Ref. Ares(2022)9005233 - 29/12/2022

EMERGE
Emerging Printed Electronics Research Infrastructure

D4.10
Open Data access Management

WP4 NA3 – Development of e-infrastructure for data and information management

The EMERGE project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement Nº 101008701

Disclaimer

Any dissemination of results reflects only the author's view and the European Commission is not responsible for any use that may be made of the information it contains.

Copyright message

The EMERGE project has received funding from the European Union’s Horizon 2020 research and Innovation programme under grant agreement Nº 101008701

Document information

Project details	
Project Acronym	EMERGE
Project title	Emerging Printed Electronics Research Infrastructure
Grant Agreement Nº	101008701
Funding scheme	RIA - Research and Innovation action
Starting date	01/07/2021
Project coordinator	Rodrigo Ferrão de Paiva Martins (UNOVA)
Work package details	
Work package ID	WP4
Work package title	Development of e-infrastructure for data and information management
Work package leader	HMU
Deliverable details	
Deliverable ID	D4.10
Deliverable title	Open Data access Management
Delivery due date	Project month 18 *(31/12/2022)*
Author(s)	Markakis Evangelos (HMU) Papatsaroucha Dimitra (HMU) Astyrakakis Nikolaos (HMU)

Responsible person for the deliverable	Evangelos Markakis, emarkakis@hmu.gr;
	Dimitra Papatsaroucha, d.papatsaroucha@pasiphae.hmu.eu;
	Nikolaos Astyrakakis, n.astyrakakis@pasiphae.hmu.eu;

Nature	Open Research Data Pilot
Dissemination level	PU - Public
Report details	
Actual submission date	29/12/2022
Number of pages	24
Contact person	Evangelos Markakis, emarkakis@hmu.gr; Dimitra Papatsaroucha, d.papatsaroucha@pasiphae.hmu.eu

Report history

Version Nº	Date	Status	Changes	Contributor(s)
0.1	16/11/2022	*Draft*	Initial Draft ToC & Requirements	Evangelos Markakis, Dimitra Papatsaroucha, Nikolaos Astyrakakis
0.2	30/11/2022	*Draft*	First Draft	Evangelos Markakis, Dimitra Papatsaroucha, Nikolaos Astyrakakis

0.3	10/12/2022	*Prefinal*	Content authorship	Evangelos Markakis, Dimitra Papatsaroucha, Nikolaos Astyrakakis
0.4	16/12/2022	*Prefinal*	Content authorship	Konstantinos Votis, Konstantina Pantelidou, Pedro Barquinha, Olivier Ronsin, Payam Hashemi, Barbara Kosednar-Legenstein
0.5	20/12/2022	*Prefinal*	Editing content issues	Dimitra Papatsaroucha
0.6	21/12/2022	*Final*	Final Polishing	Dimitra Papatsaroucha
1.0	29/12/2022	*Final*	Final check by the Coordination	Inês Cunha (UNOVA), Pedro Barquinha (UNOVA), Rodrigo Martins (UNOVA)

List of abbreviations

CA - Consortium Agreement

DDR - Distributed Data Repository

DR - Data Repository

DoA - Description of Action

EMCC - European Materials Characterization Council

EMMC - European Materials Modelling Council

EuMat - European Technology Platform for Advanced Engineering Materials and Technologies

FAIR - Findable, Accessible, Interoperable and Re-usable

FLAPEP - Flexible large-area printed electronics and photonics

FQDN - Fully Qualified Domain Name

FZJ-DCFI – Forschungszentrum Jülich / Helmholtz Institute Erlangen-Nürnberg

GDPR - General Data Protection Regulation

IP - Intellectual Property

IPR - Intellectual Property Rights

JRA - Joint Research Activity

KBest - Knowledge and Best Practice Hub

NA - Network Activity

PASETO – Platform Agnostic Security Tokens

PFSim-Prost – Phase-Field simulations of Process-Structure relationship

RDA - Research Data Alliance

R2R - Roll-to-Roll

SCB - Selection Committee Board

SEP - Single-Entry Point

UNOVA - Instituto de Desenvolvimento de Novas Tecnologias – UNINOVA

WP - Work Package

TAs - Transnational Access Activities

TLO - Technical Liaison Office

CONTENTS

The EMERGE project has received funding from the European Union's Horizon 2020 research and Innovation programme under grant agreement Nº 101008701

List of Figures

List of Tables

The EMERGE project has received funding from the European Union's Horizon 2020 research and Innovation programme under grant agreement Nº 101008701

1. Executive Summary

This deliverable (D4.10 – Open Data access Management) presents an initial version of the approach of the EMERGE project towards the Open Research Data Pilot, in which the project participates. This deliverable is based on the activities of WP4 and outlines the management of research related data produced by WP5, WP6, WP7, and WP8, which pertain to the transnational access activities (TAs) of the EMERGE project. The experiments conducted under the TAs, as well as their results, will produce the initial datasets that will populate the Data Repository (DR) of the "Knowledge and Best Practice Hub" (KBest) developed under WP4. These research data will be investigated further as the project progresses and the consortium will explore the feasibility of depositing and publishing them in EU Open Access data repositories. The information included in this document will be further elaborated and enhanced in D4.11 Open Access Data Management that is due in M30.

