep knowledge in the systems and thus able to react / plan in a more effective way.

Computing & Data Service Providers / e-Infrastructures
- Responsibilities: Provide comprehensive training on the efficient use of heterogeneous resources, covering hardware acceleration (e.g. GPUs), software best practices (e.g. containers, portability), and sustainable operations; offer secondment opportunities for Research Software Engineers to embed within operational centres. Facilitate contacts between the scientific user communities and partners from industry, in order to show further career possibilities.
- Benefits: Improved infrastructure efficiency through optimised user workflows; reduced support

burden; and a user community capable of exploiting advanced platform features.

Policy Makers / Funding Bodies

- Responsibilities**:** fund computing research positions at academia and research institutes. Ensure career paths for researchers designing / operating computing systems.
- Benefits**:** a more predictable landscape in computing allows for a long-range planning of initiatives. A solid scientific computing infrastructure in specific scientific domains allows for easier support for the long tail of research. The realization of a cohort of experts which can partially fuel innovation in the productive system.

Resource Needs

- More funds dedicated to computing experts (by formation, training or inclination) in academic and research institutes. Long-term plans guaranteeing sustained support.

Critical Dependencies

All the other priorities, but especially highlighted for

- P07 AI
- P08 Software Portability
- P09 Open Science and Reproducibility
- P10 Software Preservation and Sustainability
- P11 Community Collaboration and Co-Design

Annex B: Priority Description Template

Each of the thirteen SRIDA priorities follows a standardised template that ensures consistent treatment and enables systematic comparison across priorities. The template organises information into three sections addressing rationale, outcomes, and implementation.

Strategic Rationale and Evidence Base establishes why the priority matters for European research infrastructure. This section identifies the strategic importance, describes the current state and gaps, references community evidence from the Technical Blueprint and other sources, identifies the technical foundation within the Compute and Data Continuum, and demonstrates alignment with broader European strategies.

Target Outcomes and Impact Assessment defines what success looks like through capability-focused outcomes rather than activity metrics. This section presents strategic outcomes as capability statements, identifies qualitative success indicators, and articulates the contribution to European research competitiveness.

Implementation Pathway and Stakeholder Roles shows how the priority will be achieved and who is responsible. This section defines implementation phases across short-term (1-2 years), medium-term (3-5 years), and long-term (5+ years) horizons, identifies stakeholder responsibilities, outlines resource needs, and notes critical dependencies.

Priority Description Template

Strategic Rationale and Evidence Base Establish WHY this priority matters for European research infrastructure

Strategic Importance One-paragraph explanation of why this priority is essential for SPECTRUM's vision and European research competitiveness

Current State Brief description of existing capabilities and identified gaps in European landscape

Community Evidence For technical priorities, reference to the SPECTRUM Technical Blueprint (in particular mention the gaps/requirements in Section 2.5); reference to other relevant documents also possible

Technical Foundation Reference to supporting SPECTRUM Compute and Data Continuum capabilities

European Alignment Connection to broader European strategies (EOSC, EuroHPC, ESFRI, Digital Decade)

Target Outcomes and Impact Assessment Define WHAT success looks like through capability-focused outcomes and realistic progress indicators

Strategic Outcomes Capability statements describing what will be achieved

Success Indicators Qualitative measures showing progress toward outcomes

European Impact Contribution to European research competitiveness and strategic objectives

Implementation Pathway and Stakeholder Roles Show HOW the priority will be achieved and WHO is responsible for implementation

Implementation Phases Short-term, medium-term, and long-term actions

Stakeholder Roles Stakeholder group responsibilities and benefits

Resource Needs Investment dimensions, intensity levels, and funding alignment

Critical Dependencies Key prerequisites and coordination requirements

Annex C: Priority Evidence and Validation

The thirteen strategic priorities emerged from systematic analysis of SPECTRUM deliverables and validation by the Community of Practice and External Advisory Board. This annex documents the evidence base, validation process, and traceability to the Technical Blueprint.

C.1 Desk Research and Evidence Sources

The strategic priorities of the SRIDA emerged from systematic analysis across SPECTRUM Work Packages 3, 5, and 6:

- **D3.1 (SPECTRUM CoP Community of Practice - Early Findings)** describes the main gaps and findings from the work performed by the SPECTRUM CoP in the first period of the SPECTRUM project.
- **D5.1 (Representative Use Cases)** examined fourteen science use cases (mainly from HEP and RA), identifying common requirements for cross-facility workflow orchestration, federated data access, heterogeneous computing support, and automated resource provisioning.
- **D5.2 (Access Policy Analysis)** investigated access mechanisms and authentication frameworks across European e-Infrastructures, identifying sixteen critical gaps including lack of long-term resource planning, fragmented identity management, and barriers to unattended workflow execution.
- **D5.3 (Landscape Analysis)** assessed twenty European e-Infrastructures, characterising current capabilities, service offerings, and operational practices, whilst identifying gaps that limit effective federation and cross-infrastructure collaboration.
- **D6.1 Section 2.5 (Consolidated Gaps and Requirements)** synthesised WP5 findings into sixteen consolidated gaps ensuring traceability from identified challenges to strategic interventions.

On top of the project deliverables, feedback was also collected in 2 interactive sessions at EGI 2025 conference and at ISC 2025 and by the interviews of community experts organised by WP4. The following section provides a summary of the main findings and the relation with SRIDA priorities. For more detailed analysis, readers should consult these deliverables directly. The following sections document how this evidence was validated and mapped to priorities.

C1.1. SPECTRUM CoP - Early Findings

This section describes the main gaps and findings from the work performed by the SPECTRUM CoP as reflected in D3.1 it includes (1) the analysis of results from a survey, (2) a review of the WP3 Knowledge Hub, populated with official documents from the communities, as provided by the CoP collaborators, and (3) trends and directions as extrapolated from all the collected material. Following are the main findings:

Table C.1.1.1: Gaps and findings from the Community of Practice.

