SPECTRUM

D1.3 Data Management Plan

Status: Under EC Review

Dissemination Level: Public

Funded by the European Union

Abstract	
Key Words	Data Management, FAIR, GDPR, metadata, re-usable data
This Data Management Plan (DMP) introduces a report that specifies how research data will be collected, processed, monitored, and catalogued, following the FAIR principles. This deliverable is viewed as a living report that will advance throughout the project's life.	

Document Description			
D1.3 Data Management Plan			
Work Package Number 1			
Document Type	Report		
Document status	Under EC Review	**Version**	1
Dissemination Level	PUBLIC		
Copyright status	 		
Lead Partner	EGI Foundation		
Document link	https://documents.egi.eu/document/4063		
Digital Object Identifier	https://doi.org/10.5281/zenodo.12685887		
Author(s)	• Sjomara Specht (EGI)		
Reviewers	• Sergio Andreozzi (EGI) • Andrea Anzanello (EGI)		
Moderated by	• Sjomara Specht (EGI)		
Approved by	• Corentin Lefevre (NEOVIA) - on behalf of AMB		

Revision History			
Version	**Date**	**Description**	**Contributors**
V0.1	08/05/2024	ToC	Sjomara Specht (EGI)
V0.2	13/06/2024	Initial first input	Sjomara Specht (EGI)
v0.3	20/06/2024	Work Package & CoP dataset tables (Annex2)	Patricia Ruiz (EGI) Tommaso Boccali (INFN) Chiara Ferrari (CNRS) Hans-Christian Hoppe (FZJ) Luis Cifuentes (FZJ) Maria Girone (CERN) Xavier Salazar (EGI)
v0.4	27/06/2024	Finalised and ready for review	Sjomara Specht (EGI)
V0.5	04/07/2024	External review	Sergio Andreozzi (EGI) Andrea Anzanello (EGI)
V0.6	05/07/2024	Incorporated feedback from reviewers	Sjomara Specht (EGI)
v0.7	08/07/2024	Approval from the Board	Corentin Lefevre (NEOVIA)
V1.0	**08/07/2024**	**Final**	

Terminology / Acronyms	
Terminology / Acronym	**Definition**
CSA	Coordinated and Support Action
CoP	Community of Practice
DMP	Data Management Plan
DocDB	EGI Document Database
DOI	Digital Object Identifier
DPO	Data Protection Officer
ESFRI	European Strategy Forum on Research Infrastructures
FAIR	Findability, Accessibility, Interoperability, Reusability
GDPR	EU General Data Protection Regulation
HEP	High Energy Physics
OA	Open Access
ORE	Open Research Europe
PDI	Persistent identifiers
RA	Radio Astronomy
R&D	Research & Development
SSO	Single Sign On
URL	Uniform resource locator

Table of Contents

List of Tables

Executive summary

The deliverable D1.3 – Data Management Plan – defines the structure within which SPECTRUM will create, manage, and collect data during the project's operations. Furthermore, it specifies how the data will be used or made available for verification and re-use, as well as how the data will be curated and stored once the project is completed.

Within the SPECTRUM project there are two main datasets identified;

- **Project Results Dataset**: Includes all documentation related to the projects outputs such as reports, deliverables, presentations, and meeting minutes. These documents are essential for tracking projects progress and outcomes.
- **Community of Practice (CoP) Datasets**: Comprises data collection from the CoP, scientific communities and experts in the compute and data continuums, for instance through surveys and interviews. These datasets capture insights and feedback from the project participants and experts from the target groups, which are crucial for shaping project strategies and decisions.

Additionally, the Data Management Plan (DMP) specifies the approach for ensuring that data complies with the FAIR (Findable, Accessible, Interoperable, and Reusable) principles. It establishes comprehensive agreements on data security and addresses the ethical consideration involved in the collection and generation of the data.

SPECTRUM adheres to Horizon Europe Open Science FAIR[1] principles and strives to make data as open as possible and as closed as appropriate. The beneficiaries share the project's data in such a way that it is valuable to partners and their users outside the project while ensuring that the privacy of third parties that participated in the data collection/generation is preserved.

Under these conditions, the data will be managed and released in compliance with the certifications and safeguards of the EU General Data Protection Regulation (GDPR[2]). Every dataset is examined (in terms of sensitivity, privacy, and security) before an official decision is made on whether or not to make that specific information public.

This deliverable is viewed as a living report that will advance throughout the life of the project.

[1] https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-dissemination_en.htm

[2] EU General Data Protection Regulation (GDPR)

1. Introduction

The current initial version is composed of preliminary information and frameworks that will be followed. Hence, it is subject to updates in the future upon developments and changes during the project. In principle, the DMP describes the standards that will be used, and how the project's research data will be stored and published for verification and reuse. SPECTRUM aims for full open access to data, following the FAIR principles.