2. Introduction

The goal of EMERGE is to create an innovative information and data management platform for green Flexible Large-area Printed Electronics and Photonics (FLAPEP), defined as KBest. To facilitate the involvement of diverse actors in the domain of emerging flexible electronics, KBest establishes a platform and metadata standard for data sharing as an open collaborative initiative within the framework of the research data alliance (RDA). Thus, a proper open-data access policy can make the FLAPEP KBest repository a novel and special tool for knowledge sharing among scholars. Integration of data access policies of the DR, as defined by the project, are activities like registration of datasets by one responsible user, authentication of all users to discover and access datasets, discovery and access of selected datasets, visibility of published datasets to the public; permission of dataset access and download only for registered users. Basic components for knowledge representation, dataset registration, validation and organization are also developed to compose production services for the repository architecture. Preservation services with regular checks for data and metadata integrity are also included. Additionally, the research data generated during the EMERGE project will be examined regarding their deposition in EU Open Access data repositories in accordance with the guidelines set by the EC for projects participating in Open Research Data Pilots, such as Zenodo [1].

The EMERGE project has received funding from the European Union's Horizon 2020 research and Innovation programme under grant agreement Nº 101008701

This section presents the purpose and score of the document, the relation of the deliverable to other work packages, and the outline the structure of the document.

2.1. Purpose and Scope

The purpose of this deliverable is to present an initial version of the approach of the EMERGE project towards the open access management of the research data produced by experiments conducted under the TAs of the project. This document is related to the activities of WP4, WP5, WP6, WP7, and WP8. The information included in this document will be further elaborated and enhanced in D4.11 Open Access Data Management that is due in M30.

2.2. Relation to other Work Packages

This deliverable (D4.10) describes the open data access management that is followed to complete the KBest platform. It is correlated with other WP4 tasks, such as Task 4.1 “Requirements gather for building an Artificial Intelligence powered Knowledge Repository” (M3-M12) and Task 4.5 “Deploy the platform developed in the previous three tasks” (M3-M48). This deliverable is also correlated with WP5, WP6, WP7, and WP8, which pertain to the TAs, as well as their experiments and results that will produce an initial set of research data under the EMERGE project.

2.3. Structure of the deliverable

This deliverable is separated into the following sections:

- Section 1 introduces the executive summary of the deliverable;
- Section **Error! Reference source not found.** focuses on the introduction section and more specifically the purpose and scope of the deliverable, the relation to other Work Packages and the content structure;
- Section **Error! Reference source not found.** presents the data collection, the different types of collected data and the re-use, origin, size of data and eventually the data utility;
- Section 4 focuses on FAIR (findable, accessible, interoperable and re-usable) data principles;
- Section 5 is focused on the allocation of sources, the costs for making data FAIR and information about data management;

The EMERGE project has received funding from the European Union’s Horizon 2020 research and Innovation programme under grant agreement Nº 101008701

- Section 6 describes data security;
- Section 7 focuses on ethical issues about data;
- Section 8 concludes the deliverable.

3. Data Summary

Regarding data collection and handling for their storage in the DR and use in the KBest platform, a data structure based on the "Scrambled Eggs Experiment" scenario, which was presented in D4.1 [2] and further described in D4.4 "Internal Test of the Information System", was used. More specifically, a basic data representation is showcased below, which refers to the data structure that the DR of the KBest platform will expect to be inserted for a simple experiment. On Figure 1, it is observed that there are some input materials that are being processed by following specific experimental steps, which eventually lead to an experiment product.

