EMERGE
Emerging Printed Electronics Research Infrastructure

D1.3 Draft metadata standard for FLAPEP data

WP1 - 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	01/09/2021
Project coordinator	Elvira Maria Correia Fortunato (UNOVA)
Work package details	
Work package ID	WP1
Work package title	Management - MGT
Work package leader	UNOVA
Deliverable details	
Deliverable ID	D1.3
Deliverable title	Draft metadata standard for FLAPEP data
Delivery due date	Project month 6 (31/12/22)
Author(s)	Evangelos Markakis, Konstantinos Rogdakis, Emmanuel Kymakis, Yannis Nikoloudakis (HMU)
Responsible person for the deliverable	Evangelos Markakis, emarkakis@hmu.gr
Nature	Report
Dissemination level	Public
Report details	
Actual submission date	07/02/2022
Number of pages	19
Contact person	Konstantinos Rogdakis (krogdakis@hmu.gr)

Report history

Version Nº	Issue Date	Status	Changes	Contributors
0.1	14/12/21	Draft	Initial ToC	Evangelos Markakis Yannis Nikoloudakis
0.2	15/12/2021	Draft	Content authorship	Evangelos Markakis Yannis Nikoloudakis
0.5	5/1/2022	Draft	Editing content issues	Evangelos Markakis Konstantinos Rogdakis Emmanuel Kymakis
1.0	5/1/2022	Draft	Final polishing	Yannis Nikoloudakis
1.1	4/2/2022	Final	Quality check, template uniformization, coordination	Pedro Barquinha, Rodrigo Martins, Elvira Fortunato

List of abbreviations

AI – Artificial Intelligence
EMERGE - Emerging Printed Electronics Research Infrastructure
FAIR – Findability, Accessibility, Interoperability, Reusability
FLAPEP - Flexible Large-Area Printed Electronics and Photonics
FRBR – Functional Requirements for Bibliographic Records
IM – Information management
JSON – JavaScript Object Notation
OAIS – Reference Model for an Open Archival Information System
RDF – Resource Description Framework
W3C – World Wide Web Consortium
WP – Work package
XLM – Extensible Markup Language

Contents

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 (D1.3), which is the initial output of the preparatory procedures for task T1.8 (M13-M48), aims to provide a metadata model, which will describe the experiments and the derived data, within the context of Emerging Printed Electronics Research Infrastructure (EMERGE) allowing for well documented/defined cataloguing, access, and exchange procedures of the project's results, throughout and beyond its lifecycle. This metadata model that aims to describe all the experiments, the ontologies and their inter-relationships that partake in these experiments, aims to pave the way towards the establishment of a Flexible Large Area Printed Electronics and Photonics (FLAPEP) metadata definition. This document is the first of a series of two documents (D1.3-M6, D1.5-M30), and aims to present the initial (Draft) design of the metadata model.

2. Purpose and scope

The objective of this deliverable is to pave the way towards the definition in the data representation of the EMERGE experiments and their derived data, within the context of printable electronics, by designing and developing a metadata data model. This model will allow scientists to seamlessly produce and share data within the EMERGE scientific community. Moreover, since the purpose of this model is to target possibly diverse use cases and communities in the FLAPEP domain, it needs to be designed as such to be able to accommodate such diverse use cases and paradigms. Therefore, in this deliverable we firstly present the initial state of the art research, which constitutes the base of the derived model's design and methodology. Consequently, we present the methodology followed, which resulted in the derived draft metadata model. Given the fact that this document is the first of the two-fold series of reports, the purpose of this document is to present the initial draft design of an envisioned FLAPEP metadata model, which wishes to accommodate the representation of the experiment procedure and the exported data, within the context of EMERGE and any other FLAPEP endeavor.

