Data Management Plan

Project PID2020-116540RB-C22

financed by MCIN/ AEI /10.13039/501100011033

DOCUMENT VERSION

Date	Version	Editor	Comments	Status
28/2/2022	v1	C. Guardiola Albert		Draft
15/3/2022	V2	C. Guardiola Albert	Based on comments given by M. Gómez, M. Sanabria, M. Béjar, O. Monserrat y M. Crossetto	Final
31/8/2023	V3	C. Guardiola Albert	Update with new CSIC repositories	Final

Contents

Data Management Plan

This Data Management Plan (DMP) describes the life cycle of the management of all the data sets that will be collected, processed or generated by the research project "Towards an intelligent exploitation of ground movement data for the prevention and mitigation of geological-geotechnical risks (SARAI)". This document describes how the data will be stored and preserved during the course of the research project and even after its conclusion, describing which data will be collected, processed or generated, what methodologlies and standards will be used, how they will be shared and openly accessible, and finally how they will be conserved and preserved. The DMP is a living document, which will evolve and acquire more precision and entity during the term of the project. In the absence of a DMP template within the Spanish National Plan, an attempt has been made to adjust the DMP to the template provided by the European Commission for the management of Horizon 2020 projects. The DMP will be updated at least once halfway through the course of the project. and another at the end to make the necessary adjustments to the data generated and the uses identified by those responsible for the project, since not all the data or the potential uses that can be made of them are clear from the beginning. New versions of the DMP should be created whenever there are important changes in the project due to the inclusion of new data sets or external factors.

Project Details

- **Name of the project:** Towards an intelligent exploitation of ground movement data for the prevention and mitigation of geological-geotechnical risks (SARAI)
- **Summary:** This project aims to improve the capabilities of exploiting radar-based remote sensing techniques to measure and monitor earth deformation: Differential Interferometric SAR (DInSAR) and Persistent Scattering Interferometry (PSI). The launch of the Sentinel-1A and 1B missions of the Copernicus Program set a paradigm shift: SAR data will be available for many years from now on worldwide and free of charge. In 2016, a European initiative proposed to implement a European Earth Movement Service (EGMS) based on DInSAR to provide consistent data on a continental scale. The SARAI project wants to improve the exploitation of the forthcoming EGMS results, by developing tools and procedures to process and analyse the main products of DInSAR and PSI (i.e. linear strain rate over the observed period and time series). To facilitate the use of this data by non-expert users, SARAI will create tools to: (i) Identify deformation signals and separate them from noise; (ii) Classify these signals according to the causing phenomenon, natural or anthropic; (iii) Estimate the areas most likely to experience ground movements; (iv) Predict the temporal evolution of ground deformation
- **Start and end date of the project:** 1/9/2021 – 31/8/2025
- **ID:** PID2020-116540RB-C22
- **Funder of the project:** Spanish Ministry of Science and Innovation
- **URL:** http://maps.igme.es/sarai/

Contributors

The main researchers of the project and responsible for the correct management of the data, are, **Oriol Monserrat Hernández** (oriol.monserrat@cttc.cat) y **Michele Crosetto** (michele.crosetto@cttc.cat) by the Centre Tecnològic de Telecomunicacions de Catalunya (CTTC) **Marta Béjar Pizarro** (m.bejar@igme.es) and **Carolina Guardiola Albert** (c.guardiola@igme.es) by the National Center of the CSIC Geological and Mining Institute of Spain (IGME-CSIC) .

1. Data summary

1.1. **Purpose of data collection and generation** [What is the purpose of the data collection/generation?]

The purpose, both of the data collection that is going to be carried out in this project, and of the generation of new data, is none other than that pursued by the SARAI project itself: to improve the capacities for exploiting the data of Earth deformation measured with Differential Interferometric SAR (DInSAR) and Persistent Scattering Interferometry (PSI).

1.2. **Relation of the data with the objectives of the project.** [What is the relation of the data to the objectives of the project?]

