D3.1 Community of Practice - Interim report

Status: UNDER EC REVIEW

Dissemination Level: Public

Funded by
the European Union

Abstract	
Key Words	Community of Practice, Survey, High Energy Physics, Radio Astronomy, e-Infrastructures, EuroHPC, Quantum Computing, Cloud Computing, Knowledge Hub
The present document describes the work executed in WP3 and in particular by the SPECTRUM Community of Practice (CoP). This includes: (1) the 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.	

Document Description			
D3.1 Community of Practice – Interim report			
Work Package Number 3			
Document Type	Report		
Document status	Under EC Review	**Version**	1.0
Dissemination Level	Public		
Copyright status	This material by Contributing Parties of the SPECTRUM Consortium is licensed under a [Creative Commons Attribution 4.0 International License]().		
Lead Partner	INFN		
Document link	https://documents.egi.eu/document/4069		
Digital Object Identifier	https://zenodo.org/records/15100587		
Author(s)	• Tommaso Boccali (INFN) • The SPECTRUM CoP Members		
Reviewers	• Jeff Wagg (OCA) • Eric Wulff (CERN)		
Moderated by	• Patricia Ruiz (EGI)		
Approved by	• Xavier Salazar (EGI) – on behalf of AMB		

Revision History			
Version	**Date**	**Description**	**Contributors**
V0.1	26/12/2024	First draft	Tommaso Boccali (INFN)
V0.2	20/02/2025	Added the "findings" section	CoP members
v0.3	22/02/2025	Substituted all the pie charts	Tommaso Boccali (INFN)
v0.4	25/02/2025	Renormalized tables	Tommaso Boccali (INFN)
V0.5	14/03/2025	Internal review	Jeff Wagg (OCA) Eric Wulff (CERN)
V0.6	26/03/2025	AMB Approval	Xavier Salazar (EGI)
V1.0	31/03/3025	**Final**	

Terminology / Acronyms	
Terminology / Acronym	**Definition**
AI	Artificial Intelligence
AMB	Activity Management Board
CoP	Community of Practice
CM	Communications Manager
DB	Database
DoA	Description of Action
DocDB	EGI Document Database
DOI	Digital Object Identifier
EAB	External Advisory Board
EM	Exploitation Manager
GA	General Assembly
GDPR	General Data Protection Regulation
HEP	High Energy Physics
HTC	High-Throughput Computing
HPC	High Performance Computing
KER	Key Exploitable Result
KPI	Key Performance Indicator
MS	Milestone
PD	Project Director
PM	Project Manager
PMO	Project Management Office
PO	Project Objective
QRM	Quality and Risk Manager
RA	Radio Astronomy
RI	Research Infrastructure
SRIDA	Strategic Research, Innovation and Deployment Agenda
SSO	Single Sign On

WG	Working Group
WP	Work Package
WPL	Work Package Leader

Table of Contents

List of Figures

List of Tables

Executive summary

This deliverable describes the work executed in Spectrum Work Package (WP) 3 and in particular by the SPECTRUM Community of Practice (CoP), a place "where new ideas can be proposed and analysed, not only by official experiment representatives, but also by research groups who want to offer new ideas on the path of a viable process for the next generation of experiments, from the HEP and RA and other scientific areas".

In the first 15 months of the SPECTRUM project, WP3 successfully launched a Community of Practice among diverse communities such as High Energy Physics, Radio Astronomy and e-Infrastructure managers and administrators. The CoP is divided into 6 Working Groups, with more than 115 participants.

A survey was submitted to the enlarged domain (High Energy Physics, Radio Astronomy and e-Infrastructures) on July 18th in order to collect inputs from an enlarged community – including researchers, managers, and infrastructure stakeholders; an internal report prepared for MS5 (October 2024) presented preliminary results from the survey. SPECTRUM decided to keep the survey open indefinitely, in order to allow for late respondents[1] whose contributions will be included later in the project.

The main findings reported in this deliverable are the Working Groups discussions, the survey, and from a Knowledge Hub consisting of more than 60 documents. They include suggestions on how to address the career problem in domain-specific computing tasks, the need to evolve our software stacks towards more efficient and performing computing architectures, the need to embrace AI as a base tool for our software, and policy aspects like the need to adhere to FAIR principles and participate into inter- and extra- domains collaborations. A large focus is on the utilization of modern e-Infrastructure, including Commercial and Public Clouds, and HPC systems.

The deliverable will serve as one of the inputs to the second phase of SPECTRUM (months 16-30) and eventually for the

[1] Reports will use a snapshot of the results, taken at a given moment.

1. Introduction

The present document, edited by Work Package 3 (WP3) of the SPECTRUM project, contains the final deliverable of the WP work, in the period of January 2024 to March 2025.

It focuses on several distinct aspects:

- The organisation of the WP work via the Community of Practice;
- The building of a Knowledge Hub;
- The preparation, execution and analysis of a domain-wide survey on the expectations for future computing;
- An initial analysis of the findings and recommendations.

The work in SPECTRUM WP3 has been executed with inputs and discussions from different initiatives, all interested in the same topics on a similar time frame.

The current document, submitted as the deliverable 3.1 of SPECTRUM, is also the basis of additional documents to be submitted by those initiatives; for example:

- A report from the JENA computing group to JENA (see below);
- The input SPECTRUM wants to deliver for the European Strategy For Particle Physics Update, a process initiated by CERN to define the strategy for the next Collider Experiment in Europe.

2. Collaboration with other initiatives

2.1. JENA

Many initiatives, at this point in time, are interested in defining an affordable path to computing for "Big Science" initiatives. While SPECTRUM specifically focuses on High Energy Physics (HEP) and Radio Astronomy (RA), JENA (Joint ECFA NuPECC APPEC Activities) is an initiative from the three main committees at continental level for Collider, Nuclear and Astroparticle Physics, hence with a somewhat broader scope; JENA focuses on areas like detector development, common roadmaps, and from 2022 Computing. Working Groups (WGs) were created in JENA to discuss the computing scenario and its possible commonalities in the three scientific domains, which are very close to SPECTRUM's: HEP is in the JENA landscape, while RA is closely related.

The JENA Computing program of work foresaw a document to be submitted to the JENA plenary assembly by the end of 2024. This document includes the submission of a survey to the relevant communities. This activity is well in line with the SPECTRUM WP3, which motivated the decision to join efforts with JENA and create a single survey together. At the level of documents, it has been decided that documents from one initiative will be reviewed by the other and vice versa in order to make sure a consistent message is delivered.

A final discussion of the JENA computing reports is expected at the JENAS meeting in April 2025. Finally, WG Chairs in JENA participate actively in SPECTRUM WGs and report about the activities there. Symmetrically, SPECTRUM members are in JENA WGss.

2.2. WLCG

The Worldwide LHC Computing Grid (WLCG) foresees and coordinates the distributed computing infrastructure which processes data from the LHC Experiments at CERN; as such, it is particularly interested in the evolution of computing for HEP in the mid-to-long term future. The WLCG organises meetings with European Computing Centers (for example those from EuroHPC JU), where discussion items are also of interest to SPECTRUM. The link is maintained by the fact that many SPECTRUM collaborators are also part of

WLCG, with (notably) the WP3 Leader being the deputy-lead of the WLCG project. Three such meetings happened in 2024-2025, with the **US**, the **EU** and the **Global** communities.

WLCG is expected to submit inputs to the European Strategy for Particle Physics Update (**ESPPU**, 2024-2026), an initiative scheduled only recently that will produce a roadmap for Particle Physics in Europe. Hence, the collaboration with WLCG is important in order to submit coherent inputs.

2.3. The ESCAPE Project and the ESCAPE Open Collaboration

The ESCAPE EU Project (**link**) was active from 2019 to 2023 to address the Open Science challenges shared by ESFRI facilities (SKA, CTA, KM3Net, EST, ELT, HL-LHC, FAIR) as well as other pan-European research infrastructures (CERN, ESO, JIVE) in astronomy and particle physics. ESCAPE actions were focused on developing solutions for the large data sets handled by the ESFRI facilities.

Many members of SPECTRUM were active in ESCAPE, and are now joining the ESCAPE Open Collaboration (**link**). SPECTRUM considers the possibility to engage with a larger and more diverse community than that of ESCAPE to gather insight on science processes beyond its reference domains.

2.4. The Square Kilometre Array Regional Centre Network (SRCNet)

Construction activities have begun on the **Square Kilometre Array (SKA)**, which will consist of two radio interferometers in the Karoo desert of South Africa and Inyarrimanha Ilgari Bundara of Western Australia. The observatory is expected to generate around 700 PB of science data products each year. After initial HPC data processing at the science processing centres in Cape Town and Perth, the data will be transported to the SRCNet nodes, which will generate the advanced data products and serve as the interface with the broader scientific community.

Some members of SPECTRUM are involved in the development and prototyping of the SRCNet architecture. There are several European countries who are anticipated to host nodes of this network, for which the resources are being contributed through an in-kind model similar to that of the WLCG.

3. The Organization of the Community of Practice (CoP)

3.1. Context of the CoP

3.1.1. Definition of the CoP

The concept of a Community of Practice (CoP) was introduced by **Etienne Wenger** in the 1990s, defining CoPs as “groups of people who share a concern or a passion for something they do and learn how to do it better as they interact regularly.” The European Commission followed Wenger's ideas in 2021 by publishing the “**Communities of Practice Playbook**”. In this guide, the Commission provides a framework for structuring CoPs in the context of projects funded by the European Commission, as the SPECTRUM project. Therefore, the SPECTRUMCoP is a place “*where new ideas can be proposed and analysed, not only by official experiment representatives, but also by research groups who want to offer new ideas on the path of a viable process for the next generation of experiments, from the HEP and RA and other scientific areas*”.

The success of the SPECTRUMCoP will be measured using indicators at different times: from the start, such as the number of participants and the number of working groups; at the end of the SPECTRUM project, such as the number of entities that have contributed to the Technical Blueprint and the SRIDA (to be presented in future deliverables); and after the project, such as the number of institutions that have signed the community charter.

3.1.2. Challenges and goals of the SPECTRUM CoP

The amount of data gathered, shared and processed in frontier research is set to increase steeply in the coming decade, leading to unprecedented data processing, simulation and analysis needs. In particular, High Energy Physics (HEP) and Radio Astronomy (RA) are gearing up for groundbreaking instruments, necessitating infrastructures many times larger than the current capabilities. In this context, SPECTRUM brings together leading European science organisations and e-Infrastructure providers to formulate a Strategic Research, Innovation, and Deployment Agenda (SRIDA) along with a Technical Blueprint for a European compute and data continuum. This collaborative effort is set to create an Exabyte-scale research data federation and compute continuum, fostering data-intensive scientific collaborations across Europe.

In order to fulfil SPECTRUM goals, the project needs to have the informed opinions and plans from research communities; in principle HEP and RA are the target communities, but since (*a posteriori*) it was understood that no other domain specific project was granted in the same EU call for proposals, it was decided to enlarge the contacted domains with a wider scope, including the JENA domains (see above) and in general accepting answers from any research community.

The concept of a Community of Practice was selected when designing the project to collect the informed opinions and the perspectives from the researchers and the e-Infrastructure managers.

In our view the CoP is meant to:

1. Gather together people with an interest in discussions about the future needs and the tools required to implement them, considering also the views from the e-Infrastructures;
2. Try to reach as many potential participants / researchers as possible, by submitting a domain-wide survey to the broader community.

The intention is to have the CoP being active not only during the SPECTRUM project (30 months), but also remaining afterwards as a recognized place of communication and synchronisation among the communities.

3.1.3. Calendar of the CoP during the first 15 months of the SPECTRUM project

- May 21st, 2024: first meeting with WG co-chairs (**CoP link**)
- From May to September 2024: some 20+ meetings (see Google calendar **link**)
- September 26th, 2024: plenary WG meeting for the preparation of the SPECTRUM Meeting in Lecce (**CoP link**)
- For the period November 2024 to March 2025: Plenary meetings every 15 days (**CoP link**)

3.2. Creation of the Working Groups (WGs)

3.2.1. Creation of the Expert Group

The first step towards the creation of the CoP was the definition of the Working Groups, since the CoP would call directly for participation in them. There are many possibilities for working groups, whenever you spend some time and effort imagining them. One could imagine vertical (domain specific) WGs – for example having one WG for HEP, one for RA and one for e-Infrastructures; the approach, while solid, does not go into the direction of a better cross-community understanding. One could go horizontal (technology driven) – for example having WGs on Storage, Processing, etc. Also at the level of details, one can easily imagine tens of WGs, including subjects like training, dissemination, and social aspects like careers.

In order to facilitate a decision, a provisional expert group was established, composed of one person from each partner of SPECTRUM WP3. These people were involved in discussions on two strategic directions:

1. How to select a number of Working Groups
2. How to reach the wider community when requesting for participation

By the end of March 2024, the group was able to focus on a proposal with 6 Working Groups, as in the table below; this has been a compromise between the details one would like to discuss, and the time availability of contributors (who, when selected from the community, are unpaid volunteers).

Table 1: the SPECTRUM CoP Working Groups, with a list of potential discussion topics.

WG1: Data Management and Access • Data Management • Data Access Protocols • Data Archiving • Security	**WG4: SW tools** • Machine Learning Frameworks • Multithreading Frameworks • Multi-Node Tools • Compilers, toolchains, … • Quantum computing tools and frameworks • Code Management Practices
WG2: Workflow management and organisation • Resource Discovery and Workflow Submission • Resource Allocation • Complex Workflows	**WG5: Scientific Use cases** • Typical Use Cases • Requirements and Needs • Best Practices Collection • Data Fluxes and Paths
WG3: Compute Environment • Expected Tools and Services • Facility Expectations • Edge Services • Library Provisioning	**WG6: Facilities** • HPC Centers • Access to Quantum Computing Hardware • Access to Commercial and Public Clouds • Sustainability • Security

WG5 is highlighted since it is in connection with SPECTRUM WP5, as a means to collect additional use cases with respect to those internal to the project; hence it is to be used as an information gathering point for WP5. The results are not reported in this document. WG6 is also highlighted, since it is meant to collect feedback from a e-Infrastructures (HTC/HPC centers, Public Clouds, Quantum Computing Centers), collecting the "offer" which in the second part of SPECTRUM will be matched with the "demand" from the user communities.

3.2.2. Creation and use of a domain wide mailing list

A mailing list was hence created to inform the community about the launch of the SPECTRUMCoP, and to ask for contributions both at the level of the CoP and for the survey. By the end of April 2024, more than 100 email addresses were collected, many of which pointing to mailing lists of initiatives like experiments, instruments and boards. A full list cannot be disclosed due to GDPR, but is available if requested for project evaluation.

The mailing list was used twice: once to call for participation to the CoP, the second to ask later for the compilation of the survey.

3.2.3. Composition of the WGs

The call for CoP participation resulted in a good response from the community. By June 30th 2024, the launch date of the CoP, 76 participants were listed to the working groups; since some of them subscribed to more than one working group, the number of unique participants was 56. The number is not final: since then, more participants have been added on a weekly basis. At the beginning of 2025, the number of unique participants was 65 with more than 119 subscriptions.

We decided to appoint people linked to SPECTRUM, and hence paid (partially) by the project, as chairs in each WG (at least one, and two when possible). The current status of chairs per WG is shown in Table 2.

Table 2: the SPECTRUM CoP Working Group chairs.

WG	Chair(s)
WG1 (Data Management and Access)	BAGNASCO Stefano (INFN, ET, Virgo)
WG2 (Workflow management and organisation)	DELL'AGNELLO Luca (INFN CNAF-T1)
WG3 (Compute Environment)	BOZZI Concezio (INFN, LHCb, JENA-HPC)
WG4 (SW tools)	SWINBANK John (Astron) VILOTTE Jean-Pierre (CNRS)
WG5 (Scientific Use cases)	FERRARI Chiara (CNRS, OCA) GIRONE Maria (CERN)
WG6 (Facilities)	HOPPE Hans-Christian (FZJ)

4. The Knowledge Hub

4.1. Introduction: Definition of Knowledge Hub in SPECTRUM

The collaborative platform enabling SPECTRUMCoP members to communicate and share important documents is called the Knowledge Hub. This Knowledge Hub takes the form of a Confluence platform organised primarily around working groups. It is designed to serve as a repository and an active platform where information, experiences, best practices, and research results are gathered, shared, and organised. The Knowledge Hub plays a pivotal role in bridging the gap between different working groups of the SPECTRUMCoP and external communities (e.g. JENA). In the SPECTRUM context, the Knowledge Hub ensures that contributions and documents from various experts and organisations, especially from High-Energy Physics and Radio Astronomy, are systematically captured and made available to all members.

The Knowledge Hub helps to facilitate collaboration across the different WGs by integrating various platforms and tools for document storage, discussion, and knowledge exchange. This ultimately supports the broader goal of delivering a Strategic Research, Innovation, and Deployment Agenda (SRIDA) and a Technical Blueprint for a European compute and data continuum.

4.2. Links between the SPECTRUMCoP and the Knowledge Hub

The setting up of the WGs is closely linked to the implementation of the Knowledge Hub. In order to ensure the best possible coherence between these two SPECTRUMCoP structures, the SPECTRUM WP3 guarantees that the different WGs understand the organisation of the Knowledge Hub in order to contribute efficiently, thus avoiding working in separate silos.

4.2.1. Links with WGs

During the implementation phase of the WGs, the first few months were dedicated to organising the first meetings of each WG, ensuring that each member had received and understood all of the information related to the SPECTRUMCoP. We therefore chose to have the launch managed by each WG chair. To ensure the coherence of each WG, a person from the SPECTRUM project supervising all the WGs was present if possible.

The WG meetings stopped during the summer of 2024, after the launch of the survey. It was decided that in order to relaunch the WG meetings in the future, priority should be given to meetings bringing together all of the WG members in a plenary format. The first meeting of this type was well attended, and was followed by similar plenary CoP meetings up to the end of March 2025.

4.2.2. Regular Updates of the Knowledge Hub

In order to have a centralised place where SPECTRUMCoP members can keep up to date with the progress of the project, SPECTRUM WP3 members may occasionally produce short messages via various communication channels.
These updates are communicated both by email to all SPECTRUMCoP members and directly on the Confluence home page via the 'blog' function. The frequency of these updates is not regular, which means that they can be adapted according to the activity of the project. The form of these posts is adapted so that SPECTRUMCoP members can quickly identify the latest developments of SPECTRUMCoP. The posts are concise and focused on the essentials.

For example, we communicated during the summer of 2024 on the progress of the publication of the survey, or to inform members of an important future inter-WG meeting.
From November 2024, plenary meetings were the main place to get regular updates from SPECTRUMCoP, with presentations and minutes systematically available on the Knowledge Hub.

4.2.3. How the Knowledge Hub Will Help Write the Blueprint and SRIDA

The Knowledge Hub plays a direct role in the development of key SPECTRUM deliverables, including the Blueprint and the SRIDA. These documents will rely heavily on the organised feedback from the community, gathered through surveys, WGs, expert consultations and the document repository. The Knowledge Hub will provide the foundational knowledge required for drafting these documents, making it an indispensable tool for the project.

4.2.4. Statistics of the Knowledge Hub

Most of the modifications and updates to the Knowledge Hub have been made by people internal to the SPECTRUM project. This is primarily because the target communities are already overbusy, and that most of the activities took place during WG meetings and plenary meetings, rather than asynchronously. The high proportion of people attending the meetings (in relation to the total number of SPECTRUMCoP members) shows that synchronous work is the best way to effectively animate the CoP.It should be noted that more than 60 documents listed in the document repository (see below) were collected by the WGs.

4.3. Tools Used

Between January and May 2024, the WP3 of SPECTRUM considered the most relevant and easiest tools for future SPECTRUMCoP members to implement and use. The aim of these tools is to ensure effective collaboration and knowledge management.

4.3.1. Confluence Platform and SPECTRUMCoP

Several options were considered for the platform used to host the Knowledge Hub. One of them was to use the **Circle.so** platform, which brings together all the tools needed to run a community of practice (videoconferencing, chat, pages, events, etc.). This choice had the advantage of being quite easy to handle and corresponded to the identified objectives, but it was a new platform for the vast majority of future SPECTRUMCoP members to familiarise themselves with and, as a result, it also increased the risk of problems for the platform moderators (SPECTRUM WP3).

We therefore chose to use the Confluence platform, already used for the SPECTRUM project and already hosted by EGI, as the main platform for the Knowledge Hub. This collaborative platform serves as the backbone of the Knowledge Hub, providing an intuitive interface for organising documents, storing feedback, and managing communications across the WGs. The Knowledge Hub home page is accessible **via this link**.

4.3.2. Other tools used

For the meetings held in each working group, we have decided, for reasons of flexibility, to allow the chairs of the WGs (see list above) to use the tool they prefer, even if most of them have chosen to use the Zoom platform. For plenary meetings involving all SPECTRUMCoP members and meetings dedicated to WG chairs, we also use the Zoom platform.

For asynchronous communications, we initially envisaged using the Slack platform provided by EGI for all SPECTRUM project members. But the risk of seeing these tools rarely used by members (the application has to be opened every time if SPECTRUMCoP members want to access the discussions) was significant. As a result, we chose to communicate more conventionally by email. We therefore created mailing lists using Mailman, which is hosted by EGI. There is a mailing list for each WG, a list dedicated to chairs, a list dedicated to the expert group (see above) and a list for external people who want to become members of SPECTRUMCoP.

For the general calendar, we had begun to list all the SPECTRUMCoP meetings on a publicly accessible Indico platform, following the example of the SPECTRUM project, but the need identified was rather to have an overview of all the SPECTRUMCoP meetings accessible at the click of a button. This is why we have chosen to display a Google Calendar module on the homepage of the SPECTRUMCoP Confluence platform, which allows users to see upcoming meetings or important SPECTRUMCoP deadlines easily and to access the calendar link to integrate it directly into their Outlook or Thunderbird applications.

Finally, for the survey, we had initially considered using LimeSurvey hosted by EGI, which has the advantage of being free and open source. But as SPECTRUM is funded by Horizon Europe, we thought it would be more relevant to use the platform set up by the European Commission, EU Survey, as detailed later in this document.

4.3.3. Document repository

One of the main functions of the Knowledge Hub is the document repository which has been set up to enable the gathering of important and relevant documentation in accordance with the objectives defined in the SPECTRUMCoP charter. SPECTRUM has therefore developed a template form to be completed for each document analysed by the WGs to ensure that all key information is easily accessible. Setting up a document repository may in the future enable certain synergies with related initiatives such as JENA.

Table 3 presents the references from the Knowledge Hub. This document repository is accessible via this link.

Table 3: SPECTRUM CoP Knowledge Hub corpus.

#	Reference
KH01	The DUNE Collaboration (2022). *DUNE Offline Computing Conceptual Design Report* [Technical Report]
KH02	Dave Casper et al. (2022). *Software and Computing for Small HEP Experiments* [Workshop Summary]
KH03	ATLAS Collaboration (2022). *ATLAS Software and Computing HL-LHC Roadmap*
KH04	J. Adelman-McCarthy et al. (2023). *Extending the distributed computing infrastructure of the CMS experiment with HPC resources* [Conference Paper]
KH05	Maria Girone. (2019). *Common challenges for HPC integration into LHC computing* [Technical Report]
KH06	Fernando Barreiro Megino et al. (2022). *US ATLAS and US CMS HPC and Cloud Blueprint* [Technical report]

KH07	A Beche et al. (2018). Supercomputers, Clouds and Grids powered by BigPanDA for Brain studies [Conference Paper]
KH08	C. Acosta et al. (2023). Integration of the Barcelona Supercomputing Center for CMS computing: Towards large scale production [Presentation]
KH09	Ian Fisk et al. (2024). HEP/HPC Strategy Meeting - USA [Meeting Summary]
KH10	Maria Girone & Tommaso Boccali. (2025). HEP/HPC Strategy Meeting - All Regions [Meeting Summary]
KH11	Andrej Filipcic & Maria Girone & Tommaso Boccali. (2024). HEP/HPC Strategy Meeting - Europe [Meeting Summary]
KH12	Christopher Hollowell et al. (2023). Financial Case Study: Use of Cloud Resources in HEP Computing [Presentation, CHEP 2023]
KH13	Yonatan Kahn et al. (2022). Snowmass2021 Cosmic Frontier: Modeling, statistics, simulations, and computing needs for direct dark matter detection
KH14	Peter Couvares. (2021). Gravitational-Wave Data Analysis: Computing Challenges in the 3G Era [Technical Report]
KH15	TWG9 Open Science and Data. (2024). NuPECC TWG9 : Open Science and Data
KH16	HEP Software Foundation. (2020). HL-LHC Computing Review: Common Tools and Community Software [Technical report]
KH17	V. Daniel Elvira 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
KH18	Mohammad Atif et al. (2023). Evaluating Portable Parallelization Strategies for Heterogeneous Architectures in High Energy Physics
KH19	Max Fischer et al. (2020). Effective Dynamic Integration and Utilization of Heterogenous Compute Resources [Conference Paper]
KH20	Markus Diefenthaler et al. (2022). Future Trends in Nuclear Physics Computing [Editorial]
KH21	Tadashi Maeno et al. (2024). Utilizing Distributed Heterogeneous Computing with PanDA in ATLAS [Conference Paper]
KH22	The ALICE Collaboration (2019). The ALICE Collaboration: Evolution of the O2 system [Technical Report]
KH23	J. Wagg et al. (2021). SKA1 Scientific Use Cases [Technical document]
KH24	Tommaso Boccali et al. (2019). Summary of the cross-experiment HPC workshop [Workshop Summary]
KH25	Valerio Bertone et al. (2024). Nuclear Physics Tools – Machine Learning, Artificial Intelligence, and Quantum Computing [Report]
KH26	EuroHPC (2024). EuroHPC Summit 2024: Interconnecting EuroHPC Supercomputers for Scientific and Industrial Advancement [Presentation]
KH27	SKA Regional Centre Steering Committee (SRCSC) WG6, Task Package 1 (2021). SKA Science Regional Centres Community Input Questionnaire [Survey]