1.1. Purpose and Scope of the document

The SPECTRUM Projet is a Coordinated and Support Action (CSA), therefore the majority of results will be textual outputs such as reports, presentations, white papers, and meeting minutes. These will be primarily gathered from Working Group experts, scientific, and technical communities (primary sources) and supplemented by papers and other textual reports (secondary sources). All personal data will be managed in accordance with the General Data Protection Regulation (GDPR).

Open Science practices are fundamental to our methodology. The creation of the **Community of Practice** (CoP) fosters open, collaborative work to share knowledge across scientific communities and experts in computing and data continuums. At the project's outset, we have established Working Groups in relevant domains, which facilitates knowledge sharing and consultations with experts from **High Energy Physics** (HEP), **Radio Astronomy** (RA), **e-infrastructures**, and key **ESFRI clusters**, such as the **ESCAPE Science Cluster**. We will validate early results within broader data-intensive Research Infrastructures (RIs), both online, through targeted dissemination, and indirectly via e-infrastructure user engagement channels. All communication and dissemination outputs will be released under a Creative Commons - CC BY license[3], ensuring maximum accessibility and reuse.

1.2. Structure of the document

This document is comprised of the following chapters:

Section 1: presents an introduction to the project and the document.

Section 2: presents the purpose of data collection, type and format, origin of the data and its expected size.

Section 3: outlines SPECTRUM FAIR data strategies

Section 4: briefly describes the allocation of resources.

Section 5, 6 and 7 outline data security, ethical and other issues.

Section 8: concludes this deliverable.

[3] https://creativecommons.org/licenses/by/4.0/

2. Data Summary

This section provides a comprehensive overview of the data managed during the SPECTRUM project, including details on data types, origins, and the tools used for data collection.

Potential data assets generated by the project are:

- **Personal contact data:** Information collected during user registration and project participation.
- **Project documentation:** Includes reports, presentations, meeting minutes, strategic documents, and other communication materials.
- **Surveys, questionnaires, and interview data:** Data generated from engaging target groups collected using tools like EUSurvey[4]. This includes both structured (quantitative) and unstructured (qualitative) data

Types of data:

- **Qualitative data:** Includes open-ended survey responses, interview transcripts, detailed feedback, and narrative descriptions from meetings.
- **Quantitative data:** Collected through structured survey responses, statistical data from usage analytics and numerical data from project activities.
- **Derived data:** Processed or aggregated data such as summarised survey results, analytical reports, and synthesed information from raw datasets.
- **Multimedia data:** Audio recordings of interviews, video recordings of presentations, and images from project events.

Origin of data:

- **Primary sources:** Data directly generated through project activities, including feedback from participants and CoP members.
- **Secondary sources:** External datasets, academic articles, public documents, and technical reports used to support project objectives.
- **Data documentation:** Detailed descriptions and metadata for all datasets are provided in Annex 2. This includes structure, content, and data dictionaries for easy understanding and re-use.

Survey Tool:

- EUSurvey: We will utilise EUSurvey, an EU-hosted survey platform, for conducting large-scale surveys. This platform ensures robust data security compliance with GDPR and supports various survey types and data formats.
- **Purpose of EUSurvey:** To facilitate efficient and secure collection of quantitative and qualitative data from a broad audience, including project participants and external stakeholders.

Personal contact data are collected during the EGI SSO (Single Sign On) account creation by the users. After the access is granted, the users can manage their data in autonomy.

Lists of participants in general project meetings,work package meetings and other groups (Community of Practice Working Groups for instance) are kept as internal documents in the respective work package pages on the Project Confluence and are used for project management and internal communication.

Results from surveys of target groups are stored in the respective Project Google Drive folder. Personal data collected via EUSurvey is stored on the EUSurvey platform for 1 month after the survey's deadline expiry.

[4] https://ec.europa.eu/eusurvey/home/welcome

2.1. Purpose of the data generation and re-use, in relation to the project's objectives

The **overarching vision** of the SPECTRUM project consortium is that data-intensive scientific collaborations will need to have access to a European Exabyte-scale research data federation and compute continuum. The project's overall objective is to deliver a Strategic Research, Innovation and Deployment Agenda (SRIDA) that defines the vision, overall goals, main technical and non-technical priorities, investment areas and a research, innovation and deployment roadmap for data-intensive science and infrastructures, to be used as a stepping stone for the R&D and pilot activities which will eventually bring the systems into operation.

This project's overall objective is addressed through five specific **Project Objectives (POs).** Table 1 shows the Project Objectives (POs) and their corresponding description.

Table 1: Project Objectives - titles and descriptions.