Key Areas & Gaps	Finding Description	Evidence for SRIDA
Careers and skills. Lack of stable career paths for scientific software and data engineering roles	European research infrastructures depend heavily on specialised RSEs, data engineers, and domain-specific computing experts. However, most positions remain short-term or project-based, leading to high attrition to industry. This threatens Europe's ability to design, maintain, and scale the complex systems needed in the coming decade.	Strategic Driver: Workforce Priorities: P13: Skills, Careers and Training

Key Areas & Gaps	**Finding Description**	**Evidence for SRIDA**
Software evolution Insufficient evolution of scientific software for heterogeneous and emerging architectures	Many scientific codes are not optimised for multicore CPUs, vectorisation, GPUs, FPGAs, or heterogeneous workflows. Portability and performance engineering remain uneven across domains. The lack of modern tooling, parallelisation strategies, and portable abstractions limits the ability to exploit next-generation systems.	Strategic Driver: Preservation Priorities: P04 Heterogeneous Computing Integration, P08. Software Portability
AI Integration Limited adoption and integration of AI/ML across scientific workflows and tools	Although AI/ML is recognised as transformative, its uptake remains inconsistent across communities. There is a need to embed AI both as a user-facing capability (e.g., inference, analysis, surrogate modelling) and as an optimisation tool within infrastructures. Without structured integration, scientific domains risk falling behind global trends.	Strategic Driver: AI Priorities: P07 AI/ML
FAIR, open and cross-domain collaboration Fragmented data/software practices and insufficient adherence to FAIR principles	Data and software are often siloed, inconsistently documented, or incompatible across communities. Limited interoperability and collaboration hinder reuse, co-development, and joint innovation. Shared standards, repositories, and cross-domain toolchains are needed to reduce fragmentation and increase impact.	Strategic Drive: Preservation Priorities: P01 Cross-Infrastructure Governance, P08 Software Portability, P09 Open Science and Reproducibility
Cloud, HPC and virtualisation Uneven support for cloud-native workflows and portable deployment	Many e-Infrastructures operate cloud and virtualisation capabilities, but integration across providers varies. Software distribution systems (CVMFS, EESSI) ensure consistent execution environments across sites. Portability across heterogeneous architectures and cloud-HPC integration remain inconsistent.	Strategic Drive: Scale Priorities: P04 Heterogeneous Computing Integration, P06 Orchestration & Workflows, P08 Software Portability
Network and node connectivity Insufficient external networking capabilities for data-intensive workloads	A key bottleneck is the availability of high-bandwidth, external network connectivity from compute nodes. Centres express willingness to improve connectivity, but real-world capability must be validated against realistic scientific use cases, particularly for distributed or streaming workloads.	Strategic Drive: Scale Priorities: P05 Data Management and Transport
Long-Term Resource Allocation Short-duration compute/storage allocations that hinder sustained workflows	Many scientific programmes require multi-year continuous access to compute and data resources, but current allocations are often short-term. Although some centres can offer longer periods, dedicated policy and funding engagement is required to institutionalise extended allocation models.	Strategic Drive: Scale Priorities: P02 Federated Resource Allocation

C.1.2 Representative use cases

D5.1 (Representative Use Cases) examined fourteen use cases across HEP, Radio Astronomy, and adjacent domains, identifying common requirements for cross-facility workflow orchestration, federated data access, heterogeneous computing support, and automated resource provisioning. Use cases included major HEP experiments (ATLAS, CMS, LHCb, ALICE), AI applications in accelerator science, SKA Observatory and other radio astronomy initiatives (LOFAR, fast radio bursts), and data-intensive applications from meteorology, molecular dynamics, and neuroscience.

Table C.1.2.1: Challenges and Recommendations from Use Case Analysis.

Challenges	Recommendation	Evidence for SRIDA
Unprecedented Data Growth	Develop exabyte-scale data management architectures, expand long-term storage capacity, and invest in scalable data lifecycle management strategies to prevent bottlenecks across scientific domains.	Strategic Drive: Scale Priorities: P05 Data Management and Transport
Heterogeneous Computing Evolution	Design flexible, modular compute architectures that integrate CPUs, GPUs, and emerging AI accelerators; support software portability and optimization to fully exploit heterogeneous systems.	Strategic Drive: Scale Priorities: P04 Heterogeneous Computing Integration
Complex Workflow Requirements	Deploy advanced workflow orchestration frameworks capable of managing multi-step pipelines across HPC, cloud, and distributed sites; promote workflow co-design with scientific communities.	Strategic Drive: Scale Priorities: P06 Orchestration & Workflows
Data Federation	Establish robust, standards-based data federation frameworks enabling cross-site discovery, access, replication, and system-level metadata interoperability.	Strategic Drive: Scale Priorities: P05 Data Management and Transport, P09 Open Science and Reproductibility
Resource Allocation Planning	Adopt multi-year compute and storage allocation models that align with long-term scientific experiments and enable predictable planning for large-scale research programmes.	Strategic Driver: Scale Priorities: P02 Federated Resource Allocation
Standardisation Needs	Implement and enforce common facility, interface, and operational standards to reduce integration complexity, enhance interoperability, and lower maintenance costs.	Strategic Driver: Scale Priorities: P01 Cross-Infrastructure Governance, P04 Heterogeneous Computing Integration. P11 Community Collaboration and Co-Design
Authentication Evolution	Evolve authentication and authorization infrastructures toward federated, automation-friendly models (e.g., token-based systems) that balance security with usability for	Strategic Driver: Trust Priorities: P03 Federated Identity and Trust

Challenges	Recommendation	Evidence for SRIDA
	large, distributed workflows.	
Need for Architectural Flexibility	Build infrastructure supporting diverse workloads simultaneously (simulation, AI/ML, streaming, and data-intensive pipelines) within a unified but flexible exascale-ready architecture.	Strategic Driver: Scale, AI Priorities: P04 Heterogeneous Computing Integration P07 AI/ML
Scaling Data Management to Exabyte Levels	Invest in next-generation data management solutions with high-throughput ingest, automated distribution, and transparent multi-site access, ensuring scientific teams can operate across distributed environments.	Strategic Drive: Scale Priorities: P05 Data Management and Transport
Importance of Cross-Community Collaboration	Promote shared tools, AI models, data standards, and co-design initiatives between HEP, RA, and other scientific communities to increase efficiency and accelerate innovation.	Strategic Drive: Workforce Priorities P11 Community Collaboration and Co-Design

C.1.3. Interoperable access policies

D5.2 (Access Policy Analysis) investigates access mechanisms and authentication frameworks across European e-Infrastructures, identifying sixteen critical gaps including lack of long-term resource planning, fragmented identity management, and barriers to unattended workflow execution. This section summarises the main findings and recommendations emerging from this deliverable.

