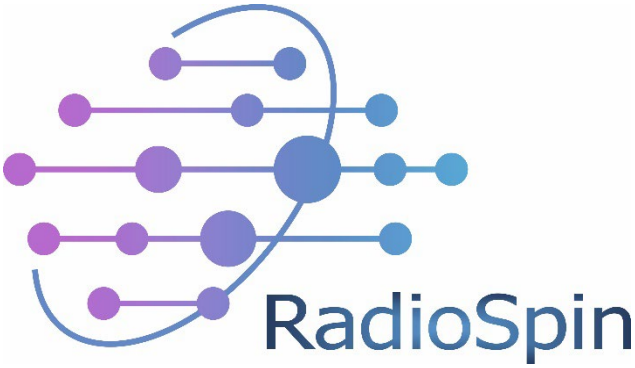


Initial Data Management Plan

RadioSpin Project

Action Number: 101017098

Action Acronym: RadioSpin

Action title: Deep Oscillatory neural networks computing and learning through the dynamics of RF neurons interconnected by RF spintronic synapses

Deliverable D8.3
Work Package 8

Date: 26th June 2022

DMP version: 1.0 Final

Deliverable No. 9.1

Dissemination Level		
PU	Public	X
PP	Restricted to other program participants (incl. the Commission Services)	
RE	Restricted to a group specified by consortium (incl. the Commission Services)	
CO	Confidential, only for members of the consortium (incl. the Commission Services)	
Nature		
R	Report	
DEM	Demonstrator	
DATA	Data sets, microdata, etc.	
ORDP	Open Research Data Pilot	X
ETH	Ethics	

Deliverable Details			
Due date:	30th June 2021		
Submission date:	28th June 2022		
Involved participants:			
	R. Clancy	Eurida	
	S. Saighi	UBx	
	A. Mizrahi	THALES	
	G. Tiberi	UBT	
	A. Jenkins	INL	
WP no. and title:	WP8 “Project Management”		
WP leader:	UBx		
Task no. and title:	Task T8.3 – Data and innovation management		
Version/Date	Final, 28th June 2022		
Keywords:	Data management, DMP, data		

Executive Summary

This document represents deliverable D8.3 – "Initial Data Management Plan" of the RadioSpin project (Grant Agreement No. 101017098), funded under the Horizon 2020 Research and Innovation Programme (H2020).
The RadioSpin consortium has identified several areas that need to be addressed in the context of data management, which are: Data protection and confidentiality, general ethics issues, personal data from working with humans and clinical trials (MammoWave use case), and possible limitations in open data sharing to avoid the risk of dual use and/or misuse of RadioSpin results.

RadioSpin activities involve human participants, as the project's technology benchmarking use case of the MammoWave technology addresses novel breast cancer diagnostics with patient data being generated in a dedicated clinical trial. This clearly poses a restriction of data collection and sharing. Data on patient collected during the trial will be used in an anonymous way; the subject will only be identified by a subject number, as indicated in the clinical protocol and in the informed consent.

For other activities carried out by the project (outside the work with humans and clinical trials), it may be necessary to collect basic personal data (e.g. name, background, contact details), even though the project will avoid collecting such data unless necessary. Such data will be protected in accordance with the EU's Data Protection Directive 95/46/EC[1] of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data. National and local legislations applicable to the project will also be strictly applied (full list described in annex 2: ethics and security).
By default, all personal data, or data directly related to the residents, will first be collected when the project has received a signed informed consent form from the subjects participating.

This is the first version of the project Data Management Plan (DMP). It contains preliminary information about the data the project will generate, whether and how it will be exploited or made accessible for verification and re-use, and how it will be curated and preserved. The purpose of the Data Management Plan is to provide an outline of the main elements of the data management policy that all consortium partners will have to comply with for all datasets that will be generated by the project. The DMP is not a final report, but represents a living document that will evolve during the term of the project.

The datasets referred to in this document have been drafted during the first project year (between M1 and M12) of the project. The document therefore reflects the strategy of the project consortium and all individual partners for the overall project's data management and already foreseeable datasets. An updated DMP will be prepared as part of the first technical periodic report (due at the end of August 2022), This follows the H2020 guidelines on Data Management Plans, and the RadioSpin actions as stated in the Grant Agreement (GA No. 101017098).

Justification of Delay

This deliverable D8.3 originally had been planned for project month 6 (due date 30th June 2021). While an initial strategy on how to handle data and how to apply the FAIR principles, but also IP and ethics restrictions to data generation, curation and sharing practices had been established and shared with the consortium within the first six months of the project, the establishment of a formal data management plan (DMP) has been delayed mainly because of the delay in appointing an Ethics expert. Ethics, however, is key especially to responsibly managing the data that will be generated as part of the RadioSpin clinical trial for the MammoWave™ technology and the data in connection with RF fingerprinting. Data and results for Artificial Intelligence and neuromorphic computing in general can also be subject to restrictions in connection with avoiding the risk of dual use, misuse or any re-use of RadioSpin outputs for purposes that can cause harm in any way. As RadioSpin is committed to civil uses of its technology and to Responsible Research and Innovation, responsible data management will have to include appropriate measures to avoid harmful, unethical and unwanted use with negative implications at societal level. The RadioSpin responsibles (Project Manager, Scientific Coordinator and Communication Officer) therefore decided to involve the external Ethics Advisors in the overall data management strategy and the DMP. The delay in

appointing the Ethics Advisors as explained in deliverable D9.4 therefore had a knock-on effect also on this deliverable D8.3. As data management and sharing becomes relevant in the following project periods the delay is not expected to result in any shortcomings or data management issues for RadioSpin, even more so as the general management requirements and key questions as outlined in this document have been communicated with all consortium partners during the first six project months. The data management where patient data and humans are concerned have been dealt with by responsible partner UBT based on their expertise in previous clinical trials, including data management principles and strategies used in the MammoWave™ project funded by the SME Instrument, supported by Ethics expert Dr. Cristina Morelli who is also External Ethics Advisor in RadioSpin.