Title	Description
PO1: Join efforts of research infrastructures and e-Infrastructures to address common research and innovation needs towards exabyte-scale computing	Bring together the computing infrastructure efforts across the EU to support the next generation of RA and HEP experiments, as well as other relevant scientific domains. By collaborating on shared challenges, the project aims to eliminate the fragmentation caused by independent work streams and promote a more integrated approach to Exabyte-scale computing.
PO2: Identify the relevant use cases, related challenges and opportunities	Many initiatives have worked to define and deploy federated systems for HPC and/or scientific data management in the last 25 years. Success was achieved for specific technology nodes and/or scientific user communities, yet all attempts to develop a common system spanning multiple scientific communities and resource providers did not result in sustainable solutions. SPECTRUM is committed to engaging the scientific and technical communities to identify general use cases across domains and related challenges and to develop opportunities for common solutions.
PO3: Understand the landscape and best practices	A landscape analysis of existing infrastructure and services is essential to help identify areas for improvement and to define effective strategies for their implementation.
PO4: Increase collaborative service delivery by e-infrastructures at national, European and international levels	The various HPC and data centres across Europe all implement their policies and mechanisms for authentication and authorization of users, for submission of workflows and tasks to compute resources, and for identification of and access to data. Such policies and mechanisms guarantee local security and safety requirements, interact with effective system management and allow efficient system operation. For the sake of pan-European RIs, it will be imperative to define higher-level interoperable access policies and facilitate uptake.

PO5: Agree on strategic action paths, specific actions and policy recommendations	Ensures that the identified use cases, challenges, service gaps, and Technical Blueprint for the compute and data continuum will translate into an agreed long-term strategy with identified actions and policy recommendations for all relevant stakeholders.

2.2. Expected size of the data

The overall volume of the data generated and managed throughout the project is expected to be less than 5 GB. This includes documents, survey responses, and multimedia files.

3. FAIR Data

The data generated will be managed according to the FAIRness in line with Open Science principles. To ensure that data and project outputs are easily discoverable, they will be made accessible through the SPECTRUM Community in Zenodo and other reliable open repositories. Access will be provided according to open-access policies. Interoperability will be promoted by including all essential information, data, and metadata along with the reports. Lastly, to ensure reusability, data and project outputs will be shared under permissive licences, such as CC BY, whenever feasible.

3.1. Making data finable, including provisions for metadata

Data will be designed to be FAIR and will be made available as openly and as early as possible. Data will only be restricted when required by regulatory or legal constraints or due to the owner's legitimate interests. Raw data from consultations will be aggregated and anonymised to ensure it can be reused for future purposes.

3.1.1. Metadata

The metadata required by data repositories will be used, as outlined in Table 2.
For documents, SPECTRUM has defined a standard set of metadata that should be used, as shown in Table 3.

Table 2 - Repository metadata

Element	Definition
Title	A name given to the source.
Upload type	e.g., dataset, workflow
Abstract	Describing the document contents and main conclusions.
Submitter	The person submitting the document to the repository
Authors	The people involved in writing significant portions of the document.
DOI	Provided by the resource
Publication date	The date of first publication.
Version	The version number generated by the document repository for the repository identifier. Versioning rule: · +0.1 – a new version of the draft · +1.0 – a new version of the approved document
Language	A language of the intellectual content of the resource.
Keywords	A list of words that will support the search within the repository service
Communities	A specific community in which the upload will appear.
Licence	Specifies the copyright status under which the upload will be licensed.
Modify	The groups can modify the document. The 'EGI office' SSO group must be marked.

Table 3 - Document metadata

Element	Definition
Title	A name given to the source. For milestones and deliverables as described in the Description of Work.
Lead Partner	The recognised short name of the lead partner within the SPECTRUM project
Authors	The people involved in writing significant portions of the document.
Reviewers	The people involved in reviewing the document.
Copyright status	The material is licensed under a **Creative Commons Attribution 4.0 International License**
Document Type	e.g., deliverable, report, white paper
Status	· **Draft** - the document is being prepared · **Under EC review** - the document is submitted to the EC portal and has not yet been approved by the European Commission · **Approved by EC** - the document is approved by the European Commission · **Final** - status of the document
Dissemination Level	· **Public** - can be shared without restrictions · **Confidential** - can be shared only with European Commission and project partners
Document link	The URL in the document repository that provides access to the document on **DocDB**.
Digital Object Identifier	An identification number is assigned through a repository service.
Keywords	A list of words that will support the search within the Zenodo service
Abstract	Describing the document contents and main conclusions.

3.2. Making data accessible

Textual outputs such as reports and presentations, except for administrative documents containing sensitive legal, personal, or financial information, will be made available to the SPECTRUM Community on Zenodo upon release. These outputs will be shared under a permissive licence, such as CC BY, and accompanied by comprehensive metadata under CC0[5]. The metadata will include details like authorship, ownership, provenance, access rights, and persistent identifiers (PIDs), all in alignment with the FAIR principles (Findable, Accessible, Interoperable, and Reusable).

