



WP1 Project Management and Coordination

D1.3

Data Management Plan

Due date

M6



PROJECT DETAILS

PROJECT ACRONYM

IMPRESS

GRANT AGREEMENT NO:

101094299

START DATE

01/02/2023

Project title

Interoperable electron Microscopy Platform for advanced REsearch and Services

Funding Scheme

HORIZON Research and Innovation Actions

WORK PACKAGE DETAILS

WORK PACKAGE ID

WP1

WORK PACKAGE LEADER

CNR

WORK PACKAGE TITLE

Project Management and Coordination

DELIVERABLE DETAILS

DELIVERABLE ID

D1.3

DELIVERABLE DESCRIPTION

The purpose of the DMP is to provide an analysis of the main elements of the data management policy that will be used by the Consortium about all the data, administrative and technical documents as well as scientific datasets, that will be generated by the project. Furthermore, this document describes the infrastructure platform and the digital ecosystem built to manage all the data.

The DMP is a living document and evolves during the lifespan of the project, thus particular attention has to be given to the version number.

DELIVERABLE TITLE

Project Data Management Plan

DUE DATE

M6 31/07/2023

Actual submission DATE

31/07/2023

Submission DATE of first update

M12 – 29/01/2024

PERSON RESPONSIBLE FOR THE DELIVERABLE

Stefano COZZINI

nature

R - Report

Dissemination level

- P - Public
- S – Sensitive
- Classified EU – Restricted
- Classified EU – Confidential
- Classified EU – Secret

REPORT DETAILS

ACTUAL SUBMISSION DATE
31/07/2023

NUMBER OF PAGES
24

document authors

BAZZOCCHI Federica
AREA

Email:
federica.bazzocchi@areasciencepark.it

LOT Ruggero
AREA

Email:
ruggero.lot@areasciencepark.it

ZAMPESE Silvia
AREA

Email:
silvia.zampese@areasciencepark.it

VERSION	DATE	AUTHOR(S)	DESCRIPTION / REASON FOR MODIFICATION
1.0	31/07/2023	Federica Bazzocchi Silvia Zampese Ruggero Lot Paul Bersans	First version of the data management plan.
1.1	27/12/2023	Federica Bazzocchi Silvia Zampese	Second version of the data management plan.
1.2	26/01/2024	Regina Ciancio and Paul Bersans	Review
2.0	29/01/2024	Paul Bersans	Second version of the data management plan (first update)

CONTENTS

LIST OF FIGURES	5
LIST OF TABLES	5
TABLE OF ACRONYMS	6
1. INTRODUCTION	7
1.1. IMPRESS PROJECT OVERVIEW	7
1.2. PURPOSE OF THIS DOCUMENT	7
1.3. SUPPLEMENTARY TOOLS AND DOCUMENTS	8
2. DATA SUMMARY	8
2.1. PURPOSE OF DATA COLLECTION	8
2.2. DATA TO BE COLLECTED, PROCESSED OR PRODUCED	9
2.3. REUSE OF EXISTING DATA	13
2.4. EXPECTED SIZE OF DATA	13
2.5. DATA UTILITY	13
3. FAIR DATA MANAGEMENT	13
3.1. MAKING DATA FINDABLE	15
3.2. MAKING DATA ACCESSIBLE	16
3.3. MAKING DATA INTEROPERABLE	16
3.4. INCREASE DATA RE-USE	16
<i>Copyrights, Intellectual Property and Results dissemination</i>	17
4. ALLOCATION OF RESOURCES	17
4.1. COST OF MAKING DATA FAIR	17
4.2. DATA MANAGEMENT RESPONSIBILITIES	18
5. DATA SECURITY	18
5.1. DATA RECOVERY	19
5.2. PERSONAL DATA MANAGEMENT	19
6. ETHICAL ASPECTS	20
7. ANNEXES	21
ANNEX I - GLOSSARY	21
ANNEX II - IMPRESS WP AND TASK LEADERS	24
ANNEX III: PRIVACY POLICY / PERSONAL DATA MANAGEMENT	24

LIST OF FIGURES

Figure 1: Operation flowchart of the FAIRcube with IPR mechanisms.	14
--	----

LIST OF TABLES

Table 1: IMPRESS sets of data	9
Table 2: IMPRESS file sharing platform tree structure	15

TABLE OF ACRONYMS

CA	Consortium Agreement	PCP	Pre-Commercial Procurement
CT	Coordination Team	PMO	Project Management Officer
EB	Executive Board	PP	Project Partners
EC	European Commission	PTC	Procurement Technical Committee
GA	General Assembly	SAB	Safety Advisory Board
GDPR	General Data Protection Regulation	SciNet	Science and Innovation Network
GrA	Grant Agreement	TEM	Transmission Electron Microscopy
IP	Intellectual Property	WP	Work Package
IPR	Intellectual Property Rights		
IPRC	Intellectual Property Rights Committee		
PC	Project Coordinator		
IdM	Identity Manager		
PEC	Procurement Evaluation Committee		

1. Introduction

1.1. IMPRESS project overview

IMPRESS is a scientific project that aims to revolutionize the field of Transmission Electron Microscopy (TEM) via the creation of an **Interoperable electron Microscopy Platform** for advanced Research and Services. The platform will be based on a cartridge concept, which will allow for flexibility and adaptability to different microscopes and instrumentation. It will facilitate a wide range of multimodal experiments, correlative workflows and methodological options, which are currently not available on commercially available electron microscopes. To achieve this, the project will co-develop new electron sources, adaptive optics techniques, event-driven detectors, and in situ/operando sample environments that are relevant to different applications. The design and development of the platform will be carried out in close cooperation with companies, which will be engaged through a **Pre-Commercial Procurement (PCP)** procedure to develop standardized hardware concepts that any researcher or operator can adapt further.