Contents

Executive Summary 3
1. INTRODUCTION 6
1.1. The overarching RadioSpin data management strategy 6
1.2. Personal data, patient data and individual participants 7
1.3. Data underlying outputs for commercial exploitation 7
1.4. Legal frameworks 8
2. Data Summary 8
2.1. Purpose of the data generation or re-use and its relation to the RadioSpin objectives and target groups 8
2.2. Types, formats, size and origin of data in RadioSpin 9
2.3. Expected re-use of existing data 9
3. FAIR data 10
3.1. Making data findable, including provisions for metadata 10
3.2. Making data accessible 11
3.3. Making data interoperable 14
3.4. Increase data re-use 14
4. Other research outputs 15
5. Allocation of resources 16
6. Data security 17
7. Ethics 17
8. Other issues 18
ANNEX – RadioSpin Datasets 19

1. INTRODUCTION

The RadioSpin initial Data Management Plan (DMP) has been developed using FAIR data principles, which means making data **f**indable, **a**ccessible, **i**nteroperable and **r**eusable. The DMP outlines which datasets the project will generate, compile and/or re-use, and how these datasets will be curated, stored and made accessible. The DMP also describes the measures that have been and will be taken to safeguard and protect sensitive data as well as the procedures that assure that the produced data and results that can be shared openly will be easy to locate and access by a wider public.

RadioSpin has chosen to use the Data Management Plan Template that is provided by the European Commission and recommended for Horizon 2020 and, as updated version, for Horizon Europe beneficiaries. The template used by RadioSpin is the latest version, intended for Horizon Europe beneficiaries.

At present, very little data has been generated and/or collected by the project. The RadioSpin DMP is designed to be a 'living' document that will, in this initial version, provide the overall strategy and processes how the RadioSpin research data will be handled during and after the project. During the course of the project the DMP will be extended, reviewed and updated whenever significant changes arise, such as (but not limited to):

- new data are being generated, re-used, collected and/or gathered in any way
- periodic reports are being developed and submitted (incl. the final project report)
- adjustments to the data management strategies become necessary
- changes in individual members' data policies occur
- changes in the consortium composition and external factors occur (e.g. new consortium members joining or existing members leaving).

In preparation for this report the RadioSpin partners considered a number of issues to be addressed, which are described in this report in sections 2 'Data Summary' and 3 'Making data accessible'.
All aspects will have to be answered to for each update and for each generated dataset or data collection that will be added during the term of RadioSpin.

EURIDA will be responsible for communicating this DMP to all project partners and for organising the regular reviews and updates of the DMP. Each project partner will be responsible for managing their data, metadata, and insuring their data meets the data quality and management standard set out in the RadioSpin DMP. RadioSpin coordinator UBx will be responsible for the overall coordination of data management and the compliance with rules, regulations and legal aspects of data management processes. External advice has been and will be given by the RadioSpin external Ethics Advisors, especially on data implications arising by working with human patients, and on the risk of dual use and misuse of the RadioSpin technology.

1.1. The overarching RadioSpin data management strategy

RadioSpin's overall data management strategy follows the principle of Responsible Research and Innovation (RRI). This entails to be as open as possible, but under the consideration of strict ethics and integrity principles and respecting the requirement to protect and exploit results and intellectual property for maximum societal benefit.

RadioSpin, with its focus on artificial intelligence and deploying a use case which involves working with human patients via a dedicated clinical trial, faces particular challenges in terms of handling sensitive data and assuring patient confidentiality.
In addition, even at low technology maturity levels, RadioSpin targets application cases with clear commercial prospects in a highly innovative technology field that could put Europe in the lead of frontier research and innovation.

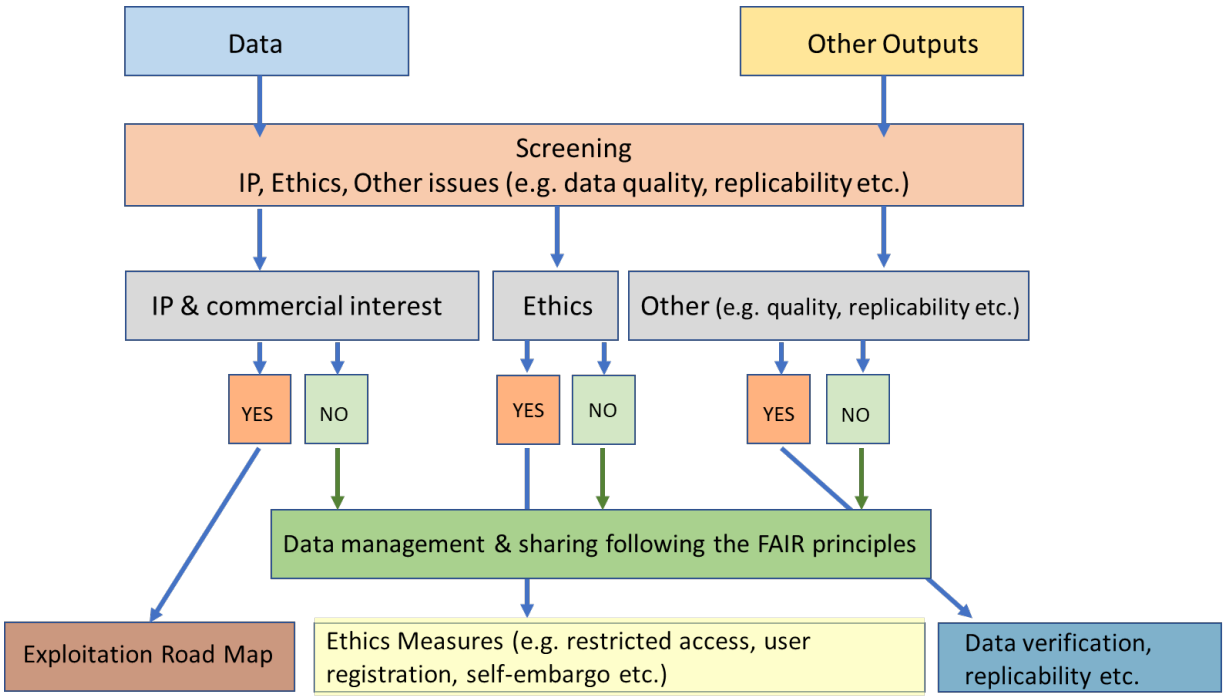
Resulting outputs and data require clear and sound management strategies to guarantee the non-disclosure of data, either for legal and/or ethical reasons or to secure maximum project impact.