Table C.1.3.1: Findings and recommendations from interoperable access policy analysis.

Key Areas	Finding Description and Recommendation	Evidence for SRIDA
Obtaining Access		
Lack of long-term, assured resource access	Grant-based allocations last 1–2 years and may not be renewed; this prevents long-term planning for experiments requiring stable 3–5 year access to compute and storage. Recommendation: Adopt longer-term and more flexible allocation processes aligned with multi-year scientific planning.	Strategic Drive: Scale Priorities: P02 Federated Resource Allocation
Allocations limited to single sites within an e-Infrastructure	Many systems (e.g., EuroHPC JU) only honour allocations on specific hosting sites, preventing users from shifting workloads for burst capacity, resiliency, or efficiency. Recommendation: Enable e-Infrastructure-wide allocation models where quotas apply across all participating sites.	Strategic Drive: Scale Priorities: P01 Cross-Infrastructure Governance
Access Tracks & Modalities		

Key Areas	Finding Description and Recommendation	Evidence for SRIDA
Limited interactive access at significant scale	Interactive workloads face delays and inefficiencies due to batch scheduling constraints; scalable interactive sessions are difficult to obtain. Recommendation: Extend orchestration/scheduling systems to support on-demand, interactive compute use with access to substantial node resources.	Strategic Drive: Scale Priorities: P06 Orchestration and Workflows
Lack of high-level, user-friendly science platforms	Most centres lack portals or workflow interfaces (like LOFAR or PanDA) that hide low-level scheduler, module, and environment complexity from domain scientists. Recommendation: Introduce high-level, general and domain-specific user interfaces enabling scientists to run workflows without deep system knowledge.	Strategic Drive: Scale Priorities: P06 Orchestration and Workflows
Access Management & Security		
Fragmented identity management across infrastructures	End-users working across multiple federated infrastructures must manage separate identities. Lack of common AAI systems limits interoperability and efficiency. Recommendation: Deploy interoperable, European-wide AAI services supporting federated identities and service accounts.	Strategic Drive: Scale Priorities: P03 Federated Identity and Trust
Difficulty running unattended long-running workflows under Multi-Factor Authentication	MFA-based authentication methods expire frequently (often daily), breaking multi-day workflows that cannot reauthenticate interactively. Recommendation: Ensure reliable, unattended execution by enabling token lifetimes, service credentials, or automation-safe authentication mechanisms.	Strategic Drive: Scale Priorities: P03 Federated Identity and Trust
Rules, Assurances & Monitoring		
Inconsistent low-level software environments across sites	Software stacks differ significantly between centres and architectures; portable builds require standardised interfaces and widely supported distribution methods (EasyBuild, Spack, containers). Recommendation: Provide efficient, standardised low-level software interfaces across sites, including shared build/distribution frameworks.	Strategic Drive: Workforce Priorities: P08 Software Portability, P10 Software Preservation and Sustainability
Lack of structured feedback and co-design loops	While some communities (e.g., WLCG, SRCNet) use collaborative planning models, most infrastructures lack systematic mechanisms for shared roadmapping and evaluation. Recommendation: Establish formal feedback, improvement, and planning loops between research communities and e-Infrastructures.	Strategic Drive: Workforce Priorities: P11 Community Collaboration and Co-Design

C1.4 Landscape, Requirements and Gap Analysis

D5.3 (Landscape Analysis) assessed twenty European e-Infrastructures, revealing persistent challenges in WAN performance, software portability across heterogeneous architectures, and environmental sustainability monitoring. This section summarizes the key findings and recommendations of the deliverable. The infrastructures studied span High Performance Computing (HPC), High Throughput Computing (HTC), dedicated Data-oriented infrastructures, and Cloud e-Infrastructures.

- **HPC-oriented infrastructures**
  The EuroHPC Joint Undertaking (JU) [80] forms the backbone of this landscape, comprising operational petascale systems [85] such as Deucalion [51], Discoverer [55], Karolina [123], Leonardo [126[, LUMI [132], MareNostrum 5 [134], Meluxina [136] and Vega [162], with exascale systems JUPITER (Germany) [121], Arrhenius (Sweden) [14], Daedalus (Greece) [50] and Alice Recoque (France) [12] planned for 2025-26. National centres such as **ICSC** (Italy) [109], **GCS** [97] / **JSC** [114] / **HLRS** [106] / **LRZ** [131] / **NHR Alliance** (Germany) [138], **RES** (Spain) [148], **CSCS** (Switzerland) [43], **EPCC** (UK) [67] and **GENCI** (France) [98] complement these resources and will support future HEP and RA experiments.
- **HTC-oriented infrastructures**
  The WLCG [164] serves the global HEP community with about 1.4 million CPU cores and 1.5 Exabytes of storage (2025). WLCG Tier-1 centres, the EGI HTC-oriented infrastructure [62] provide access for diverse scientific communities, together with upcoming HTC platforms that include regional centres such as the SKA SRCNet [152].
- **Data-oriented infrastructures**
  They include the WLCG's data lake (across the world and with CERN in Switzerland as the central hub), the SKA SRCNet [152], the EBRAINS neuroscience platform [58], the LOFAR Long-Term Archive (LTA) [129], German's ErUM4 Data Hub [69], the ICSC National Center in Italy [109], the PUNCH4NFDI **[146]** initiative and the Copernicus Earth-observation data space [41].
- **Cloud-oriented infrastructures**
  In contrast to the three categories of infrastructures above, Cloud-oriented e-Infrastructures often provide a combination of compute (HTC and increasingly HPC) and data services. This category include for example: the EGI FedCloud [61], which offers a federated cloud interface that unifies resources across Europe; SURF's Grid [161] and Spider [162] platforms mix cloud and HTC/HPC services the European Open Science Cloud (EOSC) Federation [158], which federates national nodes and thematic pilots (the first 13 pilot nodes were announced in March 2025[2]) and the Simpl Data Federation Platform [156], built under a commercial contract for the European Commission, which aims at unifying access to diverse European data spaces.

Table C.1.4.1: Requirements from landscape analysis.