KH28	I. Sfiligoi et al. (2014). *CMS experience of running glideinWMS in High Availability mode* [Conference Paper]
KH29	Antonio Pérez-Calero Yzquierdo (2020). *CMS strategy for HPC resource exploitation* [Conference paper]
KH30	Federico Stagni (2023). *Status of DiracGrid projects* [Presentation]
KH31	J. Salgado et al. (2023). *SKA Regional Centres Network (SRCNet) Software Architecture Document* [Design Description Document]
KH32	Shari Breen et al. (2021). *SKAO Science Data Products: A Summary* [Data Product Summary]
KH33	P. Dewdney et al. (2022). *SKA1 Design Baseline Description* [Design Baseline Description]
KH34	LOFAR (2023). *LOFAR2.0 compared to LOFAR: a short summary* [Technical Summary]
KH35	Roberto Pizzo et al. (2023). *LOFAR2.0 Data Management Capabilities* [Note]
KH36	Interim LOFAR ERIC Council (2021). *LOFAR ERIC Access Policy to Scientific User Services* [Policy Document]
KH37	LOFAR ERIC Council (2023). *Science Data Policy of LOFAR ERIC* [Policy Document]
KH38	Stefano Bagnasco for the Virgo Collaboration and the LIGO Scientific Collaboration (2024). *The Ligo-Virgo-KAGRA Computing Infrastructure for Gravitational-wave Research*
KH39	The ATLAS Collaboration (2024). *Total cost of ownership and evaluation of Google cloud resources for the ATLAS experiment at the LHC*
KH40	Paul Laycock et al. (2024). *Preparatory Phase for the Einstein Telescope Gravitational Wave Observatory: Computing and Data Requirements* [Deliverable]
KH41	CMS Offline Software and Computing (2020). *HPC Resources Integration at CMS*
KH42	Tommaso Boccali et al. (2021). *Enabling HPC Systems for HEP: The INFN-CINECA Experience*
KH43	Amy Roberts et al. (2019). *Dark-matter And Neutrino Computation Explored (DANCE) Community Input to Snowmass* [Meeting Summary]
KH44	Ben Bruers et al. (2023). *Resource-aware Research on Universe and Matter: Call-to-Action in Digital Transformation* [Meeting Summary]
KH45	GANIL (2020). *GANIL Data Policy* [Technical Report]
KH46	GSI/FAIR Collaboration (2023). *Instructions for uploading and linking research data/software at GSI Helmholtzzentrum für Schwerionenforschung GmbH*
KH47	GSI/FAIR Collaboration (2021). *Open source software licences at GSI/FAIR - Guidelines*
KH48	David Rohr, Giulio Eulisse (2023). *The O2 software framework and GPU usage in ALICE online and offline reconstruction in Run 3* [Presentation]
KH49	European Strategy for Particle Physics (2020). *Physics Briefing Book : Input for the European Strategy for Particle Physics Update 2020* [Strategy Report]
KH50	CMS Offline Software and Computing (2022). CMS Phase-2 Computing Model: Update
KH51	Antonio Boveia et al. (2024). Snowmass 2021 Cross Frontier Report: Dark Matter

	Complementarity (Extended Version)
KH52	M. G. Aartsen et al. (2014). *The IceProd Framework: Distributed Data Processing for the IceCube Neutrino Observatory*
KH53	T. Gal et al. (2019). *KM3NeT Report on Documentation Strategy, Environment, and Software* [Deliverable]
KH54	Stefano Bagnasco et al. (2023). *Computing Challenges for the Einstein Telescope Project* [Conference paper]
KH55	Shankha Banerjee et al. (2023). *Environmental sustainability in basic research: a perspective from HECAP+* [Technical report]
KH56	J. Reppin et al. (2021). *Interactive Analysis Notebooks on DESY Batch Resources: Bringing Jupyter to HTCondor and Maxwell at DESY*
KH57	Christoph Beyer et al. (2020). *Beyond HEP: Photon and accelerator science computing infrastructure at DESY* [Conference Paper]
KH58	Max Fischer et al. (2020). *Lightweight dynamic integration of opportunistic resources* [Conference Paper]
KH59	Michael Böhler et al. (2021). *Transparent Integration of Opportunistic Resources into the WLCG Compute Infrastructure* [Conference Paper]
KH60	Tommaso Boccali (2019). *Computing models in high energy physics*
KH61	David Abdurachmanov et al. (2014). *Explorations of the viability of ARM and Xeon Phi for physics processing*
KH62	David Britton, Simone Campana, Bernd Panzer-Stradel (2024). *A holistic study of the WLCG energy needs for the LHC scientific program* [Conference Paper]
KH63	Emanuele Simili et al. (2024). *ARMing HEP for the future Energy Efficiency of WLCG sites (ARM vs. x86)* [Conference Paper]
KH64	David Britton (2024). *Simulating the Carbon Cost of Grid Sites* [Presentation]

5. The Survey – Design and Scope

The SPECTRUM survey has been designed and validated in the WP3 WG meetings, and by design had the following specifications:

- It had to be a single survey covering all the aspects (not to annoy the community with O(10) surveys)
- It had to be well focused and avoid asking the respondents technical details they would not be qualified to answer
- It had to be compliant with GDPR, and in particular the survey data should reside on EU servers
- It had to allow for the uploading of documents
- It had to have features like a multiple stage compilation, since gathering the information to compile it could require time
- As few compulsory questions as possible: users are allowed to leave a blank answer if they are not sure (a blocking behaviour is annoying for the typical user)

Several technical solutions were considered (including open source and commercial ones). The final decision was to use EU Survey, a tool developed by EU Projects and the de-facto official surveying tool at the European Commission.
EU Survey pays a lot of attention to open licensing, and is a free tool one can also deploy on on-premise servers by downloading the code. From the main page:

- *"EUSurvey is the official online survey management tool of the European Commission"*
- *"Intellectual Property: Built by DG DIGIT and funded under the ISA, ISA² and Digital Europe Programme (DIGITAL) EUSurvey is fully open source and published under the EUPL licence. You can download the source code from GitHub: https://github.com/EUSurvey"*

The EU Survey has some limitations, in particular on the number of answers (which is not problematic on our side: we never expected hundreds or thousands of answers) and the complexity of the survey (in terms of conditional branches). Ranches are necessary to use, in order to present each submitter with the best set of questions with respect to his/her experience; EU Survey was considering the survey as "too high complexity", but in practice that never was a problem.

The survey was designed with a structure resembling the WGs, with some additional sections. The full survey has 12 sections, as detailed below:

- An introduction page detailing the scope of the survey
- A GDPR Clause which needs to be accepted
- A section in which users present themselves (name, mail, roles)
- A section in which specific questions are asked about the submitter's areas of expertise
- A section on Authentication and Authorization solutions
- A section on Computing Needs
- A section of Data Needs
- A section on the needed computing environment
- A section about Software Development
- A section about Training and Careers
- A section about Use Case Definition
- A section about e-Infrastructures

The submitter mail in particular will be used in special cases, where clarifications are needed.

Figure 1 shows the initial survey page, with all tabs activated.

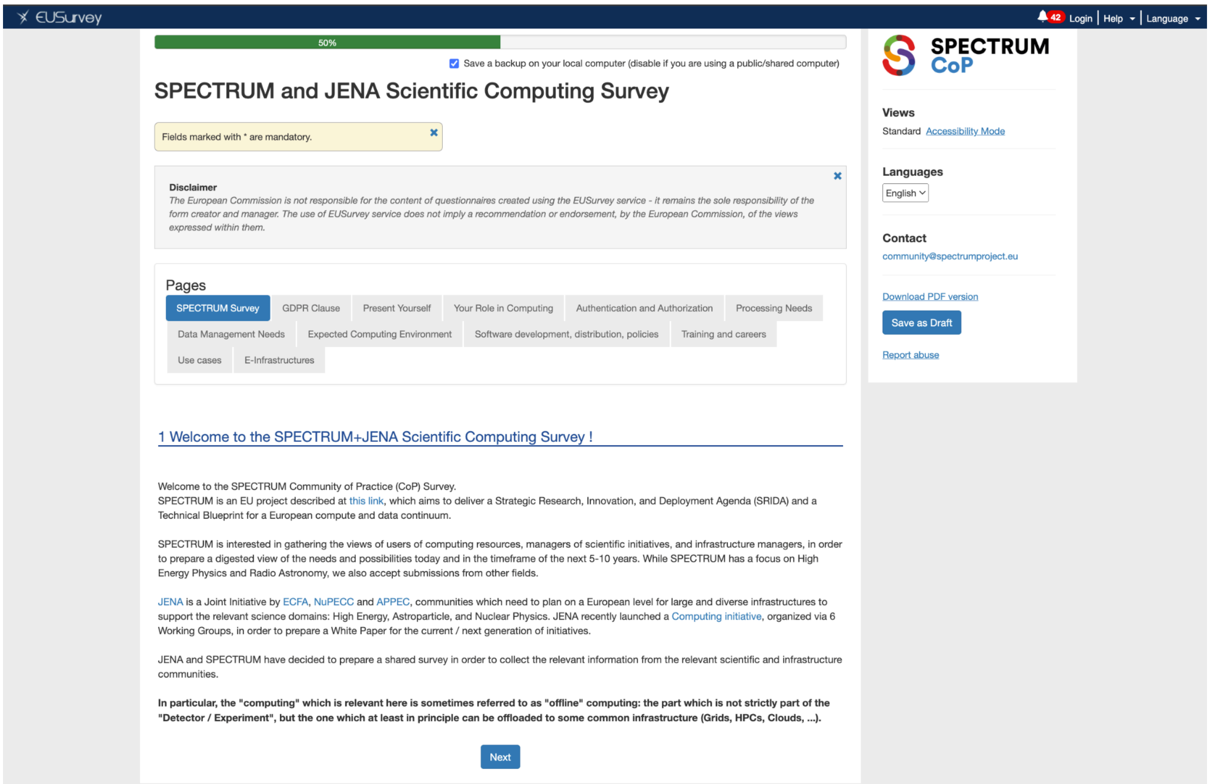

Figure 1: The initial page of the SPECTRUM survey.

Overall, the survey, when printed, is a 38 pages document[2]; but only some are presented to each submitter depending on his/her areas of expertise.

6. The Survey – Current results

In this section, we have drafted the current results from the SPECTRUM Survey. As already mentioned, these results are not final, since the survey will be kept open indefinitely. This was a decision by SPECTRUM management, and will allow us to consider late submissions, even after the end of WP3.

In the first phase of SPECTRUM it was decided not to attempt an in-depth interpretation of the results, giving just the raw results for a later interpretation (see the work plan of WPs 4, 6 and 7); still, a later section in this document presents emerging trends and directions as extrapolated from the survey and the Knowledge Hub.

In general, since the two reference communities in SPECTRUM are High Energy Physics (HEP) and Radio Astronomy (RA), it is desirable to see how the two different communities answer computing-related questions, to probe the distance between the fields. On top of that, there is interest in including different communities (for example those from JENA); hence, it was decided to show for some relevant questions 3 plots/tables: a global one, including all the answers, and two specific from Experimental HEP and Experimental RA. Only when the difference between these is minimal, a single plot / table will be shown.

The total number of submissions at the time of this draft is **78**. This number does not include failed and test submissions.

[2] Full PDF available here.

6.1. Roles

The first set of questions is about the description of the role of the person submitting the survey.

6.2. Researchers and Scientific Initiatives

We chose as possibilities:

- Researcher / user of scientific computing resources (doing analysis, R&D, operations, ...)
- Manager of a scientific initiative (for example, an experiment, an instrument, an observatory)
- Manager of an e-Infrastructure (for example a computing centre, a storage facility, a distributed computing facility)
- Research software engineer (writing, testing and managing code for an initiative)
- Other (please specify)

The distribution is shown in Table 4. From the methodological point of view, all the tables are filled taking into account that there can be multiple answers allowed (multiple checkboxes allowed in the WEB form); hence, the fractions are computed having at denominator the number of respondents who gave at least one valid answer.

Table 4: Categories of respondents to the survey.

Which are the categories which better describe your role(s)?	Answers	Fraction (%)
Researcher / user of scientific computing resources (doing analysis+ R&D+ operations+ ...)	63	84.0%
Manager of an e-Infractructure (for example a computing centre+ a storage facility+ a distributed computing facility)	16	21.3%
Manager of a scientific initiative (for example+ an experiment+ an instrument+ an observatory)	15	20.0%
Other (please specify)	4	5.3%
Research software engineer (writing+ testing and managing code for an initiative)	20	26.7%
Total valid answers	75	--

The predominance of the first category is not unexpected, since the survey was administered to the large scientific community, which consists mostly of researchers and as such users of scientific computing.

Digging further, we were interested in: (1) the scientific domain of provenance; (2) whether they participate in the survey as individuals or representing an initiative; (3) which are the area(s) of expertise and (4) whether the submitter is also a manager of a data centre.

Tables 5, 6 and Figures 2-3 show the answers we received.

Table 5: Scientific domains of the respondents.

Which is/are your scientific domain(s) of expertise (if applicable)?	Answers	Fraction (%)
Experimental High Energy Physics (HEP)	30	40.0%
Experimental Gravitational Waves (GW)	6	8.0%
Observational Astroparticle (not RA or GW)	9	12.0%
Observational Radio Astronomy (RA)	8	10.7%
Theoretical High Energy Physics (HEP)	15	20.0%
Other physics related domains (please specify below)	6	8.0%
Other non-physics related research domains (please specify below)	4	5.3%
Experimental Nuclear Physics (NP)	9	12.0%
Theoretical Nuclear Physics (NP)	12	16.0%
Theoretical Gravitational Waves (GW)	2	2.7%
Total valid answers	75	--

Even if the primary targets of the survey are the HEP and RA communities, it is interesting to see that many others participated; this is in part due to the involvement of JENA, in part to the large diffusion of the survey itself via the mailing list.

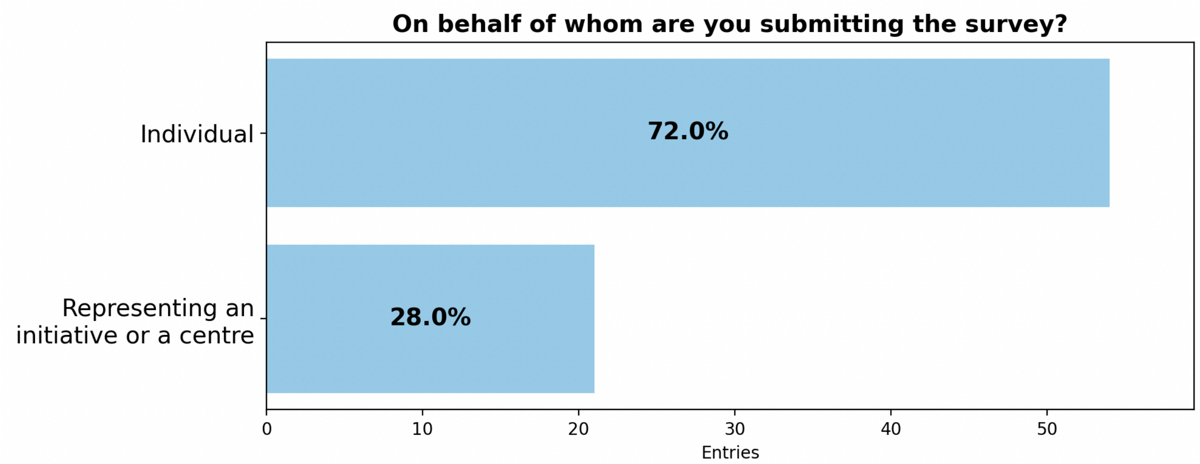

Figure 2: Type of answers (as individuals or representing a community / initiative / center).

The result is consistent with having reached the larger community, and not only selected managers.

Table 6: Areas of expertise.

Please select your areas of expertise, for which you can answer technical questions:	Answers	Fraction (%)
Workload Management	18	28.6%
Data Management	20	31.7%
Computing Environment	24	38.1%
Software, software development and software management	49	77.8%
Authentication and Authorization	9	14.3%
Training and careers	17	27.0%
Total valid answers	63	--

The result is consistent with expectations: in our fields most of the researchers are doing some kind of software development (for example when writing analysis code); on the other hand, very few experts are needed in Authentication and Authorization, probably down to 1-2 experts per initiative; for the others, AAI is just a service they need to use to connect to resources.

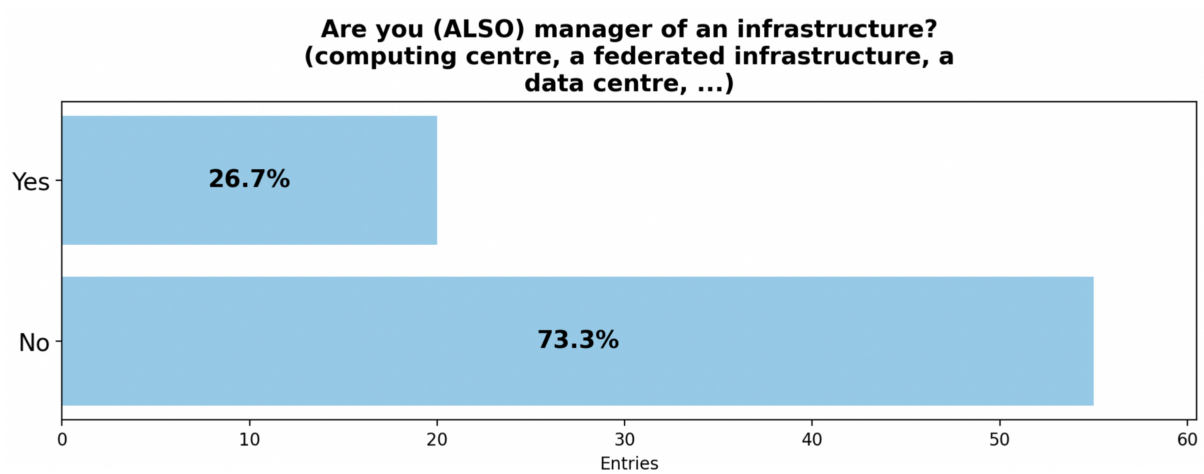

Figure 3: Fraction of respondents in charge of an e-Infrastructure.

The result shows that ~¼ of the responses we have are from infrastructure managers, which is essential so as not to have just the researchers point of view, but also the one from infrastructures – this is indeed one of the main goals of SPECTRUM: to bring together the two communities.

Finally, we collected the initiatives' names. It is difficult to report from the 75 valid answers; still we try to present the collected results as a word cloud.

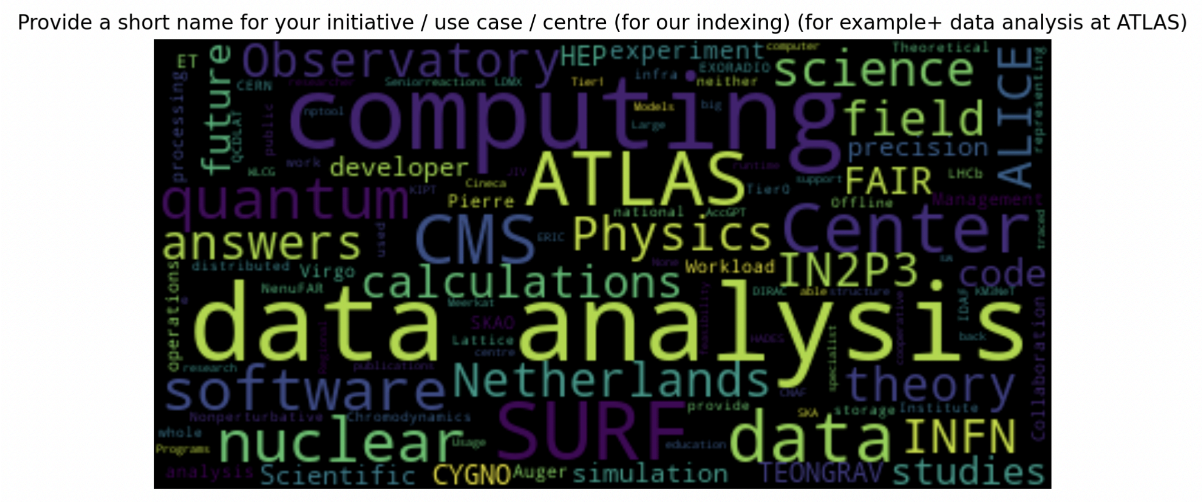

Figure 4: Word Cloud of the initiatives' names.

6.2.1. Authentication and Authorization

Authentication and Authorization Infrastructures (AAIs) are central pieces in today's computing. Users are presented with a distributed computing infrastructure, which requires some form of global identification and trust shared between the centers.
From the technical point of view, there are many solutions which respect the FAIR principles and at the same time allows for a precise auditing of activities in case of malign activities. Still, users prefer to be shielded by this as much as possible, and often see AAIs as unfriendly whenever they require methods more complicated than username/password. It is thus interesting to see which AAIs exist in real environments, being effectively used for research.

In the survey, we asked among other details: (1) which generic methods are supported for AAI; (2) which is the level of support for two/multi factor authentication and (4) which existing tools are used.

Table 7: Types of supported AAIs.

Authentication and Authorization supported method(s) [this includes Workload and Storage management]	Answers	Fraction (%)
Experiment/Institution SSO	7	100.0%
Federated (for example, via eduGain)	6	85.7%
Username / Password	3	42.9%
A locally developed system (please provide details)	1	14.3%
Other (please provide details)	1	14.3%
Total valid answers	7	--

Only 3 out of 7 still use username/password. Probably a good shift from old habits.

Table 8: Tools supported for the AAI.

Which AA tools do you support or make use of?	Answers	Fraction (%)
Indigo-IAM	6	85.7%
CERN-SSO	5	71.4%
Other (specify)	3	42.9%
EGI-Checkin	3	42.9%
Commercial Providers (e.g. Google, Facebook, Apple, Github, ...)	1	14.3%
Total valid answers	7	--

The utilisation of AA tools is sparse; possibly this means the communities could not land on a default solution. On the brighter side, most of the declared solutions are interoperable by design, for example via the AARC Blueprint [link]; this is the case for EGI-Checkin, Indigo-IAM and CERN-SSO.

Table 9: 2/Multi factor authentication capabilities.

All: Does your initiative need 2FA (for example due to some internal policies)?	Answers	Fraction (%)
No	4	57.1%
Yes (please comment how)	3	42.9%
Total	7	100%

Two/Multi factor authentication is getting traction in our communities; it is not generally a user friendly solution, but it seems to be the most adequate for matching GDPR requirements[3].

6.2.2. Processing Needs

In this section of the survey, users described their processing needs (type and amount of resources, typical workflows, etc.).

Initially, submitters were asked to identify the generic processing they need to do, and whether they work interactively, via batch systems or via other means.

[3] "Two-factor authentication should preferably be used for accessing systems that process personal data. The authentication factors could be passwords, security tokens, USB sticks with a secret token, biometrics, etc." [link]

Table 10: Typical applications.

Typical application type(s)	Answers	Fraction (%)
Offline data processing (calibration, reconstruction, ...)	13	86.7%
High-level data analysis (visualization, fitting, ...)	9	60.0%
Machine Learning/AI training/inference	11	73.3%
Near real time processing	4	26.7%
Monte Carlo simulations	13	86.7%
Model/theory calculations	5	33.3%
Other (please specify)	2	13.3%
Total valid answers	15	--

There is no clear predominance between workflow types, which means that a technical solution embracing most of them would be needed.

Table 11: Typical patterns for access to computing resources.

Typical computing access type	Answers	Fraction (%)
Interactive (shell/text)	8	53.3%
Batch (either directly submitted from a computing center, or via some experiment / initiative middleware)	13	86.7%
Via some services (for example via Web services as Jupyter notebooks)	5	33.3%
Interactive (graphical - for example using X or VNC)	3	20.0%
Other (please specify)	2	13.3%
Total valid answers	15	--

For what concerns access type, traditional batch and shell accesses are still the majority.

Users were asked about the job granularity: whether their processes can only execute in serial (single core) mode, in multi-core or multi-node modes.

Table 12: Granularity of job submission.

Granularity of job submission	Answers	Fraction (%)
Single core	12	85.7%
Multi/Many cores (on a single node)	11	78.6%
Whole node	7	50.0%
Multi node	3	21.4%
Total valid answers	14	--

Multi node is not preferred, and a further check shows that the answers in that category are not coming from a single domain.

The next question was about total needs for resources. The answers range between 200,000 core hours per year, to 5,000,000,000 core hours per year. The next plot shows the distribution in Log_10 scale, given the large range.

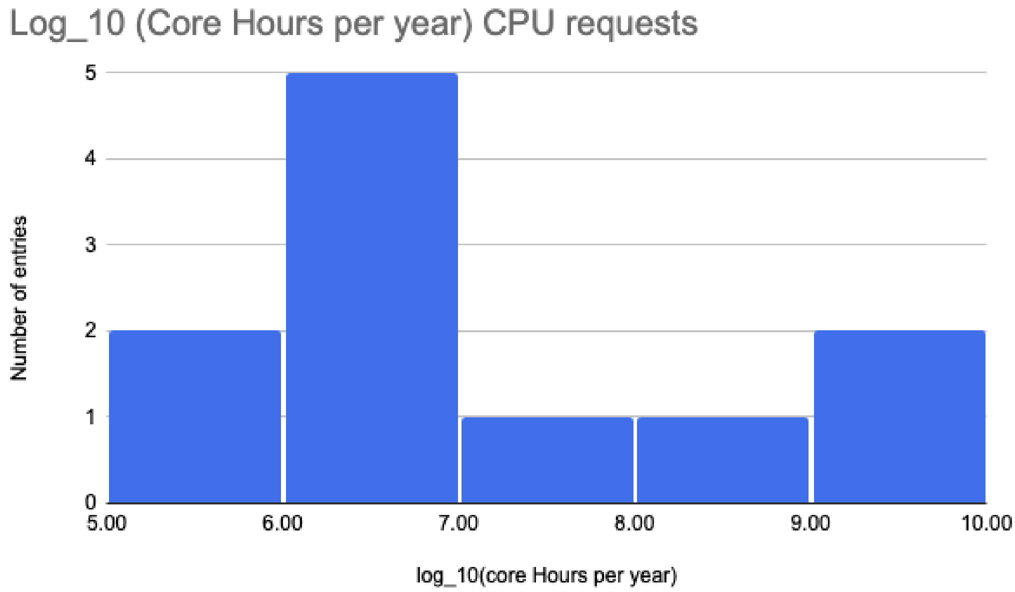

Figure 5: Size of processing requests (note the log scale for the x axis).

A similar question about GPU needs brought very few answers: most of the initiatives are NOT ready today for GPU utilisation.

But where are these resources used? Table 13 shows the most used facilities.

Table 13: Facilities used for scientific processing.

All: Which facilities are you using?	Answers	Fraction (%)
Commercial Clouds	2	13.3%
Dedicated clusters (for example owned by the initiative)	8	53.3%
Grid facilities	13	86.7%
HPC Centres	12	80.0%
Institutional Clouds	6	40.0%
Other (please specify)	1	6.7%
Total valid answers	15	--

The most used options, in both HEP and globally, are Grid facilities and HPC centers; in general, there is not an overwhelming majority for any category.

When using external resources (Grids, HPC centers, …), an important factor is the stability of the allocation: experiments are multi annual initiatives, and need to have guaranteed access to resources for a reasonable time period.

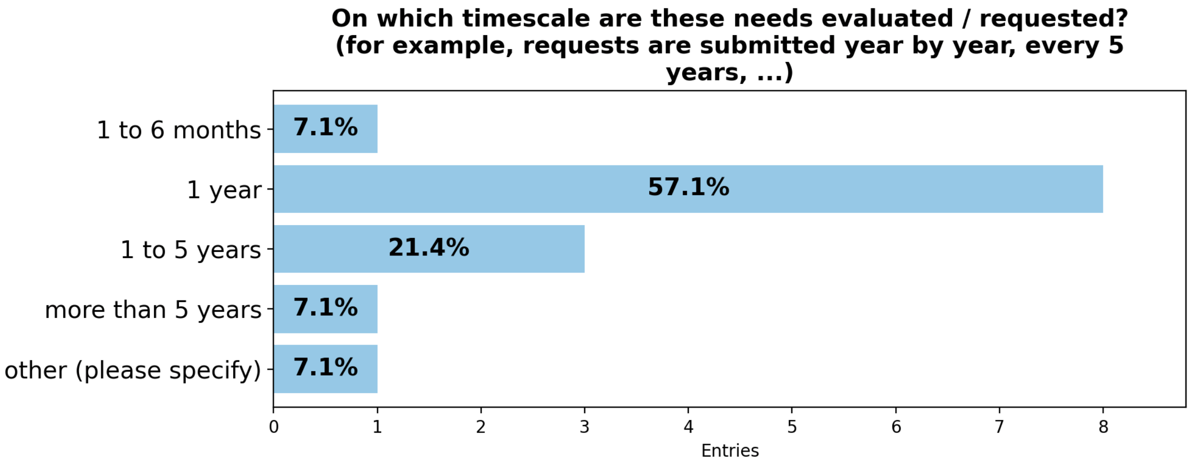

Figure 6: Time span of computing requests.