Besides those types of confidential data, RadioSpin will further generate data that will not underly any restrictions, neither for ethical, legal, nor for commercial reasons. Sharing such data openly, widely and proactively will increase the impact that RadioSpin will have in a long term. Especially the scientific community will benefit from accessing and re-using the data generated by RadioSpin.

Handling the different types of data will require a data management strategy that reflects the various needs for screening, categorising, curating, storing, protecting and/or sharing the data.

Figure 1 illustrates the overall workflow and procedures used in RadioSpin to ensure the effective protection of sensitive data and the open access to non-confidential research data.

Fig. 1 Workflow diagram for different types of data (open and confidential)



1.2. Personal data, patient data and individual participants

To protect the privacy of individual participants, especially in the RadioSpin MammoWave™ use case and dedicated clinical trials, only data that can be irreversibly anonymised to the degree that it is impossible to identify individuals will be shared publicly. Specifically, data on patient collected during the trial will be used in an anonymous way; the subject will only be identified by a subject number, as indicated in the clinical protocol and in the informed consent. Non-anonymised data will be kept internally in the Hospitals where the trials will be activated, and never shared publicly in its original format. Only anonymised data will feed into project work and provide basis for analysis in deliverables and scientific publications. If the editor of a deliverable is concerned that their deliverable contains personal information, they request a separate screening for privacy and ethics issues before submission to be sure that no personal data is included. The leader of WP9 Ethics (UBx) is responsible for performing these screenings. Public deliverables, publications and anonymised datasets will be shared openly through an open research data repository (see section 3.5).

1.3. Data underlying outputs for commercial exploitation

During the lifetime of the project, partners might discover business opportunities based on the project's results that can lead to commercial exploitation. This will be monitored by the Innovation Manager, and if cases arise appropriate steps to protect such results for exploitation purposes will be taken. As explained in the overall data management strategy (Section 1.1) and displayed in Figure 1, data underlying such results will not be openly shared.

1.4. Legal frameworks

As of May 2018, the General Data Protection Regulation (GDPR)1 is applicable in all Member States in the European Union, as well as in the countries in the European Economic Area (EEA). GDPR updates and modernises existing laws on data protection to strengthen citizens' fundamental rights and guarantee their privacy in the digital age.

GDPR regulates the processing by an individual, a company or an organisation of personal data relating to individuals in the EU[1]. It does not apply to the processing of personal data of deceased persons or of legal entities. It sets down one set of data protection rules for all companies and organisations operating anywhere in the EU and European Economic Area (EEA), for two main reasons: 1) to give people more control over their personal data, 2) level the playing field for businesses and organisations operating in the EU and EEA. GDPR grant individuals a set of rights that must be protected by any actor who processes personal data. The individual rights include the right to:

- information about the processing of your personal data;
- obtain access to the personal data held about you;
- ask for incorrect, inaccurate or incomplete personal data to be corrected;
- request that personal data be erased when it's no longer needed or if processing it is unlawful;
- object to the processing of your personal data for marketing purposes or on grounds relating to your particular situation;
- request the restriction of the processing of your personal data in specific cases;
- receive your personal data in a machine-readable format and send it to another controller ("data portability"); and
- request that decisions based on automated processing concerning you or significantly affecting you and based on your personal data are made by natural persons, not only by computers. You can also have the right in this case to express your point of view and to contest the decision.

2. Data Summary

The Annex to this report provides a list of all datasets currently envisaged to be generated and/or re-used by RadioSpin and their planned management procedures. This list will be updated and refined as the project matures and progress is achieved.

In the following the general strategies for generating and/or re-using data and managing the individual datasets as well as the rules agreed by all RadioSpin consortium partners for data curation, storage, accessibility and protection (where required for legal, ethical or IP protection requirements) are explained as far as possible at this stage of the project. Additional analyses will be performed and further plans and responses given as the project matures.

2.1. Purpose of the data generation or re-use and its relation to the RadioSpin objectives and target groups

This section of the DMP and its future updates responds to the following questions:

- *What is the purpose of the data generation or re-use and its relation to the objectives of RadioSpin?*
- *To whom might your data be useful ('data utility'), outside your project (target groups and stakeholders, e.g. which end users, policymakers, scientific community for re-use etc.?*

- **Laboratory experimental characterisation data:** The generated data includes covers the characterisation of spintronics devices/CMOS chips in a laboratory environment. Purpose of the data is to be able to characterise individual devices and arrays before and after their CMOS integration. This is essential for the circuit design and the realisation of the goals of the project.

- **Simulation data:** The generated data includes simulations of individual spintronics devices, electrical circuits and networks. Purpose of the data is to allow the system to be modelled in advance of

[1] https://eur-lex.europa.eu/eli/reg/2016/679/oj

execution to optimise relevant parameters.

- **Process flow data:** The generated data includes run sheets for the deposition and nanofabrication process for spintronics device and later stage spintronic-CMOS integration. Purpose of the data is to monitor the process steps which allows for device optimisation and analysis of process faults.

- **Clinical Trial data: U**BT Srl will create and collect technical data through phantom measurements via MammoWave. Purpose of the data is to enable the testing and optimisation of microwave imaging algorithms and the RadioSpin simulator/demonstrator.

2.2. Types, formats, size and origin of data in RadioSpin

This section of the DMP and its future updates responds to the following questions:

- *What types and formats of data will RadioSpin generate or re-use?*
- *What is the expected size of the data that you intend to generate or re-use?*
- *What is the origin/provenance of the data, either generated or re-used?*

- **Laboratory experimental characterisation data:** The data will derive from laboratory equipment at INL, for example spectrum analyser, oscilloscope and electrical source meters. Data sizes can vary from a few hundreds of KB (i.e. electrical source meters) to several GB (i.e. oscilloscopes) depending on the resolution and equipment used.

- **Simulation data:** Data includes simulations of individual spintronics devices, electrical circuits and networks. Data will be generated by INL with specialised software, for example mumax, oommf, python. Data sizes are typically in the range 1-100 MB depending on the software used.