Key Areas	Requirements Description	Evidence for SRIDA
Storage / Data Capabilities & Services		
FAIR storage for observation/experiment data across domains	Create shared interfaces and core implementations that preserve rapidly growing scientific datasets (including metadata and provenance) according to FAIR principles. Must enable discovery, search by provenance/content, public access, and seamless interoperability with HPC/HTC systems, while supporting international users.	Strategic Drivers: Preservation Priorities: P05 Data Management and Transport, P09 Open Science & Reproductibility
Support for GDPR-protected and restricted data	Provide secure, standardised frameworks for storing and processing personal or commercially sensitive data. Must include encryption, secure key management, auditing, application vetting, and possibly sandboxing. Must avoid imposing restrictive	Strategic Drivers: Trust Priorities: P03 Federated Identity and Trust

---

[2] https://www.eosc-beyond.eu/pilots

Key Areas	Requirements Description	Evidence for SRIDA
	security overheads on open scientific data and ensure separation of sensitive vs. open workflows.	
Data Transfer & Federation		
Automated and efficient data movement and staging	Automate intra-site and inter-site data transfers using high-performance, multi-stream mechanisms. Standard interfaces (e.g., FTS-like) should abstract protocol details. Third-party transfer capabilities and alignment with HPC centre security policies are required for large-scale workflows.	Strategic Drivers: Preservation. P05 Data Management and Transport
Plan transfer capacity based on community needs	Provision high-throughput data movement and ephemeral storage according to evolving domain requirements. Data movement is increasingly unpredictable; centres must be prepared to adapt transfer policies and bandwidth to dynamic workflow demands over system lifetimes.	Strategic Drivers: Preservation. P05 Data Management and Transport
Compute Capabilities & Services		
Portability of compute tasks across accelerated platforms	Ensure research codes can run across diverse accelerators (GPUs, FPGAs, vendor families). Avoid vendor lock-in (e.g. CUDA). Promote portable programming models (SYCL, Kokkos, Alpaka, oneAPI) and transparently optimised backends (PyTorch, TensorFlow). Recognise intrinsic limits in portability for architectures like FPGAs.	Strategic Drivers: Preservation. Priorities: P04 Heterogeneous Computing Integration, P08 Software Portability
Software Services, Interfaces & Stacks		
Common workflow systems supported by all e-Infrastructures	Identify and support a reduced set of workflow management systems usable across domains. Avoid duplication of domain-specific frameworks. Provide interfaces to HPC fabric schedulers (e.g., external API access to Slurm). Aim to converge toward a small, interoperable workflow ecosystem co-designed with communities.	Strategic Drivers: Scale. Priorities: P06 Orchestration and Workflows
Common software stacks for compute applications	Provide unified, container-ready software stacks optimised for combinations of CPUs, interconnects, and accelerators. Standardise low-level interfaces across heterogeneous systems to simplify portability, virtualisation, and performance optimisation.	Strategic Drivers: Preservation. Priorities : P08 Software Portability
Compute Federation Services		
Federated compute services with interoperable identity, catalogues, and access	Support federated identity (AAI), cross-site workflow execution, and simplified user interfaces across HPC sites. Current models (EuroHPC federation platform, CSCS/GENCI federations, WLCG, EGI) demonstrate partial capability but lack unified quotas, automatic workload routing, and full compute federation. Systems like Simpl provide initial catalogues, governance onboarding, secure communication, and federated data processing workflows.	Strategic Drivers: Trust Priorities: P01 Cross-Infrastructure Governance, P03 Federated Identity and Trust

C1.5 Feedback & Alignment with the Technical Blueprint (D6.1)

D6.1 Section 2.5 (Consolidated Gaps and Requirements) synthesises WP5 findings into sixteen consolidated gaps with explicit mapping to SRIDA priorities, ensuring traceability from identified challenges to strategic interventions. We report here as reference and link to SRIDA priorities.

Table C.1.5.1: Consolidated Gaps/Requirements and Recommendations.

Gap/Requirement	Recommendation	Evidence for SRIDA
Long-term resource provisioning Current grant-based allocations are short (1–2 years) and fragmented by site; HEP and RA need assured access on decadal timescales.	Establish long-term (3–5 year) flexible allocation frameworks, federated across infrastructures, with transparent accounting and monitoring. Policy activity will require dedicated funding.	Strategic Drivers: Scale Priorities P02 Federated Resource Allocation
Federated access and identity management End-users face fragmented identity systems across e-Infrastructures, complicating multi-site workflows. Authentication often interrupts unattended, long-running jobs.	Deploy interoperable AAI frameworks across European e-Infrastructures, supporting single sign-on, attribute-based authorization, and persistent tokens for long workflows.	Strategic Drivers: Trust Priorities P02 Federated Identity and Trust
Interoperable allocation and accounting Allocations and quotas are bound to specific HPC centres, limiting flexibility. Accounting practices differ across infrastructures.	Enable e-Infrastructure-wide allocations and harmonized accounting, so projects can burst across systems and providers while ensuring fair contributions and usage transparency.	Strategic Drivers: Scale Priorities P01 Cross-Infrastructure Governance, P02 Federated Resource Allocation
High-level and portable software environments Software stacks differ widely between sites; portability across CPUs, GPUs, and accelerators is inconsistent. Scientists often face brittle interfaces.	Provide standardized, efficient software stacks (containers, CVMFS) and high-level domain-specific interfaces, enabling reproducibility and accelerator portability.	Strategic Drivers: Preservation Priorities: P04 Heterogeneous Computing Integration, P08 Software Portability
Workflow execution and orchestration Workflow engines are fragmented across communities (e.g. HTCondor in HEP vs. LEXIS in EuroHPC). Cross-facility workflows lack resilience and provenance capture.	Co-design and converge on a set of common APIs for workflow systems, with support for unattended execution, provenance tracking, and cross-infrastructure orchestration.	Strategic Drivers: Scale Priorities: P06 Orchestration and Workflows
Data federation and FAIR principles No unified approach to FAIR storage, metadata, and provenance across domains; irreplaceable datasets risk fragmentation.	Establish interoperable federated data management with FAIR-compliant storage, metadata catalogs, and efficient interfaces to HPC/HTC, ensuring long-term access and reproducibility.	Strategic Drivers: Scale, Preservation Priorities: P05 Data Management, P09 Open Science

