D1.5 Data Management Plan

Lead partner:	EGI Foundation
Version:	1
Status:	Under EC review
Dissemination Level:	Public
Keywords:	Data Management
Document Link:	https://documents.egi.eu/document/3803

Deliverable Abstract
The initial Data Management Plan (DMP) introduces a report that specifies how research data will be collected, processed, monitored, and catalogued. Each dataset describes the type of data and their origin, the related metadata standards, the approach to sharing and target groups, and the approach to archival and preservation. This deliverable is considered a living document that will evolve during the project's lifespan.

EGI-ACE receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 101017567.

go.egi.eu/egi-ace

COPYRIGHT NOTICE

EGI-ACE receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 101017567.

DELIVERY SLIP

	Name	Partner/Activity
From:	Sjomara Specht	EGI Foundation/WP1
Moderated by:	Sjomara Specht	
Reviewed by:	Hien Bui Smitesh Jain	EGI Foundation EGI Foundation

DOCUMENT LOG

Issue	Date	Comment	Author
v.0.1	21/06/2022	First version request update on DMP from WPs	S. Specht, M. Krakowian
v.0.2	20/7/2022	GDPR input	B. Grenier
v.0.3	19/10/2022	Send deliverable for review	
v.1	26/10/2022	Final version	

TERMINOLOGY

https://confluence.egi.eu/display/EGIG

Contents

Executive summary

This document is an update of Deliverable D1.2, in which the Data Management Plan (DPM) for research data generated or collected by the EGI-ACE project was defined. This document provides details of each research data relating to, the type, origin and scale of data, standards and metadata, data sharing (target groups, impact, and approach) and archive and preservation.

Furthermore, the DMP describes the FAIR design of the data, agreements are established on data security, as well as addressing the moral angles related to their collection/generation.

EGI-ACE follows an Open Research Data Pilot (ORDP) and aims to make the data as open as possible and as closed as necessary. The beneficiaries distribute the project's data in such a manner that it can be worth to partners and their users outside the project while ensuring that this does not violate the privacy of the third parties who participated in the collection/generation of the data.

In this set of conditions, the data will be handled and disclosed in understanding and in accordance with the certifications and safeguards of the EU General Data Protection Regulation (GDPR). Every dataset is evaluated (in terms of sensitivity, privacy, or security angles) before an official choice is made regarding making that specific information public or not.

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

1 Introduction

Research data is defined as information, facts, or numbers, collected to be examined and considered and as a basis for reasoning, discussion, or calculation. In a research context, data examples include statistics, experiments, measurements, observations from fieldwork, survey results, interview recordings, and images[1]. The Open Research Data Pilot in Horizon 2020 focuses on research data that is available in digital form[2].

The Open Research Data Pilot applies to two types of data:

1) the data, including associated metadata, needed to validate the results presented in scientific publications as soon as possible.
2) other data (e.g., curated data not directly attributable to a publication, or raw data), including associated metadata.

The obligations regarding the digital research data generated in the action, arising from the Grant Agreement of the projects are (see article 29.3):

1) deposit in a research data repository and take measures to make it possible for third parties to access, mine, exploit, reproduce and disseminate — free of charge for any user — the following: the data, including associated metadata, needed to validate the results presented in scientific publications as soon as possible; other data, including associated metadata, as specified and within the deadlines laid down in the 'data management plan'.
2) provide information — via the repository — about tools and instruments at the disposal of the beneficiaries and necessary for validating the results (and — where possible — provide the tools and instruments themselves).

As an exception, the beneficiaries do not have to ensure open access to specific parts of their research data if the achievement of the action's main objective, as described in Annex 1, would be jeopardised by making those specific parts of the research data openly accessible. In this case, the data management plan must contain the reasons for not providing access.

In D1.2, each dataset described the type of data and their origin, the related metadata standards, the approach to sharing and target groups, and the approach to archival and preservation. As there are no significant changes to report within the DMP per WP, this deliverable will cover the implementation of FAIR & GDPR.

[1] https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/open-access_en.htm
[2] Guidelines on Data Management in Horizon 2020 http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

2 FAIR Data

EGI-ACE implemented the FAIR method, using the "Guidelines on FAIR Data Management in Horizon 2020" provided by the European Commission to help make research data Findable, Accessible, Interoperable and Re-usable (FAIR), and to ensure its sound management. These principles are essential for facilitating the exchange of scientific data and knowledge as well as knowledge discovery and innovation.