3.2.1. Repositories

All documents, presentations and other materials that form an official output of the project (not just milestones and deliverables) are placed in the document repository[6] to provide a managed central location for all materials.

In addition, public deliverables, and publications, will be shared publicly via the **Zenodo platform** to increase the discoverability of the project outputs.

All profiles, specifications, configuration files, software, workflows, and code will be deposited in Zenodo. Therefore, SPECTRUM will use DocDB and Zenodo as their standard and main repositories.

[5] https://creativecommons.org/public-domain/cc0/
[6] https://documents.egi.eu/

3.2.2. Standardised access protocol

All data will be accessible via a uniform resource locator (URL) or Document Object Identifier (DOI). There will be no restrictions on the use of the research outputs, both during and after the end of this project. People accessing the data will not need to be identified and there is no need for a data access committee.

3.2.3. Metadata availability

Metadata containing information to enable users to access the data will be openly available and published together with the data, in the same repositories as listed under **Repositories**. There is no time limit on metadata and data availability.

The SPECTRUM project acknowledges the value of documentation for interoperability purposes, increases uptake by different communities, and encourages data owners to document their research data assets. SPECTRUM does not enforce specific provisions on documentation as long as the data asset is hosted on one of the mentioned repositories and properly curated according to the repository's best practices.

Research data itself should not be considered self-documenting and each published asset must be associated with sufficient documentation resources accessible through a public URL. Documentation must be browsable and include hypertext references to facilitate its fruition. Recommended documentation formats include markdown, HTML, and other markup languages. The inclusion of machine-readable documentation such as OpenAPI where applicable is thoroughly encouraged. If a scientific publication is tied to the research data asset, the publication itself should be referenced and/or made available as part of the documentation.

3.3. Making data interoperable

The SPECTRUM consortium acknowledges the importance of data interoperability: if a data asset cannot be compared, merged, or otherwise integrated with other assets, then its publication would bring little or no value to the community. The consortium is therefore committed to support in and enforcing data interoperability.
Data will be made available in standard formats including;

- Plain text such as; .docx, .txt, .rtf, .pdf, .pptx, .xml, .html, CSV.
- Multimedia such as; jpg/jpeg, gif, png or video format.

Interoperability of the open data will be enabled via **Zenodo** platform, which will store its metadata in JSON format according to an established JSON schema.

3.4. Increase data re-use

Peer-reviewed research papers produced during the project will adhere to the European Commission's Open Access policy for publications. Preprints will be deposited in reputable open repositories, such as Arxiv or institutional archives, whenever feasible. Upon publication, these papers will be self-archived in Zenodo. Ideally, they will also be published in open-access journals or platforms, with Open Research Europe (ORE) as the preferred choice, or in selected Diamond Open Access journals. All articles will be issued with open-access licences that impose minimal restrictions. Both the papers and related materials will be made available in the SPECTRUM Community on Zenodo as needed.

4. Allocation of Resources

Any expenses associated with the collection/production of FAIR data during the SPECTRUM activities are included in the project budget. These expenditures will be required to cover a variety of particular data processing and data management operations, ranging from data collection and documentation to storage and preservation to distribution and re-utilisation.

These operations are a component of the WP that processes the relevant data, hence the needed effort will be part of the relevant WP.

The expenses of long-term data preservation are minimal, by using the EGI Document Database and Google Drive platforms. Using Zenodo (free of charge) ensures that costs for long-term preservation of the data are manageable. When applicable, a more accurate cost estimate will be provided at a later stage of the project.

4.1. Data Management responsibilities

Within the SPECTRUM project, the following roles and responsibilities are associated with Data Management, which are defined as follow:

WP leaders are in charge of organising the data processing and quality assurance that take place inside the Work Package they are leading.
Task Leaders are in charge of the data compiled/produced throughout the operation of the task that they are responsible for. In addition to that, they also make sure that the data are properly prepared to be shared among the partners, and made publicly available, when applicable.
Data Processors are consortium partners who execute processing operations on the compiled/produced data.
Quality and Risk Manager supports the WP leaders, and Task Leaders in keeping the DMP confluence pages up to date.

5. Data Security

Any gathered data will be securely handled throughout the entire duration of the SPECTRUM project, to protect it from loss and unauthorised access. Personal data is only accessible to those who are authorised to access it.

All partners/beneficiaries responsible for processing personal data have the responsibility to ensure that the data remains protected under all necessary security controls (including backup policies and integrity checks) and access controls (identification, authentication, authorization) within their infrastructure. In the unfortunate event of a personal data breach, the project partners will notify without delay their competent national supervisory authorities as well as the data subject(s) that may be affected by the breach. At the same time, they will document any personal data breaches and all related information.