Gap/Requirement	Recommendation	Evidence for SRIDA
Data transfer and staging Data movement across sites often relies on ad-hoc scripts and site-specific tools; transfer capacity planning is uneven.	Automate large-scale, third-party enabled, efficient data movement with standardized protocols, and plan data transfer/storage capacity jointly with community needs.	Strategic Drivers: Scale, Preservation Priorities P05 Data Management and Transport
Real-time and low-latency processing A small number of RA use cases (e.g. transient detection) and some HEP triggers require near-real-time data analysis via pipelines.	Extend continuum capabilities with real-time, low-latency processing.	Strategic Drivers: Scale Priorities P04 Heterogeneous Computing Integration
Security, privacy, and sensitive data handling Personal or commercially sensitive data requires stronger safeguards; GDPR compliance adds complexity.	Provide flexible architectures: robust encryption, key management, auditing, anonymization/pseudo-anymization, and possibly separate high-protection infrastructure paths.	Strategic Drivers: Trust Priorities P03 Federated Identity and Trust
Monitoring and observability Current infrastructures lack consistent, cross-site monitoring of workflows, performance, and energy usage.	Deploy monitoring and observability frameworks that capture standardized metrics describing infrastructure health, application performance, provenance, and sustainability.	Strategic Drivers: Scale Priorities P01 Cross-Infrastructure Governance
AI/ML integration AI/ML workloads are increasingly central and are being investigated in every step of the data-processing pipelines, from data-acquisition to simulation and analysis.	Enhance services in AI/ML platforms to support distributed training, inference deployment, and access to foundation models, integrated with HPC and domain workflows.	Strategic Drivers: AI/ML Priorities P07 AI/ML
Co-design and feedback loops Systematic co-design between science communities and infrastructure providers is patchy; planning often mismatches user needs.	Establish structured feedback loops (like WLCG/SRCNet) across domains, embedding co-design in infrastructure roadmaps, procurement, and evaluation.	Strategic Drivers: Workforce Priorities: P11 Community Collaboration and Co-Design
Workforce for scientific software development, funding, and sustainability Long-term sustainability depends on stable funding, governance, and skilled personnel. Current models are fragmented.	Develop coordinated European funding/governance models, align with national contributions, and invest in training and career pathways for research software engineers and data stewards.	Strategic Drivers: Workforce Priorities: P13 Skills, Careers and Training
Interactive access to significant-scale computing	Extend scheduling/orchestration to support interactive compute use	Strategic Drivers: Scale

Gap/Requirement	Recommendation	Evidence for SRIDA
Interactive access to small/medium partitions is possible. Problems include delays in availability, and potential underuse of resources.	cases (incl. those that have significant resource needs) Compromise to be reached between quick availability, performance guarantees and efficient use of resources.	Priorities: P02 Federated Resource Allocation
High-level end-user interfaces Most e-Infrastructures do not offer high-level, abstracted interfaces. Domain scientists should not need to know system details and interact with brittle application interfaces.	Introduce high-level general and domain-specific user interfaces. (like workflow portals, Jupyter, or CARTA)	Strategic Drivers: Workforce. Priorities: P11 Community Collaboration and Co-Design
Unattended execution of long-running workflows and services Two-factor authentication can impact execution of long-running workflows and services. Access established after interactive or MFA authentication has a limited period of validity. Workflow steps need to interactively re-authenticate.	Ensure reliable and unattended execution of long-running workflows and user-operated services.	Strategic Drivers: Scale Priorities: P02 Federated Resource Allocation, P06 Orchestration & Workflows

C.2 Feedback & Alignment from External events

During June 2025 public feedback was sought in external events, during the EGI 2025 conference and a BoF session at ISC 2025. The main objective was to start assessing the landscape and main challenges and gaps to be addressed in the Technical Blueprint and SRIDA. For those a brief presentation of the project was delivered backed by an interactive questionnaire organized during the event using the mentimeter platform.

The questions asked:

- What are the main challenges and gaps in compute/data intensive science?
- How current technologies and policies should evolve to meet the identified challenges/gaps?
- What are the main priorities to tackle over the next 5-7 years?

C.2.1 EGI 2025 interactive session

The interactive session at the EGI conference[3] was reasonably attended - ca. 25 participants (ca. 80% room capacity) from which over 19 responses were gathered on the mentimeter platform that were followed by a life Q&As and discussion. Next tables collect a summary of the main findings and associate them with the different SRIDA priorities.

What are the main technical and non-technical challenges in compute/data intensive science?

Table C.2.1.1: Technical and non-technical challenges from EGI2025 interactive session.

Technical Challenges	Non-Technical Challenges
Workflow Portability (P08) Heterogeneous Infrastructures (P04) Dataflow Management (P05) Software Development Practices (P10) Interface Standardization Distributed Computing Knowledge (P13) System Configuration Variability (P06) Efficient Resource Utilisation (P02)	Community Alignment (P11) Real Needs Assessment (P11) Global Coordination across communities (P11) Role separation (data generation & data usage) (P01) Collaboration dynamics (P11) Training and lack of mutual understanding across user communities and infrastructures (P01, P11) Awareness of available resources (P13) Access to funding

How current technologies and policies should evolve to meet the identified challenges/gaps?

Table C.2.1.2: Technology and policy evolution from EGI2025 interactive session.

Technology-related	Policy Related
Enforce open standards & interoperability (P09) Data-centric & FAIR infrastructures (P09) Lower technical barriers to entry (P11) Adopt proven industry solutions (P01) Update resource efficiency metrics (P12) Stay technology-agnostic & adaptable (P01)	Prioritize training & career paths (P13) Encourage collaboration & common solutions (P11) Foster EU/global coordination (e.g., through EOSC) (P01) Support inclusivity & accessibility (P02) Address cultural/organizational gaps (P11, P13) Increase transparency of metrics (P01)

What are the main priorities to tackle over the next 5-7 years?

Table C.2.1.3: Main priorities from EGI2025 interactive session.

Priority Area	Policy Related
Adopt a Data-Centric Approach	Adopt, scale, and support data-driven models
Expand and Diversify Communities of Practice	Engage diverse, real scientist input including underrepresented groups
Strengthen Collaboration and Interoperability	Build bridges, foster dual-skilled experts
Workflow Mapping & Reuse	Map workflows and promote reuse solutions to avoid redundancy
Respond to External Disruptions	Monitor and adapt to external political and technological factors (e.g. geopolitics and GenAI)
Breaking Silos	Connect science, IT, and policy communities

---

[3] EGI 2025 event page: https://indico.egi.eu/event/6638/contributions/20528/

C2.2 Interactive session at International Supercomputing (ISC 2025 BoF)

The Birds-of-a-feather interactive session[4] at the International Supercomputing Center was very well attended (over 70 people), from which 33 responses were collected in the Mentimeter platform. The following tables provide a summary of the content gathered.