Table 14: Preferred access to the resources.

Which is (are) the most common current resource allocation pattern(s)?	Answers	Fraction (%)
Fair share on batch systems	14	100.0%
Dedicated VMs or servers	6	42.9%
Opportunistic	9	64.3%
Preemptable jobs	2	14.3%
Total valid answers	14	--

The resources are best used when available as a fair share on batch systems (at Grid or HPC centers); quite surprisingly opportunistic utilisation (using external and not guaranteed pledged resources, usually for free) is very popular.

Coming to Workload Management Systems (WMSs), high level, experiment-wide WMSs are quite popular, but there is no clear winner between the choices.

Table 15:. Supported Workload Management System / Provisioning System.

All: Do you have a higher level Workload Management System (WMS:for example, an initiative-specific layer handling a distributed computing infrastructure and interactive with lower level SLURM, HTCondor or similar)?	Answers	Fraction (%)
DIRAC	7	53.8%
HTCondor	7	53.8%
Openstack	5	38.5%
Other (please specify)	4	30.8%
Panda	2	15.4%
glideinWMS	4	30.8%
k8s	2	15.4%
Total valid answers	13	--

In most cases, as expected in data-intensive sciences with distributed data silos, access to external services is needed. Comparing this with the possibility offered by some HPC centres, this is a source of concern today.

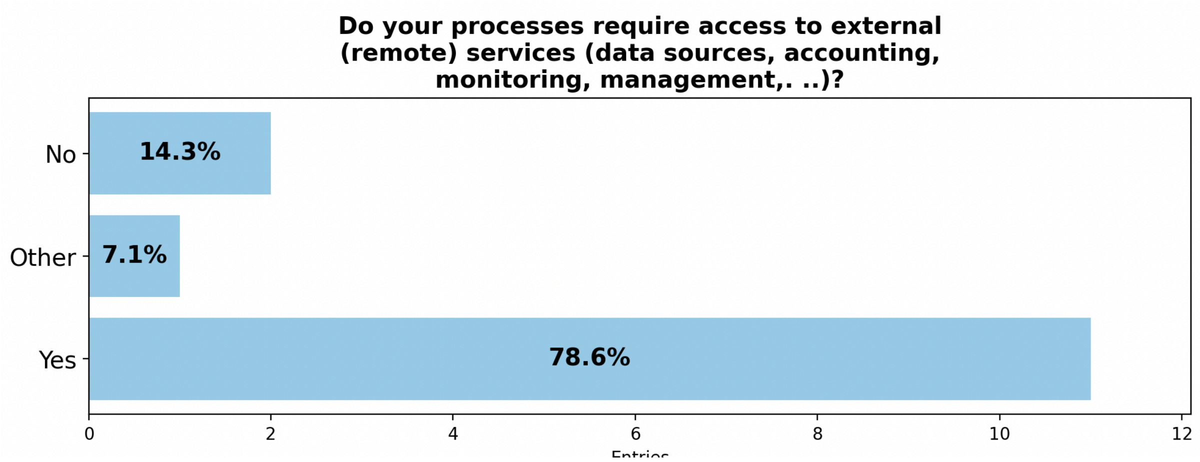

Figure 7: *Requests for remote data access during processing.*

Accessing (large) datasets in particular is a complex operation, and different initiatives are using different approaches, from local data reading to remote streaming. Resource discovery in a distributed computing environment is not a trivial task, which eventually requires a push (a managed catalogue of resources to submit to) or pull (auto registration) mode. The survey shows quite a perfect balance between these two methods.

Table 16: Supported methods for resource discovery.

All: How do you discover available resources?	Answers	Fraction (%)
Nodes / centres registering directly the computing effort (for example, HTCondor startds joining a schedd)	6	42.9%
Online resource catalogue(s) maintained by collaborators	4	28.6%
Other	2	14.3%
Rely on Cloud Provisioning	3	21.4%
Rely on GRID Provisioning	8	57.1%
Static List	4	28.6%
Total valid answers	14	--

Submitters were requested to name their initiative / use case. The next "word cloud" summarises the inputs.

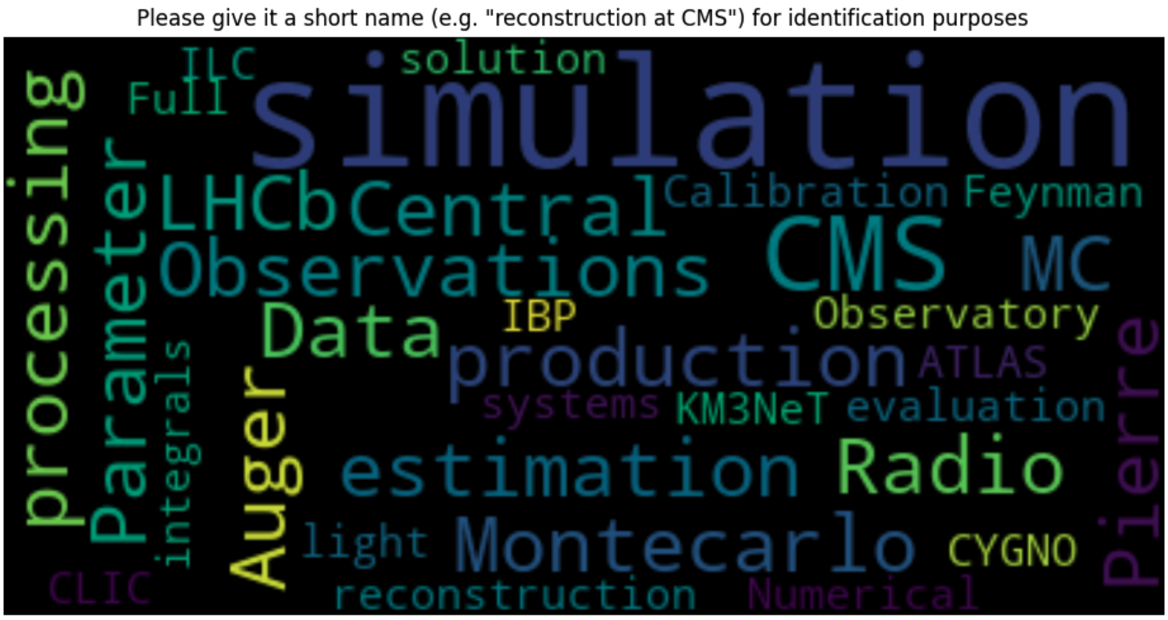

Figure 8:. Word Cloud of typical applications.

Following this part, questions were asked about the minimum resources needed to execute their workflows: memory, local disks, I/O within and outside the center.
Answers are mostly in line with typical resource acquisitions, with only a small fraction needing nodes with large memory, very large local disks or large bandwidth.

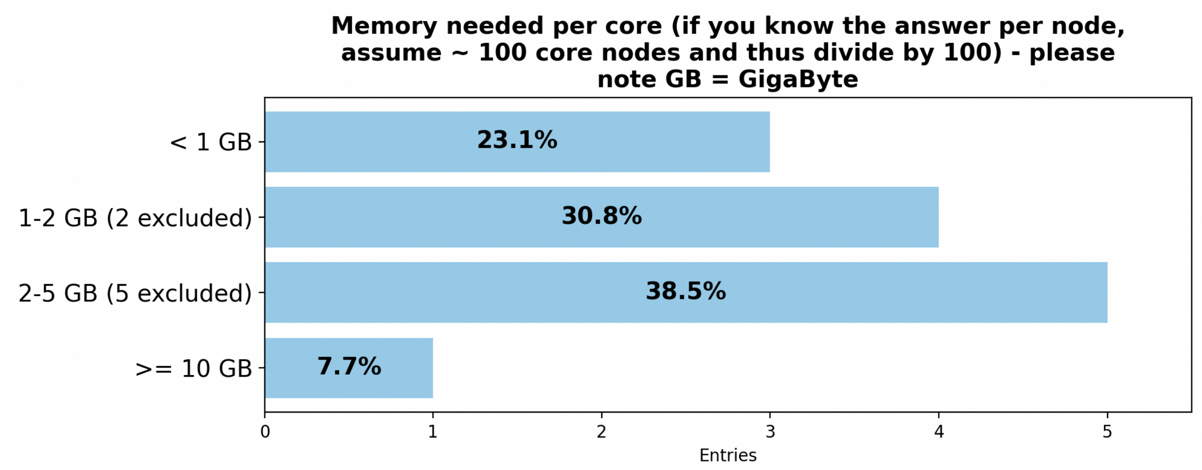

Figure 9: Memory-per-node requests.

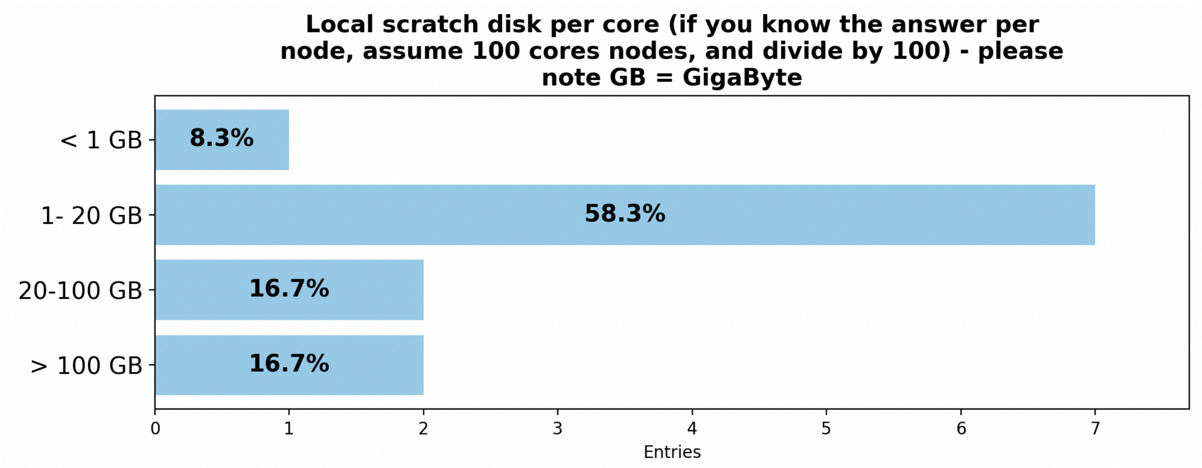

Figure 10: Local scratch disk required.

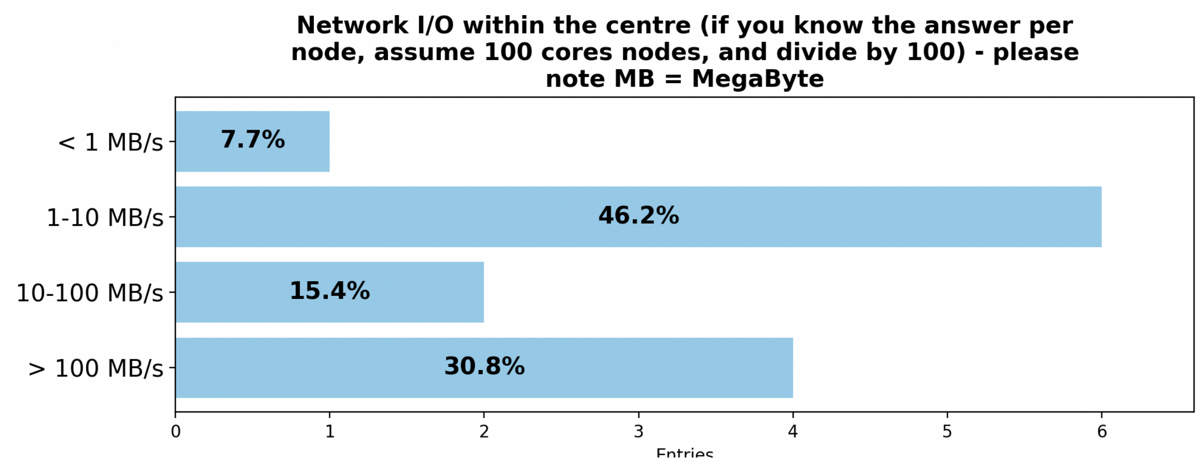

Figure 11: Internal I/O needs (node to node, node to storage).

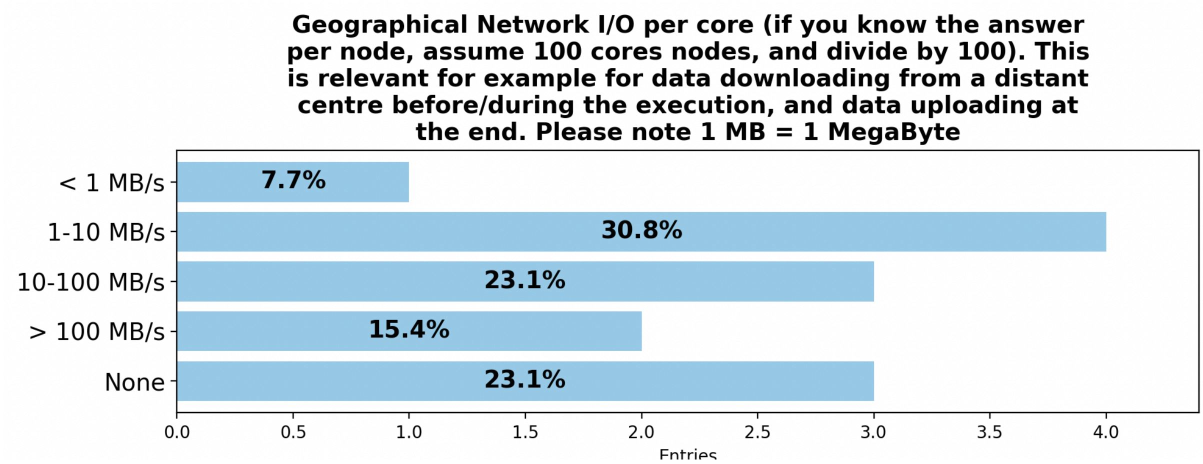

Figure 12: External I/O needs per processing core.

6.2.3. Data Management Needs

Data management is a difficult task in large, data intensive initiatives. Depending on the size, datasets up to multiple exabytes must be distributed among tens or hundreds of storage centers, by moving files on regional, continental and intercontinental routes. Data can be anything between files, DB tables, images, or a combination.

The first question on data management needs in the survey was about the data sizes acquired/generated per year; this, together with the total initiative length, can give a glimpse of the total sizes. Due to the large spread, the answers are represented using a logarithmic scale. The plots show up to Exabytes/yr, for as many as 15 years.

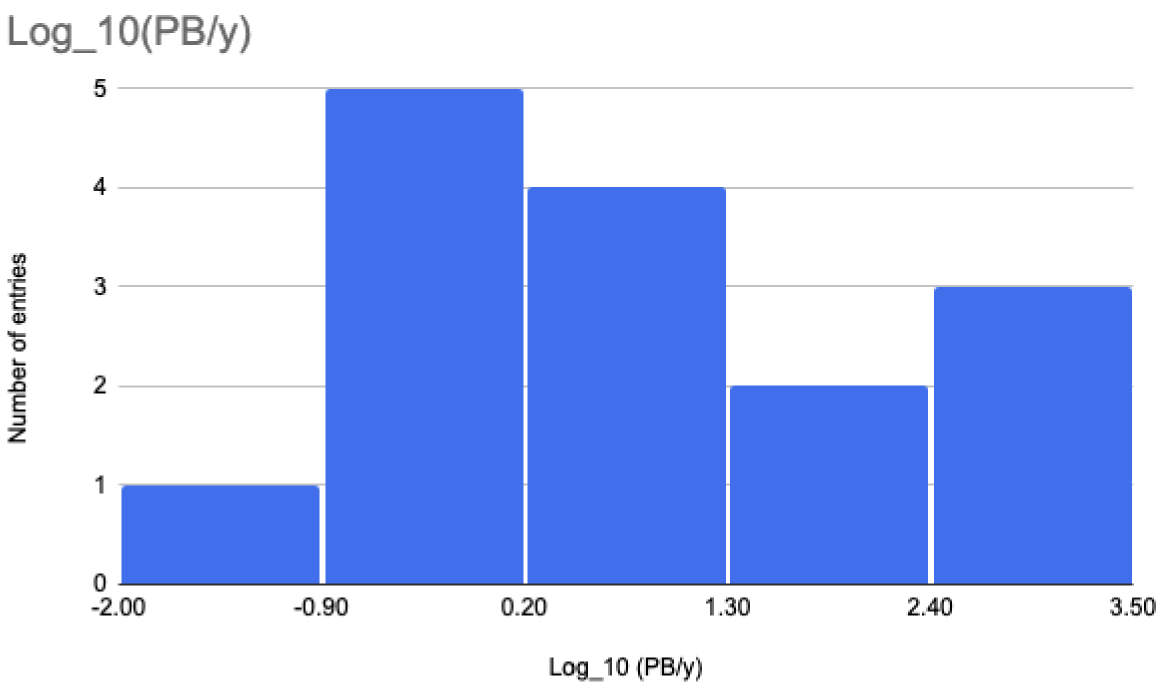

Figure 13: Storage requests (note the log scale on the x axis).

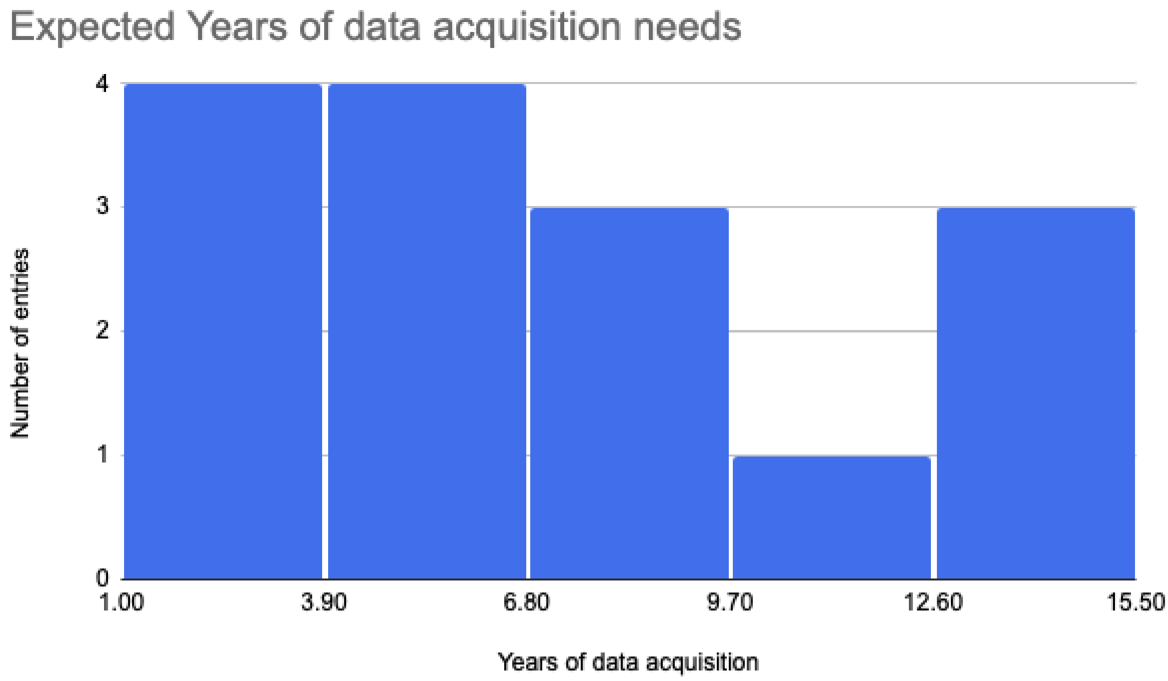

Figure 14: Time span data acquisition / generation.

Data is generated in multiple locations (supposedly simulation data) or at a single experimental site (where a detector / instrument is located).

Table 17: *Data production / generation sites.*

Where is the data produced?	Answers	Fraction (%)
At a single experimental site	10	55.6%
At multiple experimental sites	8	44.4%
At multiple computing centres (for example, in case of simulations)	12	66.7%
At a single computing centre (for example, in case of simulations)	3	16.7%
Total valid answers	18	--

Data has to be ingested into the distributed computing system, from the generation site(s). We asked the submitters which is the average bandwidth of data ingestion. In most of the cases (> 50%), it is very high, exceeding 1 GB/s and larger than 1 TB/s in 7% of the cases. Larger numbers require specifically built storage systems, and are hard to deploy at standard centres.

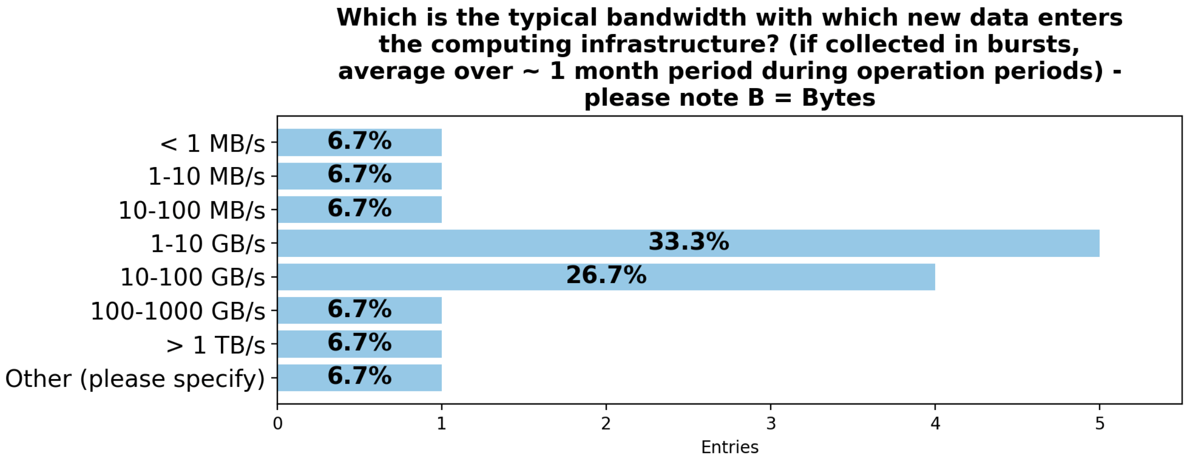

Figure 15: Bandwidth needed for the ingestion of the produced / acquired data.

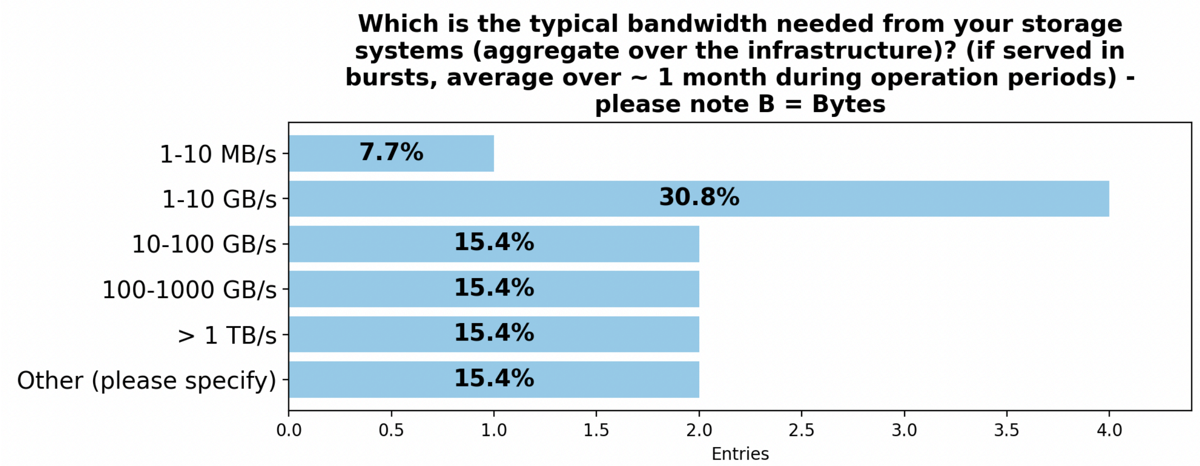

Figure 16: Typical aggregated bandwidth needed from/to storage, summed over the distributed infrastructure if any.

Data management solutions, on the initiatives' side, is essential with such large data samples. Only 20% of the initiatives, correlated with the ones with small data ingestion rates, can operate without a Data Management System. Most initiatives use open source tools for that, but in-house and commercial solutions are not excluded.

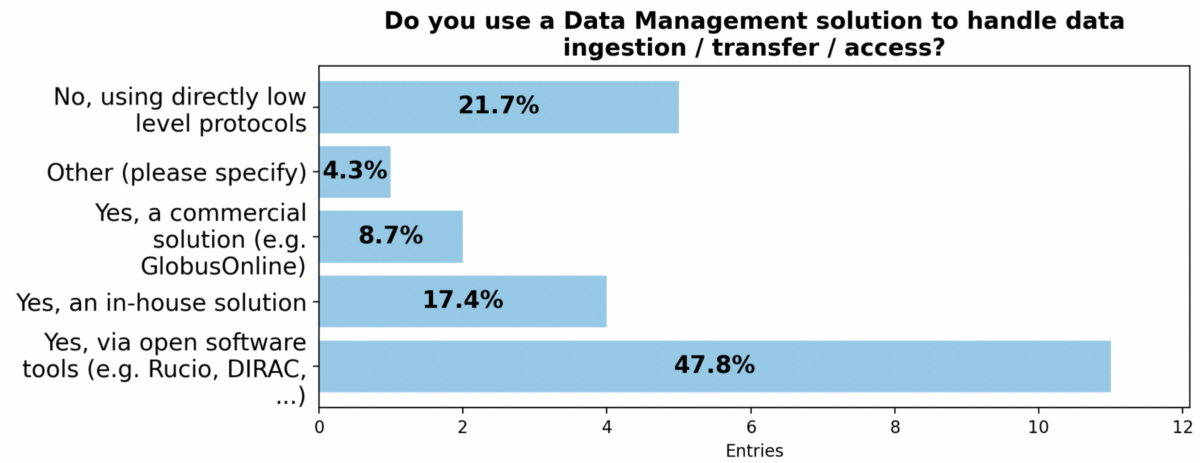

Figure 17: Data management supported solutions.

Data movers are the work-horse in data transfer: their simple but essential task is to move one unit of data (a file, typically) from storage A to storage B. FTS emerges as a clear go-to choice; initiatives working without FTS are those with smaller data movement needs.

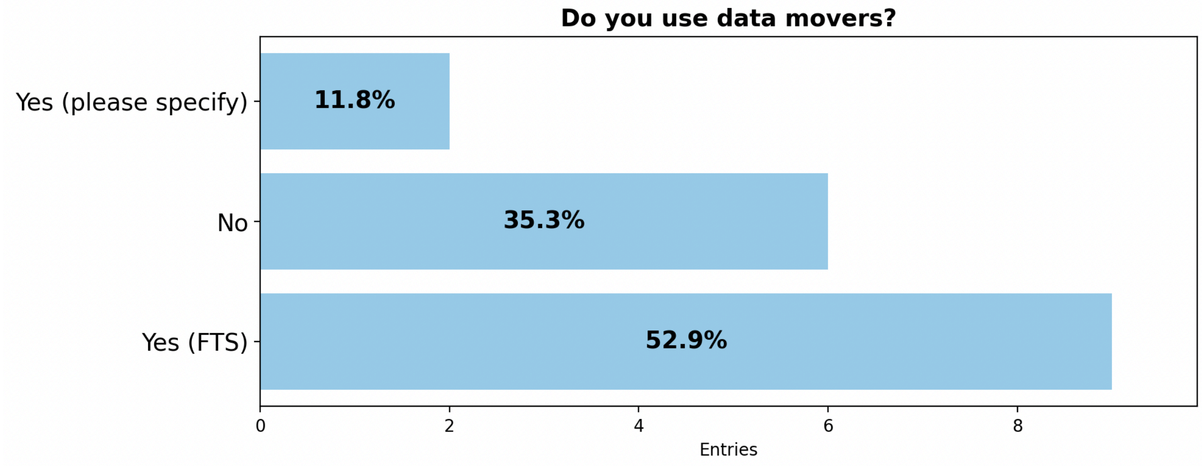

Figure 18: Need / support for data movers.