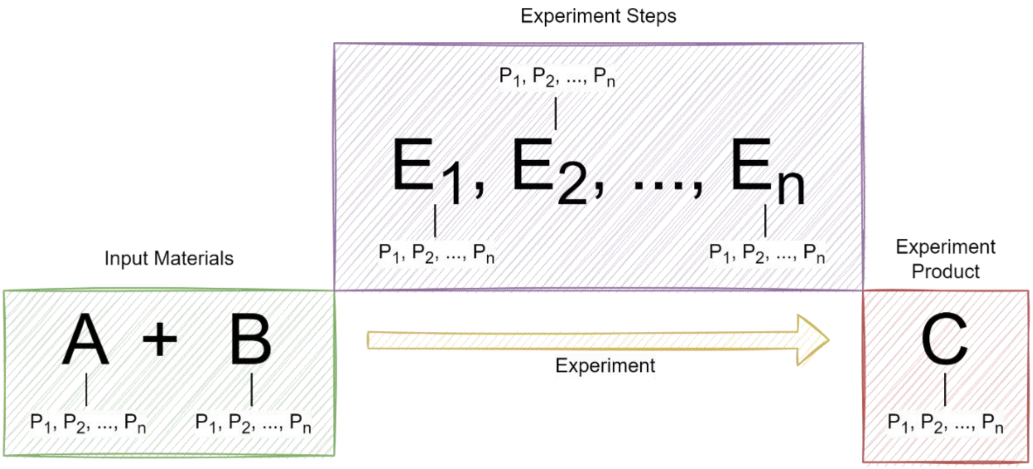

Figure 1. Scrambled Eggs Experiment Example Showcase.

The first component of this scenario is to add the input materials that refer to the different data categories than are used as an input to an experiment and to the platform in general. Additionally, the experiment steps and their properties are also recorded, to create a pipeline for each experiment and eventually the last component represents the experiment products that refer to the output data. A more detailed description of the data structure and its representation in the DR, alongside an example of experiment data inserted in the DR, can be found in D4.4 "Internal Test of the Information System".

Also, the metadata that will be collected during each experiment will be made available via the Data Discovery of the Data System.

The EMERGE project has received funding from the European Union's Horizon 2020 research and Innovation programme under grant agreement Nº 101008701

3.1. Data collection and generation

The EMERGE project will generate raw data, processed data, derived data, and metadata, resulting from experimental measurements on devices and numerical outputs from theoretical models and simulation scripts, as well as circuit designs, fabrication datasheets and processes, and software.

Metadata will contain essential details on how raw data and processed data were generated. Knowledge of the metadata will not only allow us to generate different versions of derived data (and analyse those accordingly), but it will also contain identifiers of (for example) flexible foils and type of surface's state to process devices on it, proper materials selection to build the devices to be pursued, level of integration and complexity that can be achieved allowing us to trace successful and failed samples to the original fabrication files. This also allows for systematic comparison with similar data obtained by other EMERGE partners. Our use of metadata across these data types will enable us to link numerical modelling of a device design, with the processes and datasheet regarding its fabrication, and the experimental measurement of its parameters. Such procedures are common in the development of new technologies. Here, we will use the expertise of the partners to adopt such approaches to data management and follow best practice.

All data produced through the EMERGE project will fall under three categories for the purposes of data management:

- **Partner-sensitive data** refers to data to be retained by a single partner and not shared more widely, and examples may include commercially sensitive know-how, industrial secrets and other forms of Intellectual Property (IP), for example relating to fabrication processes. We envisage the quantity of data generated under this project which falls under this category to be extremely limited.
- **Consortium sensitive data** refers the data, which are shared between two or more partners across the consortium, but not, in the first instance, made publicly available. Most of fabrication datasheets, device measurement and characterization data will fall under this category, and we will use the metadata to archive measured device and systems characteristics by platform and according to the device attributes, in a manner that enables different partners contributing to a particular platform to compare data. The database of raw, processed and metadata will be used to develop and train automation software (e.g., for proper materials and device architecture selection),

The EMERGE project has received funding from the European Union's Horizon 2020 research and Innovation programme under grant agreement Nº 101008701

including datasets for machine-learning models for tuning and device optimization. All partners will hold the intellectual property rights (IPR) for the research data they generate, and data sharing amongst the consortium will be performed in a manner compatible with IPR and legitimate non-disclosure interests of the partners, as established in the Consortium Agreement (CA).

- **Public/ user data** includes all publications and the data or meta-data presented there-in (complete data files of the results reported in the main publications will be made available through repositories or as on-line supplementary information on the journal homepage when space permits, along with scripts to facilitate further analysis and plotting of the data), as well as public talks and presentations. It also includes additional research data which has been vetted and cleared of IPR or other commercial sensitivities and released for public access in a FAIR form. Such public research data may be of use in industrial and academic research contexts to understand feasibility and reliability of devices and systems built, and as training datasets for development by additional parties not connected with EMERGE. In this respect, EMERGE will be an active contributor to the European Materials Modelling Council (EMMC) and of the European Materials Characterization Council (EMCC), with which Instituto de Desenvolvimento de Novas Tecnologias (UNOVA) has tight connections via European Technology Platform for Advanced Engineering Materials and Technologies (EuMat).