- **Process flow data:** Data includes run sheets for deposition and nanofabrication process for spintronics device and later stage spintronic-CMOS integration. These are tool-specific and are stored in either a text document, e.g., *.doc or similar – or a laboratory management tool. These typically originate from a set of process steps, which are tool-specific, e.g., dry etching, wet etching, metal sputtering or evaporation, oxide deposition, etc., and are compiled by process operators and process flow designers. Size depends on format (i.e. simple spreadsheet or including images) and is typically 1-100 MB.

- **Clinical trial data:** UBT Srl will create and collect technical data through phantom measurements via MammoWave. All of this data will be created in digital format and text files. Measurement data may have size up to few GB. The data will typically be created and stored as mat or txt formats. The data will contain information that will enable testing microwave imaging algorithms and RadioSpin simulator/demonstrator, producing as output .fig /jpg/tiff or similar.

2.3. Expected re-use of existing data

For the above listed data only newly generated data are foreseen in RadioSpin. This is mainly due to the originality of the project and the target technology. Specific cases for data re-use are:

Demonstrator evaluation on RF fingerprinting application: Data use targets the evaluation of the RadioSpin demonstrator capabilities in terms of training and performance before scale-up to the actual use case (developed in WP6). The database and all related activities are described in detail in deliverable D5.2, submitted on the 31st March 2022.

For the evaluation purposes radiofrequency (RF) signals emitted by commercial drones and their radio-controllers are re-used, previously collected by Basak et al. in “Drone classification from RF fingerprints using deep residual nets” (IEEE COMSNETS conference, 2021[2]). The database is composed of 400 signals. Contrary to the initial dataset proposed by Basak et al., RadioSpin chose to not separate “test” from “train” signals upfront. The database is stored in an h5 file, a format adapted to databases. Inside the

[2] S. Basak, S. Rajendran, S. Pollin and B. Scheers, "Drone classification from RF fingerprints using deep residual nets," *2021 International Conference on COMmunication Systems & NETworkS (COMSNETS)*, 2021, pp. 548-555, doi: 10.1109/COMSNETS51098.2021.9352891.

file there are two datasets: the signals ('Signals') and the targets ('Targets'). The data can then easily be extracted and fed to a machine learning data-loader (from the Pytorch library for instance) for further re-use. The database file is named following the RadioSpin naming convention (*RadioSpin_D52_RF_fingerprinting.h5).*

Clinical Trial: There are no plans to re-use existing data as part of the RadioSpin clinical trial. The main reason is that in a dedicated clinical trial all necessary ethics measures like patient consent, data confidentiality and full anonymisation can be guaranteed, while working with existing data from previous trials this is not possible.

3. FAIR data

The following details refer to openly shared data only. Measures for the curation and protection of sensitive data and/or data that underly IP protected results are described in *Section 6, Data Security*.

To comply with the principles of FAIR data, the RadioSpin consortium decided to use Zenodo (https://zenodo.org) as the main repository for making the project's research data and other outputs, such as scientific publications, fact sheets and other info materials findable in accordance with the requirements towards open data as stated in the Grant Agreement and the Horizon 2020 Open Data principles.

We will create a RadioSpin community in the Zenodo repository, so all open datasets, public deliverables, publications and other public outputs can be uploaded in this community by all consortium partners.

Through Zenodo, all uploads will be linked to OpenAire (https://www.openaire.eu/). This will ensure maximum visibility of RadioSpin data and results among the European scientific and expert community and make data findable via the Zenodo metadata standards.

3.1. Making data findable, including provisions for metadata

This section of the DMP and its future updates responds to the following questions:

- *How will RadioSpin data be identified, e.g. via persistent identifier?*
- *Will rich metadata be provided to allow discovery? What metadata will be created? What disciplinary or general standards will be followed? In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.*
- *Will search keywords be provided in the metadata to optimize the possibility for discovery and then potential re-use?*
- *Will metadata be offered in such a way that it can be harvested and indexed?*

Making data findable with metadata

The following principles are used by Zenodo to make research data findable (F1-F4). Those principles also apply for all open datasets shared by RadioSpin via Zenodo:

- **F1**: (meta)data are assigned a globally unique and persistent identifier
  - A DOI is issued to every published record on Zenodo.
- **F2**: data are described with rich metadata ((meta)data are richly described with a plurality of accurate and relevant attributes; each record contains a minimum of DataCite's mandatory terms, with optionally additional DataCite recommended terms and Zenodo's enrichments.
  - Zenodo's metadata is compliant with DataCite's Metadata Schema minimum and recommended terms, with a few additional enrichements.
- **F3**: metadata clearly and explicitly include the identifier of the data it describes
  - The DOI is a top-level and a mandatory field in the metadata of each record.
- **F4**: (meta)data are registered or indexed in a searchable resource
  - Metadata of each record is indexed and searchable directly in Zenodo's search engine immediately after publishing.
  - Metadata of each record is sent to DataCite servers during DOI registration and indexed there.

Metadata associated with each data set that will be published by RadioSpin in Zenodo will by default include:

- Digital Object Identifiers
- Version numbers
- Bibliographic information
- Keywords
- Abstract/description
- Associated project and community
- Associated publications and reports
- Grant information
- Access and licensing info
- Language

Project name and Grant Agreement number represent standard details as part of the grant information.

Keywords for optimised discovery

The researchers collecting the data at each organisation involved in the project will be responsible for uploading the specific datasets that they have created. All datasets will include a set of keywords associated with the data. The keywords must be descriptive to the content of the dataset. For example, a dataset containing information on the evaluation of algorithms for the biomedical application in MammoWave should be tagged with corresponding descriptive keywords like 'algorithm', 'MammoWave', 'biomedical imaging'.
Before publishing, public datasets and suggested keywords will be submitted to the Project Coordinator, Sylvain Saïghi, and the Scientific Coordinator, Julie Grollier, for review and feedback.