What are the main challenges and gaps in compute/data intensive science?

Table C.2.2.1: Technical and non-technical challenges from ISC2025 interactive session.

Technical Challenges	Non-Technical Challenges
Infrastructure reliability (P03) Data sharing & collaboration (P05) Data transport (P05) Interoperability (systems, models) (P04) Data security (P03) Data management & lifecycle Documentation (P05) Access patterns for compute/data (P03) Keeping up with technology (P04)	Training & upskilling (P13) Ethics & compliance (P03) Reproducibility & sustainability (P10) Trusted research environments (P03) Fast, pragmatic action over high bureaucracy (P01) Community awareness (P11)

How current technologies and policies should evolve to meet the identified challenges/gaps?

Table C.2.2.2: Technology and policy evolution from ISC2025 interactive session.

Technology-related	Policy Related
Federated access for data/compute (P03) Cross-platform software deployment (P08, P10) Scalable Data Management (P05) Open Federation Standards (P09) Cloud Evolution (P04)	Common European policies & harmonization (P01) Integrated, cross-border access frameworks (P02) Agility over perfection: prioritize timely policy action (P01) Foster cross-domain agreements on shared challenges (e.g. metadata, provenance) (P01) Coordinated, sustainable funding strategies

What are the main priorities to tackle over the next 5-7 years?

Table C.2.2.3: Main priorities from ISC2025 interactive session.

Priority Area	Policy Related
Advance Scientific Data & Metadata Management	Adopt robust, FAIR-aligned, long-term practices
Sustainable Data Storage	Plan and fund scalable, long-term storage solutions
AI Competitiveness	Invest in infrastructure and skills to keep pace with AI
Plan Strategic Funding & Partnerships	Strategic planning with experiment and infrastructure upgrades
Usability for Non-Experts	Simplify access and use of compute/data infrastructure
Reuse & Integration	Build on existing solutions; break down silos

---

[4] ISC BoF page: https://isc.app.swapcard.com/widget/event/isc-high-performance-2025/planning/UGxhbm5pbmdfMjU4NjExMg==

What are the main challenges and gaps in compute/data intensive science?

Table C.2.3.1: Main challenges and gaps from CoP Interviews.

Challenge Area	Related Gaps
Compute Availability & Suitability (P04, P07)	- Shortage of suitable compute resources for scientific workloads; EU investments are skewed toward AI-optimised architectures not broadly usable by most sciences - Need for general-purpose HPC, not only GPU- or AI-centric systems; risk of misalignment with scientific needs driven by political agendas (e.g., AI Act) - Operational stability: scientific digital-twin infrastructures require near-continuous operation, but supercomputers have downtime and unpredictable maintenance cycles
Data Scale, Storage, and Mobility (P05)	- Data volumes exploding (multi-exabyte scale in SKA; HL-LHC era; RA/astronomy; DestinE) causing challenges for storage, transfer, and long-term preservation - Long-term storage remains unfunded and under-prioritised despite being essential and costly - Data mobility bottlenecks, especially cross-site and cross-border transfers; lack of a mature federation framework
Fragmentation of Infrastructure & Lack of Interoperability (P01, P04, P09)	- Extreme fragmentation: many domain-specific clusters, bespoke tools, disconnected policies; a single continuum is not realistic but greater coherence is needed - Metadata gaps: many communities not ready to standardise metadata; data without structure is useless
Skills, Workforce, and Retention (P13)	- Severe shortage of RSEs, HPC experts, and cross-skilled profiles (all interviews reaffirm this) - High attrition to industry due to salary gaps and unclear career paths - Training bottlenecks: porting to accelerators, heterogeneous architectures, data science methods; many scientists lack sustainable coding practices
Misaligned Funding & Sustainability (P01, P10)	- Project-based allocations do not support environments requiring continuous access - Infrastructure sustainability—power cost, energy efficiency, operational cost escalation
Slow Adaptation to Emerging Technologies (P04, P08, P10)	- AI growing fast but unevenly adopted; many workflows not prepared for GPUs or ML - Quantum computing considered impactful only long-term - Technology lock-in (especially NVIDIA CUDA) creates major porting challenges

How current technologies and policies should evolve to meet the identified challenges/gaps?

Table C.2.3.2: Challenges and Recommendations from CoP Interviews.

Challenge Area	Recommendations
Infrastructure & Architecture Evolution (P01, P04)	- Build a federated but not monolithic European data–compute continuum, enabling 80% common services and 20% domain-specific features - Invest in data-centric design: data locality, distributed storage replication, high-speed cross-border networks - Diversify architecture choices beyond quantum and GPUs, e.g., neuromorphic, FPGAs - Increase interoperability with container-oriented, cloud-native, and Kubernetes-like environments
Policy & Governance Changes (P01)	- Shift from project-based to sustainable long-term operational funding models for research infrastructures - Align European and national strategies to avoid duplication and fragmentation - Ensure data sovereignty, avoid commercial lock-in, and keep scientific data in open-access, public infrastructures - Integrate energy-efficiency mandates into software and hardware funding criteria
Skills, Workforce & Training Evolution (P13)	- Professionalise career paths, create stable long-term contracts, and recognise computing roles - Scale training in heterogeneous computing, GPU programming, modern software engineering, metadata practices - Adopt cross-community hands-on formats (hackathons, scaling workshops, co-design labs) that have proven effective
Technology Adoption and Co-Design (P07, P11)	- Co-develop tools and workflows with communities, not impose top-down systems - Promote shared software stacks (e.g., Rucio, PanDA) across scientific domains - Boost applied AI/ML integration in simulation, modelling, anomaly detection, and digital twins

What are the main priorities to tackle over the next 5-7 years?

Table C.2.3.3: Main priorities from CoP Interviews.