Data transfers between two storage systems must use protocols understood by the systems; potentially two different protocols can be used at source and destination. It is important that the data transfer tool allows for the largest variety of protocols; it is hence even more important which are the protocols supported by the initiatives, via for example their storage sites or software stacks. Next tables show quite a flat preference between Grid, Cloud protocols, with a minority using old protocols like bare FTP.

Table 18: Supported low level protocols for data transfer and access.

All: Which low level protocols are you using for data ingestion / transfer / access?	Answers	Fraction (%)
FTP	1	6.2%
GRIDFTP	2	12.5%
Others (please specify)	2	12.5%
POSIX (via mounted storage areas)	6	37.5%
S3	4	25.0%
SCP	8	50.0%
WebDAV	8	50.0%
XrootD	10	62.5%
iRODS	6	37.5%
Total valid answers	16	--

HEP: Which low level protocols are you using for data ingestion / transfer / access?	Answers	Fraction (%)
FTP	1	12.5%
GRIDFTP	2	25.0%
POSIX (via mounted storage areas)	4	50.0%
S3	3	37.5%
SCP	3	37.5%
WebDAV	6	75.0%
XrootD	8	100.0%
iRODS	2	25.0%
Total valid answers	8	--

RA: Which low level protocols are you using for data ingestion / transfer / access?	Answers	Fraction (%)
GRIDFTP	1	33.3%
Others (please specify)	1	33.3%
POSIX (via mounted storage areas)	1	33.3%
S3	2	66.7%
SCP	2	66.7%
WebDAV	1	33.3%
XrootD	1	33.3%
iRODS	2	66.7%
Total valid answers	3	--

Regarding Data Formats, each domain has usually converged to a small number of shared solutions. They are shown in the next word cloud.

Please specify which formats are you using (for example+ CSV+ XLS+ ROOT+ HDF5+ FITS+ ...)

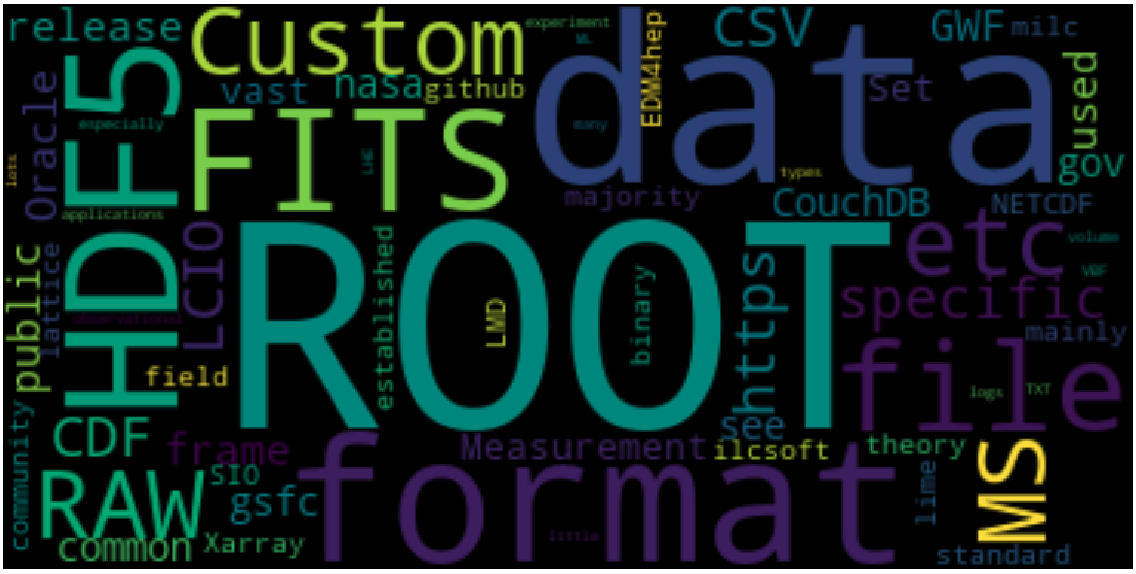

Figure 19: Word cloud of supported file formats.

How the data is read has consequences on the storage systems, and on the added capabilities from caches, non sequential storage systems etc. Table 19 shows the typical access patterns as collected.

Table 19: Typical access patterns.

All: Typical file / record access patterns	Answers	Fraction (%)
Iterated (read each file multiple times)	2	12.5%
Random (read specific parts of each file)	5	31.2%
Read multiple files per job	10	62.5%
Read one file per job	9	56.2%
Sequential (read each file from start to end)	15	93.8%
Total valid answers	16	--

HEP: Typical file / record access patterns	Answers	Fraction (%)
Iterated (read each file multiple times)	2	28.6%
Random (read specific parts of each file)	3	42.9%
Read multiple files per job	5	71.4%
Read one file per job	5	71.4%
Sequential (read each file from start to end)	7	100.0%
Total valid answers	7	--

RA: Typical file / record access patterns	Answers	Fraction (%)
Random (read specific parts of each file)	1	33.3%
Read multiple files per job	2	66.7%
Sequential (read each file from start to end)	3	100.0%
Total valid answers	3	--

Coming to the actual data, it is typically organised on files or tables in a DB. Table 20 shows how the most common file sizes are over 1 GB, as would be expected not to clog transfer systems. The objects the researchers are interested in are in most cases files, but DB records are not absent.

Table 20: Typical file / record sizes.

Typical file / record size(s) - please note B = Byte	Answers	Fraction (%)
1 GB - 10 GB	11	78.6%
10 GB - 100 GB	5	35.7%
1 MB - 1 GB	3	21.4%
> 100 GB	1	7.1%
1 kB - 1 MB	1	7.1%
Total valid answers	14	--

Table 21: Typical data type (file or DB record).

All: Typical data structures	Answers	Fraction (%)
DB records	4	23.5%
Files	17	100.0%
Total valid answers	17	--

HEP: Typical data structures	Answers	Fraction (%)
DB records	2	25.0%
Files	8	100.0%
Total valid answers	8	--

RA: Typical data structures	Answers	Fraction (%)
DB records	1	33.3%
Files	3	100.0%
Total valid answers	3	--

With exabyte-level data and gigabyte-level single objects, it is not surprising that the number of objects to be handled can be very large. Indeed, in almost ⅓ of the cases, the objects to be managed, tracked and transferred exceed 1 billion.

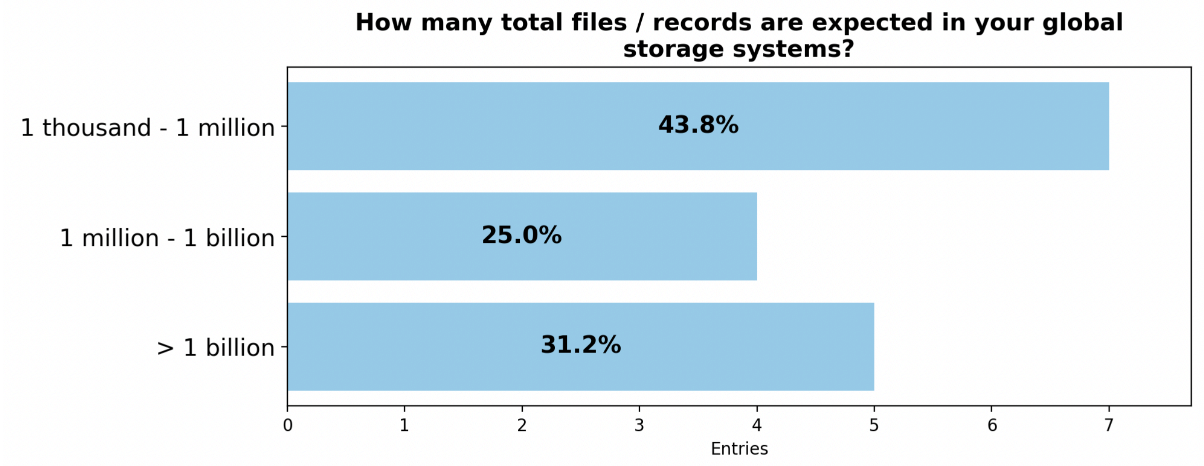

Figure 20: Total number of objects in the storage systems / DBs.

FAIR (Findable, Accessible, Interoperable and Reusable) data is now requested by most funding agencies, reflecting the need to open as much as possible data collected or created via publicly funded initiatives. While less than 50% of the initiatives are handling the data according to FAIR principles, the large majority are in the process of *transitioning* to their use, as shown in the chart.

FAIR data doesn't necessarily mean open data: the most used definition is "as open as possible, as closed as necessary". Some ¼ of the initiatives declare their data is simply "open"; in all the other cases, embargoes are used to limit initial access or the data is simply closed. In both situations, a mechanism ensuring only the authorised persons can access the data is needed.

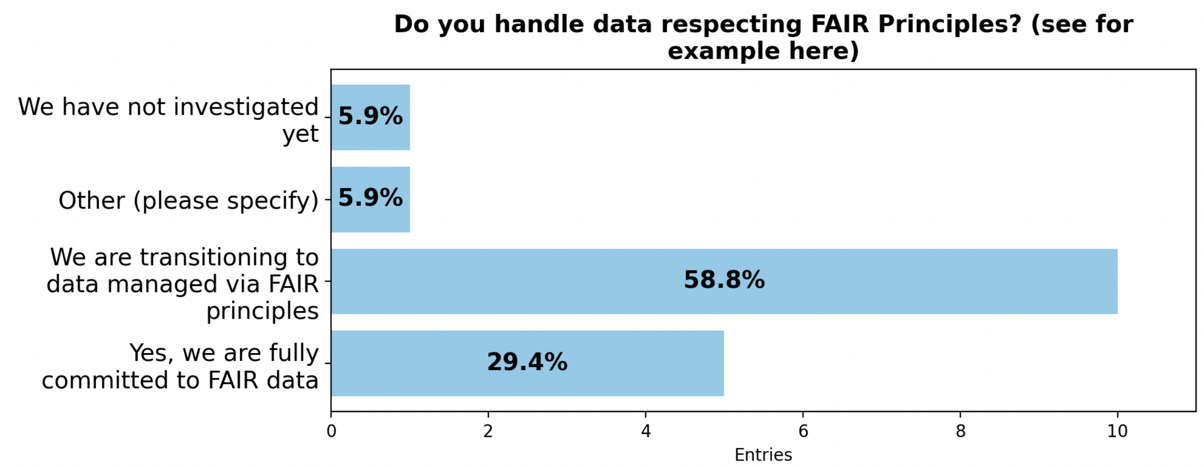

Figure 21: Adherence to FAIR principles for data.

Table 22: Need for sensitive data handling.

Needs for data confidentiality / controlled access	Answers	Fraction (%)
Needs an AAI mechanism (see the Compute Environment later)	12	70.6%
Data embargos	7	41.2%
No need (data is public)	6	35.3%
Needs data encryption on storage	1	5.9%
Total valid answers	17	--

Policies against data loss can vary between "don't care" to "multiple copies on different systems".Table 23 shows the respondents' desires.

Table 23: Need for data redundancy.

Which policies against data loss do you need?	Answers	Fraction (%)
Multiple replicas on disk/tape	12	70.6%
We rely on low level safety (RAID systems, for example)	9	52.9%
Multiple replicas on disk	6	35.3%
We accept a fraction of data loss	2	11.8%
Total valid answers	17	--

Coming to data interpretation, the situation varies between self describing data (for example text files) and simple byte streams which cannot be interpreted without a schema, external data or similar. Next plots show how self-describing the data is: in most cases external metadata is needed for utilisation. Persistent identifiers (PIDs) can be used for data reference, and catalogues are needed to match data and metadata and to allow for some levels of queries (The "F" in FAIR).

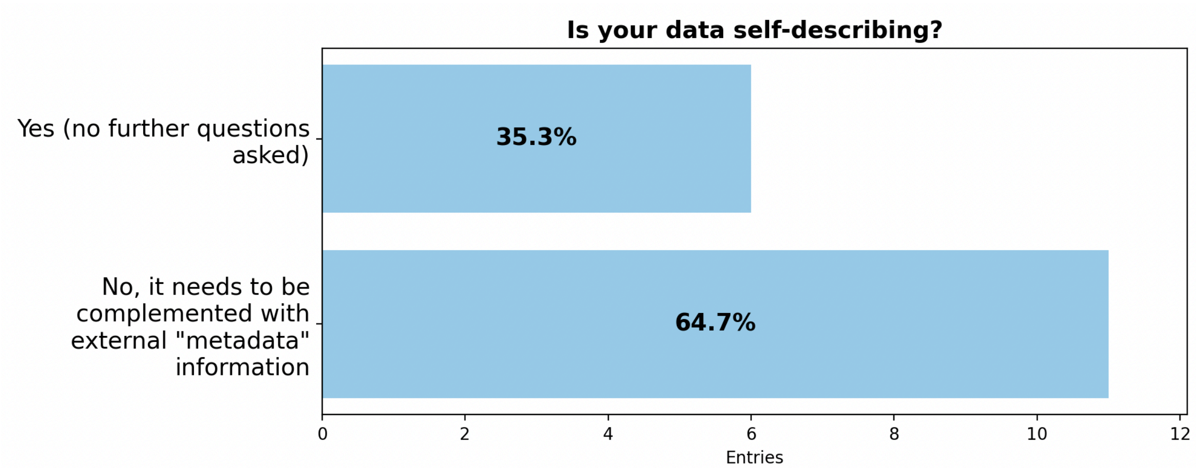

Figure 22: Self description capability of data.

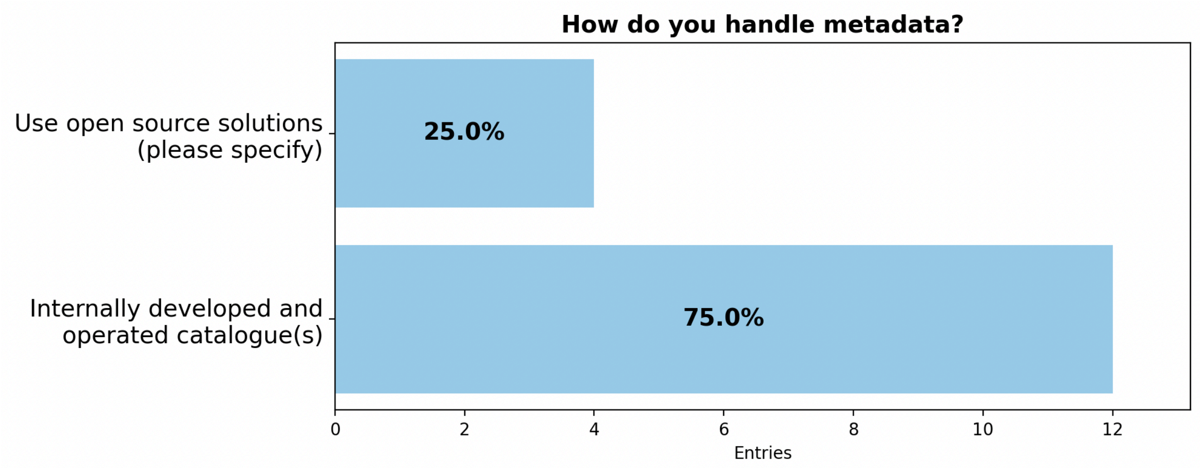

Figure 23: Modalities for handling metadata information.

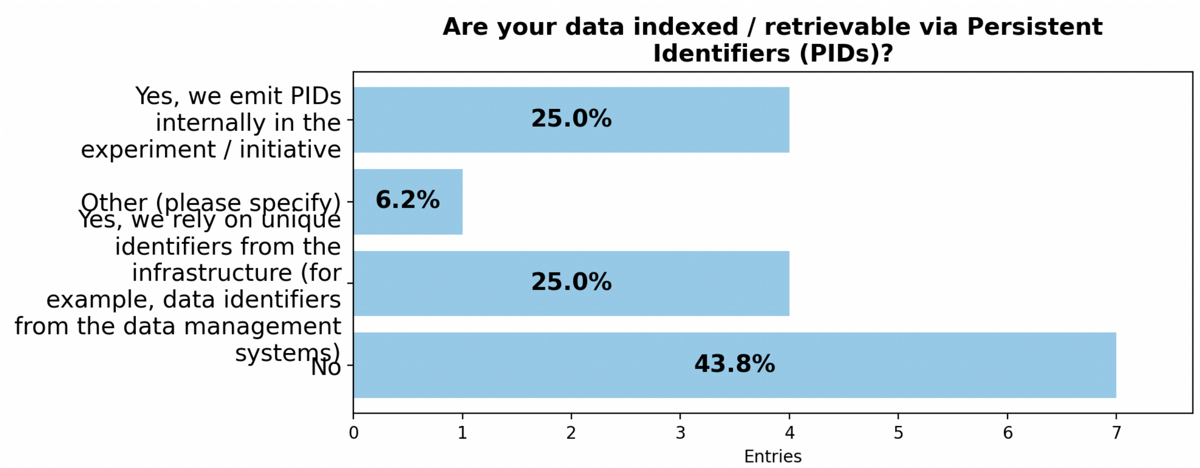

Figure 24: Usage of Persistent Identifiers (PIDs).

6.2.4. Expected Compute Environment

One of the main obstacles when trying to execute scientific workflows on a system not explicitly managed to support them lies in a series of technical details which can make the process very difficult, if not impossible. Researchers expect to find compute environments compatible with the needs of their workflows, but at times this is not compatible with a given center's hardware, software or even political configurations. This section of the survey was designed to clarify the expectations from users when (attempting to) execute a workflow, in order to be compared with site expectations as in the last survey section.

The first part of this survey section was about basic hardware and software parameters (CPU, Operating System, networking capabilities ...).
Intel/AMD x86_64 is largely the most supported architecture; there are no substantial requirements on the operating system as long as it is a modern 64 bit Linux. Since the focus of the survey is on data intensive sciences, it is not unexpected that we see a need for compute nodes to access external resources.

Table 24: Supported base computing architectures.

Supported base architectures	Answers	Fraction (%)
x86_64 (Intel, AMD, ...)	21	100.0%
aarch64	8	38.1%
Other (please specify)	4	19.0%
Total valid answers	21	--

Table 25: Supported operating systems.

All: Supported operating systems	Answers	Fraction (%)
Generic Linux 64bit	16	80.0%
MacOs	4	20.0%
Specific linux versions (please specify)	6	30.0%
Windows	3	15.0%
Total valid answers	20	--

HEP: Supported operating systems	Answers	Fraction (%)
Generic Linux 64bit	4	50.0%
MacOs	2	25.0%
Specific linux versions (please specify)	5	62.5%
Windows	1	12.5%
Total valid answers	8	--

RA: Supported operating systems	Answers	Fraction (%)
Generic Linux 64bit	3	100.0%
Specific linux versions (please specify)	1	33.3%
Windows	1	33.3%
Total valid answers	3	--

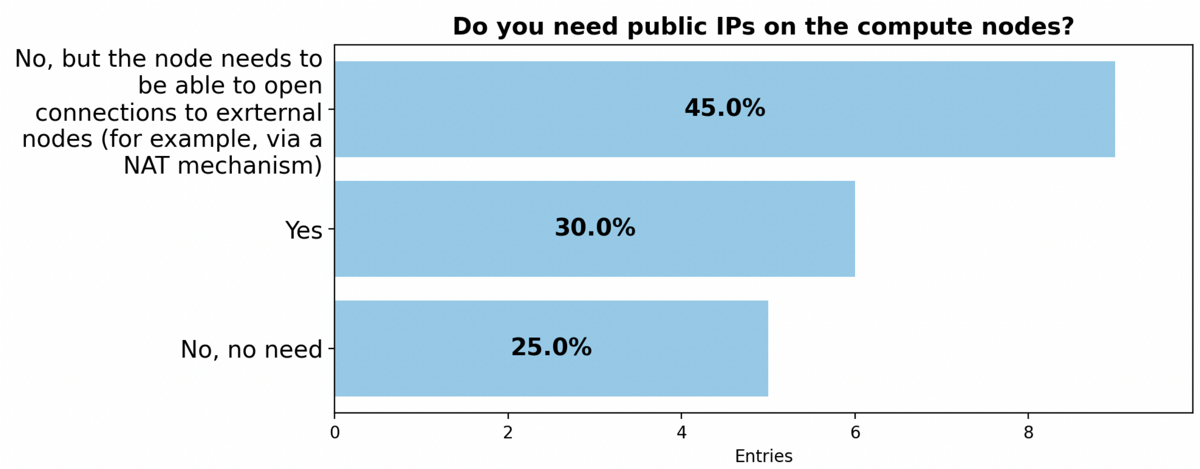

Figure 25: Need for public IPs on the compute nodes.

When moving to GPUs, NVidia is the go-to product at the moment; still, some initiatives seem to be able to utilise hardware from other vendors, like those from Intel and AMD. Initiatives able to efficiently use Quantum Systems (either emulators or real) are a surprise which is worth exploring in detail, for example contacting the submitters.

Table 26: *Capability to use accelerators (real / emulated).*

Do you expect to need accelerators / special hardware?	Answers	Fraction (%)
Nvidia GPUs	12	100.0%
AMD GPUs	8	66.7%
Quantum Computing Emulators	2	16.7%
Xilinx FPGAs	1	8.3%
Quantum Computing Hardware	1	8.3%
Intel GPUs	4	33.3%
Total valid answers	12	--

Being able to use a GPU does not indicate to which extent it is used efficiently. A question was about the total fraction of the workflows that can indeed be offloaded to GPUs. Not unexpectedly, the average is below 50%.

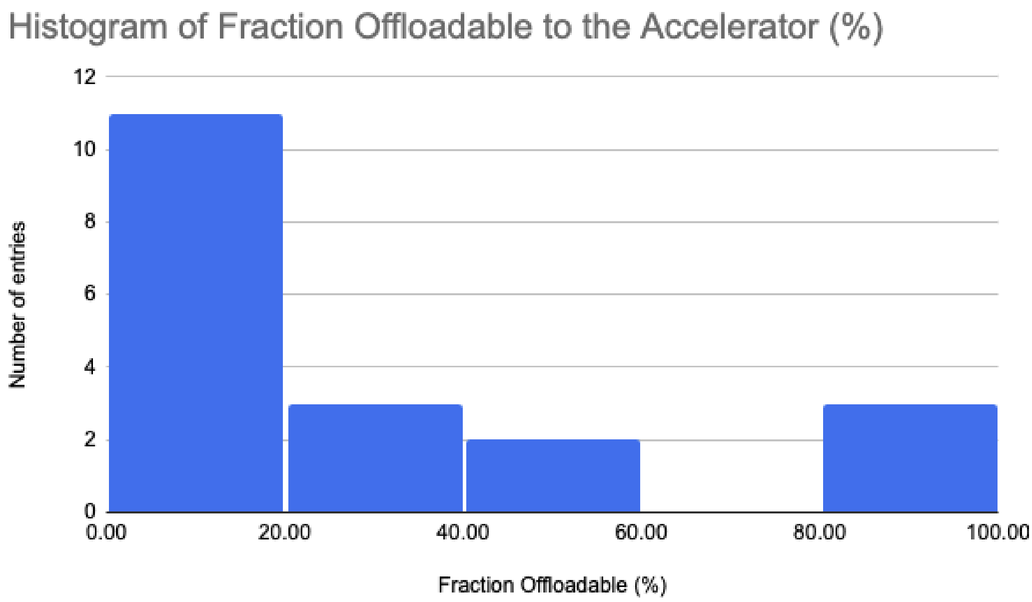

Figure 26: Fraction of payloads which can be offloaded to accelerators.

Moving to storage, data intensive workflows often need local scratch space as a buffer during execution, which is too large to reside in memory. Only a small fraction of the initiatives indeed declare they can proceed without such a need. For the others, a local disk is better but many can indeed survive with shared areas, if sufficiently performant.

Table 27: Need for shared file systems available on the compute nodes.

Do you need a local/shared disk to be used as area of work for the process (and scratched afterwards)?	Answers	Fraction (%)
Local Disk needed	17	85.0%
No disk needed	2	10.0%
Not needed, but beneficial if available	2	10.0%
Shared disk areas needed	13	65.0%
Total valid answers	20	--

Asking access as "root" to an unowned system is of course a very difficult request to be met by a site. Indeed in the vast majority of cases the workflows appear not to request that. The "Other" responses simply refer to running services, which is actually covered by another question (about edge services).

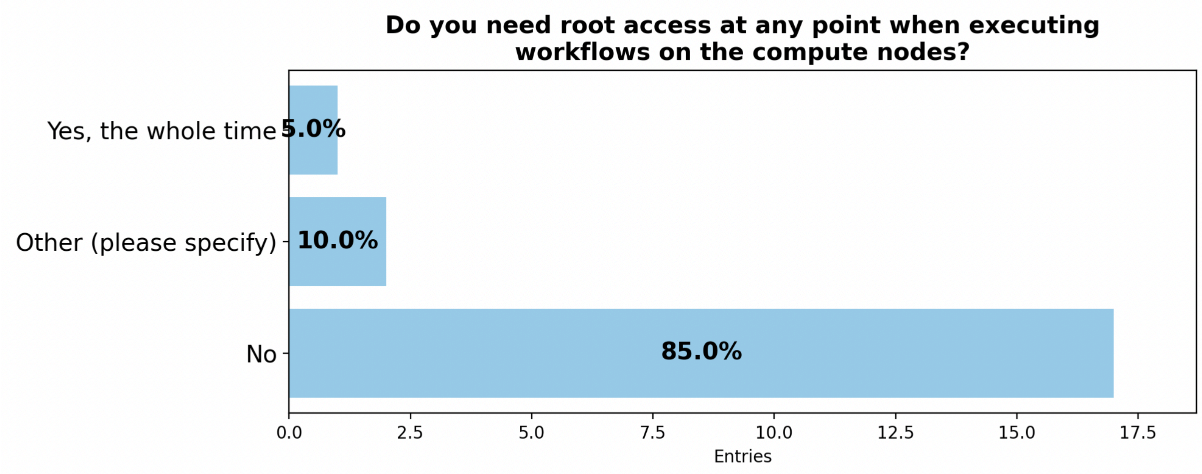

Figure 27: Need for privileged access during workflow execution.

Software installation, especially when in the presence of large software stacks, is problematic and not always possible from the user point of view (for example, some installations could require more space than the quota in the user's $HOME directory). Still, in 8% of the initiatives users are still doing that. In the large majority of cases, though, external solutions are needed, varying from "an email to the sysadmins" to the use of package managers, to the use of pre-cooked software areas like for CVMFS. Of course, the use of virtualized systems, in which a full system image is in the user's full control, is another solution to the problem.

Table 28: Preferred means of software distribution / installation.