Additionally, IMPRESS will develop software for simulating experiments and enabling remote access based on artificial intelligence.

Furthermore, the IMPRESS project aims to establish the first **Open Knowledge and Innovation Hub for TEM** named “**FAIRcube**”. The hub aims to manage and share scientific, technological, and operational data in an accessible repository based on FAIR principles and aligned with the European Open Science Cloud (EOSC) ecosystem. FAIRcube will provide a dynamic space to access and exchange information, enabling collaboration between developers, experienced researchers, and newcomers.

1.2. Purpose of this document

The DMP aims to establish guidelines and procedures for data management within the IMPRESS Consortium for the entire project lifecycle. This document has to be intended as a dynamic document that will be versioned and will evolve throughout the project's duration to accommodate for new requirements.

The guidelines in this document are specific for each type of data along the three axes of the project: document, technologies, and scientific data. Data collection, storage, publication, and security guidelines will be specified for the three axes. This document will:

1. Describe the data management lifecycle for the data to be collected and/or generated in the framework of IMPRESS.
2. Outline the methodology employed to safeguard the sound management of the data collected, and/or generated as well as to make them Findable, Accessible, Interoperable and Re-usable (FAIR).
3. Describe how the data will be made openly accessible and searchable to interested stakeholders as well as their curation and preservation.
4. Present an estimation of the resources allocated to make data FAIR, while also identifying the responsibilities pertaining to data management and addressing data security.

This is the second version of the IMPRESS Data Management Plan (DMP), which provides an overview of the kind of data to be generated and collected aggregated per WPs, along with the specific conditions associated with them. Updates to the DMP will be provided at M24, M36 and M48, detailing the data to be made accessible and the methods of achieving open access.

This document is intended for members of the IMPRESS consortium.

1.3. Supplementary tools and documents

The IMPRESS DMP is complemented by a list of other documents that provide information on the way data will be collected and managed within the IMPRESS project:

- Annex 1: Glossary containing the definitions needed to deal with data management and IMPRESS procedures. The glossary aims to provide a common language with all terms clearly defined.
- Annex 2: IMPRESS Work Packages and Task leaders
- Annex 3: IMPRESS Personal Data Management

2. Data Summary

2.1. Purpose of data collection

IMPRESS is a Research and Innovations Actions (RIA) Project, aiming primarily to establish new knowledge and/or to explore the feasibility of a new or improved technology. To this end, IMPRESS will follow the policy of *“as open as possible and as closed as necessary”* and selected data will be easily discoverable, accessible, assessable, intelligible, interoperable, and ultimately useable beyond its original purpose.

The primary objective of data collection/generation within the IMPRESS project is to effectively administer and organize project activities, as well as facilitate dissemination and communication to achieve the project's goals. The collected/generated data results from targeted research activities. In terms of data management, the project will undertake the following measures:

- Develop and maintain an up-to-date DMP, which will outline the strategies and procedures for handling the project's data throughout its lifecycle.
- Deposit the data in a dedicated research data repository, ensuring that it is securely stored and easily accessible to relevant stakeholders.
- Enable third parties to freely access, mine, exploit, reproduce, and disseminate the data, with the requirement that proper referencing of the project is included.
- Provide relevant information and identify the necessary tools for utilizing the raw data.

By implementing these data management practices, the IMPRESS project aims to ensure efficient data handling, maximize the utility of the collected/generated data, and foster collaboration and knowledge exchange within the research community.

2.2. Data to be collected, processed or produced

During the implementation of the IMPRESS project, Project Partners (PPs) will collect and manage distinct sets of data, which can be classified into three major categories as follows:

1. **Documents:** The majority of the collected data will consist of various types of documents. This includes administrative information, technical requirements, technical drawings, projects, reports, and other relevant documentation related to the project's activities and deliverables.
2. **Datasets:** Specific datasets will be collected for the purpose of testing the new instrumentations and validating numerical methods developed within the project's technical WPs.
3. **Personal Data:** Personal data will be collected through registration forms as part of the project's activities. This may include personal information such as names, contact details, affiliations, and other relevant data provided by individuals during the registration process.

Further details regarding the data collected in each category can be found in the following tables, outlining the specifics and characteristics of each dataset.