For general guidance and as part of the RadioSpin content branding, a set of general keywords that shall be used for all public datasets, scientific publications and public deliverables have been defined. These are as follows:

- Neuromorphic computing technologies
- Nano-technology
- Materials engineering
- Spintronics
- Radio-frequency devices
- Deep learning
- Non-linear dynamics
- Artificial Intelligence
- Neural networks

3.2. Making data accessible

Data Repository:

This section of the DMP and its future updates responds to the following questions:

- *Will the data be deposited in a public/trusted repository?*
- *Have you explored appropriate arrangements with the identified repository where your data will be deposited?*
- *Does the repository ensure that the data is assigned an identifier? Will the repository resolve the identifier to a digital object?*

As highlighted in *Section 3.1 Making data findable*, all public datasets, scientific publications and deliverables that are assessed as 'open' and can therefore be shared with the public, will be uploaded to Zenodo and made openly available free of charge. Publications and underlaying data sets will be linked through the use of persistent identifiers (DOI issued by Zenodo). Datasets that have been assessed as "confidential" (for personal, ethics or exploitation reasons) will not be shared. This is further explained under 'Data' below.

Zenodo takes the following measures to make all data accessible (A1-A2):

- **A1**: (meta)data are retrievable by their identifier using a standardized communications protocol
  - Metadata for individual records as well as record collections are harvestable using the OAI-PMH protocol by the record identifier and the collection name.
  - Metadata is also retrievable through the public REST API.
- **A1.1**: the protocol is open, free, and universally implementable
  - See point A1. OAI-PMH and REST are open, free and univesal protocols for information retrieval on the web.
- **A1.2**: the protocol allows for an authentication and authorization procedure, where necessary
  - Metadata are publicly accessible and licensed under public domain. No authorization is ever necessary to retrieve it.
- **A2**: metadata are accessible, even when the data are no longer available
  - Data and metadata will be retained for the lifetime of the repository. This is currently the lifetime of the host laboratory CERN, which currently has an experimental programme defined for the next 20 years at least.
  - Metadata are stored in high-availability database servers at CERN, which are separate to the data itself.

The list of expected datasets in Annex A represents a first version which will be updated and extended as the project evolves. Furthermore, not all details are already known at this stage. This includes the size of datasets or other specific information which will only become available once the data has been generated. Therefore, updated versions of the datasets listed in the Annex will be delivered during either one of the updated versions of the DMP or, the latest, for its final version at the end of the project.

Data:

This section of the DMP and its future updates responds to the following questions:

- *Will all data be made openly available?*
- *If an embargo is applied to give time to publish or seek protection of the intellectual property (e.g. patents), specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.*
- *Will the data be accessible through a free and standardized access protocol?*
- *If there are restrictions on use, how will access be provided to the data, both during and after the end of the project?*
- *How will the identity of the person accessing the data be ascertained?*
- *Is there a need for a data access committee (e.g. to evaluate/approve access requests to personal/sensitive data)?*

In accordance with the Horizon 2020 Open Access Mandate RadioSpin commits to making all project data and results openly accessible with as few restrictions as possible. The European Commission's open access principle however entails the strict protection of personal and sensitive data for reasons of personal rights, ethics and/or for commercial exploitation.
RadioSpin will fully comply with those requirements. Management strategies that are currently foreseen are listed below. Further plans and updates will be added as the project matures.

Restrictions on data from clinical trials: Working with humans in the dedicated MammoWave™ clinical trial, the possible ethics issue of dual use and misuse of the target technology as well as future commercial prospects in the highly innovative and competitive area of neuromorphic computing and deep learning requires RadioSpin to apply strict protection measures for personal and sensitive data. While some of the datasets that cannot be shared due to the protection of personal data are already known at this stage and further explained in the following, some datasets that for example may be restricted due to ethics or to protection measures for commercial exploitation will be described in detail in the updated and final versions of the DMP.

Regarding the MammoWave™ clinical trials data, data on patient collected during the trial will be used in

an anonymous way; the subject will only be identified by a subject number, as indicated in the clinical protocol and in the informed consent. Specifically, the Clinical Investigator will assign a subject number, i.e. and ID code, to the patients undergoing the trial, Non-anonymised data will be kept internally in the Hospitals where the trials will be activated, and never shared publicly in its original format. Only anonymised data will feed into project work and provide basis for analysis in deliverables and scientific publications.
Specifically, the Clinical Investigator will assign a subject number, i.e. and ID code, to the patients undergoing the trial (and only the Clinical Investigator will know such number). Next, using this ID code, the Clinical Investigator will complete the Case Report Form where the outcome of the investigation will be summarized.

More in details, subjects will be identified only through the study number, i.e. ID code, that will be attributed as follows:
- Site Code with 2 digits (01, 02, …)
- Subject Code with 3 digits (001, 002,…)

The subjects will be progressively numbered among each site.

After completion of the study and analysis of the data, the essential documents (completely anonymized) will be stored in the Trial Master File, which will be deposited in the UBT's archive for storage for at least 25 years according to new EU regulation. Participant confidentiality and privacy is strictly held in trust by the participating Investigators, their staff, and UBT. Therefore, the study protocol, documentation, data, and all other information generated will be held in strict confidence. No information concerning the study or the data will be released to any unauthorized third party. All research activities will be conducted in as private a setting as possible. The Study participant's contact information will be securely stored at each clinical site for internal use during the study.

Only the outcomes of the investigation will be made openly available.

Restrictions on data due to ethical implications (dual use/misuse or possible harmful applications):
RadioSpin is fully committed to the principles of Responsible Research and Innovation. This includes to avoid any risks of harmful use of project results, including but not restricted to possible dual use or misuse.

The Ethics Management Team, supported by the External Ethics Advisor Dr. Inga Ulnicane (Ethics expert in the Human Brain Project), will assess all data with regards to the possibility of conflicts with the RRI principle. In case any risks are detected, restrictions towards the Open Data policy of RadioSPin will be assessed. This can include the following action points:

- Restricted access to selected data for selected users only
- Registered access to selected data, combined with an end user screening
- Self-embargo on data publishing for particularly sensitive data

Restrictions on data due to commercial use and IP protection:
RadioSpin will protect all data that will be essential for commercialisation and/or the protection of intellectual property. The exact strategy as regards embargo periods, future access of data one IP has been protected, data ownership etc. will be discussed, agreed and fixed in the DMP per dataset. Updates will be included in future DMP versions and become part of the respective periodic technical report.