3. Relation to other Work Packages

This two-fold document series is a direct output of the effort within the task T1.8 (work package 1 – WP1). Nevertheless, the derived metadata model will serve as a direct input for all the tasks of WP4 ("Development of e-infrastructure for data and information management"), wherein the design and development of a distributed e-infrastructure for data and information management is envisioned, through which results derived from the experiments within the context of EMERGE will be disseminated. The results of this work also act as indirect input for tasks of WP5 (“Access to Design, Modelling and Simulation HMU Contribution”), WP6 (“Access to Materials Synthesis and Characterization”), and WP8 (“Access to demonstrators' characterization and validation”).

4. Methodology

The purpose of the envisioned FLAPEP metadata model is to accommodate not only the needs of the EMERGE project, but to be promoted into a guidance document that will be able to accommodate the representation of diverse use cases and paradigms. Therefore, the model itself must be designed in a generic and inclusive manner, ideally addressing, and incorporating all possible needs. Thus, the design methodology must include the current state of the art, concerning best practices, frameworks, and metadata standards, on the on hand, but also take under consideration the needs of the FLAPEP community’s stakeholders’ requirements. The outcome of this methodology will be the amalgamation of these diverse inputs from different sources and groups. Nevertheless, the design effort must be iterative and largely inclusive, as much as possible within the context of EMERGE project. Taking the above under consideration, the metadata design will be the result of an iterative process, wherein data and information will be gathered from i) empirical data, ii) existing standards and best practices, and iii) data from EMERGE consortium, extracted through questionnaires. Since the authorship of this document takes place on Month 6 (December 2021) of the project’s lifecycle, for the purposes of the prompt design of the draft version of the metadata model, part iii of the required inputs was omitted, and we only requested the unanimous consensus of the EMERGE project. The iterative information and extraction procedure through questionnaires, will be employed during the lifecycle of Task T1.8, which will produce the final metadata model and the final report (D1.5-M30). Nevertheless, Figure 1 presents the envisioned flow of information that will feed the design and development of the FLAPEP metadata model.

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

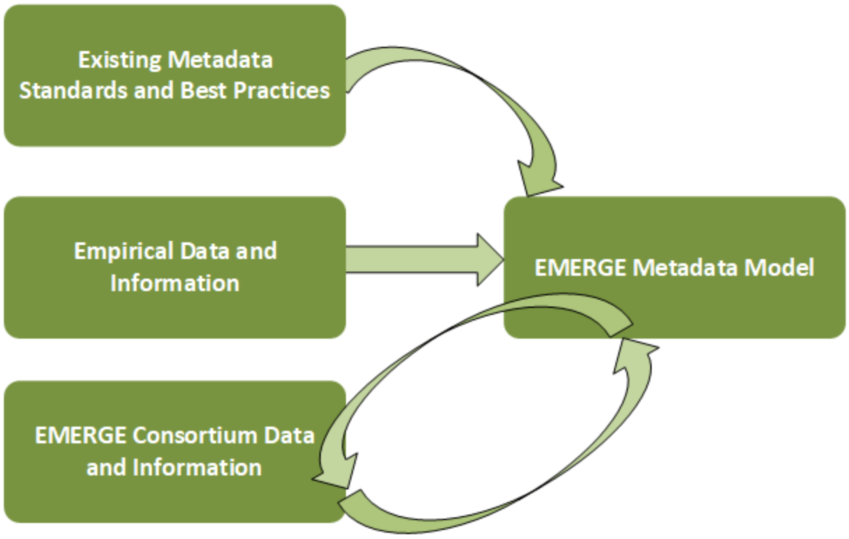

Figure 1 – Methodology Metadata Flow.

5. State of the Art

In this section, we present the current state of the art for best practices, frameworks, standards, and recommendations for Information Management (IM), and particularly for the FLAPEP domain. For the purposes of this document, we conducted initial research of the current state of the art. The findings were assessed and considered for the design and development of the envisioned FLAPEP metadata model.