Table 1: IMPRESS sets of data

Project Documentation, Reports and technical data	
Purpose of the data collection/generation	Administrative information, technical requirements, technical drawings, projects, reports, and other relevant documentation related to the project's activities and deliverables. Project's documents are shared between the individual components of the different axes and work packages through a NextCloud environment embedded in the FAIRcube ecosystem. Permissions are managed in a granular way on individual folders and on individual documents, to ensure data access consistent with the roles assigned within the IMPRESS project.
Origin of the data	Project documentation is produced by scientists and technical/administrative staff involved in the project.
Data formats / types	Open files formats, although some files could be in proprietary formats such as .docx, .xlsx.
Expected size of the data	Tens of GB.
Data utility	These data are the core data of the hub communities because they will allow sharing of information within the consortium partners in the different tasks

Research Data	
Purpose of the data collection/generation	Research Data collected from and generated in research activities will be used to test, improve, and validate the solutions developed within the IMPRESS project.
Origin of the data	Research Data will be collected within the research and technical Work Packages from manual and automated processes.
Data formats / types	Binary, HDF5, Text files, instrument file.
Expected size of the data	Few TB
Data utility	The data could be used for scientific research and further validation, peer reviews, reproducibility and education, in the spirit of FAIR Data

Personal Data	
Purpose of the data collection/generation	Supporting the creation of the IMPRESS community of stakeholders (including TEM manufacturers, TEM scientists and users)
Origin of the data	Personal data is collected directly from individuals who voluntarily provide personal data through registration forms on the FAIRcube platform. Personal data may also be collected from individuals who express their interest in receiving updates on project activities or subscribing to the project's newsletter.
Data formats / types	Text format
Expected size of the data	Hundreds of MB
Data utility	<p>Personal data are collected to:</p> <ul style="list-style-type: none"> - Facilitate the organization of project events. - Organize and segment stakeholders into a series of categories (scientific, industrial, policy maker, investor communities and public) to facilitate communication efforts. - Ensure project visibility, promote awareness of the project and its potential among the scientific community, different stakeholder groups, policymakers, and the public. - Inform about IMPRESS' achievements, progress, and results

An online form has been implemented on FAIRcube to allow WP leaders to provide comprehensive information about the data collected, processed, and produced within their respective WPs. This form served the purpose of obtaining a clear understanding of the types of data being collected and generated throughout the project. Main information collected through this form have been reported in the following tables.

Table 2a e 2b: Data details WPs breakdown

Work Package (WP)	Kind of data produced	Formats	Data source
WP1	-reports -presentations	docx, xlsx, ppt, pdf	Windows Office
WP2	Documents	docx, xlsx, ppt, pdf	Windows Office
WP3	Documents, CAD designs, prototypes, scientific data	.step, hdf5, Tex, .jpg, .csv, .py	CAD programs, Python, Pantarhei, Digital Micrograph, PCB design software
WP4	-scientific data and documents, -images, -computer codes, -experimental designs	-RAW formats: .emi, .ser, .dm3, .dm4, .bcf, .mrc,.emd; -Analysed formats: .tif, .png, .jpeg, .rec, .txt; hdf5/hpsy .docx -Computer codes: c++, matlab, python	JEOL TEM and Timepix 4 detector, simulations,ext editor and nionswift for data acquisition
WP5	-Documents, - Images, - CAD designs, - scientific data	.docx, .pptx, .xlsx, jpeg, tiff, csv, txt, dm3, dm4, dm5, emi, ser, .emd, cad, JPEG, TIFF, docx, .dmx, .mrc, CAD files, .h5, .pptx	Thermo Fisher TEMs, Gatan spectrometers, office documents generated in PCs, CAD and other design files, Electron microscopes and related detectors, computer simulations
WP7	-documents, -diagrams, -logs	SVG, PDF, TXT	Adobe Suite, IIS webserver, SQLServer DB, Matomo

Table 2b

Work Package (WP)	Purpose of the data and its relation to the objectives of the project	Utility of the data outside the project? (if any)	Do you think your data will qualify for IP?	Total data size
WP1	-Data collected and generated in T1.1 are needed for the coordination of the project -Data generated in T1.4 will be needed to define the strategy of the consortium after the end of the project.	None	no	< 500 MB
WP3	-Hardware designs contribute to D3.3. -Software generated in T3.4 corresponds to deliverable D3.4	Scientific Community, Manufacturers	yes	< 100 GB
WP4	Data from TP4 detector	Collaborators	yes	> 100 GB

		Owners of electron microscopes: RIs, academic laboratories and R&D intensive companies. Companies: innovative developers of technical instrumentation. End users: scientists studying materials.		
WP4	-Data generated inside tasks T4.1 and T4.2 are needed for the development of novel technologies - Adaptive optics and detectors, test data, experiment in task 4.3.		yes	> 100 GB
WP5	the objective is to design and develop a photocatalytic cell. The data will include its designs, presentations done, and the test data from the instruments used to validate the designs.	Could be used for publications or for possible commercial private companies.	yes	< 100 GB
WP5	Designs for the MEMS cell development, analytical Data obtained from the tests	research applications	yes	> 100 GB
WP7	Logs and matomo statistics are used for tracing website activities and for analyzing communication strategies	no	no	< 500 MB