3.3. Making data interoperable

This section of the DMP and its future updates responds to the following questions:

- *What data and metadata vocabularies, standards, formats or methodologies will you follow to make your data interoperable to allow data exchange and re-use within and across disciplines? Will you follow community-endorsed interoperability best practices? Which ones?*
- *In case it is unavoidable that you use uncommon or generate project specific ontologies or vocabularies, will you provide mappings to more commonly used ontologies? Will you openly publish the generated ontologies or vocabularies to allow reusing, refining or extending them?*
- *Will your data include qualified references[3] to other data (e.g. other data from your project, or datasets from previous research)?*

Zenodo uses the JSON schema as the internal representation of metadata and offers export to other formats such as Dublin Core, MARCXML, BibTeX, CSL, DataCite and export to Mendeley. The data record metadata will utilise the vocabularies applied by Zenodo. For certain terms, these refer to open, external vocabularies, e.g.: license (Open Definition), funders (FundRef) and grants (OpenAIRE). Reference to any external metadata is done with a resolvable URL.

3.4. Increase data re-use

This section of the DMP and its future updates responds to the following questions:

- *How will you provide documentation needed to validate data analysis and facilitate data re-use (e.g. readme files with information on methodology, codebooks, data cleaning, analyses, variable definitions, units of measurement, etc.)?*
- *Will your data be made freely available in the public domain to permit the widest re-use possible? Will your data be licensed using standard reuse licenses, in line with the obligations set out in the Grant Agreement?*
- *Will the data produced in the project be useable by third parties, in particular after the end of the project?*
- *Will the provenance of the data be thoroughly documented using the appropriate standards?*
- *Describe all relevant data quality assurance processes.*

The Zenodo digital repository uses the following principles to assure maximum re-usability of open data:

- **R1**: (meta)data are richly described with a plurality of accurate and relevant attributes
  - Each record contains a minimum of DataCite's mandatory terms, with optionally additional DataCite recommended terms and Zenodo's enrichments.
- **R1.1**: (meta)data are released with a clear and accessible data usage license
  - License is one of the mandatory terms in Zenodo's metadata, and is referring to an Open Definition license.
  - Data downloaded by the users is subject to the license specified in the metadata by the uploader.
- **R1.2**: (meta)data are associated with detailed provenance
  - All data and metadata uploaded is tracable to a registered Zenodo user.
  - Metadata can optionally describe the original authors of the published work.
- **R1.3**: (meta)data meet domain-relevant community standards
  - Zenodo is not a domain-specific repository, yet through compliance with DataCite's Metadata Schema, metadata meets one of the broadest cross-domain standards available.

[3] *A qualified reference is a cross-reference that explains its intent. For example, X is regulator of Y is a much more qualified reference than X is associated with Y, or X see also Y. The goal therefore is to create as many meaningful links as possible between (meta)data resources to enrich the contextual knowledge about the data. (Source: https://www.go-fair.org/fair-principles/i3-metadata-include-qualified-references-metadata/)*

4. Other research outputs

Overall strategy for managing other research outputs
RadioSpin targets a breakthrough technology in the area of neuromorphic computing with the overall goal to radically change the way brain-inspired computing and neural networks will be realised in the future.

During its research and development RadioSpin will create novel concepts, processes, methods, technology solutions, and scientific know-how to progress the entire field of neuromorphic computing and will set future benchmarks and standards which will live on after the project ends.

It is essential to ensure that the widest possible group of stakeholders is reached with the results of RadioSpin. That way the biggest possible scientific, economic and societal impact in Europe is secured. As Intellectual Property Rights (IPR) and Ethics issues will have to be respected, dissemination activities will be designed and implemented in close cooperation with the exploitation partners who are responsible for implementing the RadioSpin use cases MammoWave™ (UBT) and RF fingerprinting (Thales), and the external Ethics Advisors where working with humans (clinical trial) and the risk of dual use/misuse are concerned.
While RadioSpin is committed to openly sharing all results with its target communities, these two key issues pose restrictions towards the full disclosure of results, which is coherent with data sharing and disclosure strategies.
As an example, no dissemination of results may take place before a decision is made regarding its role in the exploitation plan and the possible protection through IPR on the one hand, and regarding possible ethics risks on the other hand. In advance of any disclosure all project partners and the Ethics Advisors have therefore to be contacted for their authorisation. Communication Officer Rita Clancy, supported by the project Ethics Management Team will oversee the action.

Results sharing (aka dissemination, exploitation and communication activities) in RadioSpin are based on the principles of Responsible Research and Innovation (RRI). This is one of the key priorities in the 'Innovation Union' Flagship Initiative of the European Commission which aims to maximise projects' impacts by engaging the civil society in Research and Innovation activities and making know-how openly accessible to wide user groups. Further key topics in RRI are research ethics and the consideration of gender aspects. Both are in the focus of RadioSpin, based on the intrinsic gender aspect of the innovative breast cancer detecting MammoWave™ technology as use case and the ethics aspects during clinical trials as well as strategies to avoid possible dual use or misuse of RadioSpin technology and will have effects on results sharing and disclosure.

Applying the FAIR principles to other research outputs
RadioSpin will for all results sharing activities assure that the FAIR principles will be used as widely as possible, which mainly means making other (non-data) results **f**indable, **a**ccessible, **i**nteroperable and **r**eusable.

Efforts to apply those principles are briefly summarised below and explained in detail in the RadioSpin Advanced Dissemination, Communication and Exploitation Plan (Advanced DEP, deliverable D7.2 and regular updates thereof as part of the technical periodic reports).

- ***Making outputs and results findable and accessible:*** To make research outputs findable, suitable **keywords** will be used through which interested parties can easily find them either via Google, social media channels or other channels used by RadioSpin. In addition, common channels established and widely used by the neuromorphic computing community and the application sectors 'biomedical imaging' and 'health' and 'RF fingerprinting' will be used as multipliers to increase the findability of outputs.
  Reports, fact sheets and policy briefs will be used as proven formats for highlighting outputs of relevance. All related materials that target an improved findability of outputs will include the grant information and keywords like 'Horizon 2020', 'H2020', which will further improve the findability of outputs for the Horizon 2020 community and other EU funded projects.
  For social media posts, respective keywords will be used and target community representatives tagged in posts to assure key stakeholders will find outputs and results.

All outputs, associated info materials and posts will be made accessible through the project website as single access point for project outputs.