Regarding open data, for security and long-term preservation, SPECTRUM relies on EGI Document Repository, wiki.egi.eu, Zenodo and Google Drive platforms.

6. Ethical Aspects

SPECTRUM processes data that is classified as non-sensitive personal information. This data is gathered or produced solely for specific, clear, and valid purposes related to the project's goals.

Personal data processed in the context of the SPECTRUM project are handled in accordance with the **EU General Data Protection Regulation 2017/679 (GDPR)** and any applicable local legislation. Every project partner that may process personal data as part of the SPECTRUM project will do this in compliance with their national and EU regulations.

Without the clear informed consent of the persons concerned, no data will be gathered or utilised. It will be made unequivocal on the consent form, who has access to the gathered data as well as for how long the data will be safely preserved and then erased. Participants have complete control over the data. This will enable the individuals participating to make an informed voluntary decision about whether to engage in the project based on their understanding of the research's aim, processes, and results. In addition, they will also be able to acquire extra information regarding the processing of their personal data as well as to withdraw their consent at any point. In advance, information will be supplied on how the data will be handled.

In particular, within the CoP the Information, which will be used in the implementation of the project's activities, will be prepared in compliance with the General Data Protection Regulation.
The selection of interviewees or consultation participants is done in an impartial and non-discriminatory way as well as adequate publicity will be given to the measures to be taken to ensure transparency in research process, including how results will be shared.

It is important to highlight that each partner implementing the activity is responsible for ensuring that the templates for the Information are appropriately adjusted according to the context and needs of the activity for which they will be used and to reflect the privacy policy (**Annex 1**) when using the collaborative tools of the project.

By addressing these ethical aspects, we believe we will conduct the activities in a responsible and ethical manner.

7. Other

No other national/funder/sectoral/departmental procedures for data management are currently used in the framework of SPECTRUM.

8. Conclusion

SPECTRUM's DMP is a complete data management approach that complies with Horizon Europe recommendations and that aims to make data as findable, accessible, interoperable, and reusable (FAIR) as feasible.

Relying on robust technological solutions and standards, such as the EGI Document Repository, wiki.egi.eu, Zenodo, and Google Drive for the execution of these processes. Additionally, this will ensure that the data created or compiled throughout the SPECTRUM project, including open data and public publications, will be kept and continue to be usable once the project is completed.

The plan is intended to safeguard the analysis of compiled/created data based on the privacy level and to use an alternate sharing methodology relying upon this level. Confidential information or information that raises ethical problems will not be released.

Finally, the DMP is built on guaranteeing appropriately informed consent, and protecting each participant's zone of privacy, while adhering to GDPR guidelines

Annex 1. Privacy Policy

EGI Privacy Policy for the SPECTRUM project.

Name of the Service	Collaboration tools for the **SPECTRUM project**
Description of the Service	The EGI Collaboration Tools services for the **SPECTRUM** (hereinafter referred to as: “the service” or “Collaboration Tools”) support the **SPECTRUM** project's activities. Personal data is used to provide access to the service with the proper access levels. Personal data is collected as part of the project's activities. This privacy notice describes how we, the EGI Foundation (hereinafter referred to as "we" or "the Data Controller"), collect and process data by which project members can be personally identified (“Personal Data”) when the service is used.
Data controller	The EGI Foundation Science Park 140 1098 XG Amsterdam The Netherlands.
Data protection officer	The EGI Foundation Data Protection Officer Science Park 140 1098 XG Amsterdam The Netherlands E-mail: **dpo@egi.eu**
Jurisdiction and supervisory authority	Jurisdiction: NL, The Netherlands EGI Foundation's lead supervisory authority is the Dutch Data Protection Authority. They can be contacted at https://autoriteitpersoonsgegevens.nl/en/contact-dutch-dpa/contact-us

Personal data processed	The service may process the following personal data: **Identification data:** • Name • Identification number • E-mail address • Phone number • Address • Bank details • Other: affiliation, IP address **Behavioural data:** • Usage Data • Data on purchase or payment transactions • Working time data • Other: technical logs with timestamps, attendance at meetings **Data allowing conclusions on the personality:** • Other: membership information on groups, roles, and communities **Biographical data:** • CV data **Sociodemographic data:** • Gender
Purpose of the processing of personal data	The purpose of the collection, processing, and use of the personal data mentioned above is: • To provide the service functions, to coordinate and manage the project according to applicable requirements including projects' contracts, guidelines, funding policies, and legal requirements. • To keep evidence for audit needs. • To monitor and maintain service stability, performance, and security.
Legal basis	The legal basis for processing personal data is compliance with a legal obligation or legitimate interests pursued by the controller or by a third party according to Art. 6 (1) (f) General Data Protection Regulation (GDPR).