Priority Area	Policy Related
Skills, Careers & Human Capital (P13)	• Establish permanent RSE/HPC/Data Science career tracks at national & EU levels • Provide sustainable funding for long-term staff, not only project contracts • Create European training programmes on accelerators, heterogeneous architectures, software sustainability • Encourage mobility networks, shared curricula, and incentives for technical career recognition
Data Infrastructure: Storage, Federation & Mobility (P05)	• Create a federated European data framework, not a monolithic system • Implement policies for long-term, funded scientific data archiving • Support cross-border high-speed networks and guaranteed data-mobility capabilities • Enforce metadata and interoperability standards across domains
Software Portability, Modernisation &	• Direct EU investment toward software & workflow optimisation, not only hardware procurement

Priority Area	Policy Related
Sustainability (P08, P10)	• Incentivise multi-architecture portability (avoid CUDA lock-in) and support for compiler/translation frameworks • Fund shared software stacks (e.g. Rucio, PanDA) as European public goods.
Heterogeneous Compute Architectures (P04)	• Balance EuroHPC policy to include general-purpose systems, not only AI/GPU-heavy ones • Create procurement guidelines to support diverse architectures (e.g., FPGA, neuromorphic) • Require that hardware investment includes associated data, software, and operational budgets • Ensure continuous-access policies for infrastructures needing predictable uptime
Large-Scale AI Capacity & Integration (P07)	• Fund EU-wide AI training facilities for scientific datasets • Accelerate adoption of AI-driven workflows in scientific codes • Establish ethical and open-access frameworks for scientific AI datasets and models • Promote cross-community AI co-design labs to produce reusable tools
Funding, Governance & Sustainability (P01)	• Shift from project-based HPC access to stable operational funding framework • Mandate energy-efficiency requirements at both software and facility levels. • Create mechanisms for EU–national co-funding to maintain shared infrastructures • Limit fragmentation by restricting funding for isolated, domain-specific clusters
Cross-Community Collaboration and Co-Design (P11)	• Promote RI–HPC co-design programmes (workshops, pilots) over top-down solutions • Support European initiatives that integrate multiple communities (e.g. ESCAPE-like models) • Build policies for standardised APIs, shared workflow tooling, and federated access models • Encourage joint applications and coordination between RIs to avoid parallel infrastructures

C.3 Community of Practice and Advisory Board Validation

SPECTRUM CoP, comprising approximately 65 participants across six Working Groups, provided continuous validation throughout 2024-2025. Each priority was assigned to one or more Working Groups who developed the priority specifications, ensuring domain expertise and operational knowledge informed priority content. Validation occurred through regular plenary meetings, Working Group sessions, and a community survey capturing quantitative data on resource utilisation patterns, technical requirements, and policy preferences.

The External Advisory Board provided strategic guidance refining the priority framework. Key recommendations from the second EAB meeting (post-M18) included: focusing on 2-4 main innovation avenues capable of delivering concrete services within 2-3 years (identifying P02, P05, P08, P09 as strategic focus priorities); creating priorities addressing reproducibility over 5-10 year timescales (P09) and code survivability across 20-30 year instrument lifetimes (P10); adding heterogeneous computing integration as a foundational priority (P04); and refining priority scope to distinguish ESFRI pan-European infrastructure needs from generic simulation projects. These recommendations shaped the final 13-priority framework.

Table C.3.1: Assignment of SRIDA Priorities to CoP Working Groups.

Working Group	SRIDA Priorities Developed
WG1: Data Management and Access	P05 Data Management and Transport, P07 AI/ML & Accelerated Computing (data aspects)
WG2: Workflow Management and Organisation	P02 Resource Allocation, P06 Orchestration & Workflows
WG3: Compute Environment	P04 Heterogeneous Computing Integration, P07 AI/ML & Accelerated Computing (compute aspects)
WG4: Software Tools	P08 Software Portability, P10 Software Preservation and Sustainability
WG5: Scientific Use Cases	P11 Community Collaboration and Co-Design, P13 Skills, Careers and Training
WG6: Facilities	P01 Governance & Funding Coordination, P03 Access, Security & Trust, P12 Environmental Sustainability

P09 (Open Science and Reproducibility) emerged from cross-cutting discussions spanning WG1, WG4, and WG5, receiving particular emphasis following External Advisory Board feedback.

P12 Environmental Sustainability had an extended contribution on expected policy implementation. Following is the summary:

We describe two sets of measures, ones which can to some extent already be applied and one which need to be developed, tested and then gradually applied. The latter is presented in order of complexity, the first being susceptible to be applied on the short term while the last is for a longer term.

The environmental impacts below are understood to result from multi-criteria attributional (when evaluating the impact of a system) and consequential (when evaluating changes) life-cycle analyses and apply at the local and global scales. As such, they are not limited to the current phase in the life of the resource but consider all of them.

Applicable measures :

- Train users to impact-oriented software development with an introduction to life-cycle analysis and environmental impacts, together with hardware and software levers to minimise them.

- Favour the emergence of eco-design communities for hardware and software solutions :
    - create a software eco-design CoP including users and datacentre support teams as a forum to discuss good practices and share experience, as well as foster collaborations on code to reduce their environmental impacts,
    - create a similar community to address hardware aspects at the datacentre level, including how to design, operate, maintain and procure infrastructures with minimal environmental impacts,
    - have datacentres provide resources to facilitate hardware-software co-design (software environments automating fine monitoring of execution in a transparent manner to point to portions of software requiring improvement, share hardware characteristics with users as well as platform-specific good practices),
    - propose access to experimental partitions composed of older nodes (whose environmental impacts have been amortized), of diverse types and with elevated privileges, to foster exploration of bottlenecks and software fragilities which are more obvious in such contexts for development and testing.
- Require users to provide an experimentation plan as part of their resource request with a level of detail matching its approximate expected environmental impacts.
- Power manage infrastructures :
    - concentrate use of resources on fewer nodes and switch off nodes made idle,
    - finely manage the powering of components in nodes (eg. unused CPU, accelerators, USB subsystem),
    - cool to higher temperatures,
    - power cap datacentres globally and compare strategies for power distribution to nodes (eg. selectively switch off nodes versus reducing consumption via dynamic voltage and frequency scaling).
- Require reporting IT resource usage in scientific publications (eg. as an extension to the required acknowledgements to resource providers).
- Institute sustainability governance at the datacentre and governing bodies levels through dedicated committees and people in charge of developing and evaluating sustainability in an independent way across all 3 tiers (datacentres, network, user terminals) to avoid impact transfer. This governance will also be in charge of continuously evaluating sustainability practices versus the evolving technological landscape and environmental crises.