All: How would you like to have software made available to you?	Answers	Fraction (%)
I will install myself	8	38.1%
Other (please specify)	2	9.5%
Preinstalled by the sysadmins	6	28.6%
Via CVMFS	10	47.6%
Via Module	6	28.6%
Via Spack	1	4.8%
We use containers with preinstalled SW	10	47.6%
Total valid answers	21	--

We hence asked users whether they can support different virtualization layers; ***apptainer***, despite being quite new and developed by the scientific community, is the preferred choice, probably due its lower impact on system policies and requirements.

Table 29: Workflow support for typical virtualization tools.

All: Do you need access to virtualization software?	Answers	Fraction (%)
Docker	8	38.1%
KVM	2	9.5%
No	3	14.3%
Not needed, but we can use them	4	19.0%
Other (please specify)	1	4.8%
Shifter	1	4.8%
Singularity / Apptainer	16	76.2%
Total valid answers	21	--

We then inquired about the need for edge services. Edge services are used as "connectors" between site resources and external initiative-level services, for example to provide access to calibrations, connection to external Data and Workload management systems, and for monitoring purposes. Most initiatives declare they can live without, which is contrary to our experience; this will be followed in the detailed analysis.

Table 30: Need for edge nodes.

Do you need particular services to be deployed on special machines in the centre in order to operate? (edge services, proxies, connectors ...)	Answers	Fraction (%)
No	10	50.0%
Yes, we need services executing on special nodes	8	40.0%
Yes, if there in no outgoing connectivity from the compute nodes	3	15.0%
Other, please comment	1	5.0%
Yes, if there in no incoming connectivity to the compute nodes	1	5.0%
Total valid answers	20	--

As a final question, we inquired about the need to have graphics access to the resources. More than 50% of the respondents seem to have answered positively; this is interpreted as an effect of the success of Jupyter notebook based workflows, especially in the realm of (interactive) analysis.

Table 31: Need for graphic access to compute nodes.

All: Do you need X11/Wayland/VNC/... graphical access to the nodes?	Answers	Fraction (%)
No	11	57.9%
Other	2	10.5%
VNC	4	21.1%
X11	8	42.1%
Total valid answers	19	--

HEP: Do you need X11/Wayland/VNC/... graphical access to the nodes?	Answers	Fraction (%)
No	4	57.1%
Other	1	14.3%
VNC	1	14.3%
X11	3	42.9%
Total valid answers	7	--

RA: Do you need X11/Wayland/VNC/... graphical access to the nodes?	Answers	Fraction (%)
Other	1	33.3%
VNC	2	66.7%
X11	3	100.0%
Total valid answers	3	--

6.2.5. Software, software development, software distribution, policies

6.2.5.1. Programming Languages & Libraries

We started by investigating the programming languages most prevalent amongst the SPECTRUM community.

Table 32: Most used programming languages.

All: Which are the typical programming languages that you use in your initiative?	Answers	Fraction (%)
C#	1	2.1%
C/C,,	41	87.2%
Fortran	20	42.6%
Go	2	4.3%
Java	2	4.3%
Javascript (or TS or NodeJS)	1	2.1%
Julia	4	8.5%
Other (please specify)	7	14.9%
Perl	1	2.1%
Python	40	85.1%
R	3	6.4%
Total valid answers	47	--

HEP: Which are the typical programming languages that you use in your initiative?	Answers	Fraction (%)
C#	1	6.7%
C/C,,	14	93.3%
Fortran	2	13.3%
Go	2	13.3%
Java	1	6.7%
Javascript (or TS or NodeJS)	1	6.7%
Julia	1	6.7%
Other (please specify)	1	6.7%
Python	15	100.0%
R	1	6.7%
Total valid answers	15	--

RA: Which are the typical programming languages that you use in your initiative?	Answers	Fraction (%)
C/C,,	3	100.0%
Fortran	3	100.0%
Go	1	33.3%
Java	1	33.3%
Javascript (or TS or NodeJS)	1	33.3%
Other (please specify)	1	33.3%
Python	2	66.7%
R	1	33.3%
Total valid answers	3	--

Perhaps unsurprisingly, the results show a fairly even split between C/C++ and Python as the most commonly used languages. The former presumably for high-performance codes, while the latter is popular for user data analysis, scripting and AI/ML. The high usage of Fortran demonstrates that the language is not abandoned as many had expected.

When split by community (HEP vs RA), it is clear that C/C++ and Python have the dominant role in each community. Small number statistics in the RA community make it hard to derive other conclusions.

Table 33: External dependencies in scientific software stacks.

All: Please list the main software dependencies in your code	Answers	Fraction (%)
AI toolkits / frameworks (keras, pytorch, tensorflow, …)	14	30.4%
Commercial tools (oracle bindings, labview, data analysis tools….)	1	2.2%
External generators or general physics tools (pythia, madgraph, sherpa, fastjet, hepmc ..)	19	41.3%
Generic C,, libraries and compilers (boost, gcc, llvm, cling, clang)	35	76.1%
HEP specific software (ROOT, dd4hep,…)	23	50.0%
Hardware oriented libraries (CUDA, sycl, opencl, openMPI, alpaka, …)	23	50.0%
Other (please specify)	5	10.9%
Python ecosystem (NumPy, scikit-learn, …)	32	69.6%
Quantum computing tools and libraries (kiskit, cirq, …)	4	8.7%
Radio Astronomy specific software (CASA, WSClean, PRESTO…)	3	6.5%
Simulation toolkits (geant, fluka, corsika, …)	22	47.8%
Total valid answers	46	--

HEP: Please list the main software dependencies in your code	Answers	Fraction (%)
AI toolkits / frameworks (keras, pytorch, tensorflow, …)	9	60.0%
Commercial tools (oracle bindings, labview, data analysis tools….)	1	6.7%
External generators or general physics tools (pythia, madgraph, sherpa, fastjet, hepmc ..)	11	73.3%
Generic C,, libraries and compilers (boost, gcc, llvm, cling, clang)	11	73.3%
HEP specific software (ROOT, dd4hep,…)	13	86.7%
Hardware oriented libraries (CUDA, sycl, opencl, openMPI, alpaka, …)	9	60.0%
Other (please specify)	1	6.7%
Python ecosystem (NumPy, scikit-learn, …)	14	93.3%
Quantum computing tools and libraries (kiskit, cirq, …)	1	6.7%
Radio Astronomy specific software (CASA, WSClean, PRESTO…)	1	6.7%
Simulation toolkits (geant, fluka, corsika, …)	11	73.3%
Total valid answers	15	--

RA: Please list the main software dependencies in your code	Answers	Fraction (%)
AI toolkits / frameworks (keras, pytorch, tensorflow, …)	2	66.7%
Generic C,, libraries and compilers (boost, gcc, llvm, cling, clang)	3	100.0%
HEP specific software (ROOT, dd4hep,…)	1	33.3%
Hardware oriented libraries (CUDA, sycl, opencl, openMPI, alpaka, …)	3	100.0%
Python ecosystem (NumPy, scikit-learn, …)	2	66.7%
Quantum computing tools and libraries (kiskit, cirq, …)	1	33.3%
Radio Astronomy specific software (CASA, WSClean, PRESTO…)	3	100.0%
Simulation toolkits (geant, fluka, corsika, …)	1	33.3%
Total valid answers	3	--

Having established which programming languages are predominant, we inquired about the libraries and other tools upon which our community depends. We see significant usage of standard C/C++ libraries and tools (LLVM, GCC, Boost, etc.) and the scientific Python ecosystem (NumPy, etc). There are also a significant number of users of libraries supporting HPC and accelerator hardware (MPI, CUDA, etc.), as well as off-the-shelf AI toolkits (Keras, PyTorch, scikit-learn, etc.). Wide use is made of HEP tools (ROOT, Pythia and other generators, HepMC), but this has no obvious analogue in the RA community with only a relatively small number of users of tools like CASA — but this may be due to small statistics.

6.2.5.2. Hardware

Table 34:. Software supported architectures.

All: CPU architectures	Answers	Fraction (%)
Other	1	2.4%
aarch64	13	31.0%
ppc64le	1	2.4%
riscv64	2	4.8%
x86_64	41	97.6%
Total valid answers	42	--

It is perhaps unsurprising that x86_64-based systems dominate in the SPECTRUM community, although the use of aarch64 is also significant. Some of this may be personal use on Apple laptops; other, low power systems in data centres which are especially energy footprint aware.

Table 35: Support for vector extensions.

All: CPU vectorization	Answers	Fraction (%)
AVX	10	52.6%
AVX2	11	57.9%
AVX512	13	68.4%
MMA	1	5.3%
NEON	3	15.8%
SSE	14	73.7%
Total valid answers	19	--

The prevalence of the x86_64 architecture is reflected in the widespread availability of the AVX/AVX2/AVX512 and SSE instruction sets.

Table 36: Support for GPU-type accelerators.

All: GPU Accelerator	Answers	Fraction (%)
AMD	10	43.5%
Intel	8	34.8%
Nvidia	22	95.7%
Total valid answers	23	--

While NVIDIA-based GPUs predominate, the combined total of AMD and Intel is supported in more than half of the total — perhaps a surprising result.

6.2.5.3. *Software*

The use of accelerators implies a proper software process, in which either the accelerators are directly programmed, or where a high level software framework shields the programmers from (multiple) accelerator hardware details.

We started asking whether the software supports heterogeneous hardware (including multiple CPU architectures, GPUs, etc). Most of the initiatives unfortunately seem limited to a single reference platform.

Table 37: Capability to use accelerators.

Do support heterogeneous computing? (executing software on different platforms, possibly including accelerators)	Answers	Fraction (%)
We use a single / few platforms we program directly	27	67.5%
We use an internally designed system (for example, macros)	8	20.0%
We use frameworks / toolkits (please specify)	8	20.0%
Other (please specify)	2	5.0%
Total valid answers	40	--

Table 38: Main drivers behind the choice of architectures and accelerators.

All: What is the driving force behind the architecture / accelerators decisions?	Answers	Fraction (%)
Capability to execute of more centres	8	20.5%
Cost/performance optimization	25	64.1%
Decision coming from resources deployed in the centre we need to use	9	23.1%
External decision (funders, reviewers , ..)	6	15.4%
Need for more performance than simple CPUs could provide	13	33.3%
Other	1	2.6%
Previous experience in the developers' base	7	17.9%
Total valid answers	39	--

The next series of questions was about software release models and openness.
Some initiatives do not seem to have an initiative-wide licensing model at all (probably the smallest ones); most who do are opting for some variants of Open Source licences.

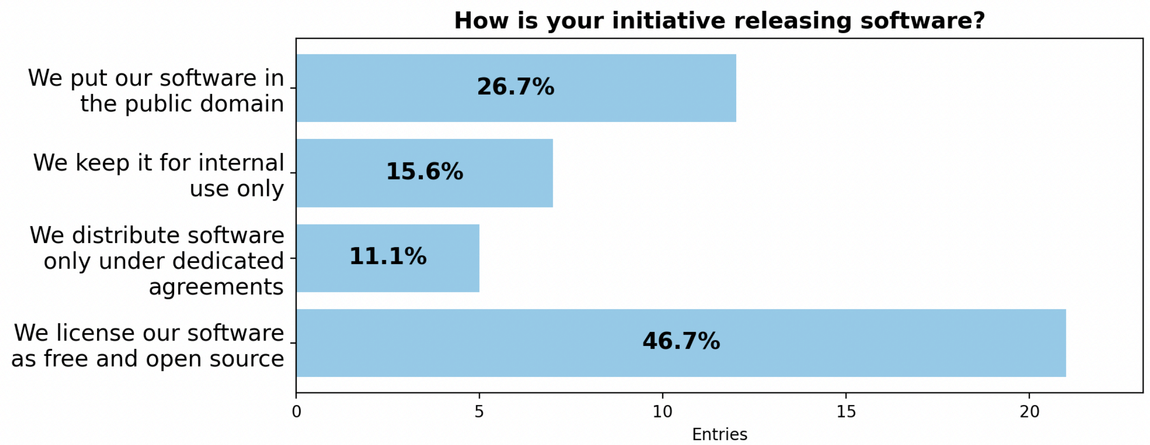

Figure 28: Software release policies.

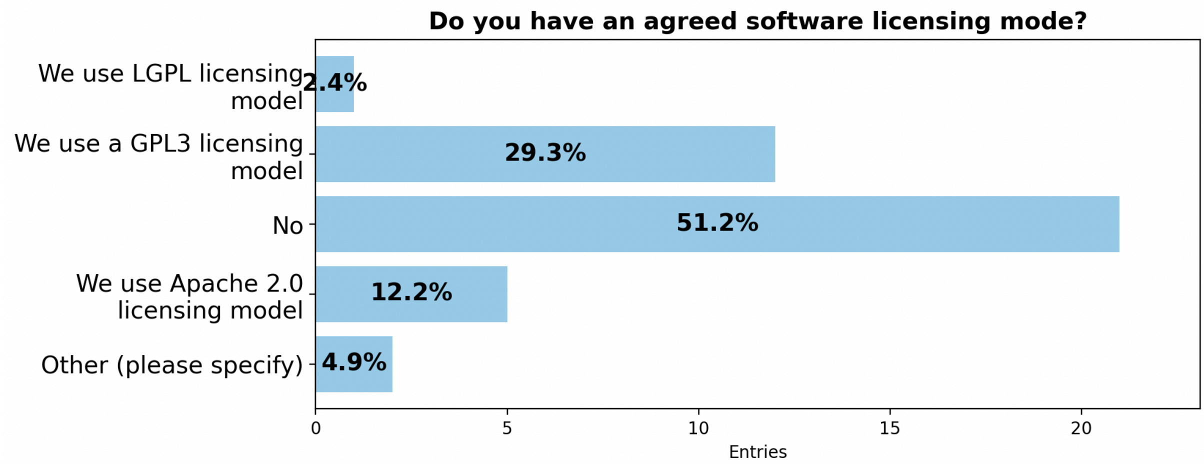

Figure 29: Initiative position about software licensing.

Software development can follow FAIR principles, for example as described in "FAIR Principles for Research Software (FAIR4RS Principles)" [link]; the situation seems already in good shape and still evolving positively.

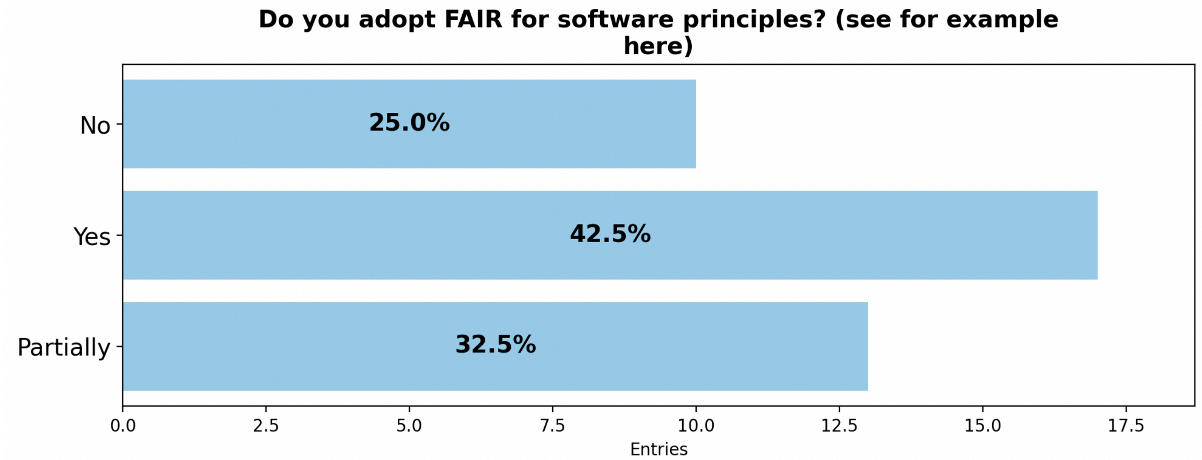

Figure 30: Adherence to FAIR principles for software.

When looking at future needs of research software, the general tone seems to go into the direction of more complex and performing software, mostly due to an increased complexity in scientific initiatives (more data, more complex instruments, …).

This can eventually push towards using more efficient / less expensive architectures. The following word cloud shows the most common answers about this; GPUs and low power architectures like ARM seem the most used solutions.

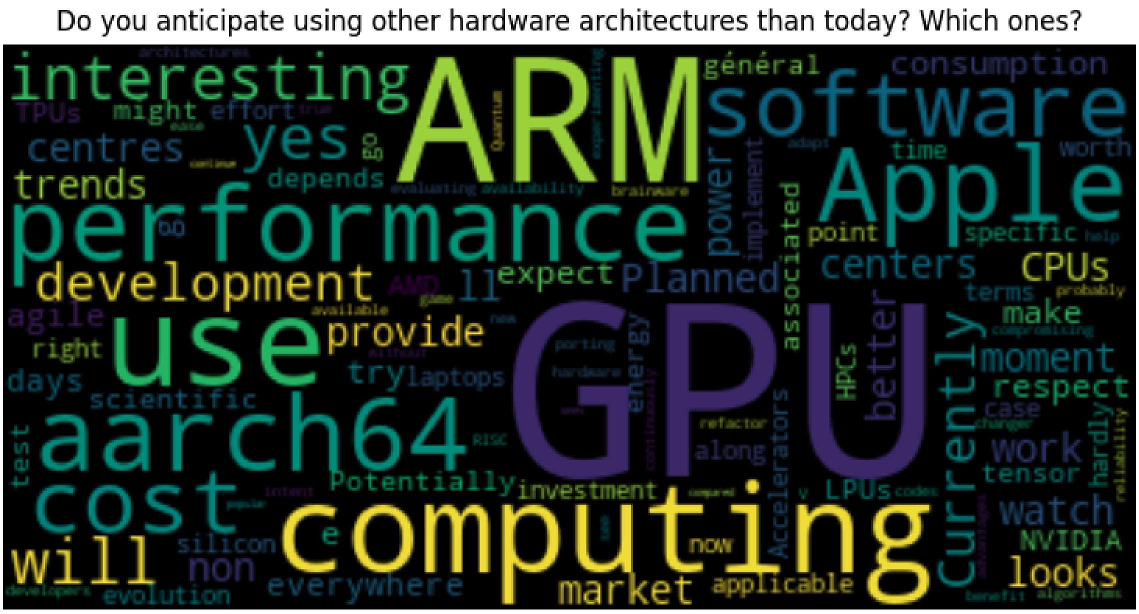

Figure 31: Word cloud on other hardware architectures.

The last section is about the trends expected by the initiatives, in terms of complex computing (more aggressive utilisation of multi core processes, vector programming, accelerators). Apart from the vector/SIMD case, all other options are expected to be used by the majority of initiatives.

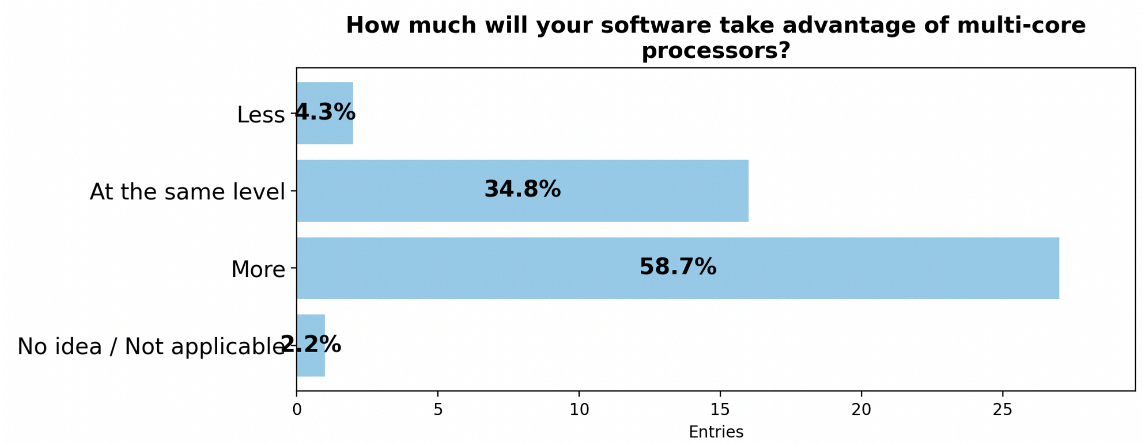

Figure 32: Software capability to use multicore CPUs.

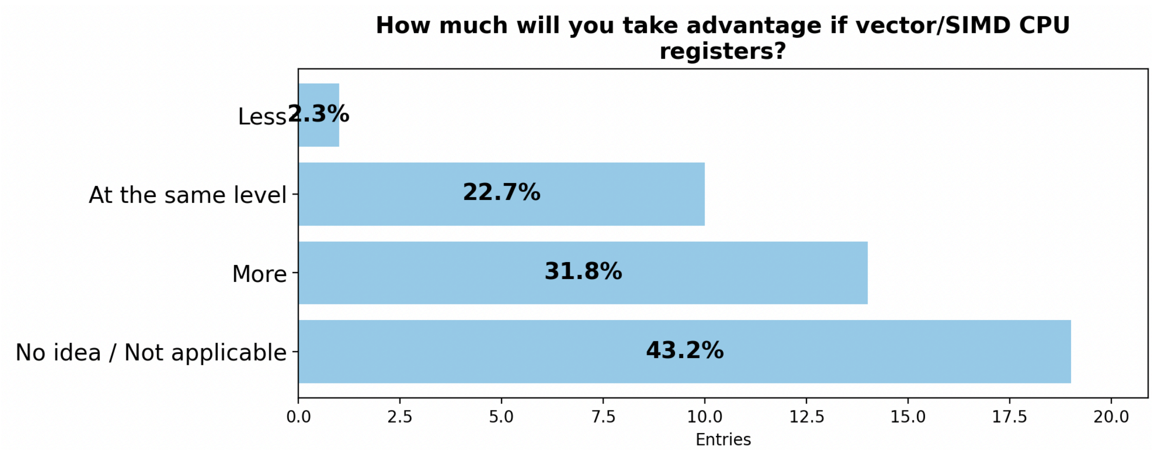

Figure 33: Software capability to use vector instructions.

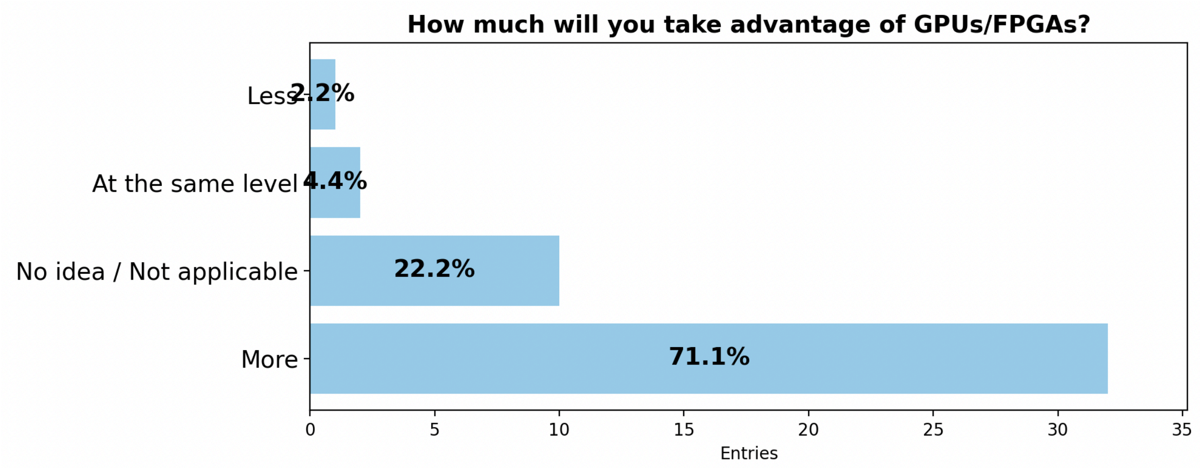

Figure 34: Software capability to use GPGPUs and FPGAs.

Finally, the community thinks the resources (human, financial) needed for such changes in the software are not available, a point to be followed closely.

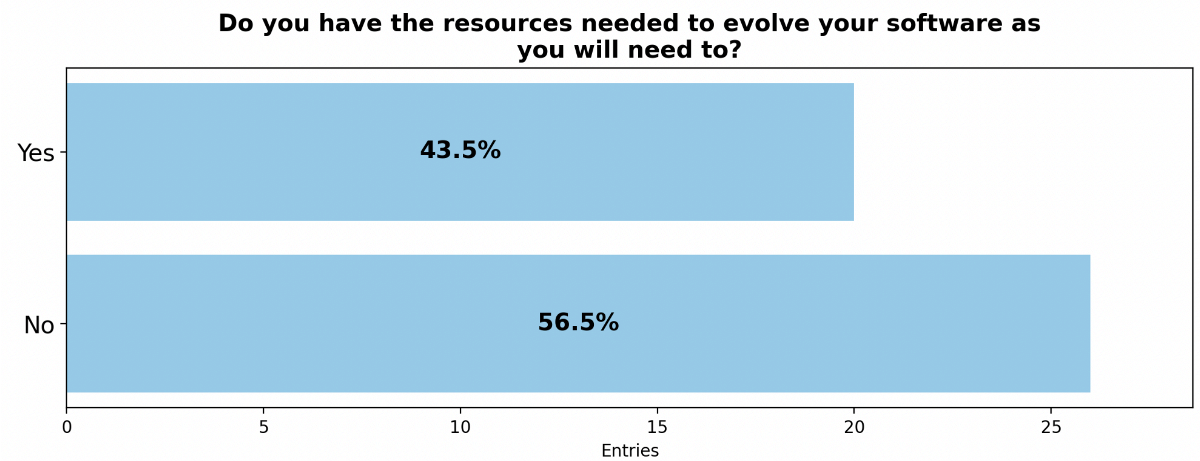

Figure 35: Extrapolation of future software development needs and their personpower coverage.

6.2.6. Training and careers

Training on computing-related matters is important in our field: most of the software developers (particularly those closer to the physics, like analysts and algorithm authors) come from a physics background, where in most cases the training received on modern computing systems is scarce / absent. In the previous generations of researchers, getting the correct competences for handling computing systems was mostly a voluntary / private path, which eventually was reasonable when the complexity of those systems was low (for example, FORTRAN on single CPU machines).
As we have seen in the previous sections, current computing systems as used in our research domain have evolved to use distributed facilities, accelerators, multiple storage systems, and software tools like virtualization, multithreading etc; assuming that these skills can be autonomously and efficiently learnt is extremely optimistic. As such, our researchers need post graduation and post PhD training on computing matters, and more generally continuous training since computing evolves at a very fast pace.

The community was asked, via the survey, who is writing software in their initiatives (either for the core scientific software or the infrastructure related one). The answers are somewhat aligned with expectations: researchers are still the largest part of the contributors to software, with the appearance of a non-negligible fraction of professionals.

Table 39: Typical developers for core software.

Who designs and maintains the software in your initiative? (analysis, reconstruction, simulation, ...)	Answers	Fraction (%)
Researchers dedicating a fraction of their time to software	14	93.3%
Professional software engineers hired for this	5	33.3%
Other (please specify)	1	6.7%
Total valid answers	15	--

Table 40: Typical developers for computing tools.

Who designs and maintains the computing tools needed by your initiative? (operations, monitoring, accounting, deployment, WMS and data management, ...)	Answers	Fraction (%)
Researchers dedicating a fraction of their time to software	12	80.0%
Professional software engineers hired for this	8	53.3%
Procured from external non-research entities	4	26.7%
Open source enthusiasts willing to collaborate	1	6.7%
Other (please specify)	1	6.7%
Total valid answers	15	--

Next, we asked if these researchers arrive with the correct skill set from their education path? The answers give a better landscape than expected.