Table 3: Metadata details WPs breakdown

Work Package (WP)	Purpose of the data and its relation to the objectives of the project	Utility of the data outside the project? (if any)	Do you think your data will qualify for IP?	Total data size
WP1	The data collected and generated in T1.1 are needed for the coordination of the project. The data generated in T1.4 will be needed to define the strategy of the consortium after the end of the project.	None	no	< 500 MB
WP3	Hardware designs contribute to D3.3. Software generated in T3.4 corresponds to deliverable D3.4	Scientific Community, Manufacturers	yes	< 100 GB
WP4	Data from TP4 detector	Collaborators	yes	> 100 GB
WP4	Data generated inside tasks T4.1 and T4.2 are needed for the development of novel technologies (2) Adaptive optics and detectors, test data, experiment in task 4.3.	Owners of electron microscopes: RIs, academic laboratories and R&D intensive companies. Companies: innovative developers of technical instrumentation. End users: scientists studying materials.	yes	> 100 GB

WP5	the objective is to design and develop a photocatalytic cell. The data will include its designs, presentations done, and the test data from the instruments used to validate the designs.	Could be used for publications or for possible commercial private companies.	yes	< 100 GB
WP5	Designs for the MEMS cell development, analytical Data obtained from the tests	research applications	yes	> 100 GB
WP7	Logs and matomo statistics are used for tracing website activities and for analyzing communication strategies	no	no	< 500 MB

2.3. Reuse of existing data

PPs agreed on sharing already existing data, know-how or information that is needed to implement the Action or exploit the achieved results (defined as “Background”) under the terms and conditions laid down in Section 9 of the IMPRESS Consortium Agreement (CA).

The specific Background identified and agreed upon for the Project is listed in the Attachment 1 of the CA.

A state-of-the-art analysis of the technical implementation studied by other communities is implicit, thus any information at disposal would be re-used in order to avoid document duplication.

2.4. Expected size of data

The exact size of the data to be managed by the IMPRESS project is currently unknown. However, dealing mostly with documents and technical requirements the expected size of the data is not huge. As anticipated the overall data size will be approximately a few TBs, with individual files expected to be of the order of the MB. Outliers will only be the sample dataset used for research data.

Data size will be monitored along the project progressing in such a way to assess correctly any adjustment needed to their management.

2.5. Data utility

Project activities data and research results with public dissemination or publishable (green or gold) level are planned to reach all the scientific worldwide community, in particular the TEM one, including academic researchers and companies’ engineers and developers interested in this kind of technical solutions.

3. FAIR Data management

According to the philosophy ‘as open as possible as closed as necessary’, the IPR Committee (IPRC) is responsible to decide which data must be protected and kept private (see section 3.4.1) and which data may

be disseminated or published (see Figure 1). The latter will be managed in accordance with the principles of FAIR data management (Findable, Accessible, Interoperable and Re-usable data).

One of the IMPRESS project goal is the creation of the first Open Knowledge and Innovation Hub for TEM, named FAIRcube that focuses on FAIR data, FAIR technology, and FAIR documentation.

This FAIRcube ecosystem will be a collection of different tools and solutions where scientific, technological, and operational data can be managed and shared among the PPs and with the stakeholders. This ecosystem is composed by a NextCloud instance for document management and will be expanded with an Object Storage solution. The Object Storage will be primarily used for scientific data. In addition to data management, the NextCloud service will also contain documentation about the prototypes that have been assembled and tested within IMPRESS. This will facilitate exchanges between developers, experienced researchers, and newcomers.

All the documentation has to be in open-source formats and equipped with a complete set of metadata providing the relevant information.

In the following sections by 'internal use' we will refer to those data shared among hub partners and stakeholders regularly registered, by 'external use' to those data disseminated or published (green or gold) on open access repository.