5.1. Information Management Best Practices

IM is an iterative, possibly never-ending, organizational process (Figure 2), within which acquisition, minimization and contextualization, distribution, and finally archiving or deletion procedures take place, to create frameworks or infrastructures through which data is gathered and disseminated to relevant communities [1].

5.1.1. Design

The design process of IM systems, adheres and often overlaps the software design principles. Therefore, the requirements extraction is of paramount importance. Following this notion, the stakeholder's definition is the first thing researchers must undertake. Further on, the meticulous and systematic analysis of the stakeholders' requirements will lead to the extraction of the actual functional, non-functional, and system requirements.

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

The requirements extraction will eventually reveal a common vocabulary, pertaining ontologies and relationships between them. This vocabulary will be the base for the development of the envisioned metadata model.

5.1.2. Representation

The digital representation of data has seen a tremendous uptake during the recent years, due to the enormous digitalization of services and eruption of data-driven Artificial Intelligence (AI) paradigm. The different representation standards used throughout the scientific communities, depend on the utilized platforms and communication mediums and protocols, and the requirements or limitations they impose. The most prevalent standards used for the digital representation of data are the Extensible Markup Language (XML)[2], the Resource Description Framework (RDF)[3], and the JavaScript Object Notation (JSON) Data Interchange Format[4]. Each of the above standards is commonly used in separate use cases and occasions. Namely, XML is a markup language created to accommodate both human and machine readability, defined by a set of rules, to enforce simplicity, usability, and generality across the internet and any digital communication medium. RDF is a World Wide Web Consortium (W3C)[5] standard, aimed at tackling the digital representation of metadata. The RDF standard is optimized for graph-based data, often used to describe knowledge, and diverse serialization formats (e.g., Turtle[6]). Finally, the JSON standard was developed to enforce human-readability, by introducing a key-value pair and array structure, allowing for seamless serialization of data, making it a perfect candidate for web-based applications.

5.1.3. Distribution

Scholarly data, after their generation and storage, are rarely utilized from other scientific communities, due to inherent impediments to discoverability, and usability of such data. Therefore, the FAIR data principles were presented by Wilkinson et.al. in [7], and recently officially promoted and supported by the European Commission [8]. The FAIR (Findability, Accessibility, Interoperability, Reusability) data principles represent a set of rules and guidelines for data/knowledge producers to follow, in order to render the data seamlessly findable, accessible, interoperable, and reusable[7]. Following such largely adopted and promoted guidelines, allows the generated data/knowledge to contribute to the scientific community.

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

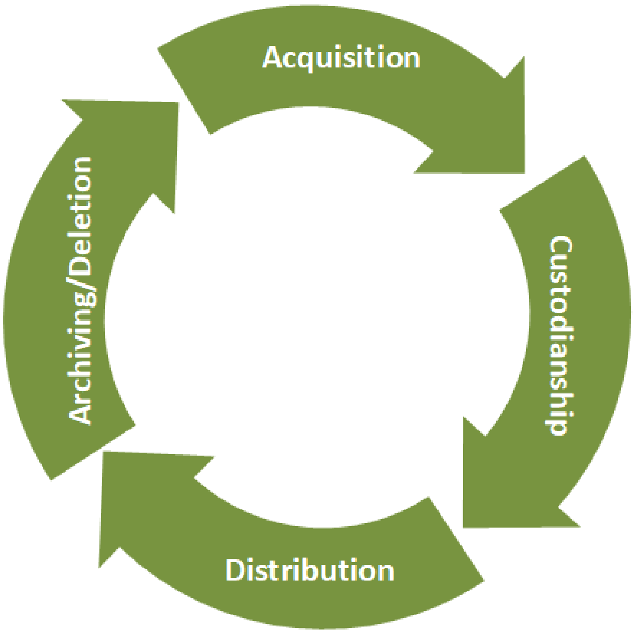

Figure 2 – Information Management Lifecycle.