The TAs under WP5, WP6, WP7, and WP8 will produce various, heterogenous types of research data that will, eventually, be stored in the DR of the KBest platform. The project will use several open-source software in different joint research activities (JRAs) and TAs. The theoretical facility will feature several open-source packages for:

- FLAPEP operation and fabrication process simulations (like for instance multi-scale process simulation including film formation, deposition, drying kinetics to get insight and guidelines for material growth, functional design, and fabrication);
- Establishing design rules for printed integrated electronic circuitry for new systems with novel properties that can be integrated on free-form objects at low costs and high volumes;
- Establishing protocols for industrially compatible printing setups easy to transfer to factory and manufacture industry;

The EMERGE project has received funding from the European Union’s Horizon 2020 research and Innovation programme under grant agreement Nº 101008701

- Developing of novel interconnection technologies for flexible hybrids;
- Establishing data sheets for quality and reliability assessment of the produced materials;
- Standardization of methods for the evaluation of the nanomaterials / device quality, among others, all actively developed and used by several partners.

Regarding the activities of the TAs in the project, WP5 in EMERGE and TA1 “Access to design, modelling, and simulation” cover a broad spectrum of computer assisted design tools (ranging from 2D/3D design of frames, device or full systems, of the manufacturing process) as well as modelling / simulation tools (ranging from simulations of single printed layer deposition to electrical circuit simulations of full printed circuit boards). All data of TA1 are generated through these software tools.

In addition, the activities of WP6 in EMERGE, which refer to TA2 “Access to material synthesis and characterization”, cover a broad spectrum, namely:

- Synthesis of materials for application in printed electronics, and corresponding characterizations;
- Chemical and physical formulation of inks/pastes for printing, and corresponding characterizations;
- Printing of electronics, and corresponding characterizations;
- Characterization at each step, synthesis, formulation, printing, and performance.

Regarding WP8 and TA4 “Access to demonstrator characterization and validation”, the Raman measurement data support the project in terms of material characterization of the used materials (inks, pastes, polymers). The raw data will be generated during the measurements and data evaluation afterwards.

Regarding WP7 and T3 “Access to prototype fabrication”, the data stemming out of these activities will be elaborated more in future versions of this document, such as D4.11 “Open Access Data Management” that is due in M30.

3.2. Types and formats of generated/collected data

For WP5, due to the variety of software used, the data (nature of information, data format, data type) is very heterogeneous. To start with, the focus will be set on a single software for the sake of building the database approach. For example, for the “Scrambled Eggs

The EMERGE project has received funding from the European Union’s Horizon 2020 research and Innovation programme under grant agreement Nº 101008701

Experiment" scenario, WP5 provided the data generated with the in-house code "PFSim-Prost" (Phase-Field simulations of Process-Structure relationship) of FZJ-DCFI (Forschungszentrum Jülich / Helmholtz Institute Erlangen-Nürnberg), which are more detailly described in D4.4 "Internal Test of the Information System". All data are available in Matlab/Octave ".mat" format. However, the "input data" and "experiment step" data are provided as .csv files.

Regarding WP6, due to the wide variety of methods applied (i.e., synthesis, formulation, printing characterization), various types of data are collected, namely:

- Synthesis: starting materials, synthesis method, work-up, products;
- Formulation: chemical formulation and added additives, types and parameters of the applied physical formulation, properties of the obtained products;
- Printing: types of the printing methods applied and the parameters, properties of the printed patterns and electronics;
- Characterization: used techniques, samples preparation, analysis parameters, operators, calculation method of the acquired raw data.

In brief, some data are conditions and parameters of the applied methods and recorded in log files of the devices which are later, usually manually, exported to already-designed tables in excel files. Additionally, some of the acquired raw data are in the form of thousands of data points, from which graphs or tables are manually developed by the user via their preferred software (e.g., software of each specific instrument, Origin, Matlab, Python, Excel, etc.) and, finally, an interpretation or decision, based on these data points, is made by an expert.

Moreover, WP8 produce measurement data from Raman measurements in .csv or .txt format as well as Raman images (.jpg or bmp) depending on the chosen measurement type. The types and formats of the data produced under WP7 will be further elaborated in future versions of this document, such as D4.11 "Open Access Data Management" that is due in M30.

3.3. Re-use of data

KBest platform is deployed with the long-term preservation and curation of all data in mind. It can also provide all the tools required for users to search, discover, and retrieve data. The

The EMERGE project has received funding from the European Union's Horizon 2020 research and Innovation programme under grant agreement Nº 101008701

online data analysis tools will allow quality to be maintained while also adding value to the data and providing the means for data re-use over time. In general, data that will be stored in the DR of the KBest platform will be able to be accessed by other users of KBest, who, according to their access privileges (described in D4.4 “Internal Test of the Information System” in detail), will be able to either view only or even download the data of the experiments. The DR of the KBest platform will operate as an open access infrastructure; however, several access granting procedures will be followed, for respecting the IPRs of data owners: when registered users require access to experiment data, an email is sent to the partner that is the data owner and the user that is granted access to these data will be required to provide proper acknowledgement to the data owner in case of using these research data. In addition, the deposit of research data in EU Open Access data repositories will be investigated further throughout the course of the project following the directions of the Open Research Data Pilot.