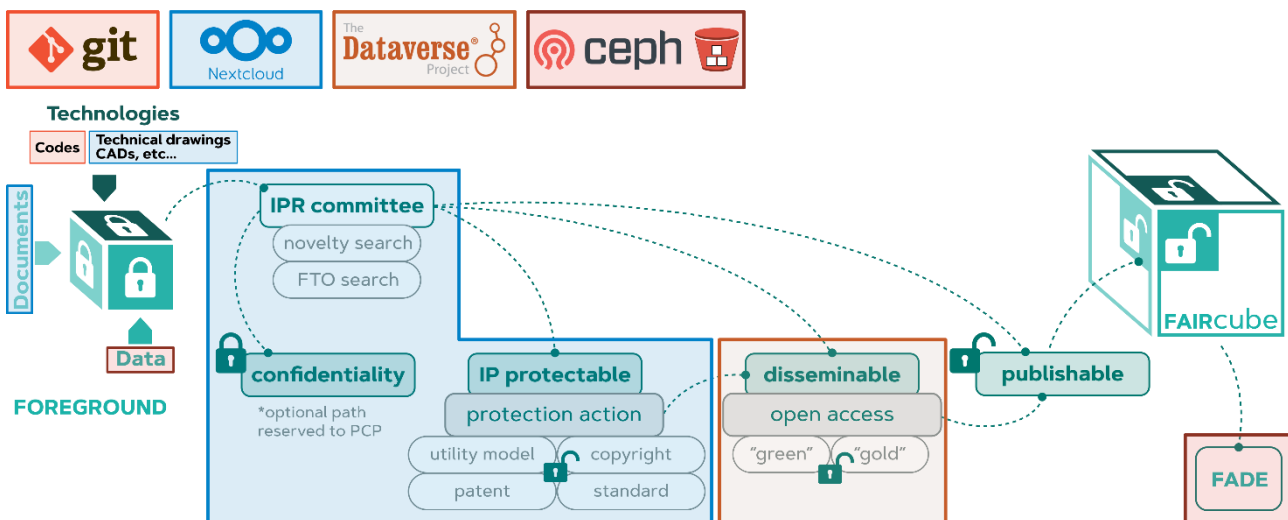


Figure 1: Operation flowchart of the FAIRcube with IPR mechanisms.

The figure above shows a collection of software we have selected to develop the FAIRcube ecosystem is in the legend at the top left of the figure. Git is a perfect solution to create and share code between PPs. We have not indicated a specific online sharing platform for Git because all the developers inside the consortium already use one of their choices; if we advise any need for a self-hosted solution, Gittea or GitLabCE will be considered. Nexcloud is at the core of documental and technical management. On NextCloud, committees that need to view and decide the level of confidentiality of a specific piece of information can be implemented as groups that can review the involved documents and constantly check the visibility of such documents. Once the committee grants a file the label of disseminable, the file owner can place it in open-access repositories. We have indicated Dataverse as a self-hostable example of one such repository that will be

considered during the platform's evolution. Finally, an Object Storage stores all the sample data shared among different PPs for research purposes.

3.1. Making data findable

INTERNAL USE:

The FAIRcube platform developed in IMPRESS is a file-sharing and collaboration tool designed to securely store and facilitate collaboration on various documents and files related to the project. Such files include administrative records, deliverables, working papers, reports, templates, technical documentation, and more.

The platform comprises several areas with different levels of confidentiality depending on the information available. One area is strictly reserved for consortium members. Another area will be open to stakeholders, in accordance with EU guidelines for open science.

The consortium area of the platform is structured with a data management system illustrated by several indicative folders detailed in the following table.

Table 2: IMPRESS file sharing platform tree structure

Category	Content	Accessible by
IMPRESS Shared	Information shared with the whole consortium. Can be used as a collaborative platform.	Participants of the project tasks.
IMPRESS Admin	Contractual information. Information linked to the organisation of events.	Partners' administrative staff.
IMPRESS CT	Repository of working documents for CT members.	CT members.
IMPRESS PCP	Repository of working documents related to the PCP.	Staff involved in the PCP including LP, PCT, PEC, IPRC and SAB' members.
IMPRESS SciNet	Information shared with SciNet members.	SciNet Members.
IMPRESS WP	Data produced by WP and shared among the consortium and stakeholders	Partners and Stakeholders

In addition, within each WP main folder, there is a dedicated folder for each type of document, such as deliverables, technical requirements, etc. These documents are clearly marked with version numbers to ensure proper version control and tracking.

EXTERNAL USE

All the data that are considered disseminable and/or publishable are uploaded in open access repository, such as Zenodo or Dataverse. In this way they are identified by a unique Persistent Identifier (PID) and

described according to a suitable metadata schema. The specific details and guidelines for the metadata schema will be defined in future versions of this document.

3.2. Making data accessible

INTERNAL USE:

NextCloud access is regulated by the Project Management Officer (PMO). PPs have access by default while, external stakeholders can request access by filling out an online form. The PMO receives an email notification with the request, which may be accepted at granular level. In this way the user may have access to a subset of folders according to the stakeholder and IPR committee decisions.

During the registration process, the personal data collected include name, surname, institution/company, and other relevant information. No personal data collected during the registration form will be shared outside the Consortium. The data remains confidential and is used for the purpose of managing access to NextCloud and other online services offered within the FAIRcube. They could also be used to facilitate collaborations within the project Consortium.

EXTERNAL USE

The data that are considered disseminable and/or publishable are uploaded in trusted open access repository with standardized access protocol, such as Zenodo. Access to these data will be granted according to the IPRC decisions that would arrange for embargo periods as well as access restrictions based on specific policies.

3.3. Making data interoperable

INTERNAL USE

Currently, the internal data are not intended to be interoperable. However, this approach may be subject to future revisions or changes as the project progresses.

EXTERNAL USE

Data that are uploaded on external open access repository will be enriched by metadata organized according to a declared metadata schema that would allow them to be interoperable. As the project progresses and data is identified, collected and generated, further information on making data interoperable as well as on metadata schema will be outlined in further versions of this document.

3.4. Increase data re-use

INTERNAL USE:

NextCloud platform is organized at a granular level and data sharing among PPs and stakeholders is intended to encourage data re-use among the hub, translating in time saving and project processing optimization. All data will be enriched by correlated documentation such as readme files, methodology information and so on, to enhance the understanding and usability of the shared data.