5.2. Information Management Frameworks

The initial state of the art research revealed two major frameworks that provide guidelines and rules as for how to manage data and knowledge in general. These two frameworks are similar and, in some cases, overlapping. Nonetheless, they both can contribute to the overall strategy towards the development of a robust FLAPEP metadata model.

5.2.1. Functional Requirements for Bibliographic Records (FRBR)

The FRBR [9] framework elaborates on the functional requirements for relevant stakeholders, meaning the functions (actions) end users utilize to interact with inter-corelated data, in the case of FRBR, bibliographic records. The presented user actions are:

- **Find**: Users utilize the offered data to find a set of objects, depending on their search criteria.
- **Identify**: Through the discovered objects, users identify their desired object.
- **Select**: Users can select one or more objects they think matches their needs.
- **Obtain**: Users can gain access on the selected objects' information.

As a preferred methodology, Madison et.al. propose to first identify the key objects of interest, within the context of a particular domain that the data will describe. This will lead to a comprehensive Entity Relationships Diagram.

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

5.2.2. Reference Model for an Open Archival Information System (OAIS)

The OAIS reference model [10] describes the procedures and actions required to design and develop a long-term IM system. The OAIS reference model mandates the actions an organization must perform to conform with the OAIS strategy and create an OAIS-compatible IM system [11].

Negotiate: *"Negotiate for and accept appropriate information from information Producers."*

Obtain: ***"****Obtain sufficient control of the information provided to the level needed to ensure Long Term Preservation."*

Determine: "*Determine, either by itself or in conjunction with other parties, which communities should become the Designated Community and, therefore, should be able to understand the information provided, thereby defining its Knowledge Base."*

Ensure: "*Ensure that the information to be preserved is Independently Understandable to the Designated Community. In particular, the Designated Community should be able to understand the information without needing special resources such as the assistance of the experts who produced the information."*

Follow: *"Follow documented policies and procedures which ensure that the information is preserved against all reasonable contingencies, including the demise of the Archive, ensuring that it is never deleted unless allowed as part of an approved strategy. There should be no ad-hoc deletions."*

Make Available: *"Make the preserved information available to the Designated Community and enable the information to be disseminated as copies of, or as traceable to, the original submitted Data Objects with evidence supporting its Authenticity."*

5.3. Discussion

As one can infer, creating a FLAPEP metadata model is a rather complex procedure and the roadmap cannot be a straight line. In this notion, no one standard or framework can be suitable for this achievement. Thus, an amalgamation of best practices, frameworks, and standards, towards meeting our requirements as accurately as possible, is the preferable approach. Therefore, as discussed in Section 4 (Methodology), we will approach this endeavour following the software design principles and choose the guidelines that are suitable for our use domain, from the above-mentioned frameworks and standards [12].

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

6. Common Vocabulary

We approach the design of the metadata model by identifying the entities (ontologies) that partake in a FLAPEP ecosystem/lifecycle. Thus, we create the data structs and the relationships between them that will be the base of our envisioned model. This information will constitute a common vocabulary that will allow for a common understanding between scientists of the same or different disciplines. To create a basis for our design, we utilised the outcomes of a similar EU-funded INFRAIA project (**NFFA-Europe**[1]) [12], which was a good candidate for EMERGE. The base model was extended to match the FLAPEP domain requirements and create the initial (draft) model, which will be updated, finalized and reported on D1.5, M30.

Research User: A person, a group of people, or an organization who wants to conduct one or more experiments, on one or more FLAPEP facilities, using one or more FLAPEP instruments to collect and analyse raw data/simulation data, or is interested in data collected or analysed by other research users on the same or other facilities.

Project: An activity, or a series of activities performed by one or more research users, on one or more facilities, using one or more instruments to take one or more measurements of one or more samples, during one or more experiments.