- ***Making outputs interoperable and reusable:*** The RadioSpin technology has many appealing characteristics for future uptake by various target sectors and communities. Interoperability in the sense of non-data research outputs can be understood as interoperability and integrability with future industrial applications. In the centre of RadioSpin are applications that require ultrafast and embedded classification of RF signals. The most interesting ones have already been chosen as target applications within the project, which are RF fingerprinting (for example for commercial drone identification) and biomedical electronics (for example for X-ray free mammography).
  To assure the highest levels of interoperability and re-use of RadioSpin outputs, we utilise an End User Panel, which is a group of selected end users from the possible future application fields for the RadioSpin technology. This Panel will give insight into industry requirements and needs, including the viability and usability of technology outputs and the specifics for smooth technology integration in existing devices or processes.
  At the end of the project, the Panel members will get the chance to participate in a demonstration round of the lab prototype and provide feedback for example on application-specific requirements towards RadioSpin technology including usability, costs and other parameters that are critical for target applications and markets. Follow up plans will be discussed for the maturing of TRLs and the upscaling of RadioSpin technology. Potential members of the Panel will be contacted through the wider consortium partner networks and client bases. Furthermore, the panel may include representatives of standardisation bodies or expert groups, above all CENELEC as the European official authority, to receive input on the latest developments. By engaging end-users already at low TRL levels, we will be able to increase and maximise the compatibility, integrability, feasibility and viability of RadioSpin outputs, hence their interoperability and re-usability.

5. Allocation of resources

Costs

RadioSpin uses standard tools and a free of charge research data repository. The costs of data management activities are limited to project management costs and will be covered by allocated resources in the project budget.

Long-term preservation of the public data is ensured through Zenodo. Other resources needed to support reuse of data after the project ends will be solved on a case-by-case basis.

Data Manager

The overall responsibility for data management lies with the project coordinator, Sylvain Saïghi from University of Bordeaux, supported by RadioSpin's scientific project coordinator, Ms. Julie Grollier from THALES.

Additional support on necessary data confidentiality due to IP protection needs will be given by the IP expert office AST – Aquitaine Sciences Transfer. This will guarantee the tracking of IP sensitive data that underly patentable results (in accordance with IP Management procedures detailed in the Consortium Agreement).

Coordinator for data generated in the clinical trials for the MammoWave use case is partner UBT, supported by the assigned Data Protection Officer, Martina Paoli (m.paoli@ubttech.com), and the external Ethics Advisor, Ms. Cristina Morelli from the MTA Group, who is specialised in ethics and data issues in the health sector.

This data management team is further supported by the RadioSpin Communication Officer, Rita Clancy from EURIDA, who will assure that no confidential data will be disclosed during dissemination, communication and exploitation activities, and secure the systematic and timely release and proactive sharing of open, non-confidential project data to wide user groups for maximum accessibility and re-use.

6. Data security

Data security – The RadioSpin internal repository

RadioSpin uses MS Teams as single Sharepoint for all internal project resources, including data and other project outputs (e.g. reports, deliverables). Members to the RadioSpin MS Teams are individually invited with their email addresses.

Microsoft Teams is built on the Microsoft 365 and Office 365 hyper-scale, enterprise-grade cloud, delivering advanced security and compliance capabilities.

Teams enforces team-wide and organization-wide two-factor authentication, single sign-on through Active Directory, and encryption of data in transit and at rest. Files are stored in SharePoint and are backed by SharePoint encryption. Notes are stored in OneNote and are backed by OneNote encryption. The OneNote data is stored in the team SharePoint site. The Wiki tab can also be used for note taking and its content is also stored within the team SharePoint site. Therefore, all shared and stored content is subject to the two-factor authentication.

Security protocols for the Teams channels follow the recommended security road-map provided by Microsoft[4]

Data security – The Zenodo digital repository

The following list describes the security settings for Zenodo:

- Versions: Data files are versioned. Records are not versioned. The uploaded data is archived as a Submission Information Package. Derivatives of data files are generated, but original content is never modified. Records can be retracted from public view; however, the data files and records are preserved.
- Replicas: All data files are stored in the CERN Data Centres, primarily Geneva, with replicas in Budapest. Data files are kept in multiple replicas in a distributed file system, which is backed up to tape on a nightly basis.
- Retention period: Items will be retained for the lifetime of the repository. The host laboratory of Zenodo CERN, has defined a lifetime for the repository of the next 20 years minimum.
- Functional preservation: Zenodo makes no promises of usability and understandability of deposited objects over time.
- File preservation: Data files and metadata are backed up nightly and replicated into multiple copies in the online system.
- Fixity and authenticity: All data files are stored along with an MD5 checksum of the file content.
- Files are regularly checked against their checksums to assure that file content remains constant.
- Succession plans: In case of closure of the repository, Zenodo guarantees to migrate all content to suitable alternative institutional and/or subject based repositories.

7. Ethics

RadioSpin involves working with humans and patients as part of the MammoWave clinical trials. As regards data management, all ethics issues are dealt with in accordance with the Ethics requirements identified and explained in the project's Grant Agreement.

RadioSpin further includes a separate Ethics Work Package (WP9) and four Ethics deliverables, either in relation to ethics issues associated with the RadioSpin MammoWave clinical trail or the dual use/misuse potential of the technology.

The precise measures that will be taken to avoid any ethics risks are listed in the GA or explained in detail in the project deliverables D9.1 to D9.4.

Below, the data management activities that are connected to ethics issues are summarised in brief:

[4] https://docs.microsoft.com/en-us/microsoft-365/security/office-365-security/security-roadmap?view=o365-worldwide

MammoWave Clinical Trial. Prior to execution, approval should be obtained by the relevant Ethics Committee(s) or Institutional Board(s). A dedicated protocol and informed consent sheet will be submitted.

The study will be carried out in accordance to the protocol and with principles enunciated in the current version of the Declaration of Helsinki, the guidelines of Good Clinical Practice (GCP) issued by ICH, in case of medical device: the European Directive on medical devices 93/42/EEC and the ISO Norm 14155 and ISO 14971, the Regulation (EU) 2017/745 on medical devices (MDR) the European Law and regulatory authority's requirements. The Ethics Committee(s) and regulatory authorities will receive annual safety and interim reports and be informed about study stop/end in agreement

Informed consent for data sharing and long term preservation: The investigator or a person designed by him/her has to collect written informed consent form from each patient before her participation in the study. This version has to be previously approved by relevant ethics committee and should include all the elements required by law according to the ICH-GCP recommendations. Prior to this, the investigator has to inform each participant of the objectives, aims and details of the study. The eligible patient will be provided with an information and consent form in clear, simple language. The investigator must allow enough time to all participants to let them decide whether or not they are willing to participate in the study.