EXTERNAL USE:

All the documents disseminated and published are intended to encourage re-use and improvements of the technical achievements reached by the TEM experts hub involved in the project. IPRC could decide for an embargo period before granting full re-usability of the data. All data will be enriched by correlated documentation such as readme files, methodology information, codebooks as well as complete metadata.

Copyrights, Intellectual Property and Results dissemination

IMPRESS results will be extremely variegated, including innovative technical solutions, scientific results, and publications. Thus, deep care has been given in editing the regulations.

According to IMPRESS CA (Art. 8.1) Results are owned by the Party that generates them.

Any Intellectual Property Rights (IPRs) that arise from the PCP (in WP2) are owned by the respective company that generated them. This grants companies the freedom to commercialize and further develop their solutions. In return, the IMPRESS procurers will receive a non-exclusive, free license to use the results internally, and the right to require participating companies to license the results to other third parties under fair, reasonable, and non-discriminatory terms. However, if the results are not exploited, abused against the public interest, or fail to be commercialized, then the IMPRESS procurers may demand the transfer of IPR generated by the company under the PCP (call back clause).

Furthermore, in keeping with the open science approach advocated by IMPRESS, companies will always have the option to make their solutions, or parts of them, publicly available in the FAIRcube ecosystem.

With regard to the results, products, and use cases produced in other WPs of IMPRESS, the IPR strategy is established based on the principles laid out in the GrA and CA. The objective of the IPR strategy is to safeguard the IP that arises from the project activities, reduce potential risks, and prevent conflicts related to IPR matters when sharing results in the FAIRcube ecosystem.

For what concerns own Results dissemination by one or several Parties like publications and presentations, it shall be governed by the Grant Agreement and during the project and for a period of 1 year after the end of the project it will be regulated by the following restrictions:

- *Prior notice of any planned publication, including a draft of the publication, shall be given to the other Parties at least 45 calendar days before the publication. Any objection to the planned submission shall be made in accordance with the GrA by written notice to the Coordinator and to the Party or Parties proposing the dissemination within 30 calendar days after receipt of the notice. If no objection is made within the time limit stated above, the publication is permitted.*

4. Allocation of resources

4.1. Cost of making data FAIR

To make data FAIR is one of the project objectives and appropriate resources were allocated by each PP to cover costs for data harmonisation, integration, and publication.

The project has a lifespan of 4 years from February 2023 to January 2027.

As foreseen in the GrA, the PPs agreed on maintaining the Platform running for a minimum of 5 years after the end of the project and to continue to offer the developed prototypes as a service to users at Research Infrastructures (RIs). Costs associated for long-term preservation will be defined in task 1.4.

4.2. Data management responsibilities

The responsibility for data management in the IMPRESS project lies with Area Science Park, supported by the IMPRESS PMO. They are responsible for initially drafting the DMP in collaboration with the PPs and for making any necessary updates to it throughout the project.

However, it is the collective responsibility of all partner organizations to ensure that the data they generate and collect within the project adheres to the guidelines outlined in the DMP and complies with relevant legal and ethical standards. Each partner is expected to allocate appropriate resources and effectively implement data management activities in accordance with the DMP.

By distributing the responsibility among all PPs, the IMPRESS project aims to foster a collaborative approach to data management, ensuring that data is handled consistently, securely, and in line with the established standards and practices outlined in the DMP and all the other official agreements. In particular, Area Science Park will be responsible for running the infrastructure. All the partners have the responsibility to ensure that data is shared through the FAIRCube ecosystem in a consistent manner accordingly to the guidelines defined by this document and the IPR committee.

5. Data Security

As stated in section 2.2, IMPRESS will generate three types of data: documents, datasets, and personal data. All these types of data can be either intended for internal use or both internal and external use.

Documents and datasets intended for external use are not subject to any security policy since they are publicly available. Instead, data intended for internal use are subject to software and hardware security policies to ensure confidentiality of data.

All data transfers outside the data center will be encrypted at the software level. Personal data will be stored and managed only through a centralized Identity Management (IdM) service; all other services will only receive single-use authentication tokens to access a given service. A two-factor authentication Time-based one-time password (TOTP) will be implemented as an extra security step to access those services. A regular update policy will be in place to apply bug fixes and patches to Common Vulnerabilities and Exposures (CVEs). Within each service, data sharing will be directly under the control of the single user. Users could define granular access to data with different partners through access control lists (ACLs) or equivalent mechanisms.

A single object-storage distributed storage system service located in the same data center where the services are hosted is responsible for storing all the data. Communication between the storage system and the devices that offer the services that compose the FAIRCube ecosystem happens on a dedicated network.

At the hardware level, the location is secure from external unauthorized access and under constant video surveillance. An access control policy is also in place. Data locally is not ciphered. When external parties service hardware containing data, they are never given access to a portion of the infrastructure that would allow them to recover the data coherently and readably at a practical cost. This intrinsic level of security is

thanks to the distributed nature and configuration of the storage infrastructure. Remote backups of the data are encrypted and moved through a secure channel.

5.1. Data recovery

For the NextCloud service, an automatic backup procedure is performed every seven days. AREA Science Park will grant backups for two years, with the following granularity: up to 8 weeks weekly, up to 6 months bimonthly, and up to 2 years half-yearly. There will be two remote encrypted repositories to ensure data backup redundancy. The same applies to the IdM solution of choice.