3.4. Origin of data

The origin of data, produced under the WPs that pertain to the TAs of the EMERGE project, varies due to the several design and simulation software, applied methods, and procedures followed in each different TA. The origin of TA data will be further elaborated in future versions of this document, such as D4.11 “Open Access Data Management” that is due in M30.

3.5. Size of data

There are different data types and categories that are used and implemented in the knowledge repository (DR of the KBest platform). Thus, there is an initial estimation of the size of data that will be produced under the WPs that pertain to the TAs of the EMERGE project, as it can be seen in the Table 1. The size of data produced under WP7 will be further elaborated in future versions of this document, such as D4.11 “Open Access Data Management” that is due in M30.

The EMERGE project has received funding from the European Union’s Horizon 2020 research and Innovation programme under grant agreement Nº 101008701

Table 1. Estimated Size of data per WP.

	Work Package		
	WP5	WP6	WP8
Size of data	Depending on the type of simulation, the size of the experiment product data ranges from ~1MB to several GB (currently) and potentially up to 1TB in near future.	The size of Data points, graphs, and tables is 1-10 MB each.	The size of the measurement raw data can go up to more than 100 MB.
		The size of images, for example obtained from SEM, TEM, AFM, etc., is expected to be 1-200 MB each	The provided csv or txt files have a few KB.

3.6. Data utility

The storage of TA-related experiment data on the KBest platform will assist users and experts from different fields to compare experiments and results and get to know all the details about different data categories. As previously said, the goal of the KBest platform is to provide different experiments to users, so that they can achieve the highest possible results and performance, based on their own equipment, and needs. Furthermore, the potential deposit of research related data of the EMERGE project in EU Open Access data repositories, while respecting IPRs of the data owners, will further assist the scientific community in the field to investigate and explore FLAPEP data.

4. FAIR Data

As explained in D4.1[2], the data that are used in KBest platform should follow all the FAIR data principles [3]. The implementation of the FAIR guiding principles can be considered with reference to the four higher principles of FAIR, namely: findability, accessibility, interoperability, and reusability.

4.1. Making data findable

The necessity to make data findable is addressed in the first higher principle of FAIR and is the first step towards realizing the rest of the other three FAIR principles. Making data findable through the KBest platform is considered that will be accomplished by providing uniquely identifiable datasets and the option for the datasets to be linked. The findability of research related data of the EMERGE project will be enhanced by their potential deposit in EU Open Access data repositories, always respecting IPRs of the data owners. Both actions will be investigated more and further elaborated in D4.11 “Open Access Data Management” that is due in M30.

4.2. Making data openly accessible

The necessity to improve data and other digital assets' accessibility is emphasized by FAIR's second higher principle. For integrating FAIR data into a domain with clear access constraints, such accessibility is a crucial higher principle. The three crucial elements are access protocol, access authorization, and metadata lifespan. While metadata should ideally be kept around continuously to preserve the scientific record of the original data, data often have a set and limited lifespan.

Regarding the EMERGE project, research related data that will be stored in the DR of the KBest platform are going to be accessible under certain rules, such as the user’s registration and a request for access to the respective data owner. Some research related datasets are also foreseen to be deposited in the EU Open Access data repositories. However, the latter is going to investigated by the consortium and the partners involved in the TAs of the project and will be further elaborated in D4.11 “Open Access Data Management” that is due in M30.

4.3. Making data interoperable

Improved interoperability of data and other digital assets is emphasized in the third higher principle of FAIR. The various interoperable strategies are built on FAIR principles and connected open data. According to FAIR, data and metadata should be expressed in a formal, open, transferable, and widely applicable format. By adhering to these standards, a semantic data layer that spans unprocessed data to highly processed data will be created. To improve the quality and depth of the data and metadata for interpretation and analysis, it will also be crucial to provide data provenance by creating and preserving links to underlying publications or raw data, as well as other sources. The methodology for the development of the metadata model for FLAPEP data under the EMERGE project, which will describe the experiments and the derived data, is thoroughly explained in D1.3[4] and will be finalized in D1.8 “Final Metadata Standard for FLAPEP data” tha is due in M30.