Experiment: Identifiable activity with a clear start and finish time, conducted by research users who use instruments to investigate or produce samples and collect raw data. An experiment consists of one or a series of measurements and may also include one or a series of data analyses. An experiment can be a computer simulation or a combination with physical measurements.

Facility: An organization, or a division of it that operates on one or more FLAPEP instruments for research users. For software simulations, the facility may include hardware,

[1] https://cordis.europa.eu/project/id/654360

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

and/or software platforms or services that allow to schedule and manage computational experiments.

Instrument: Identifiable equipment that allows to conduct an independent FLAPEP research. The instruments are hosted by a facility and are operated by the research users. Instruments may be used for sample production. Measurements conducted on instruments result in raw data during an experiment. Instruments can also be software for computer simulations. Alternatively, an instrument can produce measurements, from which samples can be extracted from further analysis.

Instrument Scientist: A person, or a group of people who is responsible and operates one or more Instruments.

Measurement: A measurement is a data collection during an experiment using a particular or a set of instruments. Measurement can be a computer simulation. The measurement is specific to an instrument and depending on the research context measurement may involve measuring the same sample under different conditions.

Sample: Identifiable piece of material with distinctive properties collected during an experiment. A sample can be a model, a configuration, or data input in a computer simulation.

Raw Data: Identifiable unit of data collected by a research user during an experiment. Raw data is the result of a measurement. Typically, a data file or a data stream, but can be found in other forms of data, relevant in a particular data management context. Raw data can also be a result of a computer simulation. The raw data is a part of the data asset, which may bear some semantics of what the data is and its origin of it.

Data Asset: A combination of data units that derive from analysed data, or data analyses. Depending on the context, data asset most of the time is a dataset, or a collection. Data units remain identifiable within data asset. The data asset allows capturing relationships between data units or/and their origin, or/and data curation operations performed on data

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

units. Data Asset may also serve as a "container" for different representations of the same data. Finally, data asset can be used to express an accumulated result of measurement.

Data Analysis: The identifiable action of processing raw data or/and analysed data, with a data analysis software. As analysed data is an outcome of Data analysis, one can combine data analyses in chains or workflows. The definition of workflows and means of modelling them, however, is beyond the scope of this deliverable.

Data Analysis Software: Software used for raw data analysis that generates analysed data as an output. If a software is used for simulation, it is considered an instrument and should be described as such.

Analysed Data: Identifiable unit of data which is a result of raw data processing obtained with the use of data analysis software. Unit of data is typically a data file and can be potentially a data stream or other form of data relevant in the context. Analysed data can be a part of data asset, which may bear some semantics of what the data is and its origin of it.

7. Implementation

After establishing the common vocabulary, we are able to proceed with the implementation of the envisioned FLAPEP metadata model. In this section we present the step-by-step design and development process and its outcomes.

7.1. Entity Relationships Diagram

This common vocabulary constitutes the baseline for our envisioned metadata model since it illustrates the main ontologies that participate in a scientific experiment within the printable electronics domain. Thus, is easy to create an entity relationship diagram that will illustrate the ontologies and the relationships between them (Figure 3).

One thing that must be considered is that all the information pertaining an experiment (stakeholders and produced data/outcomes) constitute pieces of knowledge. Therefore, it is obvious that the representation of this information cannot follow simplistic or information-dismissive data-storage guidelines. Thus, to represent the given information as accurately and comprehensive as possible, we utilized a graph-based representation, which dictates that entities (ontologies) in the graph have specific, identifiable relationships, as opposed to

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

classic 1-1, 1-n, n-1 relations. In more details, relationships are described by their direction (inward or outward) and description (i.e., a verb that best describes that relationship). That kind of relationships can also contain extra metadata that provide extra context and comprehensiveness. The diagram below is a high-level illustration of the proposed graph.