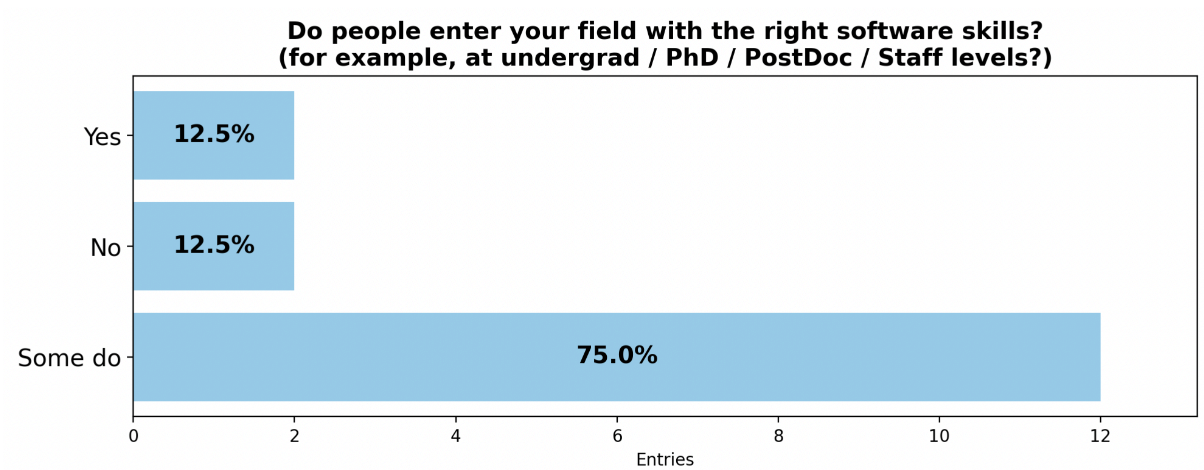

Figure 36: Evaluation of typical computing skills of collaborators.

Still, continuous training is needed in most cases. The next question was about which subjects they could get training for.

Table 41: Training opportunities offered to collaborators.

On which categories could you get training?	Answers	Fraction (%)
Basic programming (languages, development tools, ...)	10	76.9%
Distribute computing (access to Grids, storage systems, AAI, ...)	7	53.8%
Advanced programming (languages, development tools, ...)	9	69.2%
Efficient programming (profiling, energy measurement, ...)	5	38.5%
Heterogeneous computing (GPUs, FPGAs, Quantum Computing, ...)	8	61.5%
Principles of FAIR development	3	23.1%
Total valid answers	13	--

Finally, and perhaps most interesting, is the perception of the computing community on how their contribution to the initiatives is recognised, for example in terms of career advancements / paths. The initial perception from the survey authors was pretty negative on this, and unfortunately the results confirm it: more than 90% of the answers are not happy / completely unhappy about the recognition they receive.

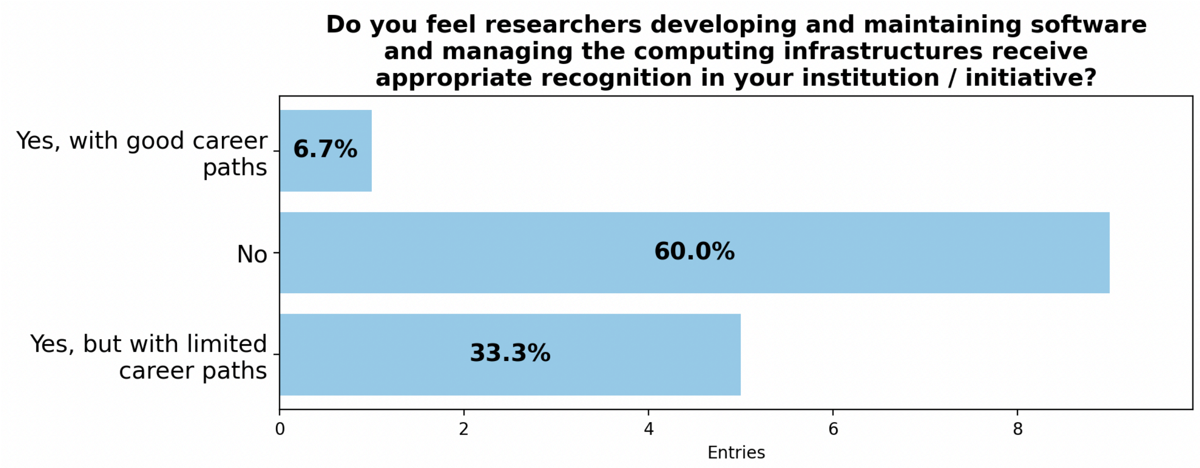

Figure 37: Perception of software-related career opportunities in the domain.

6.3. E-Infrastructures

E-infrastructure managers are the third community in the SPECTRUM project, together with HEP and RA researchers. The survey wanted to collect typical e-Infrastructure views from the managers, regarding the process and the modality in which they, today and in the next future, intend to offer computing resources to users.

The first question concerned the type of resources offered. The most common are CPU and GPU centers, either classified as HTCs (Grid facilities and computing farms) or HPCs (supercomputers). While the precise definition is left to the submitter, the latter usually include high speed networks among computing nodes. Storage is also generally offered, while Quantum Emulators are rare. The Survey also offered the possibility to declare quantum hardware, but none of the submitters did.

Table 42: Deployed architectures and services.

Which type(s) of resources does your e-Infrastructure provide today?	Answers	Fraction (%)
HPC (CPU only)	6	37.5%
HPC (CPU and GPU)	10	62.5%
HTC CPU only (Grid-style, batch, ...)	7	43.8%
HTC CPU,GPU (Grid-style, batch, ...)	9	56.2%
Storage systems (disks)	14	87.5%
Storage systems (tapes)	11	68.8%
Cloud (CPU and GPU)	6	37.5%
Quantum Emulators (classical systems emulating quantum hardware)	3	18.8%
Other (FPGAs, TPUs, ...) - please specify	2	12.5%
Cloud (CPU only)	3	18.8%
Total valid answers	16	--

Centers can either be managed directly, or be part of a larger infrastructure, like WLCG or EuroHPC JU or EGI. The vast majority of submissions are of the former type.

Table 43: Participation in larger e-Infrastructures.

Is the centre part of a larger e-infrastructure?	Answers	Fraction (%)
It is a WLCG Tier0/1/2/3	10	62.5%
Part of a national infrastructure (please specify)	9	56.2%
Part of a domain infrastructure (please specify)	4	25.0%
EOSC	2	12.5%
Other (please specify)	1	6.2%
EGI Federation	4	25.0%
No	4	25.0%
Total valid answers	16	--

Access to resources can be offered in various manners: as response to bids, by funding the center, using pay-per-use. The majority of answers hint at local / national / institutional use as the most used method.

Table 44: Access policies and grant attribution.

Which are the main mechanisms to get access to your infrastructure?	Answers	Fraction (%)
National level grants	8	53.3%
EU/US/... level grants	3	20.0%
Open only to some institutions' users	6	40.0%
Other (please specify)	5	33.3%
Pay per use	1	6.7%
Open to communities which fund the acquisition of resources	3	20.0%
Only open to some domains' users	3	20.0%
Total valid answers	15	--

Communities which require large computing efforts use in general initiative-side services, in order to manage the resources. These can either be deployed centrally, or on the centers where processing happens; in this latter case, the communities need to deploy services on the computing center. The next question is about whether this is allowed, and how. The answers show that more handshaking is needed to allow the diverse centers to collaborate with our scientific initiatives.

Table 45: Availability of user-operated edge nodes.

Do you allow users to deploy services on login / edge nodes?	Answers	Fraction (%)
No	9	69.2%
Other	2	15.4%
Yes+ upon verification and test from the site admins	2	15.4%
Yes+ via containers	3	23.1%
Total valid answers	13	--

A series of questions were about the total resources the centers can offer. In Figure 38 CPU cores, number of GPUs, TBs of disk and tape are shown.

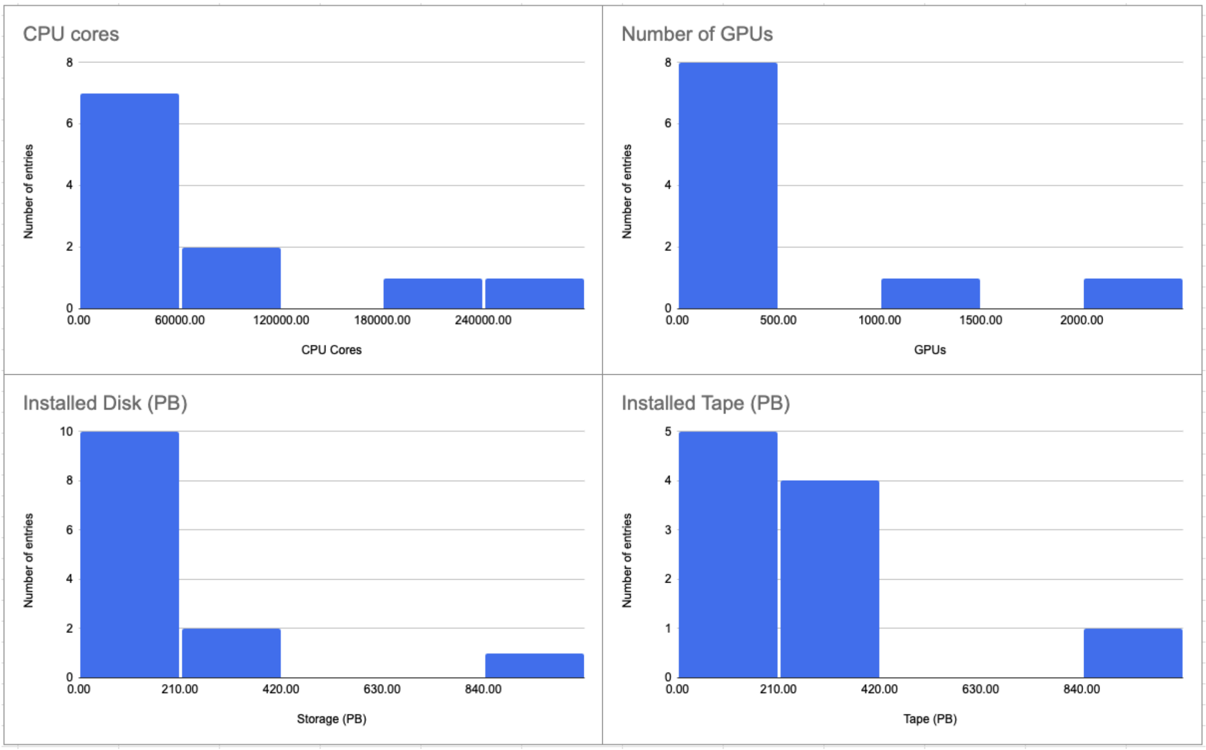

Figure 38: Resources operated on site.

Internally to the centers, resources must be allocated to the users, using some form of resource sharing. Tables 46 and 47 show how old-fashioned batch systems are still the preferred option.

Table 46: Resource provisioning methods.

Resource allocation methods	Answers	Fraction (%)
Via batch system (SLURM, LSF, PBS, HTCondor, ...)	13	86.7%
Via Cloud interfaces (k8s, Openstack, ...)	7	46.7%
Via external services (dask, jupyter, ...)	6	40.0%
Static allocations	3	20.0%
Other	1	6.7%
Total valid answers	15	--

Table 47: Allowed access patterns.

Which access patterns to the system you support?	Answers	Fraction (%)
Batch queues (for example SLURM, HTCondor, PBS, LSF, ...)	13	92.9%
Interactive "shell only"	9	64.3%
Interactive "graphical" (for example via X, VNC, ...)	7	50.0%
Via a Cloud Middleware / Stack	4	28.6%
Via Kubernetes	3	21.4%
Interactive "graphical" (Web based, for example via Jupyter notebooks, ...)	7	50.0%
Others (please specify)	1	7.1%
Total valid answers	14	--

Allocations, seen as a collaboration between a center and a community, can have different durations in time. Some can be very long term (for example when a community directly funded a center), while typical EU grants are for 12 months or less.

Table 48: Typical allocation length.

How long are typical resource allocations? (project / grant length)	Answers	Fraction (%)
Other	2	15.4%
2-6 months (6 excluded)	1	7.7%
7-12 months (12 excluded)	2	15.4%
1-3 years (3 excluded)	6	46.2%
Longer	5	38.5%
Total valid answers	13	--

Table 49: Typical processing slot length.

How long can the single processes be executed for?	Answers	Fraction (%)
1-12 h (12 excluded)	2	18.2%
12-24 h (24 excluded)	3	27.3%
24-48h (48 excluded)	2	18.2%
48-72h (72 included)	1	9.1%
> 72h	8	72.7%
Total valid answers	11	--

Getting to the technical part, it is important to know which precise type of resources the centres offer, with respect to CPUs, GPUs, Storage Systems and networks. Tables 50-52 and figures 39-44 summarise what the submitters declared, going into the details of hardware configurations.

Table 50: Available base architectures.

Which base system architecture is your centre supporting?	Answers	Fraction (%)
x84_64	12	100.0%
aarch64	4	33.3%
riscv64	1	8.3%
Total valid answers	12	--

Table 51: Deployed / supported GPGPus.

Which GPGPU architectures are you supporting?	Answers	Fraction (%)
Nvidia	10	90.9%
AMD	4	36.4%
Intel	1	9.1%
Total valid answers	11	--

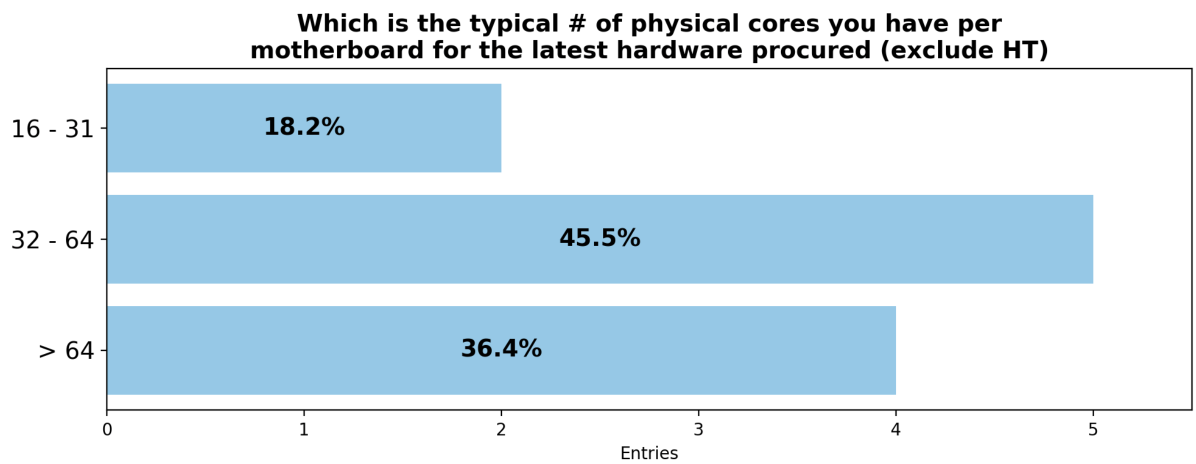

Figure 39: Core density per node.

Table 52: Typical memory per node (in GB/core) on deployed nodes.

Which is the typical memory per core you deploy, in GB? Please note 1 GB = 1 GigaByte	Answers	Fraction (%)
2-4	7	63.6%
4-8	2	18.2%
1-2	1	9.1%
>8	1	9.1%
Total valid answers	11	--

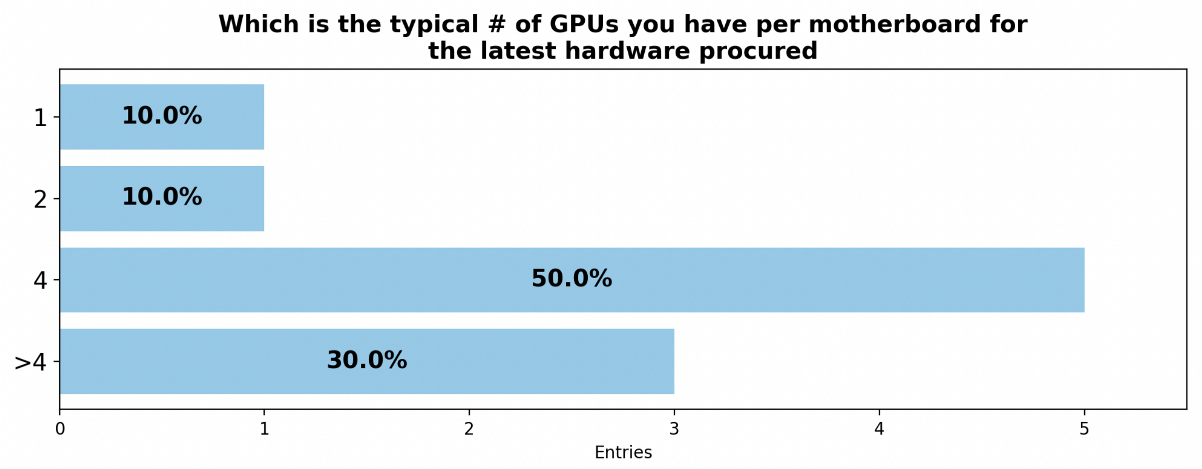

Figure 40: Typical number of GPUs per node.

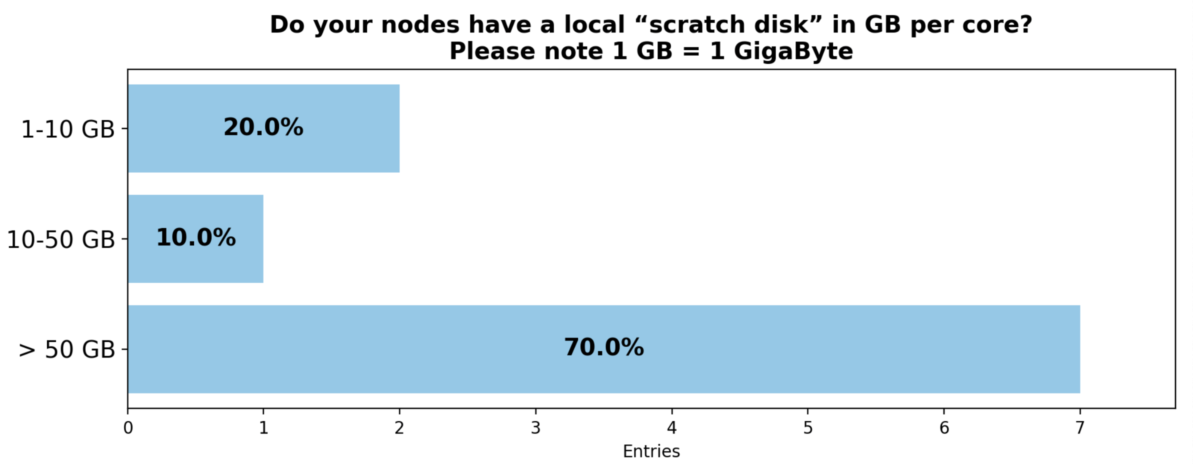

Figure 41: Availability of local scratch space.

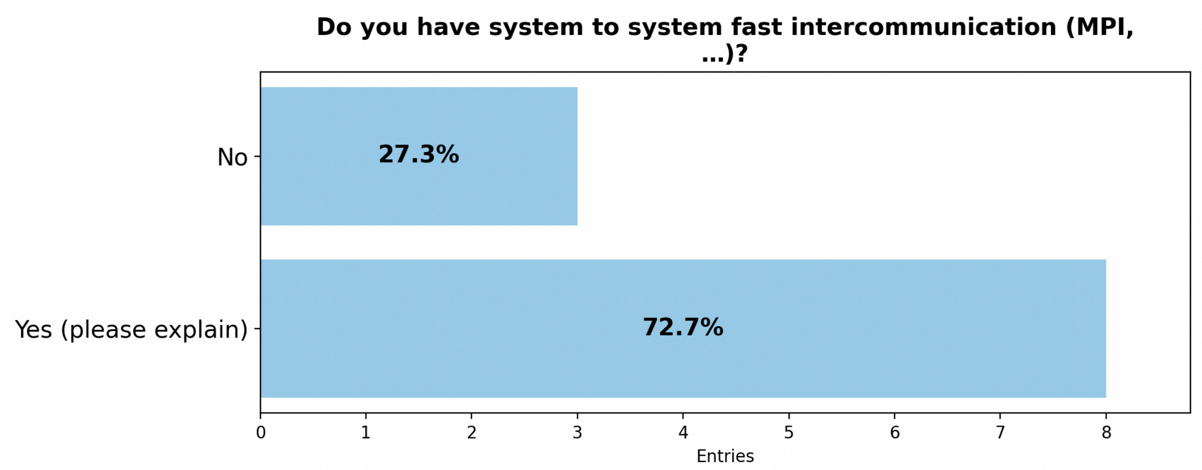

Figure 42: Availability of high-speed node-to-node interconnection.

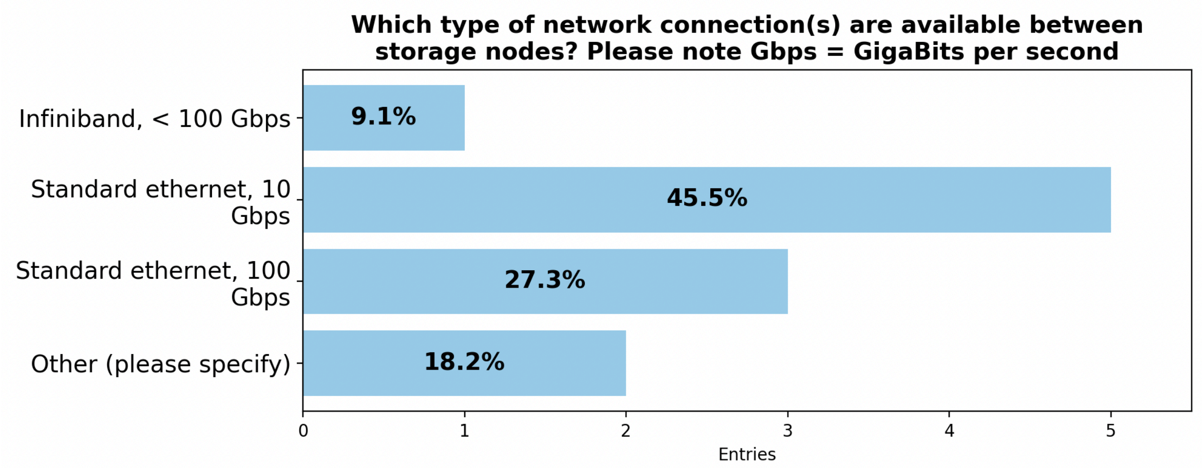

Figure 43: Available bandwidth between compute nodes.

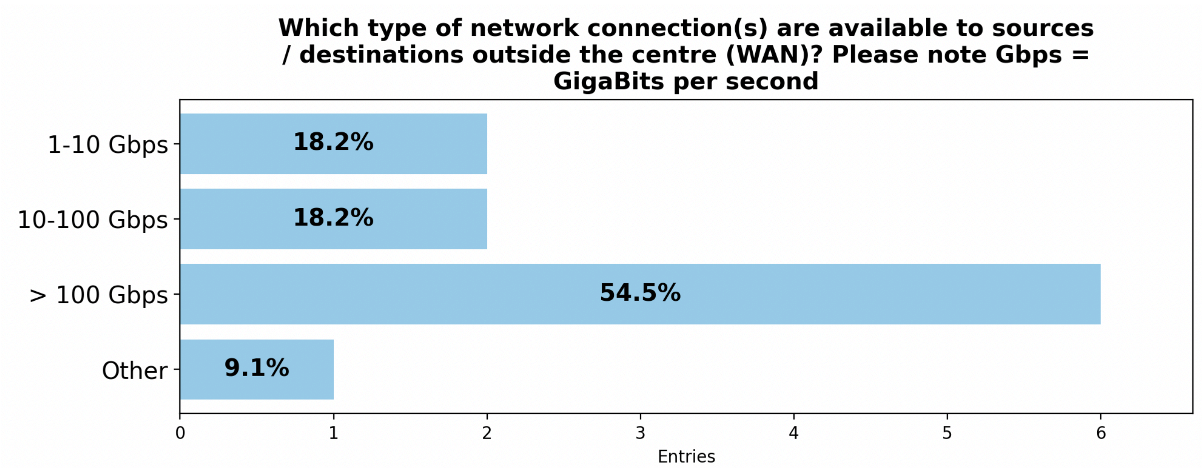

Figure 44: Availability and bandwidth of remote .

Centers with large storage deployments can offer official tools for data movement, to be offered (at least) to the communities lacking an internal mechanism. From table 53, it seems most indeed do.

Table 53: Availability of site-operated data management tools.

Does your centre use / support data management tools to move / access / manage data?	Answers	Fraction (%)
Yes+ via internally developed tools (please specify in the next question)	4	30.8%
Yes+ via open source tools (please specify in the next question)	6	46.2%
Yes+ via commercial tools (please specify in the next question)	1	7.7%
No+ all left to the users	5	38.5%
Total valid answers	13	--

A particularly important issue when dealing with our scientific domains is network connections. As we learned from the previous sections, data intensive initiatives can work in modes which imply remote data access, either via data downloaded before the start of the computation, or in streaming. Due to this, it is essential to have outgoing connections to at least a selected number of subnets/destinations, e.g. those needed to connect to the remote data silos.

Table 54: Routing options for incoming connections to compute nodes.

Which routing options are available from your compute nodes? INCOMING CONNECTIONS	Answers	Fraction (%)
Connection from onsite storage nodes	5	55.6%
Other	1	11.1%
Connection from other onsite compute nodes	5	55.6%
Connection from special onsite services	4	44.4%
Connection from a selected list of outside locations	2	22.2%
No connection	1	11.1%
Total valid answers	9	--

Table 55: Routing options for outgoing connections to compute nodes.

Which routing options are available from your compute nodes? OUTGOING CONNECTIONS	Answers	Fraction (%)
Connection to onsite storage nodes	7	77.8%
Connection to a selected list of outside locations	7	77.8%
Connection to other onsite compute nodes	5	55.6%
Connection to special onsite services	4	44.4%
Other	3	33.3%
Total valid answers	9	--

From the survey, the situation seems encouraging, at least for the subset probed by the respondents.

Specify the data management tool(s) your site uses and put a link to its documentation if available?

Figure 45: Word cloud on data management tools/protocols available to users.

Moving data requires more than existing network connections. For the internal storage, it is also a matter of how it is handled and which are the protocols offered at user jobs for connection. The situation shown by the survey shows that “industry standard” (for example WebDAV, S3, ..) and “domain specific” (XrootD, ...) are available at a subset of the centers. The existence of a proper “user to center” matched transfer protocol still needs to be checked case by case; but the fact that each of them are existing in at least one case, allows for some optimism.

Table 56: Storage management systems.

How do you manage disk-based storage?	Answers	Fraction (%)
Xrootd	4	36.4%
dCache	6	54.5%
EOS	4	36.4%
GPFS	5	45.5%
Lustre	2	18.2%
NFS	7	63.6%
Other	2	18.2%
Total valid answers	11	--

Table 57: Tape management infrastructure.