The EMERGE project has received funding from the European Union’s Horizon 2020 research and Innovation programme under grant agreement Nº 101008701

4.4. Increase data re-use

The fourth higher principle of FAIR addresses the need to make data reusable. Reusability of data is FAIR's main goal. Additionally, it sets FAIR data governance apart from conventional data management. Implementing reuse calls for a varied approach as reusability allows data to be recycled for new user communities, demands, and applications. Through this, data can gain greater value. Regarding KBest platform, once users have been granted access, they are able to download experiment data and re-use these research data and experiments to extract knowledge through the Data Analytics component of the KBest platform or by utilizing these data to finetune their own experiments based on best practices. Moreover, as research related data of the EMERGE project will be investigated regarding their deposit in EU Open Access data repositories, research that data will eventually reside in these repositories will be able to be explored and re-used by the scientific community.

5. Allocation of resources

In this section, the costs for making data follow the FAIR data principles are mentioned in detail. Also, information about the data management handling is given.

5.1. Costs for making data FAIR

EMERGE is organized in 11 WPs, each chaired by a different leader from the Consortium, and the Management team (WP1) is responsible for coordinating the common actions of the WPs and run the overall administration and communication with the EC and partner institutions.

Considering the overall budget of 6.18M €, the distribution of budget and responsibilities has followed the rationale of maximizing the interaction among the Consortium. The main scientific and technical objectives of the EMERGE proposal are linked to the development of network activities (NAs) (budget allocation ≈24%), TA activities (budget allocation ≈45.5%) to support scientific communities (network users) in their access to the identified key research infrastructures and finally to the organization of JRA (budget allocation ≈24.5%) among network's members to improve in quality and/or quantity the integrated services provided at European level by the infrastructures. The budget of management is kept low (budget allocation ≈6%) as the e-infrastructure will be pervasive and much of the

The EMERGE project has received funding from the European Union's Horizon 2020 research and Innovation programme under grant agreement Nº 101008701

communication and project control will be done by the Single-Entry Point (SEP) portal and KBest and by videoconferencing, therefore reducing travel costs and time.

The FAIR data is mostly related to WP4 activities and requires the allocation of 75 person months and ≈7.6% of the total budget. Besides, part of the budget distributed among partners (43299.2 €) is allocated to costs for publishing in open access journals and patents. A dynamic access scheme is built into EMERGE management to adjust budget needs according to effective demand as the project runs.

5.2. Data Management

During the execution of EMERGE project, the data generated by the experimental/ computational work activities will be mainly used by the EMERGE Consortium, just like the data collected on the mainframe of exploitation/ dissemination, to ensure smooth cooperation between partners with the purpose of advancing the project activities, documentation, and exploitation/dissemination.

This deliverable regulates data collection, storage, backup, and archiving, and it is based on the FAIR principles.

5.3. Data storage, backup, and archiving

After exploitation in patents, peer-review journal publications and potential company spinoffs, EMERGE will provide a repository where all data recorded will be stored and could be retrieved through semantic search tools and/or directly analysed on-line by means of advanced data analysis services based on the KBest platform.

The EMERGE data policy will be proposed to the users who, on a voluntary basis, may agree to allow for open access under the proper rules (e.g. agreeing on an embargo time), as developed under the dedicated JRAs. The project will use the appropriate standards to represent scientific experimental data coming from FLAPEP experiments and theoretical analysis. At present there is no unique standard for such heterogeneous environments. A dedicated effort is foreseen within the project to properly identify and define a successful standard for these data and metadata description: it will be developed on the top of the currently available standards including the results of JRAs and TAs. Such a standard will be implemented, as it will become available and used, on an experimental basis for all scientific data produced within the project, with the agreement of the users. An initial version of the

The EMERGE project has received funding from the European Union's Horizon 2020 research and Innovation programme under grant agreement Nº 101008701

metadata standard developed under the EMERGE project for FLAPEP data is described in D1.3[4]. Additionally, an initial version of the experiment data structure of the TAs and its representation in the DR of the KBest platform is detailly described in D4.4 “Internal Test of the Information System”. It is foreseen that all data acquired within the project, from proposal submission to final analysis and or calculation, will be stored in the KBest platform.

The data policy will be designed to warrant open access and IPR in a compatible way by agreeing on embargo periods and selecting with the users the complete sets of data and metadata for access. Data produced by the Partners, as results of JRA, will also be made available to open access after IPR issues, if any, will be addressed by the project IPR management. Within the limitations of the IPR policy all information, including the relevant technical drawings and calibration data, where relevant, will be made public via regulated access to the repository. These aspects will be managed by the CA.

KBest will be specifically deployed keeping in mind long term preservation and curation for all the data. It will enable all the needed tools to allow user to search, discover and retrieve data. The online data analysis tools will allow quality to be maintained and add value to the data and provide tools for re-use over time.

6. Data Security

This chapter analyses the procedures followed for the data storage, the security of the storage, the backups, the restoration, and curation of the data by the infrastructure management and maintenance team. The current version of the first testbed infrastructure of the KBest platform is deployed in the premises of HMU and is further described in D4.6 – First Testbed available [M18].