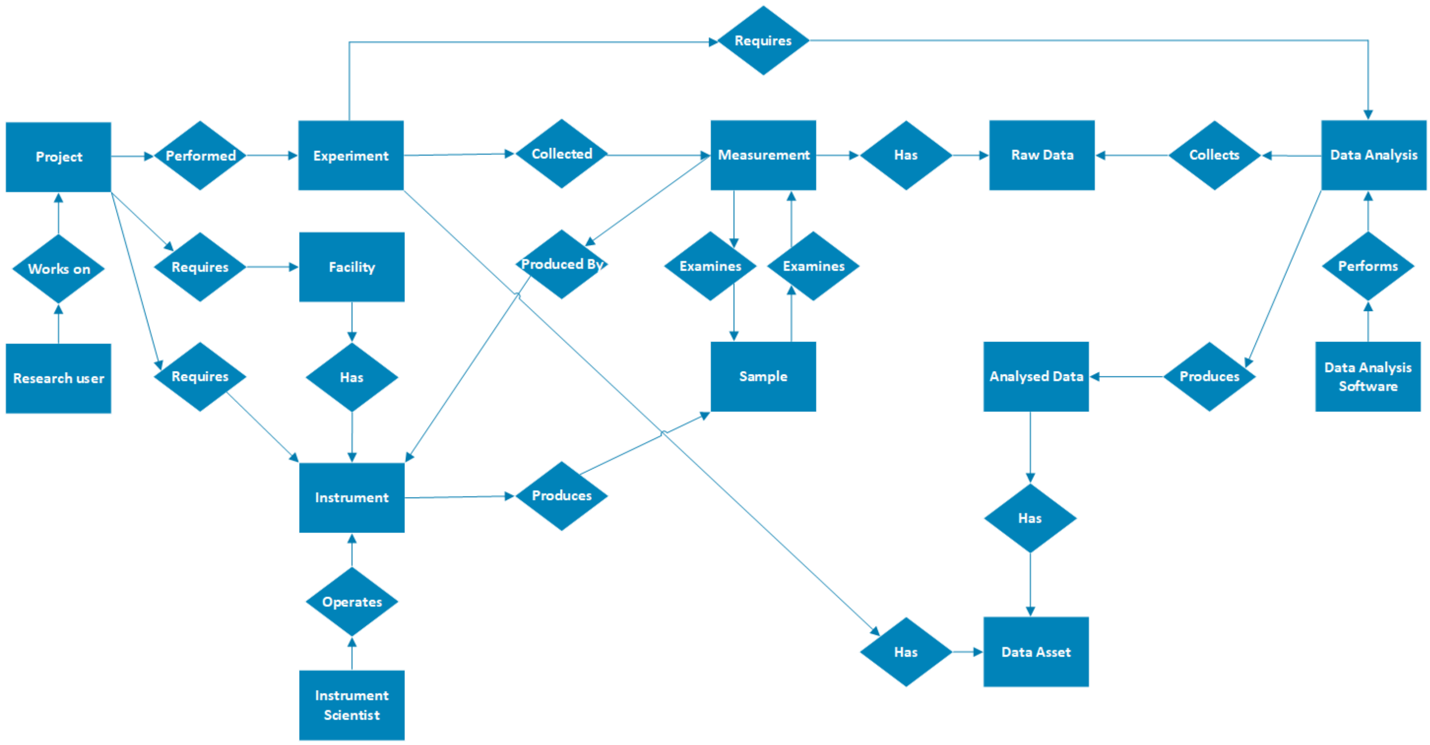

Figure 3. Entity Relationship Diagram.

7.2. Proposed Workflow

As stated above, the requirements of the partaking stakeholders in a specific use case are the main drivers for the design of a data model. Similarly, for the design and implementation of our envisioned model, we must specify a common workflow, generic and inclusive enough, to remain relevant for the entire domain of printable electronics. Therefore, we propose the common workflow that will drive our model.