How do you manage tape-based storage?	Answers	Fraction (%)
Tivoli/TSM	3	33.3%
dCache	5	55.6%
CTA	4	44.4%
Other	2	22.2%
Total valid answers	9	--

Table 58: Local storage access protocols.

Protocols to access the storage systems, from internal hosts (for example compute nodes)	Answers	Fraction (%)
Xrootd	9	69.2%
Other	3	23.1%
POSIX (for example, NSF mount or any other fuse mount)	11	84.6%
WebDAV	6	46.2%
S3	3	23.1%
Total valid answers	13	--

Table 59: Remote storage access protocols.

Protocols to access the storage systems, from external hosts (for example storage to storage geographical transfers)	Answers	Fraction (%)
Xrootd	9	81.8%
Other	3	27.3%
WebDAV	6	54.5%
S3	2	18.2%
Total valid answers	11	--

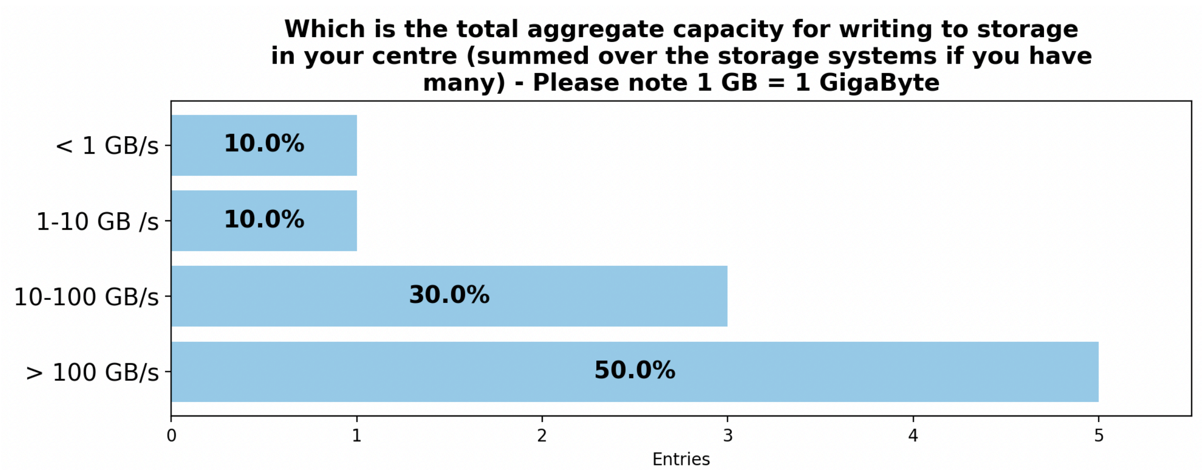

Figure 46: Aggregate storage write bandwidth.

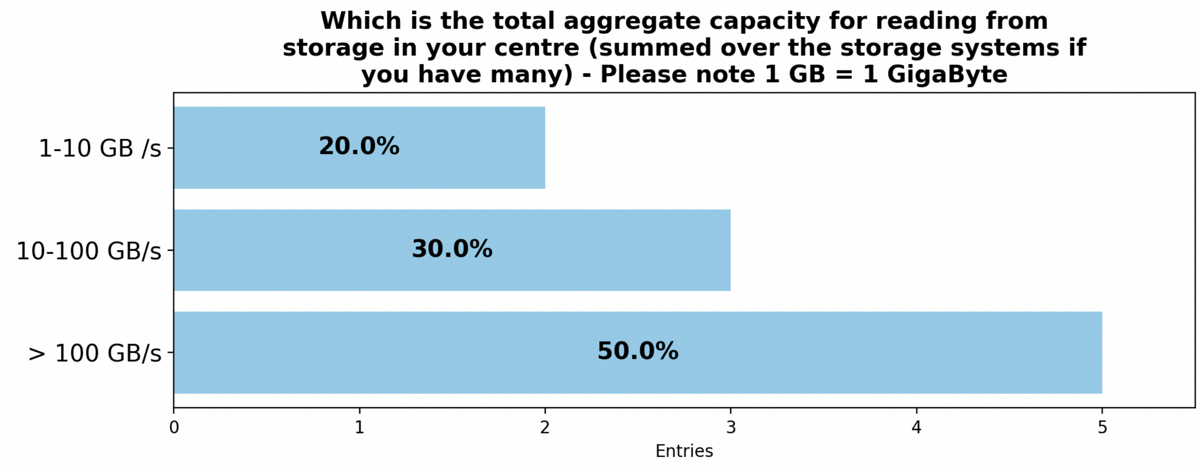

Figure 47: Aggregate storage read bandwidth.

Software distribution is a central subject when dealing with large codebases, which cannot realistically be installed by each user. Solutions like CVMFS are typical in HTC centers (for example those in the WLCG hierarchy), while they are reportedly not a supported solution in HPC centers. A good aspect in the received responses is a general tendency towards virtualization, which (also) goes a long way into a general solution of the problem. Tools like Module, Guix and similar remain a typical feature at HPC centers.

Table 60: Software installation / virtualization options.

Which of the following features do you support?	Answers	Fraction (%)
CVMFS	7	58.3%
Virtualization (singularity, apptainer, docker, udocker, shifter, ..)	12	100.0%
SW installation tools (Module, Conda, Guix, APT, RPM, ...)	5	41.7%
Other (please specify)	1	8.3%
Total valid answers	12	--

A "hot topic", in all senses, for computing centers is the energy footprint, a complex issue which can be defined in many ways (minimization of energy consumption and minimization of carbon footprint, for example). The survey is not the proper place to ask for precise details. At this level, we mostly wanted to check whether the problem is considered at all when deploying resources, and whether it is considered in the acquisition / operation processes. Only a small fraction of the centers seem to operate without implementing / putting in the roadmap solutions with energy/carbon footprint considerations in mind.

Table 61: Climate-sensitive center-level policies.

Is carbon/energy footprint a relevant factor when designing / operating your centre?	Answers	Fraction (%)
Yes	9	75.0%
No	3	25.0%
Total valid answers	12	--

Table 62: Climate-positive actions.

If yes: how do you address it?	Answers	Fraction (%)
Utilization of low carbon energy from the provider	4	50.0%
Energy production in situ / in a connected site via wind, solar, ...	2	25.0%
Re-utilization of excess heat (for example for house /office heating)	4	50.0%
Choice of cooling solution (adiabatic vs chillers, ...)	5	62.5%
Choice of low power/performance machines at procurement	3	37.5%
Operational choices (node shutdown when not needed, lower clock, job packing to maximise systems which can be shut down, ...)	6	75.0%
We continuously monitor the PUE of the site	5	62.5%
Total valid answers	8	--

Table 63: Climate-positive architecture choices.

Have you considered deploying more efficient / different architectures in order to improve power optimization?	Answers	Fraction (%)
Yes, we are doing it (please specify below)	5	50.0%
Yes, we are tracing technologies to possibly do it in the near future	6	60.0%
No, since user application needs specific architectures	1	10.0%
Other (please specify)	1	10.0%
Total valid answers	10	--

The next part of the survey was about the allowed Authentication and Authorization methods, which obviously should match the ones supported by the large initiatives' software stacks. Results are in line with the tools declared by the initiatives.

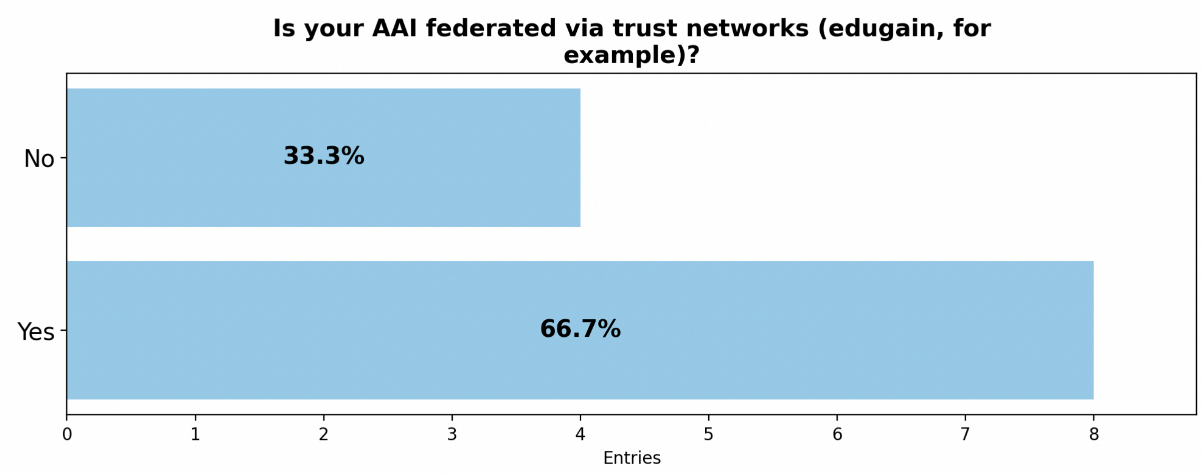

Figure 48: Access possible via federated AAIs.

Table 64:. Supported AA methods.

Which authorization and authentication methods do you support?	Answers	Fraction (%)
An open source infrastructure	9	69.2%
Username and password	10	76.9%
2FA	6	46.2%
An in-house developed system	2	15.4%
Total valid answers	13	--

Table 65: AA supported solutions / tools.

Which authorization and authentication technical solutions do you support?	Answers	Fraction (%)
X509 proxies	8	61.5%
tokens	9	69.2%
macaroons	5	38.5%
Other	1	7.7%
Username and password	10	76.9%
Total valid answers	13	--

Sensitive datasets are rare in our environment, if you consider this at the level of medical data or that containing personal data. Still, data from initiatives is not to be considered "public" at the time of acquisition: most initiatives apply periods of embargo, in which data is accessible only to formal collaborators. After the embargo, data is in principle made public, even if in practice the technical difficulties can be large[4]. Proper handling of sensitive data requires the certification of at least the storage systems where it is hosted; three sites answered positively to that.

[4] There are many difficulties in making large datasets publicly available: complexity of data formats, need for a deep-enough understanding of the content, serving TB or PB sized data outside research networks, huge processing resources to analyse them; these are just a few hurdles.

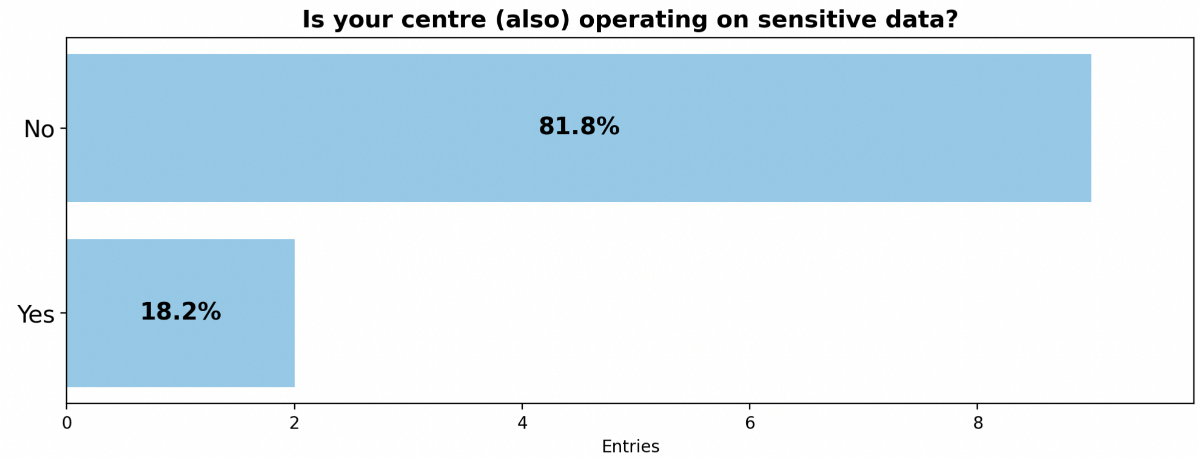

Figure 49: Availability of a certification for handling sensitive data.

Table 66: Climate positive actions.

If yes+ how do you operate it?	Answers	Fraction (%)
We have a fraction of the system with ISO 2700x certification	3	100.0%
Total	3	100%

The last part of the survey was about the availability of Quantum hardware/emulators. The answers seem to say that it is too early for centers to deploy such systems; this is possibly due to the timing of the survey: later in 2025, systems funded by the EuroHPC JU should start to appear [link, link, link, link].

Table 67: Available quantum software stacks.

Which quantum emulation stacks are you supporting?	Answers	Fraction (%)
N/A	4	57.1%
Other (specify)	2	28.6%
ProjectQ	1	14.3%
Total valid answers	7	--

Table 68: Available quantum hardware.

Which quantum hardware solution have you deployed?	Answers	Fraction (%)
N/A	4	66.7%
IBM	1	16.7%
Quantera	1	16.7%
Total valid answers	6	--

Table 69: Exploitation means for quantum hardware.

How do you support access to quantum hardware?	Answers	Fraction (%)
N/A	4	80.0%
As a remote accelerator	1	20.0%
Total valid answers	5	--

6.4. Future Trends

In most of this survey section, we asked (as free text) for the expected changes over the time scale of the next 5-10 years, if known or estimated.

6.4.1. Authentication and Authorization - "What changes do you expect in the next 5-10 years?"

Going forward we intend to move away from user certificates and do AAI with tokens	Move further towards tokens and away from certificates. Increase in federated identity based access to facilities.
VOMS[5] will be decommissioned	Better alignment between AAI systems (AARC BPA[6] is insufficient). Better granular support for Compute in AAI.
We plan to implement AAI for all services, and to be able to delegate invitations to virtual-orgs sub-groups to proposal's PIs.	I hope no more (voms-admin to IAM was bad enough), I guess there will be more use of tokens for data transfers

6.4.2. Processing Needs - "Future (5-10y) changes to the current model"

In the next future (5-10y), the Collaboration will define and evolve its computing model, while the experiment is being designed and built. Mock Data Challenges computing needs will ramp up, and likely several tools will be tested as components of the final framework and environment.	CMS' computing needs will increase dramatically for HL-LHC
For O5[7] we (Virgo) expect an increase in the rate of events of a factor of up to 10, proportionally increasing the needed computing power for PE. Computing power used for searches or detector characterization should grow less (but estimates are still vague)	We are exploiting ways to speed-up simulation with e.g. ML techniques or by using accelerators (GPUs). We foresee an increased usage of ML/AI algorithms and methods for physics analysis, but we cannot quantify the need at this time.
Integration of HPC, HTC and AI workloads into a	The basic scenario stays the same. Involving more

[5] "Virtual Organization Membership Service", a system to manage and grant authorization rights to users [link].

[6] AARC Blueprint Architecture is a set of software building blocks that can be used to implement federated access management solutions for international research collaborations. The Blueprint Architecture lets software architects and technical decision makers mix and match tried and tested components to build customised solutions for their requirements [link].

[7] One observation period in Virgo, specifically O5 covers approximately 2027-2030.

single physical system supporting high performance data analytics (HPDA)	resources of HPC type with or without accelerators, running multi-core/multi-node jobs.
The research infrastructure is currently under construction and is currently at about 15% of its final size. We are taking data while constructing the infrastructure, the numbers given above represent the current situation (thus the 15% detector configuration).The infrastructure will be completed in the next 5 years, after which it will remain operational for another ~15 years. The requirements will scale more or less linearly with the size of the infrastructure. We are in the transition to grid processing and are setting up our workflows in Dirac. We expect that this can go into production within the next year or 2.	I expect little changes to the model, but increased need for computing time and data storage when certain studies ramp up activities
Steady, but if multi-core job + high memory is available it can be useful for some models	The requirements always grow.
Development of more efficient parallelized GPU computing (OpenMPI, etc)	raw data throughput grow up of six time in next two year
I expect the number of cores per job to increase. General changes to the model are described in https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/UPGRADE/CERN-LHCC-2022-005/	

6.4.3. Data Management Needs - "Please briefly explain how you think Data Management will evolve in 5-10 years"

We are looking into adopting RUCIO to handle Data Management for our data.	As mentioned, everything is being defined.
Moving towards wider adoption of Rucio (and other mainstream software such as Kafka) for data management needs, minimizing use of self-supported code.	We will have an explosion of data during HL-LHC we will need many of our R&D effort to succeed
We expect that the current data management infrastructure will still be adequate.	We are building our infrastructure, so this is ongoing work.
More FAIR compliant storage and DM. Increased network capabilities will show that the compute to data model is perhaps less urgent than people think.	There are ongoing community efforts to relaunch an international lattice data grid (ILDG) that was founded in 2002 but fell into disuse during the 2010s. This may encourage more streamlined procedures and easier adherence to FAIR principles. The overall amount of data I expect to continue growing roughly linearly in the long term.
Included in SKA Regional Center	the biggest challenge is the management of

	Auxiliary Data (logs, machine state,...) in a consistent way
The research infrastructure is currently under construction and will be completed in the next ~5 years. The data volume will increase more or less linearly with the number of operational modules. At the moment we are at about 15% of the size of the final infrastructure. The numbers in 7.1 represent the situation for a 15% infrastructure size. When the infrastructure is complete, we expect to require 2.5 PB storage per year. Data taking will continue for ~15 years after completion. The setup of the Rucio data management system is currently in the setup phase, we expect to commission it in the next 1-2 years.	In the process of developing a data management plan for VERITAS, as observatory will shut down in the next 3-6 years; requires sunset procedures to ensure long-term availability and usability of data
Follows the WLCG and CERN ALICE requirements	DIRAC should be able to scale in our use case for the next 5 to 10 years, as we are still far below LHCb scale
Funding agencies will require data to be public but no funding will be allocated no nothing will happen	There are lots of predictions; the only clear thing is that we'll have more data. See https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/UPGRADE/CERN-LHCC-2022-005/

6.4.4. Expected Compute Environment - “About "Computing Environment - Basic architecture", do you expect changes in the next 5-10 years?”

No *[reported by many answers]*	We expect the ability to distribute and process streaming data for low-latency analyses to be developed for the distributing computing infrastructure
The use case is being developed, so that's our goal	Shared (posix-like) filesystems are overloaded on large systems, other solutions (e.g., object-like) are needed.
I don't see significant changes, but more flexibility as we adapt to diverse resources	Evolution but not revolution.

6.4.5. Expected Compute Environment - “About "Access to resources", do you expect any changes in the next 5-10 years?”

No *[reported by many answers]*	The use case is being developed, so that's our goal
We might be using accelerators e.g. GPUs for simulation, but it depends on external packages (Geant4); we might be using GPUs for ML in	Hopefully more accelerators use and better use of existing CPU, network and hierarchical storage. This is basically a skills and capacity question for the

analysis and reconstruction but no clear plans.	research and research-IT community
the data providers (radio telescopes) will continue to provide the same data, only a lot more (2x, 3x as much as now)	More use of X11 or graphical UI
Use of GPUs	More pressure for public data
There is potential for a growing hunger for GPUs in the coming years.	

6.4.6. Expected Compute Environment - "About "Computing environment on the nodes", do you expect any changes in the next 5-10 years?"

The use case is being developed, so that's our goal	No *[reported by many answers]*
Not mentioned above, but python environments (e.g., Conda) are slowly being prevented on large systems due to overload on filesystems. More energy/cost aware computing mechanisms to help instruct users	Depending on the efficiency gain from moving to GPU (if any; currently under research, expected outcome 2.5 years from now) it may or may impact our future needs
The use of accelerators with high memory might render a lot of multi-node calculations to single-node.	Everywhere Container support, IPFS file access
I don't see significant changes, but more flexibility as we adapt to diverse resources	Possible whole-node

6.4.7. E-Infrastructures - "If available, please elaborate on the future (5-10y) plan of the infrastructure In particular: any relevant change with respect to the numbers you inserted any document to be uploaded"

The amount of resources is expected to grow by ~15% per annum. The architecture is adjusted based on market trends and economical/environmental considerations.	High Luminosity LHC will bring "a factor of 10" increase in resource requirements, precise details are an area of active study.
Due to ambitious goals in the future (massive increase in data taking by experiments and collaboration partners, planning of novel generation of accelerator technology Petra IV on campus, etc.), DESY strives to significantly grow computational resources and storage capacities over the next decade. This shall be complemented by joint scientific computing competence activities and the establishment of a research group on computational methods (algorithms, HPC, data processing) for matter (the latter is already in progress). This will also come with demands on energy/compute and storage efficiency.	More quantum, more accelerators and more low-power devices could be desirable if it fits with the bulk of the workloads
Very uncertain and probably we will taper down	We have to follow the requirements of the WLCG with about 10-15% linear increase

7. Findings

In this section, preliminary findings from the various activities are listed. As mentioned in the initial part of this document, the Deliverable D3.1 represents the bottom-up phase in SPECTRUM, in which material is collected via various means from the base of researchers, initiatives, and all involved parties. The second part, from month 16 to month 30, will consider this material during the writing of the Technical Blueprint and the SRIDA. Still, some findings have emerged clearly from the collected material, and are reported in the following sections.

7.1. Findings from the Knowledge Hub

The SPECTRUM Knowledge Hub document repository has grown to be an important corpus of material of various genres (documents, presentations, meetings, etc.), in the domain of SPECTRUM studies.

The CoP was asked to review the corpus of documents, with the aims to:

- Improve the cross-domain understanding of the computing landscape,
- Identify recurrent themes and problems,
- Identify common lines of action.

Among the many proposed for the latter point, the project has identified a smaller number to be proposed for further analysis.

7.1.1. Position towards the utilization of Cloud and HPC systems

The need for extending the resource utilization to resources outside the direct control of the scientific initiative or domain is present in many documents from the Knowledge Hub (for example from KH01 to KH12[8]). Among the various resource providers, High performance Computing Centers (HPCs) and Cloud Providers (academic or commercial) are the most interesting due to their large resource installations, many orders of magnitude larger than the domain-hosted centers. A fruitful utilization of these resources poses non-trivial problems, and as such the integration needs to be prepared well before the actual need.

The corpus of documents shows aligned positions on:

- The need for a political level interaction: current HPC centers have policies (access, data management, network, etc.) that are quite restrictive with respect to more typical centers used by the SPECTRUM scientific domains centers. A technical handshake would be needed in order to be able to insert these systems in our computing infrastructures. The situation with Clouds seems more versatile and less problematic.
- HPC systems (and Clouds, to a smaller extent) are orienting themselves towards Artificial Intelligence workflows, by acquiring resources with GPU accelerators; these are with current scientific codes/workloads in HEP and RA. A future utilization of such systems requires actions on the software development strategies.
- HPC systems are "free" for us (they are paid for by governments and eventually offered for scientific purposes), but there is the "hidden cost" of ancillary services that must be maintained to integrate them into the initiatives' computing infrastructures, in terms of capital expenditures as well as employee time. (Commercial) Clouds are pay-per-use, with cost schemes depending on the SLA[9] requested. A recent study (KH12) has shown how the Cloud cost model is not attractive at the moment for durable computing efforts at large scale, especially when large outgoing data movements ("egress") are expected.
- There is currently small or non-existing standardization among HPC centers, which need to be integrated one by one in experiments' computing infrastructures, with technical solutions varying case by case. This requires much effort from the initiatives, which can overrun the advantages. More standardization would be helpful, and the need for this should be communicated to funding

[8] *References can be found in the Document repository section of this document.*

[9] Service Level Agreement

agencies. Cloud providers, even when using different solutions, have in general much more standardized interfaces, which reduces the integration costs.

The documents and the discussions align on the fact that experiments and initiatives should strive to enable operations on Cloud and HPC systems, in order to maximise the physics output; at the same time, though, funders of HEP/RA and of HPC centers should impose the named centers to realize more friendly interfaces towards the processing systems used in HEP/RA. Missing that, the HEP/RA community cannot guarantee a proper utilization of such systems, at the price of an inefficient utilization (or even, no utilization).

7.1.2. Position about the career problem in computing related tasks

The difficulty of rewarding careers as "computing experts" in the SPECTRUM scientific domains has been cited in many documents in the corpus (KH02 and from KH13 to KH17). The domain transition from a mode of scientific computing, in which every researcher had to contribute to the data collection and analysis system, to large and specialised computing systems, which need dedicated professionals and many years of development, pose a serious risk to the initiatives. Computing professionals (with or without a physics background) need to be stably available for these decennial endeavours, and as such need to have rewarding careers and opportunities. Missing these, these experts, formed in our labs, will find an easy transition path towards the richer industry of large software and infrastructure providers.

In order to maintain a functional environment, it is recommended to (extracted from KH17 for HEP, but assumed valid also for RA, after careful discussion) :

- Maintain a strong investment in career development for HEP Software and Computing researchers to ensure future success.
- Maintain sustainable efforts in computation via continual recruitment and training of the HEP workforce. We need to create an environment that is inclusive, supportive, and welcoming in order to integrate diverse skill sets and experiences.
- Organize training events through HEP experiments, institutes or organizations, and growing numbers of university courses. We need to continue and grow these efforts for documentation and training at multiple levels.
- Push for faculty/stable positions for physicists with expertise in Software and Computing for HEP. These are scarce and the person-power shortfall in this area is endemic. Funding agencies can catalyze faculty-level appointments in S&C with joint appointments at national laboratories.

7.1.3. Position about the needs for an evolution of programming techniques

Cost-effective scientific computing needs the capability to be executed on the best performance / cost platforms, at any time. LHC computing was lucky to start in a very favourable time period, where Linux PCs with Intel processors constituted the fastest expanding market, with strong competition keeping prices down.xw

Today's landscape is far more fragmented (see in the corpus, for example, KH03, KH06, from KH09 to KH11, from KH18 to KH21, and KH63):

- Intel processors have smaller market share, surpassed by "mobile" architectures like ARM and in the near future RISC-V (see **link**);
- The Artificial Intelligence revolution is pushing for accelerators like GPUs, with a decreased interest in standard CPUs;
- No platform (CPU or GPU) is expected to dominate for a long period, asking for an expensive continuous code rewriting to stay on top of the technology.

The scientific domains need to find a reasonably comfortable position in this turmoil of competing architectures. Our software stacks are large (up to 10s of million lines of code for the LHC experiments, as in KH60), and have a lifetime comparable to the lifetime of the initiatives. The field cannot afford a continuous rewriting of the software stack to adapt to the changing scenarios, nor is it big enough to ask for a standardization even on scientific centers like the HPCs.

The most viable solution seems to be a transition to heterogeneous programming models and frameworks, which shield the users from the concrete execution architecture, thus providing an abstraction layer. Studies in this respect have analyzed the existing software solutions, with an evaluation of their usability in our field. Figure 50 shows a comparison between the most used ones from the point of view of the execution on different platforms, as in Figure 51, from [link].

	CUDA	HIP	OpenMP Offload	Kokkos	dpc++ / SYCL	alpaka	std::par
NVidia GPU					codeplay and intel/llvm		nvc++
AMD GPU				feature complete for select GPUs	via openSYCL and intel/llvm	hip 4.0.1 / clang	
Intel GPU		CHIP-SPV early prototype		native and via OpenMP target offload		prototype	oneAPI::dpl
multicore CPU							g++ & tbb
FPGA						via SYCL	

Figure 50: Comparison of available processing backends supported by popular heterogeneous frameworks.

More in depth studies go as far as compiling lists of features, and can be the guide towards the choice of one, or a few preferential solutions.