Datasets will not be subject to backup, for what concerns data stored on trusted repositories they have their own recovery strategy.

5.2. Personal data management

Managing personal data within a Data Management Plan (DMP) involves implementing specific measures to ensure the privacy, security, and compliance of personal data throughout its lifecycle. In the case of the IMPRESS project, these measures will be elaborated by the Area Science Park legal office and reported in Annex III which will be appended to the upcoming version of the DMP (at M24). In particular, a Joint Data Processing Agreement, pursuant to Article 26 of Regulation (EU) 2016/679, will be signed by the PPs handling personal data. Through this Agreement, the Parties aim to specifically identify the personal data processing activities they will undertake as joint controllers, transparently defining and allocating their respective responsibilities regarding compliance with obligations arising from the GDPR, with particular emphasis on the exercise of data subject rights. The content of the Annex III will summarize the main points included in the Joint Data Processing Agreement, including information on the following points:

- **Data protection:** detailed information on how personal data will be collected, processed, and stored within the IMPRESS project. This section will outline the specific procedures and safeguards in place to protect the confidentiality and integrity of personal data. It will also clarify the legal basis for processing personal data, such as explicit consent or compliance with relevant data protection regulations like GDPR.
- **Data minimization:** the project will adopt a data minimization approach, collecting and retaining only the necessary personal data essential for the implementation of project activities. Whenever feasible, the collection of sensitive personal data (if any) will be minimized to mitigate privacy risks.
- **Data retention and deletion:** Guidelines for determining appropriate retention periods for personal data, considering legal requirements and the specific needs of the IMPRESS project. The Annex will outline procedures for securely deleting or anonymizing personal data once it is no longer necessary.
- **Data sharing and transfer:** Clarity on how personal data will be shared (if applicable), including the mechanisms and safeguards in place to protect the privacy and security of the data during transfer. This section may cover data protection agreements, encryption standards, and secure communication channels.
- **Consent and data subject rights:** Procedures for obtaining and managing consent from data subjects within the IMPRESS project. It will also address how data subjects' rights, such as the right to access, rectify, or delete personal data, will be handled and respected.

6. Ethical aspects

Data collected or generated by PPs within the IMPRESS project do not raise ethical questions for their sharing to the scientific community. All personal data collected for the implementation of Project activities comply with the principles of purpose limitation, data minimization, accuracy, storage limitation, integrity and confidentiality. Personal data are retained with the scope of delivering the service and the project follows the Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 – general data protection regulation (GDPR).

7. ANNEXES

Annex I - Glossary

Glossary containing the definitions needed to deal with data management and IMPRESS procedures. The glossary has the aim of providing a common language with all terms clearly defined.

Access Rights: it means rights to use Results or Background under the terms and conditions laid down in the CA.

Background: any data, know-how or information - whatever its form or nature (tangible or intangible), including any rights such as IPRs - that: (a) is held by the beneficiaries before they acceded to the CA, and (b) is needed to implement the action or exploit the Results.

Cartridge: key part of the interoperable platform, is termed “e-CAT cartridge”, where the letters stand for “e-DREAM”, “Correlative”, “Adaptable” and “Transferable”, respectively. The same e-CAT cartridges will be compatible with different electron optical planes on existing commercial microscopes at the partners’ laboratories, as well as with other equipment.

Consortium Agreement (CA): a document, signed by all partner organizations involved in a project, that governs the relationship between the coordinator (lead applicant) and PPs. It specifies the roles, responsibilities, and rights of each consortium member, including the management of IP, financial arrangements, decision-making processes, and dispute resolution mechanisms.

Consortium: it means any management body described in the CA.

Data Management Plan (DMP): a formal document that outlines what to do with data during and after any project that involves data. It describes the type of data collected, used, organized, and stored, and in which formats. It details how data will be accessible and documented for sharing and reuse during and after the project is finished.

Data Repository: an operational information system for managing and organizing digital resources, particularly suitable for Datasets or Publication Data which are not likely to be altered again. The Data Repository contains Metadata about the Datasets, as well as given rules for data access. A Data Repository may be associated with a certain Institution or a group of them, or a certain Instrument or a group of them, or may be run by a third-party. A Data Repository may or may not be directly used by Research Users.

Dataset: collection of scientifically related research data which can be raw data, analyzed data, or other datasets, each described by their related Metadata. The components of a Dataset remain individually identifiable within the Dataset.

FAIRCube / Open Knowledge and Innovation Hub: An ecosystem of software that will be deployed and personalized to allow for easy communication of the different PPs within the the project.

FAIR Data: data which meets the FAIR principles of findability, accessibility, interoperability, and reusability. The FAIR principles emphasize machine-actionability, i.e. the capacity of computational systems to find, access, interoperate, and reuse data with no or minimal human intervention.

FADE: A software storage solution for scientific data that will be deployed and tested within the WP6.

Gold Publication Open Access (OA): publication that can be accessed at no cost on the journal's website, is licensed under Creative Commons Licenses (CC).

Granting Authority: the body awarding the grant for the Project.