Third parties to whom personal data is disclosed	Personal data will not be used beyond the original purpose of their acquisition. If forwarding to third parties should be necessary to answer an inquiry or to carry out a service, the consent of the data subject is considered to have been given by entering a contract when using the respective function or service. In particular, the data provided will not be used for advertising purposes. For the purpose given in this privacy policy, personal data may be passed to the following third parties: **Within the European Union (EU) / European Economic Area (EEA):** • CESNET: resource provider, sub-contracted data processor of EGI Foundation • Google Ireland Limited: resource provider, sub-contracted data processor of EGI Foundation • EUfin: financial reporting • Zenodo: public deliverables and publications • European Commission • Project's partners • Individuals responsible for managing projects, Work Packages, and tasks. • The records of your use and technical log files produced by the service components may be shared for security incident response purposes with other authorised participants in the academic and research-distributed digital infrastructures via secured mechanisms, only for the same purposes and only as far as necessary to provide the incident response capability were doing so is likely to assist in the investigation of suspected misuse of Infrastructure resources. **Outside the EU / EEA:** • Project's partners • Individuals responsible for managing projects, Work Packages, and tasks. Any data transfer to a third country outside the EU or the EEA only takes place under the conditions contained in Chapter V of the GDPR and in compliance with the provisions of this privacy policy and any related policies adopted by the EGI Federation.
Your rights	You can exercise the following rights at any time by contacting our Data Protection Officer using the contact details provided in the Data Protection Officer section: • Information about the data stored with us and their processing • Correction of incorrect personal data • Deletion of the data stored by us • Restriction of data processing, if we are not yet allowed to delete the data due to legal obligations • Objection to the processing of the data by us • Data portability Project members can complain at any time to the supervisory data protection authority (DPA). The responsible DPA depends on the country and state of residence, of the project member's workplace, or of the presumed violation. A list of the supervisory authorities with addresses can be found at https://edpb.europa.eu/about-edpb/board/members_en. You can contact EGI Foundation's lead supervising authority using the contact details provided in the Jurisdiction and Supervisory Authority section.

Data retention and deletion	As per EC Grant agreements' requirements, data should be kept for at least 5 years after the end of the project. The data are deleted or anonymised as soon as retention periods have passed, and the data are not required anymore for any of the purposes listed above. The records of the project member's use and technical log files produced by the service components will be deleted or anonymised after, at most, 18 months as documented in EGI-doc 2732: Policy on the Processing of Personal Data.
Security	We take appropriate technical and organisational measures to ensure data security and protection against accidental or unlawful destruction, accidental loss, alteration, unauthorised disclosure, or access. A comprehensive overview of the technical and organisational measures taken by EGI Foundation can be found at EGI Document 3737: EGI Foundation Technical and Organisational Measures (TOM)
Data Protection Code of Conduct	EGI Foundation is conforming to the GEANT Code of Conduct and project member's personal data will be processed in accordance with the Code of Conduct for Service Providers and the EGI-doc-2732-v3: Policy on the Processing of Personal Data.
	This policy is based on AARC Policy development kit (licenced under CC BY-NC-SA 4.0)

Annex 2. Data sets

Data sets Work Packages

Data Summary	
Data description: Types of data	1. Project Documentation • Metrics • Risks • Procedures • Plans • Meetings agenda • Meetings participation list • Meeting minutes • Presentations • Deliverables • Mailing list archive • External Feedback (surveys, events, etc) • Promotional material (flyers, posters, branding materials, etc.) • Other communication & dissemination material (multimedia, video, etc)
Data description: Origin of data	All the data will be produced and provided by project members.
Data description: Scale of data	<1GB
Standards and metadata	plain text such as .docx, .txt, .rtf, .pdf, .pptx, xml, .xls, .html . Multimedia such as jpg/jpeg, gif, tiff, png or video formats
Data sharing: Target groups	All project members and the EC Project office. Communication activities; will be publicly available focusing on the target audiences of the project: including users, technology providers and infrastructure providers
Data sharing: Scientific Impact	Scientific Publications in peer-reviewed journals, conferences & events aiming to engage stakeholders