6.1. Backup

The backups of the cluster and services of the KBest infrastructure are executed daily, in an automated way, every night at 23.59 and 01.00, respectively. The backups are performed without the disruption of services and in real time. The captured backups are stored following the 3-2-1 standardized backups strategy [5]. This strategy defines that backups shall be stored as follows: out of the three (3) total backups that are taken, two (2) of them are stored on-site (locally) and in two (2) different storage mediums (devices). Finally, one (1) backup

The EMERGE project has received funding from the European Union’s Horizon 2020 research and Innovation programme under grant agreement Nº 101008701

out of the total three (3) is stored remotely on another infrastructure - site - that contains storage servers. The backups are targeting the whole site infrastructure and specific services, such as the Distributed Data Repository (DDR) of the KBest platform. Moreover, Figure 2 depicts the backup strategy.

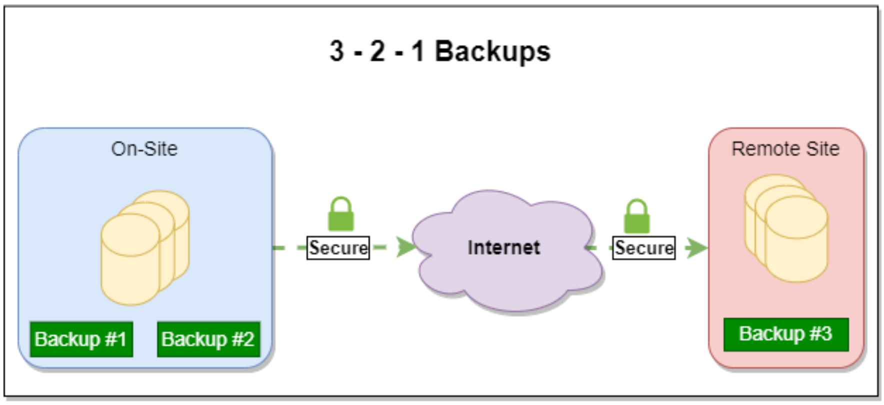

Figure 2. Backups Strategy.

6.1. Restore

The restoration procedure of these backups is still under development and therefore this deliverable section will be updated in the next iteration of this deliverable (D4.11 - Open Access Data Management [M30]).

6.2. Security

The main network security of the infrastructure where KBest is deployed, hosted by HMU, is based on digital SSL certificates [6]. Physical access to servers / storage media is prohibited for everyone and access to such machines is permitted only to maintenance and backup teams.

The backups are transferred to a remote site, over a secure and encrypted communication channel, utilising TCP/IP Over SSL (TLS) [7] and Hypertext Transfer Protocol Secure (HTTPS) [8]. Moreover, the local (on-site) backups are not accessible through the internet and the backups are securely stored in the server room, in a storage cluster, in an encrypted format.

Moreover, the access to KBest is secured with digital SSL certificates and the KBest service is only accessible with a specific fully qualified domain name [9] (FQDN) and by utilising a Platform Agnostic Security Tokens (PASETO) [10] web token on every request. The

The EMERGE project has received funding from the European Union’s Horizon 2020 research and Innovation programme under grant agreement Nº 101008701

PASETO tokens are encrypted and are used instead of user credentials for security measures.

6.3. Curation / Preservation

Hence far, the data collected in the infrastructure's servers do not include any personal information about users, so there are no particular protocols in place for data curation and data preservation. Furthermore, the data stored in the servers is only utilised for this project and research purposes. Moreover, the backups are preserved for a maximum time of 30 days and secure deletion procedures are performed daily on backup storage Medias to erase old backups.

7. Ethical Aspects

Considering the EMERGE activities, the ethics issues are primarily related to data protection and privacy, such as those associated with WP3 (Data communication), WP4 (Development of e-infrastructure for data and information management), and possible use of research results, such as those associated with the development of sensors and electronic security systems. The activity also addresses issues of research integrity, such as fabrication, falsification, and plagiarism in proposing, conducting, or reviewing research, or in reporting research results; this includes misrepresenting credentials and authorship improprieties. Any personal data processing activities carried out during the execution phase of the research and development and support are in accordance with European data protection legal requirements and ethical standards, namely the Regulation (EU) 2016/679 (General Data Protection Regulation - GDPR), national laws implementing the ePrivacy Directive 2002/58/EC, and the EU Charter of Fundamental Rights. Furthermore, an Ethics Management Plan is developed to align ethical awareness, ethical decision making, and transparency among the EMERGE project partners. The Ethics Management Plan is planned to align ethical awareness, ethical decision making, and transparency among the EMERGE project partners in accordance with European Regulations.