Green Publication OA; publication that is archived on website controlled by authors, or their funder, or on an independent repository. The deposited version of the article may or may not be final, may or not be the accepted and reviewed version by a journal (pre or post-print).

Interoperable Platform: a platform based on common interfaces and data formats and is highly flexible and adaptable to meet the scientific needs of individual users, who will be able to define and develop innovative experiments to carry out research using different instruments and techniques across the spatial and temporal scales. The constitutive element of the interoperable platform is a set of at least eight prototype e-CAT cartridges and associated technologies, which will be designed and then developed at technology readiness level (TRL) 8 through a pre- PCP procedure and made available to the PPs for experiments with users from the participating RIs during the project.

IPR Committee: the dedicated body or group formed within the IMPRESS project that oversees and manages matters related to IPR (IPR). It comprises IPR specialists from the IMPRESS consortium and operates in consultation to define and evaluate the IPR rules in the project. It will define technical data and documentation and technologies to be archived on the FAIRcube, upon evaluation of the IPR issues to be compliant with the IMPRESS GrA and the PCP IPR strategy.

License: official permission or permit to do, use, or own something (as well as the document of that permission or permit)

Metadata: set of descriptive, structural and contextual information describing the context, content and structure of Research Data and/or Datasets and their management through time. It describes information pertaining to research projects, including (but not limited to) the context of the experiment, the Research Users, the Data Analysis methods, and other logistical information. Metadata may include descriptions of how data and files are named, physically structured, and stored.

Metadata Schema: a logical plan showing the relationships between Metadata, normally through establishing rules for the use and management of Metadata specifically as regards the semantics, the syntax and the cardinality (mandatory, optional, recommended) of values. It can be written, e.g., in XSD (XML Schema Definition) or in JSON format, and may be implemented as machine actionable through consistent data entries and the inclusion of access points using controlled vocabularies. A Metadata Schema that gains wide acceptance from a reference user community and has been formally approved by Standards organisations, becomes a Metadata Standard.

Metadata Standard: a Metadata Schema that fulfils the needs of a scientific community, has obtained consensus, and has been ratified as a standard by some official bodies, such as the National Institute of Standards and Technology (NIST), the Dublin Core Metadata Initiative or the NeXus Data Format. A Metadata Standard describes the information and the terms needed to properly define specific data and it favours interoperability. In the NeXus Data Format, a Metadata Standard is called "Application Definition".

Open Access (OA): practice of providing online access to scientific information that is free of charge to the end-user and reusable. 'Scientific' refers to all academic disciplines. In the context of research and innovation, 'scientific information' can mean: peer-reviewed scientific research articles (published in scholarly journals),

or Research Data (data underlying publications, curated data and/or Raw Data). Open Access is granted by providing a suitable open Licence such as Creative Commons Licences (CC BY or CC0).

Open Format: open standard which specifies a file format. An Open Format is a file format for storing digital data, defined by a published specification, usually maintained by a standards organization, and which can be used and implemented by anyone. Open Formats are also called free file formats if they are not encumbered by any copyrights, patents, trademarks or other restrictions so that anyone may use the format at no monetary cost for any desired purpose.

Project partners (PPs): within this project that signed the CA.

Persistent Identifier (PID): long-lasting, actionable reference to a Research Digital Object.

Pre-Commercial Procurement (PCP): an approach that allows public procurers to buy R&D from several competing suppliers in parallel, to compare alternative solution approaches, and to identify the best value-for-money solutions that the market can deliver to address their needs. R&D is split into phases, such as solution design, prototyping, original development, and validation/testing of a limited set of first products. In PCP, there is a risk-benefit sharing under market conditions between the public procurer and the suppliers, and a clear separation between the PCP and the deployment of commercial volumes of end-products.

Raw Data: Research Data which is the primary output of a measurement, generated by a particular instrument and collected during an experiment by a research user or instrument scientist, before any subsequent processing. Raw Data is typically in the form of a data file but it can be potentially a data stream, or other form of data relevant in a particular data management context.

Result: any (tangible or intangible) output of the Project such as data, knowledge, or information — whatever its form or nature, whether it can be protected or not — that is generated in the IMPRESS project, as well as any rights attached to it, including IPR.

Scientific Publication: any of the following contributions, peer-reviewed or not: article in a scientific journal (and related supporting information), monograph, book or book chapter, conference proceedings and 'grey literature' (i.e. informally published material not having gone through a standard publishing process, e.g. reports and highlights).

Stakeholders: individuals, groups of people, institutions, private entities who in any way can be positively or negatively influenced by the intervention proposed by the project and whose action and interests can favor or hinder the achievement of the project objectives. Within IMPRESS project the main stakeholders are: RIs, TEM owners, companies and user communities.

Third Party: any legal entity that has not signed the Grant Agreement.

Annex II - IMPRESS WP and Task Leaders

IMPRESS: Work Packages and Task Leaders are reported and continuously updated in D1.1. We will report them only in the last version of this DMP at M48.

Annex III: Privacy Policy / Personal Data Management

The Personal Data Management will be provided in the next version of the DMP foreseen at M24.