1. **Create a project:** A *Research user* creates a project, which represents a scheduled scientific experiment with specific goals, and requires specific instruments and software for its realization.
    a. The scheduled *project* requires a scientific *instrument*, operated by an *instrument scientist*, located in a laboratory facility.
2. **Perform an experiment:** During the *project* lifecycle, *experiments* will be performed
3. **Collect/Produce measurements:** During the *experiment* measurements are produced

In order to remain inclusive for several diverse use cases, we propose two alternative sub-flows

a. Sub-Flow #1:
    i. The *instrument* produces *measurements* (e.g., continuous frequency measurements)
    ii. Examine a *sample* from the *measurements*
b. Sub-Flow #2:
    i. The *instrument* produces *samples* (e.g., printed nanomaterial)
    ii. Take *measurements* on specific *samples*

4. ***Collect raw data:*** After the end of the experiment raw data have been produced. The *Research user* collects this *raw data.*
5. **Analyse data:** The collected *raw data* are sent to data analysis
    a. Data analysis is performed by specific *Data analysis software*
6. **Collect analysed data:** The *data analysis* procedure produces a set of analysed data
7. **Create data assets:** The final outcome of the experiment is the production of *data assets* which will be added to the *experiment* entity.

7.3. Data Structure

Metadata Elements	Related Information Entity	Value Type	Required	Description
User ID	Research user	Identifier	Y	Unique identifier for the user
User Name	Research user	Text	Y	Name of user
User Identifier	Research user	Text		PID
User Affiliation	Research user	Text		Affiliation of the user
Instrument Scientist ID	Instrument Scientist	Identifier	Y	Unique identifier for the instrument scientist
Instrument Scientist Name	Instrument Scientist	Text	Y	Name of scientist
Instrument Scientist Identifier	Instrument Scientist	Text		PID
Instrument Scientist Affiliation	Instrument Scientist	Text		Affiliation of the scientist
Project ID	Project	Identifier	Y	Unique identifier for the project
Project Name	Project	Text	Y	Name of the project

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

Project Description	Project	Text		Textual description of the project
Facility ID	Facility	Identifier	Y	Unique identifier for the facility
Facility Identifier	Facility	Text		PID
Facility Name	Facility	Text	Y	Name of the facility
Instrument ID	Instrument	Identifier	Y	Unique identifier for the Instrument
Instrument Identifier	Instrument	Text		PID
Instrument Name	Instrument	Text	Y	Name of the Instrument
Instrument Type	Instrument	Text	Y	Class of instrument
Experimental Technique	Instrument	Text		Class of experimental technique supported by Instrument
Experiment ID	Experiment	Identifier	Y	Unique identifier for the Experiment
Experiment Title	Experiment	Text		PID
Experiment Start Time	Experiment	DateTime		Date and time of the Experiment started
Experiment End Time	Experiment	DateTime	Y	Date and time of the Experiment completed
Experiment Description	Experiment	Text		Description of the Experiment
Sample ID	Sample	Identifier	Y	Unique identifier for the Sample
Sample Identifier	Sample	Text		PID
Sample Name	Sample	Text	Y	Name of the Sample
Sample Description	Sample	Text		Textual description of the Sample
External metadata reference	Sample	URL		Reference for more detailed metadata
Data ID	Raw Data	Identifier	Y	Unique identifier for Raw Data
Data Identifier	Raw Data	Text		PID
Data Name	Raw Data	Text	Y	Filename or stream name of Raw Data
Data Format	Raw Data	Text		Format of the Raw Data
Data Format Identifier	Raw Data	Text		Identifier for the data format as assigned by an external organization