No other serious or complex ethical issues are foreseen, namely the ones impacting on life, since no human or animal experiments are addressed.

During registration on EMERGE website, users (from academy, SMEs, Industry) will be asked for some basic personal, socio-demographic and professional information that will be

The EMERGE project has received funding from the European Union's Horizon 2020 research and Innovation programme under grant agreement Nº 101008701

added to their profile and will be asked to accept the privacy terms of EMERGE (in accordance with the GDPR law), since the data will become part of the project's dataset. After activation of the new user account by the Coordination, who is responsible for defining the type of access for each type of user (subscriber; general consortium member; technical liaison office – TLO; Selection Committee Board – SCB; external reviewer; and administrator), the user will have selective access to the private area of the EMERGE website upon login to start preparing the submission of a proposal, consult submitted proposals, evaluate proposals, provide user surveys ("User Report" and "User Feedback"). At any time, the user can change the account settings and may cancel the newsletter subscription and/ or the account by accessing the private area ("Account settings").

The access to EMERGE installations may be granted to participants who submit a feasible proposal with positive feedback by the external evaluators and fulfil the established eligibility criteria. Since non-European users are allowed to participant in EMERGE initiatives (except of those with Russian nationality or working in Russia institutions), transfer of personal data from a non-EU country to EU within EMERGE may occur.

The TLOs, SCB, members and management team have access to all proposals submitted through the SEP-portal, while the external reviewer has access only to the one that was assigned to him/her. The SCB and external reviewer are required to keep confidential all details of the administrative and peer review process on submitted proposals.

Users with granted projects must sign a "Collaboration Agreement" prior access to EMERGE host institution, within the scope of the travel and subsistence for transnational access users and confidentiality of all information acquired within the scope of the activities carried out under this agreement. For all the dissemination material originating fully or partially from the transnational activity implementation, the user is asked to acknowledge the EMERGE project and partner(s) hosting the transnational activity implementation.

Even though the EMERGE developments will be open-source, user cooperating with industry can choose to apply for free-of-charge access, in which the results are published, or fee-based access, in which all work and results are confidential. In the latter case, the users should directly contact the project coordination through the contact form or by sending an email to info@emerge-infrastructure.eu. In the specific case of users working for SMEs, they are allowed to use results obtained within EMERGE for proprietary research.

After completing their research activities, the users are asked to send to the Coordination two surveys: i) feedback about the overall experience in the realization of the project (User

The EMERGE project has received funding from the European Union's Horizon 2020 research and Innovation programme under grant agreement Nº 101008701

Feedback), revealing optionally the user's interest in using the Knowledge & Best Practice Hub that is currently under development; ii) a "User Report", describing the scientific and technical content of the work developed. These results are shared at the level of Coordination, TLO and respective members that assisted in the project activities of the participant.

8. Conclusion

This deliverable refers to the open data access management and all the appropriate components that should be taken into consideration. Data handling, and more specifically information about data generation, types and size of data, re-use of data and data utility are mentioned to provide a clear view of the data that are used in the KBest platform. In addition, to provide a successful open data access policy, data follows all the FAIR data principles, which are explained as well. Allocation of resources, such as costs for making data FAIR and data management policies are examined. Ethical and security issues are mentioned to follow all the appropriate rules and restrictions and provide safety to the entire network of users.

9. Bibliography

[1] "Zenodo." https://zenodo.org/ (accessed Dec. 23, 2022).
[2] EMERGE Consortium (2022), "D4.1 Requirement of Knowledge Repository of EMERGE," 2022.
[3] European Commission. Directorate-General for Research and Innovation. and PwC EU Services., *Cost-benefit analysis for FAIR research data : cost of not having FAIR research data.* doi: 10.2777/02999.
[4] EMERGE Consortium (2022), "D1.3 Draft Metadata Standard for FLAPEP data," 2022.
[5] "Backup." https://en.wikipedia.org/wiki/Backup (accessed Dec. 23, 2022).
[6] "Public Key Certificate." https://en.wikipedia.org/wiki/Public_key_certificate (accessed Dec. 23, 2022).
[7] "Transport Layer Security." https://en.wikipedia.org/wiki/Transport_Layer_Security (accessed Dec. 23, 2022).
[8] "HTTPS." https://en.wikipedia.org/wiki/HTTPS (accessed Dec. 23, 2022).
[9] "Fully Qualified Domain Name." https://en.wikipedia.org/wiki/Fully_qualified_domain_name (accessed Dec. 23, 2022).
[10] "PASETO." https://paseto.io/ (accessed Dec. 23, 2022).

The EMERGE project has received funding from the European Union's Horizon 2020 research and Innovation programme under grant agreement Nº 101008701