Data sharing: Approach to sharing	1. Shared within the consortium and European Commission • **Presentations**: Public presentations are made public via indico portal or external conference pages • **Deliverables**: All deliverables are shared within the consortium and also with the European Commission. Public deliverables are accessible to everyone via the project website and Zenodo portal. • **Mailing list archive**: only accessible by the mailing list members. • **Publications** will be available via the project website and SPECTRUM community on Zenodo Repository. • **Promotional and other audio-visual material** will be available via the project website. 2. Shared with the Project office and management boards to support work, as well as with the European Commission. Unless otherwise stated all content will be available under the CC BY 4.0 licence and metadata under the CC0 licence. Any consortium-restricted content is shared via access-protected confluence space
Archiving and preservation	Once the project is finished, all the information will be preserved by the EGI Foundation for at least 5 years as well on the EC funding portal. Publications will also be kept in the Zenodo Community.
Allocation of resources	
Who will be responsible for data management in your WP/Task?	Work Package Leaders and Task Leaders.
How will long-term preservation be ensured?	Long-term preservation is not needed, except from the contractual 5 years after the project. A copy of all the documentation of the project is kept by the European Commission in the funding portal. Deliverables, publications and other dissemination material are shared via the Zenodo portal, which grants long-term preservation
Data Security	
What provisions are or will be in place for data security (including data recovery as well as secure storage/archiving and transfer of sensitive data)?	To access the data shared only within the consortium, an EGI SSO account is required. Accounts and access management is the responsibility of the coordinator.
Will the data be safely stored in trusted repositories for long-term preservation and curation?	For security and long-term preservation, SPECTRUM relies on EGI Document Repository, Zenodo and Google Drive platforms
Other issues	

Do you, or will you, make use of other national/funder/sectorial/departmental procedures for data management? If yes, which ones?	EGI Foundation will take care of the data according to the ISO 27000 standard for Information security management and GDPR.

Data Set Community of Practice

WP/Task	WP3
Contact	Tommaso Boccali
Established a DMP, addressing important aspects of RDM.	☐ In place ☐ In progress ☑ Non
Data Summary	
Will you re-use any existing data and what will you re-use it for?	We will reuse public documentation existing at the level of initiatives, to collect the state-of-the-art. It will be used mostly to foster discussions and understanding within the different (sub)domains in SPECTRUM.
Will you re-use any existing data and will this generate new data?	The task will both reuse existing data (e.g. analysis of IT requirements from specific communities) and create new data (e.g. via survey).
What types and formats of data will the project generate or re-use?	PDF files, PowerPoint Presentations, discussion during the CoP meetings captured in minutes, survey, CoP Knowledge Hub (based on Confluence)
What is the purpose of the data generation or re-use and its relation to the objectives of the project?	The main objective is to consolidate the understanding and the survey in a final document to be used internally (for WP4-5) and for the project deliverables
What is the expected size of the data that you intend to generate or re-use?	overall, 1 GB or less
What is the origin/provenance of the data, either generated or re-used?	public data, presentations at the CoP, outcomes from the survey
To whom might your data be useful ('data utility'), outside your project?	To the scientific initiatives (for better planning of their future computing systems), to the funders (to direct future initiatives), to infrastructure managers (to better align with user's communities' desires)

FAIR Data	
1.) Making data findable, including provisions for metadata	*Will data be identified by a persistent identifier?* The final report will be made findable; a DOI will eventually be used. The knowledge basis will be public and searchable. It will not have a DOI since it consists of many documents, most of which already have a DOI.
	Will rich metadata be provided to allow discovery? What metadata will be created? What disciplinary or general standards will be followed? In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how. We plan to tag documents with domain ("HEP, "RA", ...), temporal validity ("current", "next 5 years", "next 10 years", ...), technological domain ("data management", "quantum computing", ...)
	Will search keywords be provided in the metadata to optimize the possibility for discovery and then potential re-use? We will provide a title, author name, doi when existing, year of publication and the tags as per the above paragraph
	Will metadata be offered in such a way that it can be harvested and indexed? Yes, as in the paragraph above.
2.) Making data openly accessible	
a) Repository:	*Will the data be deposited in a trusted repository?* Not needed, since all the data will be public and/or collected via means which specifically require it (for example the survey)
	Have you explored appropriate arrangements with the identified repository where your data will be deposited? We use Confluence from EGI Foundation, already used for similar purposes in EU projects.
	Does the repository ensure that the data is assigned an identifier? Will the repository resolve the identifier to a digital object? No, the original objects (papers, for example) will have DOI tags from the entity producing this at first.

b) Data:	*Will all data be made openly available? If certain datasets cannot be shared (or need to be shared under restricted access conditions), explain why clearly separating legal and contractual reasons from intentional restrictions. Note that in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if opening their data goes against their legitimate interests or other constraints as per the Grant Agreement.* The knowledge basis and the survey results will be made publicly available. We asked the CoP members whether they would provide non-public documents for example via NDAs. If anyone requests such an arrangement, we will use EGI Foundation Confluence, protected by the EGI SSO system.
	If an embargo is applied to give time to publish or seek the protection of the intellectual property (e.g. patents), specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible. We are not dealing with scientific data, which can be embargoed. The scientific documents we will use are public domain, with the caveat as per the previous paragraph.
	Will the data be accessible through a free and standardised access protocol? Via public web pages, with links for the direct download of papers etc.
	If there are restrictions on use, how will access be provided to the data, both during and after the end of the project? We do not expect any of our results to be restricted. As per the inputs, if NDA-like solutions are requested, we will protect the documents via EGI SSO.
	How will the identity of the person accessing the data be ascertained? The persons accessing data under NDA will need to be facilitated via EGI SSO.
	Is there a need for a data access committee (e.g. to evaluate/approve access requests to personal/sensitive data)? No, in case data under NDA is requested, the project Activity Management Board will decide.