The FAIR elements can be summarized as follows:

- Findable data: Making data findable, including provisions for metadata (clear naming and versioning of (meta-) data, use of search keywords and DOI)
- Accessible data: Making data openly accessible (specification of how data are made available – is there a need for a committee and what tools are needed to access data)
- Interoperable data: Making data interoperable allowing data exchange and re-use (use of standards and vocabularies for (meta-)data and datatypes)
- Re-usable data: Increase data re-use (Specification of when - and for which duration - data are made available – specifically after the project, licensing of data)

The data will be handled and made openly accessible in line with and in accordance with the EU General Data Protection Regulation (GDPR), as specified in Section 5 of this deliverable. To ensure the confidentiality of the data subjects, any personal data gathered or created shall be deemed closed data before the anonymisation and aggregation.

Furthermore, each unique group of data will be evaluated (in terms of sensitivity, privacy, and security) before a final decision is made on whether to make that specific set of data public or not.

This approach is used in the datasets of WP-5 (described in D1.2[3]) to make data findable, accessible, and interoperable, and allow for the largest possible re-use.

[3] https://zenodo.org/record/6602163

3 Data Security

Any gathered data will be securely handled throughout the entire duration of EGI-ACE, to protect it from loss and unauthorized access. Personal data is only accessible to those who are authorized to access it.

All partners/beneficiaries responsible for processing personal data[4] have the responsibility, to ensure that the data remains protected under all necessary security controls (including backup policies and integrity checks[5]) 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 for long-term preservation EGI-ACE relies on EGI Document Repository[6], Zenodo[7] and Google Drive platforms.

[4] *Processing, according to Regulation (EU) 2017/679 of the European Parliament (GDPR), means any operation or set of operations which is performed on personal data or on sets of personal data, whether by automated means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure, or destruction*

[5] *The integrity check is the process of comparing the current state of stored data and/or programs to a previously recorded state to determine any alteration or change.*

[6] https://documents.egi.eu/

[7] https://zenodo.org/

4 Ethical Aspects

EGI-ACE carries out operations that entail the processing of data that does not fall into any uncommon classification of personal information (i.e., non-sensitive data). Only for determined, unequivocal and valid purposes related to the project's goals, these data will be gathered/produced.

Personal data processed in the context of the EGI-ACE 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 EGI-ACE 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 utilized. 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.

5 GDPR

Personal data processed in the context of the EGI-ACE 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 EGI-ACE project will do this in compliance with their national and EU regulations.

5.1.1 Data management responsibilities

Given that EGI-ACE is a project and consortium of parties and not a legal entity by itself, EGI-ACE is not a data controller according to Art. 24 GDPR.

Every partner/beneficiary in the project can be a data controller for activities related to the project. It is the responsibility of each partner to ensure compliance with GDPR when processing personal data in the context of the project.

It may be required for a controller to record their data processing activities according to Art. 30 GDPR, to document a privacy statement for a given processing activity and make it available to the data subjects. In some situations, project partners acting as controllers may use other parties including project partners as data processors on their behalf according to Art. 28 GDPR. In this case, it is the responsibility of the controller to put in place a data processing agreement (DPA) with each processor. In other situations, two or more project partners may also act as joint controllers according to Art. 26 GDPR. In these cases, it is the responsibility of the joint controllers to establish a joint controller agreement.

Nevertheless, due to the very specific nature, and scope of the EGI-ACE project, and its focus on the EGI activities, EGI Foundation is the project coordinator, and is having the role of data controller for many data processing activities taking place in the context of the project. In general, EGI Foundation is usually the Data Controller for activities falling under the **WP1 Project Management**, **WP2 Coordination and Cooperation**, **WP3 Infrastructure Services**, **WP4 Platform Services**, **WP6 Federated Access Services** and **WP7 Service Delivery and Planning**. Nevertheless, depending on the activity, another partner can be the controller.
For the personal data processing activities under the **WP5 Federated Data Spaces**, it's usually the partner providing the related service that is acting as a data controller. Those partners are usually treated and recorded as EGI-ACE customers in the EGI-ACE Customer database[8].