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

Data Type	Raw Data	Text		Type of the data in Raw Data
Data Size	Raw Data	Integer		Size of the data in Raw Data in bytes
Data Checksum	Raw Data	Integer		Calculated checksum of the Raw Data
Date of Collection	Raw Data	DateTime	Y	Date and time of the completion of the collection of Raw Data
Intellectual Property Rights	Raw Data	Text		Licensing information
Data ID	Analysed Data	Identifier	Y	Unique identifier for the Analysed Data
Data Identifier	Analysed Data	Text		PID
Data Name	Analysed Data	Text	Y	Filename or stream name for the Analysed Data
Data Format	Analysed Data	Text		Format of the Analysed Data
Data Format Identifier	Analysed Data	Text		Identifier for the data format as assigned by an external organization
Data Size	Analysed Data	Integer		Size of Analysed Data in Bytes
Data checksum	Analysed Data	Integer		Calculated checksum of the Analysed Data
Date of Creation	Analysed Data	DateTime	Y	Date and time of the completion of the collection of the Analysed Data
Intellectual Property Rights	Analysed Data	Text		Licensing information
Software ID	Data Analysis Software	Identifier	Y	Unique Identifier for the Data Analysis Software
Software Package Name	Data Analysis Software	Text	Y	Name for the Data Analysis Software
Software Version	Data Analysis Software	Text		Specific version of the Data Analysis Software
Software Package Identifier	Data Analysis Software	URL		Link to more information regarding the software package

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

8. Discussion-Conclusion

This report is the first of a two-fold series document. It elaborates on the research, design, and development of a FLAPEP metadata model. This version of the document presents a draft implementation of that model, since the authorship of this report takes place rather early of the EMERGE project's lifecycle (M6) and the relevant task (T1.8), for which the model will be a direct input, has not started yet (M13). The goal of this document is to report on the preparatory research that has been performed, towards the creation of the envisioned model. This is an iterative process, which requires continuous input from the project's consortium, who are active experts in the field of printable electronics, and external advisory board. The second and final version of this report (D1.5-M30) will present the whole effort and the final results of this research endeavour.

9. References

[1] A. Taylor and S. Farrell, "Information management in context," *Aslib Proc.*, vol. 44, no. 9, pp. 319–322, 1992.

[2] "Extensible Markup Language (XML) 1.0 (Fifth Edition)." [Online]. Available: https://www.w3.org/TR/xml/. [Accessed: 05-Jan-2022].

[3] "RDF - Semantic Web Standards." [Online]. Available: https://www.w3.org/RDF/. [Accessed: 05-Jan-2022].

[4] "rfc7159." [Online]. Available: https://datatracker.ietf.org/doc/html/rfc7159. [Accessed: 05-Jan-2022].

[5] "World Wide Web Consortium (W3C)." [Online]. Available: https://www.w3.org/. [Accessed: 05-Jan-2022].

[6] "Turtle (syntax) - Wikipedia." [Online]. Available: https://en.wikipedia.org/wiki/Turtle_(syntax). [Accessed: 05-Jan-2022].

[7] M. D. Wilkinson *et al.*, "The FAIR Guiding Principles for scientific data management and stewardship," *Sci. Data*, vol. 3, no. 1, p. 160018, Dec. 2016.

[8] E. Commission, "Final Report and Action Plan from the European Commission Expert Group on FAIR Data TURNING FAIR INTO REALITY 2018 Research and Innovation."

[9] O. M. A. Madison, "The IFLA Functional Requirements for Bibliographic Records," *Libr. Resour. Tech. Serv.*, vol. 44, no. 3, pp. 153–159, Jul. 2000.

[10] Consultative Committee For Space Data Systems, "Open Archival Information System (OAIS) Reference Model," in *Encyclopedia of Library and Information Sciences, Third Edition*, vol. 7, no. January, CRC Press, 2009, pp. 4020–4030.

[11] CCSDS, "Recommendation for Space Data System Practices MAGENTA BOOK REFERENCE MODEL FOR AN OPEN ARCHIVAL INFORMATION SYSTEM (OAIS) RECOMMENDED PRACTICE," 2012.

[12] nfaa.eu, "D11.2 Draft metadata standard for nanoscience data," 2016.

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