If the eligible patient is unable to read, an impartial witness should be present during the entire informed consent discussion in order to sign and date it together with the person responsible for collecting the informed consent form. The eligible patient will be given one signed original information and consent form; the second original will be kept by the investigator.

All patients/legal guardians or representatives must sign and date the informed consent. No data will be transferred without patient's prior consent.

Any change to one of the information and consent form constitutes an amendment to this document and must be, if applicable, submitted for approval to the ethics committee.

The investigator affirms and upholds the principle of the participant's right to privacy and that they shall comply with applicable privacy laws. Especially, anonymity of the participants shall be guaranteed when presenting the data at scientific meetings or publishing them in scientific journals.

Individual subject medical information obtained as a result of this study is considered confidential and disclosure to third parties is prohibited. Subject confidentiality will be further ensured by utilising subject identification code numbers to correspond to treatment data in the computer files.

Dual Use/Misuse

The Ethics Management Team, supported by the External Ethics Advisor Dr. Inga Ulnicane (Ethics expert in the Human Brain Project), will assess all data with regards to the possibility of conflicts with the RRI principle. In case any risks are detected, restrictions towards the Open Data policy of RadioSPin will be assessed. This can include the following action points:

- Restricted access to selected data for selected users only
- Registered access to selected data, combined with an end user screening
- Self-embargo on data publishing for particularly sensitive data

8. Other issues

No other data management issues have been identified so far. Should organisational, national or other data management rules or standards be considered in addition to those underpinning this initial DMP, they will be included in the next version of the DMP.

ANNEX – RadioSpin Datasets

WP	Origin of Data	Name of dataset	Description	Format	Responsible for data management	Classification	Timeline	Comments
WP 1	Experiments on physical devices	Exp_full_spin_NN	Dataset resulting from the experimental realization of the fully spintronic neural network	Raw and processed data in text (.dat), code in Python (.py)	CNRS (Julie Grollier)	Public	At publication of paper, expected late 2022 – early 2023	This dataset will be made available following publication of a scientific paper
WP 4 and WP 2	Numerical simulations and models	Simu_NN	Dataset resulting from the use of the neural network simulator with semi-independent oscillators, using the behavioural models of spintronic devices	Resulting data in text (.dat), codes and models in Python (.py)	Thales (Alice Mizrahi)	Public	With Exp_full_spin_NN dataset	This dataset will be made available following publication of a scientific paper
WP 6	Modification of an existing database	Drones_RF	Dataset of RF signals emitted by drones. This database is a modification of a public database, to suit the requirements of our system	Large file in .h5 archive	Thales (Alice Mizrahi)	Public	With Exp_full_spin_NN dataset	This dataset will be made available following publication of a scientific paper
WP 4	Numerical simulations	Simus_NN_dynamical	Dataset resulting from the use of the neural network simulator with coupled oscillators	Resulting data in text (.dat), codes in Python (.py)	CNRS (Julie Grollier	Public	2023 at earliest	This dataset will be made available following publication of a scientific paper
WP 5	Designed database of signals	RF_fingerprinting_demo	Database of RF signals designed to mimic an RF fingerprinting task. This database is designed using insight from Drones_RF.	Text (.dat) or large file in .h5 archive if needed	Thales (Alice Mizrahi)	Public	End of project, along with Exp_test_demo	To be used on demonstrator. This dataset will be made available following publication of a scientific paper
WP 5	Experiments on demonstrator	Exp_test_demo	Dataset of results from the experiments on the RadioSpin demonstrator	Raw and processed data in text (.dat), code in Python (.py)	CNRS (Julie Grollier	Public	End of project	This dataset will be made available following publication of a scientific paper
WP 6	Numerical simulations	Simus_NN_Mammo	Dataset resulting from the application of the neural network simulator to the MammoWave biomedical application	Resulting data in text (.dat), code in Python (.py)	Thales (Alice Mizrahi)	Public	End of project	This dataset will be made available following publication of a scientific paper

WP 1 WP 3	Device and circuit characterisation	Laboratory experimental characterisation data	Data characterisation of spintronics devices/CMOS chips in lab environment to characterise individual devices and arrays before and after CMOS integration essential for circuit design and the realisation of the goals of the project.	Text files (*.txt) or reports (*.doc, *.pdf)	INL	Public	M1-M46	tba
WP 1 WP 3	Device modelling	Simulation data	Data from simulations of individual spintronics devices, electrical circuits and networks for system modelling in advance of execution to optimise relevant parameters.	Text files (*.txt) or reports (*.doc, *.pdf)	INL	Public	M1-M46	tba
WP 1 WP 3	Device fabrication and integration	Process flow data	Data include run sheets for deposition &nanofabrication process for spintronics device and later stage spintronic-CMOS integration to monitor process steps for device optimisation and analysis of process faults.	Text files (*.txt) or reports (*.doc, *.pdf)	INL	Confidential	M1-M46	tba
WP 5	Technical data through MammoWave™ phantom measurements.	MammoWave™ phantom data	The data will contain information that will enable testing microwave imaging algorithms and RadioSpin simulator/ demonstrator, producing as Measurement data may have size up to few GB.	Data output format: .fig /jpg/tiff or similar. Data will typically be stored as mat or txt formats.	UBT (Gianluigi Tiberi)	Public	M18-M50	Data management tasks will be supported and supervised by the External Ethics Advisor Dr. Cristina Morelli (MTA Group).
WP 6	Technical data through MammoWave™ clinical trial measurements.	MammoWave™ Clinical Trial data	The data will contain information that will enable testing microwave imaging algorithms and Radiospin simulator/ demonstrator. Data will be created in digital format and text files. Measurement data may have size up to few GB.	Data output format: .fig /jpg/tiff or similar. Data will typically be stored as mat or txt formats.	UBT (Gianluigi Tiberi)	Confidential	M18-M50	Data management tasks will be supported and supervised by the External Ethics Advisor Dr. Cristina Morelli (MTA Group).