EGI Foundation, as project coordinator, is supporting the partners in identifying the various roles and responsibilities. EGI Foundation also provides templates (privacy notice, data processing agreements, subcontractor agreements) that can be used to put in place the required agreements. The EGI Foundation Data Protection Officer role is defined and can be contacted for support at dpo@egi.eu.

[8] https://confluence.egi.eu/display/EGIACE/Customers+DB

5.1.2 Specificities for activities with EGI Foundation as the data controller

For personal data processing activities where EGI Foundation is acting as a data controller, and especially for EGI Central services, the following measures are put in place:

- The personal data-related activities are supervised by the EGI Foundation Data Protection Officer and actioned as part of the Information Security Management (ISM) process of the EGI Foundation Information Management System (IMS), certified against ISO 9001:2015 and ISO/IEC 20000-1:2018[9].
- The personal data processing activity is recorded as part of the EGI IMS. This allows for identifying and documenting the personal data processing and to consider any specifics (like the processing of sensitive data or the transfer to third parties).
- When the data processing is done by a partner on behalf of EGI Foundation, a Data Processing Agreement is put in place between EGI Foundation as the Data Controller and the partner(s) providing the service as Data Processor. This is also documented in the Operational Level Agreement (OLA) agreed by EGI Foundation with the partner[10]. EGI Foundation is providing templates that should be used for those purposes.
- A privacy notice, meant to inform the service's users about how their personal data may be processed, is put in place. This task is also managed as part of the EGI IMS.
    - The privacy notice is based on the AARC Policy Development Kit (PDK)[11] and covers the requirements of the GDPR regulation. The EGI Check-in privacy notice[12] is an example of such a privacy notice.

5.1.3 For any service being integrated with EGI Check-in

When any service is integrated with EGI Check-in, its privacy policy is reviewed to ensure it contains minimal required information to inform the users about how their personal data may be processed.

In relation with the requirements set out by the GDPR Article 13,[13] and similarly to the REFEDS Data Protection Code of Conduct[14], the following points are reviewed:

- The identity and the contact details of the Data Controller, including their name and address
- The contact details of the data protection officer, if applicable
- The jurisdiction and supervisory authority relevant for the Data Controller
- The list of personal data processed
- The purposes of the processing of the personal data
- The legal basis for the processing of the personal data
- Information about third parties to whom personal data is disclosed
- Information about data retention and deletion

[9] https://www.egi.eu/egi-foundation/certifications/
[10] https://documents.egi.eu/public/ShowDocument?docid=3672
[11] https://aarc-project.eu/policies/policy-development-kit/
[12] https://aai.egi.eu/privacy.html
[13] https://gdpr.eu/article-13-personal-data-collected/
[14] https://wiki.refeds.org/display/CODE/Data+Protection+Code+of+Conduct+Home

- Information on the rights the user can exercise, as documented in GDPR Chapter 3[15]

For services for whose EGI Foundation is the data controller, those requirements are enforced and agreed upon between EGI Foundation and the service provider.

For services for whose EGI Foundation is not the data controller, this review is done to ensure that services integrated with EGI Check-in all share a common baseline, allowing users to have a certain level of confidence in the service, and can make an informed decision when deciding to access a service via Check-in.

[15] https://gdpr.eu/tag/chapter-3/

6 Other issues

EGI-ACE does not make use of other national/funder/sectoral/departmental procedures for Data Management.

7 Conclusions

EGI-ACE Data Management Plan (DMP) represents a comprehensive Data Management strategy in line with Horizon 2020 Open Data requirements as well as the strategy to make data findable, accessible, interoperable, and re-usable to the widest possible extent (FAIR).

For the realization of these processes, DMP depends on technology solutions and standards such as the OpenAIRE initiative, EGI Document Repositor, Zenodo and Google Drive. This will also guarantee that the produced/compiled data throughout the EGI-ACE project, including public data and open publications, will be retained and will remain accessible and usable when the project is completed.

The DMP is intended to protect the analysation of compiled/created data based on the level of their privacy and to use an alternate sharing methodology relying upon this level. Confidential data or data that raises ethical concerns, will not be distributed.

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