Metric	Kokkos	alpaka	SYCL	std::par/nvc++	OpenMP
Ease of Learning	Similar to C++ and CUDA, optimization more challenging	Very verbose API and sparse documentation make for steep learning curve	Similar to C++ and CUDA. Lots of documentation.	is C++	C++ with extra pragmas. Sparse documentation/examples for offload.
Serial Code Conversion	Similar to CUDA, though different syntax. Specialized optimizations not straightforward	Work needed to wrap kernels in callable objects. Many typedefs in API benefit from layer of template functions.	Similar to CUDA, though different syntax	very simple	Easy incremental porting mechanism from serial code (add pragmas, get offload working, performance tuning). More work to port from existing CUDA code.
Code Modification	Can be used in an existing complicated application without changes elsewhere. Kokkos runtime needs explicit init/fini. Can take and interpret command line arguments	similar to CUDA	Parallelize loops with command group handlers and parallel_for. If using USM, necessary to orchestrate kernel calls explicitly via waiting on events	Memory accessible on device must be allocated/freed in a file compiled by nvc++, and on the heap. This may require some copying of data.	Special memory allocation and transfer APIs. Can operate device and host parallel simultaneously.
Data Model Modification	Views can be used as a smart pointer to 1D data Crafting a SoA with Views tedious. Jagged arrays (Views of Views) not gracefully supported	Buffers used to wrap existing objects tedious to use. Alpaka managed memory buffers can lead to unexpected behavior	Buffers can be instantiated only from device copyable types. USM is most compatible with current EDM and custom data types.	May need to copy data to make it visible to USM	In general simpler than CUDA
Build System Integration	Can choose at most one backend for each execution space type. Choice must be done at the time of configuring the Kokkos runtime build.	Extensive configurability via CMake	Mostly seamless integration with CMake and make. Depending on the target platform/backend, additional Clang command line arguments are needed.	Need compiler wrapper to filter out options that CMake adds that break compiler. some bugs with gcc lib compatibility	Good integration with CMake and make.
Hardware Mapping	Supports multiple host-parallel backends, and NVIDIA, AMD, and Intel GPUs. Kokkos developers have been pro-active in supporting new hardware architectures as they emerge	Supports multiple host parallel backends, and NVIDIA and AMD GPUs	Can build and run now on any major vendor CPU and GPU. Third-party libraries can be called through interoperability. Backed by Intel, Codeplay, academic institutions and labs.	nvc++ supports CPU serial, CPU multicore, NVIDIA GPU. CPU multicore doesn't work reliably.	Supports multiple host-parallel backends, and NVIDIA, AMD, and Intel GPUs.

Figure 51: Evaluation of popular heterogeneous frameworks.

A single (or a few) domain-level choices among the possible tools would be beneficial for the field, allowing a simpler support and the exchange of expertise between initiatives. The communities should be encouraged (via projects, grants …) to identify and support such solutions.

7.1.4. Position about the use of Artificial Intelligence

Artificial Intelligence is a disruptive technology, pervasively in research, industry and social sectors. Virtually all contributions in the Knowledge Hub mention AI as a present or future key technology. Some of them (K025) explicitly cite HPC systems available to research as the main means to increase the efficiency of tasks which today consume a large fraction of the CPU cycles:

- The possibility to substitute some specific parts of theoretical calculations (for example in Lattice QCD) with automatic differentiation in Physics-driven ML.

- The possibility to substitute CPU-consuming parts in codes like Geant4 with an ML equivalent. In specific cases, it has been proven how nearly indistinguishable results from the physics point of view can be obtained with ML, with a gain of several orders of magnitude in speed[10].
- The emergence of Generative Networks able to take the place of complex algorithms or even chains of complex algorithms, in simulation and reconstruction. A recent example can be found in[11].
- The trend, initiated some 5 years ago if not more, to substitute punctual algorithms with ML solutions, either for regression or categorization.

The use of LLMs is not currently reported as a primary use case in our domains; when mentioned, LLMs are used for ancillary tasks as documentation or first-level user support, but don't impact directly the physics output. On top of this, LLMs have a large impact on the coding activities via tools like Copilot [**link**], but without the immediate need for specific HEP or RA tools. This does not rule out a breakthrough in this respect, with some ideas on both the theory and the experimental sides [**link**].

There are two other points that appear in the corpus, in different sections:

- The need for the availability of an AI-optimized computing environment. This goes hand in hand with the HPC discussion above;
- The need for AI-specific training opportunities, on top of the generic need for computing training.

7.1.5. Position about computing environmental sustainability

Already today, computing infrastructures are among the "top energy spenders" worldwide, with an estimate of "beyond 1%" in 2024[12]. Their impact is expected to increase, doubling by 2026[13]. While the environmental problem is clear, the economic impact is also worrying, especially when linked to international instabilities such as the war in Ukraine. The documents KH55, KH61, KH62, KH63, KS64 in the corpus specifically address the problem of computing sustainability.
There are many aspects to be considered:

- Attempt to increase computing efficiency by writing more efficient codes or using languages known to be more power effective. Figure 52 shows how the same algorithm, when coded in different languages, can have a completely different energy footprint[14]:

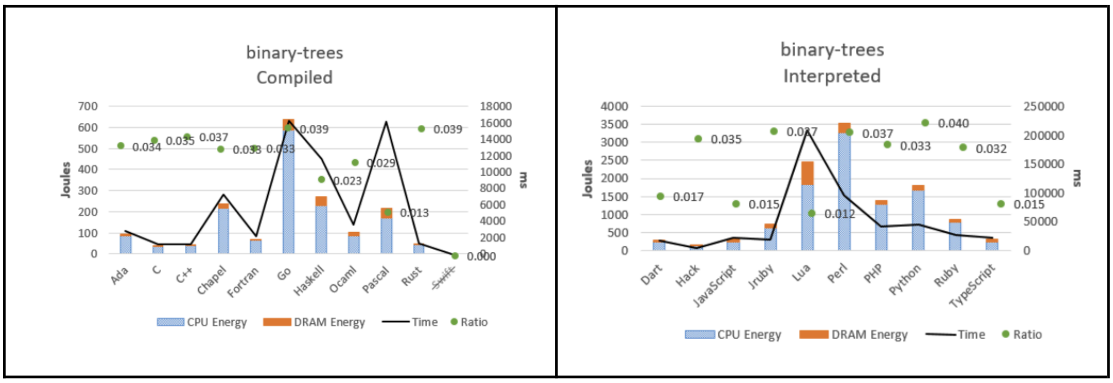

Figure 52: energy consumption for the same algorithm in compiled (a) and interpreted (b) languages. From [**link**].

- The intrinsic energy needs for RISC and CISC computing architectures. In particular, ARM and RISC-V architecture have been designed with mobile-like applications in mind, and have generally a better performance/power footprint. Transitioning to these architectures can spare up to 50% of the energy used for computing.

---

[10] See for example F. Carminati, A. Gheata, G. Khattak, P. Mendez Lorenzo, S. Sharan, and S. Vallecorsa. Three dimensional Generative Adversarial Networks for fast simulation. Journal of Physics: Conference Series, 1085(3):032016, sep 2018

[11] Francesco Vaselli et al., End-to-end simulation of particle physics events with Flow Matching and generator Oversampling, **https://arxiv.org/abs/2402.13684**

[12] **https://www.iea.org/commentaries/what-the-data-centre-and-ai-boom-could-mean-for-the-energy-sector**

[13] **https://www.iea.org/reports/electricity-2024/executive-summary**

[14] **https://sites.google.com/view/energy-efficiency-languages/results?authuser=0**

- GPU programming yields a better performance per node, at the price of a higher energy consumption. Still, with fewer nodes required for any given task, the global Performance per Watt can be better.
- Computing centers use a substantial fraction of the energy to cool the hardware; the Power Usage Effectiveness (PUE) factor accounts for the extra cooling power: a PUE of 1 denotes a perfect cooling system, while a larger number denotes excess energy utilized. Technologies like adiabatic cooling allowed in the last 20 years a substantial reduction of the PUE, and is Figure 53. Still, even better values can be reached as demonstrated by the new data center at CERN [link], with a PUE of 1.1. Most
  HEP computing centers date back to the start of LHC operations (15 years ago) and need a renovation to reach similar optimal levels.
- An optimization of HEP/RA hardware can result in better power/performance ratios, for example altering the default running CPU clocks. A summary of current studies can be found in KH64.

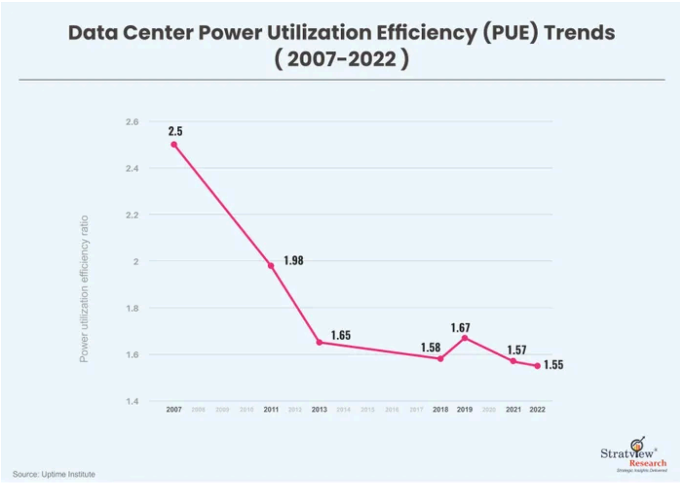

Figure 53: Data Centre PUE trends since 2007 (from [link]).

- Hardware lifetime is a key variable for sustainability. On the one hand, replacing hardware frequently with more efficient purchases reduces the electricity running costs, on the other one must offset the cost of producing the new hardware and of the disposal of the old one. Further analyses are suggested in KH55, to understand an optimal working point.
- The use of Renewable Energy can offset the carbon footprint of large installations. The document KH55 explores, as a test case, solutions – from practical to extreme – for an ideal utilization of renewable energy to power the CERN complex:
    - Solar power: CERN has 653 buildings with a total roof area of 421,000 m^2, which amounts to approximately 80 GWh annual electricity generation potential. A comparison with the electricity consumption in 2019 of 428 GWh, when the LHC was not in operation, shows that around 18% of CERN's basic (non-LHC) electricity could be produced locally with solar power.
    - Clean power from the desert: CERN is ideally placed to spearhead a project to import energy from countries rich in renewable energy sources, and transport it across international boundaries. A scenario for connecting, e.g., Morocco, Algeria or Tunisia to Southern France, Spain or Italy by sub-sea cable is plausible from a technological point of view. Costs are estimated to be around 0.06–0.07AC/kWh for a year-round power supply of 3.6 GW in the daytime and 2.2 GW at night. Electricity supplied on this scale would exceed the power needs of CERN, and surplus power could be returned to the European electricity grid to power other research institutions and universities that join the initiative.

Figure 54: importing solar power from northern Africa to CERN. From [link].

It is to be expected that the power needed by SPECTRUM's reference domains will increase with a similar trend observed in general computing, given the initiatives which will be launched from now the next 5-10 years. The Computing for LHC, for example, is expected to increase at least 10-fold on that time scale, even in the most optimistic option (see Figure 54).

Hence, it is recommended to increase the effort (from the research community, but also from the funders) towards a more sustainable computing, attacking the problem from the code stack and the infrastructure sides.

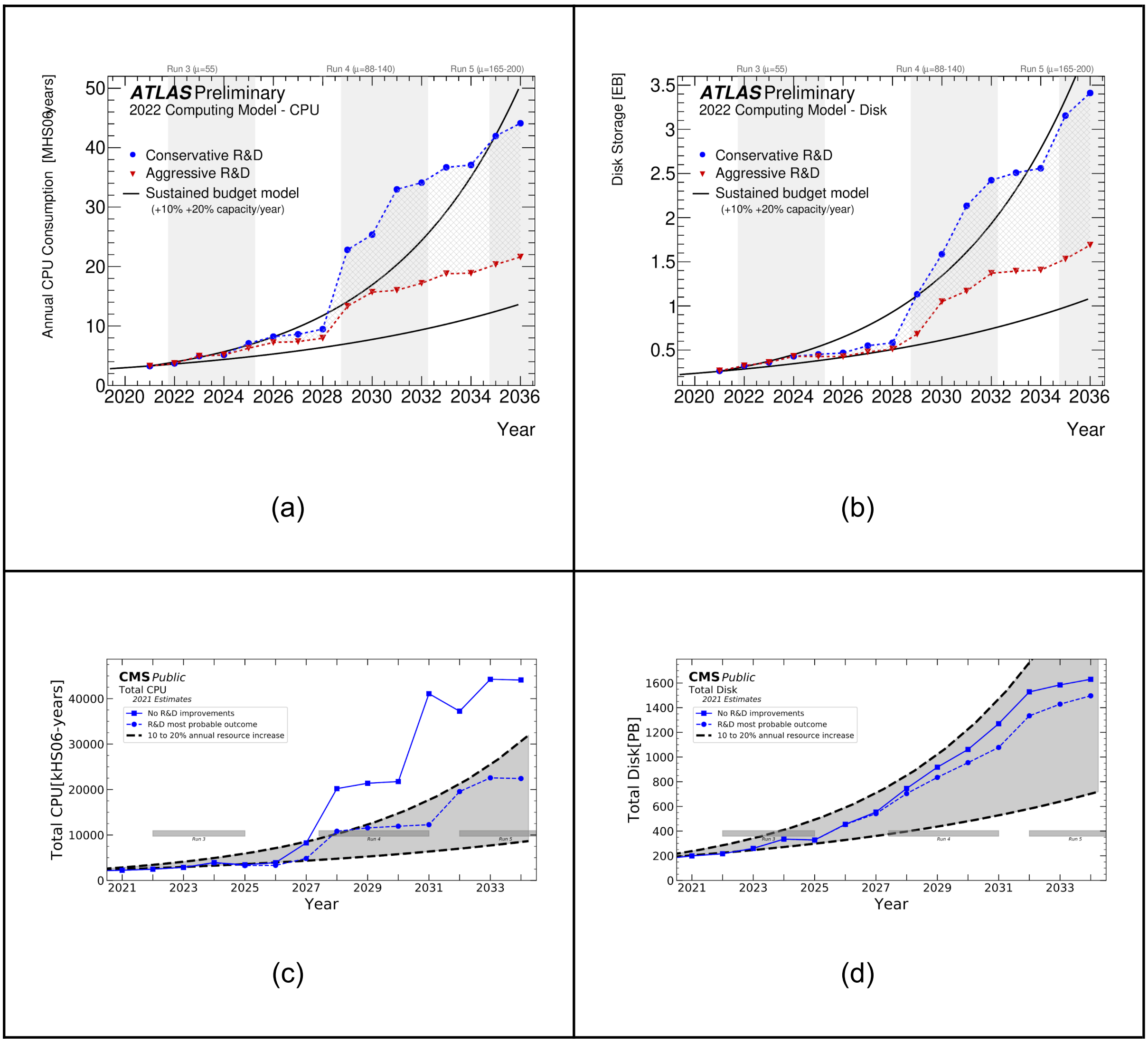

Figure 55: expected evolution of computing resource needs ((a) and (c): CPU; (b) and (d) disk) for the ATLAS ((a) and (b)) and CMS ((c) and (d)) experiments at CERN, for the next decade. [link, link].

7.1.6. Position about inter and extra domain collaborations

As much as instruments are generally "prototypes in production", built by research scientists using generally purpose-built parts , computing in the HEP / RA domains is internally planned, deployed and executed. Even now, with the availability of commercial / public clouds and regional level HPC centres, the computing model (KH60) of most initiatives is one-of-a-kind, specifically designed and tailored to the specific initiative. Examples from the Knowledge Hub corpus are in KH01, KH03, KH04, KH22, KH31, KH38, KH43, KH52, KH59, and more.

Only quite recently efforts have started to harmonize the computing of multiple initiatives, either covering very closeby domains, or attempting a cross domain vision. The most advanced is the Worldwide LHC Computing Grid (WLCG), active since 2005 and covering initially the needs of the four major LHC experiments, and by now including as partners also Belle-2, DUNE, JUNO, and Virgo. The transition from LHC-only happened thanks to the recognition that some computing tools (for example data management, workload management, authentication and authorization) are indeed generic for a broad range of close-enough domains.
Stemming from experiences like WLCG, various EU funded initiatives have broadened the inter-domain collaboration: the ASTERICS project has developed a Virtual Observatory framework for astroparticle initiatives, in particular SKA, CTA, KM3Net & E-ELT. Its successor, ESCAPE, has further expanded the development and deployment of computing tools to CTAO, EGO-Virgo, EST, ESO, FAIR, HL-LHC, JIVE,

KM3NeT, LSST, SKAO and demonstrated the viability of common solutions. OSCARS is exploring other STEM and not-STEM domains.

It is by now clear how a large part of basic computing tools are domain-independent, and are valid in most cases where the computing infrastructure must (by internal or external decision) be large, distributed, heterogeneous, and open to a large number of collaborators – eventually with highly granular access patterns. This directly aims for more strict connections to (at least) EuroHPC JU and the EOSC.
The evaluation of initiatives bringing together close or not-so-close scientific domains results extremely positive in the Knowledge Hub corpus, and seems a very reasonable, cost-effective and efficient way to deploy complex computing infrastructures for the initiatives of the next decade. To this extent, their funding should continue or accelerate.

7.1.7. Position about Open Data and FAIR data and software

The High Energy Physics and Astronomy/Astroparticle domains come from very different experiences about data open-ness. In the HEP domain, before the end of the last century, data was private, accessible to the initiatives' members or via special agreements. In the Astro world, mostly thanks to the initial commitment of NASA, the situation is drastically different, with data made open soon or close to data acquisition, coming from considerations on:

- Data being acquired in publicly funded initiatives, and thus to be returned to the public;
- A better standardization of data formats (for example, the FITS format), increasing an immediate utilization of experimental results.

On the other hand, HEP data is (still today):

- Complex to interpret and dependent on a deep knowledge of the experimental apparatuses and on the data taking conditions. A simplified analysis of data missing these would generate wrong scientific results;
- Large, at the scale of Petabyte / Exabyte, and CPU-expensive to be analysed.

These two reasons, apart from any political / ethical consideration, reduced the interest in an open distribution of HEP data. The landscape is changing, with HEP getting closer to the Astro position. This stems from multiple factors:

- The interest in the availability of small chunks of data, for education purposes;
- The availability of small and curated data formats, which require less insight on the experiment details;
- Resolutions requesting "FAIR data by default" for EU-funded scientific initiatives (see for example this link)[15];
- An increased public interest in HEP experiments and their data, even as a playground.

In the Knowledge Hub corpus, documents KH15, KH32, KH35, KH44, KH45, and KH49 explicitly mention the data FAIR-ness as a much needed feature for current and future initiatives. The current HEP situation has indeed already improved with respect to the scenario described at the start of this section: LEP experiments, but much more substantially LHC experiments, started deploying data and software with open licenses. In particular CERN has announced a policy in support of Open Science [link], and a portal to host and distribute Open Data [link].

The corpus shows that in several cases an embargo period is used by experiments, which can vary from less than 1 year for space data, to 5-10 years for LHC data (depending on the experiment and on the data tier). All documents agree that releasing open data (after possibly embargoes) and open software should be an ab-initio decision when planning new initiatives; the necessary funding should be included into the planned budget.

7.2. Findings from the Survey

The SPECTRUM survey is a tool designed to collect "bottom-up" input from the community, in order to guide discussions and in-depth analyses of interesting parts; it was not designed as a tool to directly infer the

[15] Please note that FAIR data does not necessarily imply open data: "The A of FAIR – As Open as Possible, as Closed as Necessary" [link].

directions to be included in the Blueprint and the SRIDA. Still, some messages emerge quite clearly and are worth reporting in the following sections.

7.2.1. Authentication and Authorization

Most of the initiatives declare they use some sort of "modern" AA Infrastructure, either provided by the hosting institution or via available federated tools; only a small fraction of the initiatives appear to remain attached to old style mechanisms like local systems or username/password.

There is a lot of fragmentation in the actual tools used, almost equally shared between Indigo-IAM, CERN-SSO and EGI-checkin. The situation is probably acceptable, up to the level where these tools are interoperable.
The message one gets from the survey is, still, quite confused, with no clear path highlighted. This could impact future initiatives, which do not have a clearly defined go-to tool for the implementation of the AAI.

7.2.2. Processing Needs

The section on processing needs shows quite a varied landscape of applications our domains need to execute, varying from interactive / graphical analysis to batch submission.
Also on the typical resources needed, the landscape varies a lot: from single process CPU workflows, to multi-node parallel workflows, to large allocations on grids lasting several years. The main finding is that whatever computing system the domains will have to manage in the next decade, it needs to be designed with diverse utilization patterns in mind, and deployable of heterogeneous resources like HTC, HPC and Cloud centers. The advent of AI and in general of GPU programming increases the difficulty to maintain such diverse systems under the same workload management.

Future scientific computing will be more and more data intensive, which poses serious requests to the centers storage systems and even more on the networking capabilities, internal and external. A high level agreement should be established with the funders to guarantee the usability of future large scale computing systems from our domain sciences.

7.2.3. Data Management Needs

Data needs from the survey respondents vary by 4 orders of magnitude, from the <10 TB to the thousands of PBs per year. Data is collected at multiple sites and simulation centers. Data needs to be ingested at a rate up to TB/s or more, in most cases 24x7 along many years.

Rucio, from CERN, emerges as the most used data management tool, in the majority of cases with **FTS** used as the transfer management tool. The apparent uniformity is indeed not real: FTS can handle multiple transfer protocols, and indeed when looking at the supported protocols the landscape is varied, with almost equal utilization of XrootD, S3, GRIDFTP and WebDAV (among the "modern" tools).

Still, the situation is to be considered almost optimal: a thin management layer (Rucio and FTS) is able to shield the complexity of multiple low level data movement tools, which are unavoidable and optimized for different use cases. A different dimension is the need for databases instead of "files"; for that, a more in-depth analysis is needed to understand the level of integration possible between domains, not apparent from the survey.
Metadata handling is needed when data is not self-descriptive (most of the cases); metadata management seems still to be an experiment-internal matter. It would be interesting to see whether this can be evolved into tools which can support multiple initiatives.

7.2.4. Expected Compute Environment

As in the previous sections, the landscape which emerges from the survey is quite varied, with different expectations on technology, policies, and libraries. About the Operating System, Linux has the greatest share, with some answers considering Windows and MacOS relevant platforms, probably for interactive activities.

About the compute architectures, Intel 64 bit ("x86_64") is supported practically everywhere, with a few initiatives able to validate at least a fraction of the code on ARM. When considering GPUs, Nvidia absolutely dominates, with AMD and Intel appearing as possible solutions in a quite large share of the use cases. As expected,, the fraction of code offloaded to GPUs is low, with only a few cases above 50%. This calls for an increased effort on porting, as already seen in the Knowledge Hub findings.

The most used software installation methods are (1) containers, (2) CVMFS, (3) do-it-yourself. We find the trend positive, since:

- 5 years ago containers would not have been an option, and now are the relative majority;
- CVMFS was limited to High Energy Physics, while now is the go-to solution for a large number of sciences and initiatives.

The use of containers AND CMVFS at the same time is in our opinion the solution to be suggested to new initiatives: containers allow a stable and uniform computing environment (at the level of base OS, system libraries and resolved dependencies), while CVMFS allows for rapid deployment of software packages and "small scale payloads" on that stable environment.

Network connectivity from/to the compute nodes is becoming a necessity for distributed computing systems in which data is located on sparse storage systems. The capability to connect from compute nodes to "the internet" is going to become more and more a necessity, and should be taken into account when planning the deployment of new systems.

7.2.5. Software, software development, software distribution, policies

On the programming languages, it is clear that our "long range initiatives" have decade (or more) long legacy codes, with FORTRAN still being high on the preference list. As expected, C/C++ for performance critical code and Python for analysis code are the most popular options. The appearance of Julia is in our opinion the start of a trend which can become relevant on a very short time scale.

One aspect in which we would like to push for larger and more organized efforts is the one of accelerator programming: most initiatives focus on a single technology (for example, CUDA on Nvidia hardware). This will be detrimental in the medium-long time scale: any porting effort done now to port from serial code to a specific accelerator technology might need to be repeated when/if another accelerator technology would become mainstream. Our environment does not have, by far, the personpower needed to port, validate and maintain software on different platforms every few years, hence it is somehow unsatisfactory to see the low use of frameworks / toolkits. This should be encouraged via projects, initiatives, and grants.

When considering career and training opportunities, the situation described in the survey is negative. On the one side career opportunities are "limited at best", on the other even for the few which participate actively in the software development process the training opportunities are not sufficient. This probably stems from the consideration that in previous initiatives computing was not an essential task, or at least one which could be fixed after the preparation of the instruments and the data taking. In today's and tomorrow's initiatives, the complexity of the computing system is such that it needs a long range planning, and human resources at least on par with instruments' design and preparation. The survey results clearly express the worry that in a short time scale we will not be able to design and operate the large computing systems needed for our initiatives, for scarcity of career paths and for the concurrent increasing opportunity for data/software scientists in the private market. No definite solution is hinted at, but the problem should reach the highest level of our domains in order to find countermeasures.

7.2.6. E-Infrastructures

The survey section on e-Infrastructures was designed to collect the view of managers of HTC, HPC, Cloud and Quantum Computing centers. Its main use is to collect the "offer" part of the scientific computing, to be compared with the "needs" part from the scientific communities. The landscape discovered is a varied one, with small / medium / large centers participating in the survey. Most of them are part of a larger computing effort, with only 4 cases stating they are funded specifically for a local purpose.

Apart from the resource deployment and type, the most critical aspect when trying to match needs and resources at HPC centers has been the need for external networking from the compute nodes (see also

KH41 from the Knowledge Hub). The survey results are unexpectedly positive, with most centers declaring their availability to have at least some connections enabled – and with good bandwidth; this will need to be checked for real / realistic use cases.

Virtualization is supported by 100% of the respondents, which is a huge improvement over just a few years ago. In practice, virtualization is expected to solve a large series of typical problems (software installation, use of commercial software, isolation, etc.) and its use should be increased as much as possible. CVMFS is the second in the list of supported features, which again bodes well for our domains; eventually this is understood as a consequence of diverse sciences asking for it.

A relevant political discussion is the one about the resource granting systems. In order to positively impact the computing of scientific initiatives, the resources have to be made available on a longer time scale than a few months / one year. The results show that this is a possibility in the majority of cases, which in our opinion is contradicting the current experience. In any case, it is important to initiate steps towards the e-Infrastructure funders in order to guarantee a long-term recognition towards the major centres, for SPECTRUM scientific domains.

8. Conclusions

The present document is the input of WP3 and the CoP to the SPECTRUM Project, to be used for further evaluation in WP 5, and for the realisation of final project reporting in WPs 4, 6 and 7.
It contains:

1. The description of the CoP process
2. The links to external projects
3. The description of the Knowledge Hub
4. The description of the SPECTRUM Survey
5. A light section of findings, from the Knowledge Hub and the Survey

Many findings and recommendations are proposed in the Findings section; as explained, they are not meant to constitute a final SPECTRUM output, but to guide the discussions and the definition of the Technical Blueprint and the SRIDA in the second phase (Months 16-30).

A very concise list of the findings and coherent positions from the documents and the survey includes suggestions on how to address an evident career problem in domain-specific computing tasks, the need to evolve our software stacks towards more efficient and performing computing architectures, the need to embrace AI as a base tool for our software, and policy aspects like the need to adhere to FAIR principles and participate into inter and extra domains collaborations.

Part of this document will be submitted as SPECTRUM input to the Update of the European Strategy for Particle Physics (**ESPP2024-2026**), an initiative that will produce a roadmap for Particle Physics in Europe.