Measures to be developed :

- Develop, test, evaluate and collect feedback from users for the tools proposed in the Technical Blueprint to consolidate them and de-risk transition to impact-driven resource management.
- Promote the evaluation of the environmental impacts of computing, transfers and storage in scientific publications.
- Develop a process for accessing resources to frame use and software quality in accordance to environmental impacts (tools are addressed in the Technical Blueprint):
    - regulate compute usage by enforcing use of selected partitions for selected categories of work (eg. development/prototyping, interactive, optimisation, test runs, production, long runs) and define criteria enabling categories of work related to functional, performance, scaling, sufficiency and reliability assessments,
    - manage hardware allocations to partitions depending on distribution of categories of work,
    - regulate storage usage based on performance of storage media versus software I/O performance and define criteria related to I/O usage and behaviour of software (file formats, access patterns, data volumes) versus storage technology,
    - develop active monitoring of execution through live tracking of progress and quality and adaptively control it based on its usefulness to adjust or interrupt it,
    - require crash testing of applications : handling of exceptions and errors, exceeding available resources (eg. memory usage, time or impact allocations) and hardware failures (eg. through checkpointing) to trigger adaptive software behaviour or graceful termination of execution,
- Enable impact-oriented datacentre-level management of resources :
    - given the importance of environmental impacts during fabrication, use hardware for extended durations mainly by using failure-prone hardware for functions which are not mission or impact critical and increasing the resilience of execution,

    - track datacentres' environmental impacts,
    - develop comparative life-cycle analyses for existing resources to estimate usage, maintenance, disposal, replacement and extension impacts in a prospective manner,
    - include environmental impacts as a central component of decision-making,
    - include sustainability requirements in tenders as environmental impacts should condition feasibility as cost and performance do.
- Extend the impact-oriented management of the previous two items from the datacentre level to the compute and storage continuum.
- Current evaluation of the use of IT resources focuses on resource usage both at the datacentre (pressure on resources, occupancy of partitions) and end-user (core hours) levels. This encourages growth while the relevance of such use is not precisely evaluated. Develop and evaluate metrics rewarding sufficiency and favouring minimal use for a given purpose.
- Transform resource requests to allow allocating environmental impacts :
    - require a description of the work's relative contribution to the wider scope of the scientific project,
    - require an experimentation plan with a level of detail matching the requested environmental impacts,
    - guide users in identifying resources meeting their technical needs at the datacentre and continuum levels (compute, storage) and in forecasting the environmental impacts related to different usage scenarios (in relation to the experimentation plan),
    - allow users to request access to technical resources and obtain an overall environmental impact credit for the proposed work (in replacement for core hours on predefined partitions),
    - support staged accesses (to different partitions depending on the categories of work) and rewarding sufficiency,
    - ensure user-allocated environmental impacts fit in the datacentres or continuum's overall impact credits,

    - monitor environmental impacts at the user, datacentres and continuum levels when operating under this paradigm to support them in managing their environmental credits and build a detailed accounting of the corresponding expense.

C.4 Priority to Technical Blueprint Mapping

The table below provides traceability between SRIDA priorities and D6.1 Technical Blueprint components, showing which consolidated gaps each priority addresses, which capabilities each priority implements, and which technical activities provide implementation guidance.

Table C.4.1: Mapping of SRIDA Priorities to D6.1 Technical Blueprint.

Priority	D6.1 Gaps Addressed	D6.1 Capabilities	D6.1 Activities
P01 Governance & Funding Coordination	#2 Coordinated governance, #16 Sustainability models	4.6 Resource Federation, A.8 Security & Trust	5.11 Governance & Sustainability
P02 Resource Allocation	#1 Long-term provisioning, #3 Allocation & accounting	4.6 Resource Federation, 4.7 Monitoring & Accounting	5.8 Resource Provisioning & Accounting
P03 Access, Security & Trust	#5 Identity management, #6 Unattended workflows, #7 Cross-border data protection	4.8 Security & Trust, 4.6 Resource Federation	5.10 Security, Trust & Data Protection

Priority	D6.1 Gaps Addressed	D6.1 Capabilities	D6.1 Activities
P04 Heterogeneous Computing Integration	#13 Heterogeneous computing	4.1 Compute Resources, 4.4 Orchestration & Workflows	5.1 Heterogeneous Computing Integration
P05 Data Management and Transport	#4 WAN performance, #8 Intelligent data placement, #10 Data mover capabilities	4.2 Data Resources, 4.7 Monitoring & Accounting	5.3 Data Federation & Transport, 5.4 Intelligent Data Management
P06 Orchestration & Workflows	#9 Cross-facility orchestration, #11 Resource-aware scheduling	4.4 Orchestration & Workflows, A.1 Compute Resources	5.5 Workflow Orchestration & Automation
P07 AI/ML & Accelerated Computing	#12 AI/ML with HPC, #13 Heterogeneous computing	5.5 AI/ML, A.1 Compute Resources	5.6 AI/ML Infrastructure Integration
P08 Software Portability	#14 Software portability, #15 Domain repositories	4.3 Software Distribution & Execution	5.2 Software Portability & Distribution
P09 Open Science and Reproducibility	Cross-cutting: #8, #14, #15	4.3 Software Distribution & Execution, A.2 Data Resources	3.1 Data & Software Preservation, 5.9 Open Science Practices
P10 Software Preservation and Sustainability	#15 Domain repositories	4.3 Software Distribution & Execution	3.1 Data & Software Preservation, 5.2 Software Portability
P11 Community Collaboration and Co-Design	#2 Coordinated governance	4.6 Resource Federation	5.2 Co-Design Processes
P12 Environmental Sustainability	Cross-cutting: sustainability	4.7 Monitoring & Accounting	3.2 Environmental Sustainability, 5.7 Energy-Aware Operations
P13 Skills, Careers and Training	#13 Hybrid computing expertise	4.6 Resource Federation	5.2 Co-Design Processes, 5.13 Workforce Development

This mapping serves multiple audiences: technical implementers read D6.1 for architecture then consult SRIDA priorities for stakeholder coordination; policymakers read SRIDA for strategic rationale then reference D6.1 for technical validation; research communities use both documents together to understand what infrastructure can provide and how to influence its development.