c) Metadata:	*Will metadata be made openly available and licensed under a public domain dedication CC0, as per the Grant Agreement? If not, please clarify why. Will metadata contain information to enable the user to access the data?* We plan to prepare a public Confluence page per each document, with a direct download link. The same page can be used for harvesting metadata, for example via the "tagging" system described above.
	How long will the data remain available and findable? Will metadata be guaranteed to remain available after data is no longer available? Data remains available for at least 5 years after the project ends.
	Will documentation or reference about any software be needed to access or read or process the data be included? Will it be possible to include the relevant software (e.g. in open-source code)? Our data type doesn't need any software, apart from standard PDF and PowerPoint readers.
3.) Making data interoperable	*What data and metadata vocabularies, standards, formats or methodologies will you follow to make your data interoperable to allow data exchange and re-use within and across disciplines? Will you follow community-endorsed interoperability best practices? Which ones?* We will use metadata fields like; 1. author name(s) 2. year of publication 3. doi 4. title 5. a tagging system based on our internal domain definition
	In case it is unavoidable that you use uncommon or generate project-specific ontologies or vocabularies, will you provide mappings to more commonly used ontologies? Will you openly publish the generated ontologies or vocabularies to allow reusing, refining, or extending them? This does not apply to SPECTRUM CoP. The CoP is based on HTML and media in open standard formats, which consist mostly of PDFs.

	Will your data include qualified references to other data (e.g. other data from your project, or datasets from previous research)? Possibly, being a scientific report from existing initiatives. This is outside our control.
4.) Increase data re-use	*How will you provide the documentation needed to validate data analysis and facilitate data re-use (e.g. readme files with information on methodology, codebooks, data cleaning, analyses, variable definitions, units of measurement, etc.)?* Since we only collect public documentation, this does not apply to SPECTRUM CoP.
	Will your data be made openly available in the public domain to permit the widest re-use possible? Will your data be licensed using standard re-use licenses, in line with the obligations set out in the Grant Agreement? Under which license? Since we only collect public documentation, this does not apply to SPECTRUM CoP. However, the SPECTRUM CoP Knowledge Base is publicly accessible and can be consulted trough this link; **https://confluence.egi.eu/display/SPECTRUMCoP/**
	Will the data produced in the project be useable by third parties, in particular after the end of the project? Third parties (in particular scientific initiatives, funders, policymakers, and infrastructure managers) are expected to consider the output of SPECTRUM for their future initiatives.
	Will the provenance of the data be thoroughly documented using the appropriate standards? Each input paper/document will point to its source, when available.
	Describe all relevant data quality assurance processes. Not applicable to SPECTRUM CoP.
	Further to the FAIR principles, DMPs should also address research outputs other than data, and should carefully consider aspects related to the allocation of resources, data security, and ethical aspects. Not applicable to SPECTRUM CoP.
Other research outputs	

In addition to the management of data, are you also considering and planning for the management of other research outputs that may be generated or re-used throughout the projects?	*Such outputs can be either digital (e.g. software, workflows, protocols, models, etc.) or physical (e.g. new materials, antibodies, reagents, samples, etc.) Are those also following FAIR principles?* Not applicable to SPECTRUM CoP.
Allocation of resources	
Who will be responsible for data management in your WP/Task?	The WP leader Tommaso Boccali and the co-leader Corentin Lefevre.
How will long-term preservation be ensured?	Long-term preservation is not needed, except from the contractual 5 years after the project. A copy of all the documentation of the project is kept by the European Commission in the funding portal. Publications and other dissemination material are shared via the Zenodo portal, which grants long-term preservation
Data Security	
What provisions are or will be in place for data security (including data recovery as well as secure storage/archiving and transfer of sensitive data)?	To access the data shared only within the consortium, an EGI SSO account is required. Accounts and access management is the responsibility of the coordinator.
Will the data be safely stored in trusted repositories for long-term preservation and curation?	For security and long-term preservation, SPECTRUM relies on EGI Document Repository, wiki.egi.eu, Zenodo and Google Drive platforms
Ethical Aspects	
Are there, or could there be, any ethics or legal issues that can have an impact on data sharing?	*Yes or No. (If relevant, include references to ethics deliverables and ethics chapter in the Description of the Action DoA).* No
Will informed consent for data sharing and long-term preservation be included in questionnaires dealing with personal data?	Yes
Other issues	
Do you, or will you, make use of other national/funder/sectorial/departmental procedures for data management? If yes, which ones?	No