The relationship between the specific objectives of the SARAI project and the data collected and generated during the project, managed according to this DMP, is explained below:

- O1. Development of DInSAR and PSI techniques (responsible O. Monserrat). The objective is to improve the existing chain of the CTTC for the processing of this type of data. The data collected will be the pre-processed DInSAR and PSI data itself, which through different techniques developed by the CTTC will reduce the errors of the resulting DInSAR and PSI processing (generated data).
- O2. Exploitation of the results of DInSAR/PSI (responsible O. Monserrat). There is a need to improve automatic data analysis tools to exploit the DInSAR and PSI datasets mentioned in objective O1.
- O3. Exploitation of EGMS data (responsible M. Crosetto). The EGMS will provide European coverage of DInSAR derivative products. These data are not generated or compiled by the SARAI project, but will be used based on queries made by the EGMS, to develop tools that improve the interpretation of these data.
- O4. Generation of an InSAR deformation database (responsible M. Béjar). A database will be created containing classified ground motion (PS) measurement points with the associated deformation (mean velocity and time series) and a set of environmental covariates, with two main objectives: 1) Create an inventory of classified InSAR deformation based on all the processing results obtained by the IGME and CTTC teams in previous projects; 2) The database will be used as a training dataset to develop the artificial intelligence (AI) algorithms for O5. The number and type of covariates will evolve as the project progresses.
- O5. Development of tools based on artificial intelligence to classify active deformation (responsible C. Guardiola). Different machine learning tools will be used. They will be trained with the database created according to O4. The PS classified according to the developed algorithm will be new data generated in the project, which will be stored as maps, polygons or classified points based on the efficiency obtained,
- 06. Development of tools based on artificial intelligence to generate ground motion susceptibility maps (responsible M. Béjar) The probabilities resulting of the machine learning classifier obtained in the O5 objective, as well as the weights of the covariates studied, will be used to develop Susceptibility Indices for different geological risks at a national scale. The results will be susceptibility maps of different geological risks on a national scale.

- O7. Prediction of terrain deformation using AI (C. Guardiola). In certain pilot areas, future deformation time series will be generated with an uncertainty associated with the prediction.

1.3. **Format and origin of data generated / collected.** [What types and formats of data will the project generate/collect? What is the origin of the data?]

Table 1 shows the data formats, where we have tried to always use standard formats, since it ensures long-term data usability.

As described in the previous point, the O4 objective of the SARAI project is the construction of a database, for which a significant amount of data will be collected in Table 1. The evolution of the project activities will entail a modification in this Table 1, since there will be data that could be discarded if they have little relation to the measurements of ground movements, or their impossibility to use them. **Likewise, there may be data not considered in this initial phase of the project, which later proves to be necessary. For all these reasons, Table 1 should be reviewed in the next updates of the DMP**.

Tabla1. Data used (collected and generated) in the SARAI project.

Data	**Type/Format**	**Origin**
Pre-processed InSAR/PSI data	Satellite images: S1 images in SAFE format (image data in various binary formats, 'quicklooks' in PNG format, Google Earth overlays in KML format and HTML preview files, product metadata and other information in XML) DEMs (ArcInfo ASCII and GeoTiff format), orbits (ascii)	CTTC
IGME and CTTC InSAR results: *ID PS* *PS lon, lat* *PS mean_VLOS* *PS time series (dates and displacement)* *PS time series type (linear, etc)* *PS time series metadata (ini, end, #)* *PS deformation range* *Error, std, etc* *Spatial resolution* *Metadata (asc/desc, sat, incidence, azimuth, software, track)* Interferograms	Vector (shape, csv, ascii) / raster, Data base (Timescale / hyper table)	IGME's and CTTC's InSAR published results
Sentinel-1 SAR data	Satellite images, vector (shape, csv)	https://scihub.copernicus.eu/dhus/

Data	Type/Format	Origin
Paz SAR data	Satellite images, vector (shape, csv)	Paz data proposal (AO-0001-026) - IGME User License for Scientific Use
EGMS results	Vector (shape, csv, ascii), raster	https://land.copernicus.eu/pan-european/european-ground-motion-service
GEP/EPOS InSAR results	Vector (shape, csv)	https://geohazards-tep.eu/#!
Geology/Lithology	Vector (shape)	https://info.igme.es/cartografiadigital/geologica/Geode.aspx?language=es
Active faults nearby	Vector (shape)	http://info.igme.es/qafi/
MDE	ASCII matrix ESRI (.asc)	http://centrodedescargas.cnig.es/CentroDescargas/index.jsp
Slope	ASCII ESRI (.asc)	http://centrodedescargas.cnig.es/CentroDescargas/index.jsp
Land cover	Vector (shape)	Corine CLC 2018 https://land.copernicus.eu/pan-european/corine-land-cover/clc2018
Aquifer beneath the PS (Masa de Agua Subterránea)	Vector (shape)	https://www.miteco.gob.es/es/cartografia-y-sig/ide/descargas/agua/masas-de-agua-phc-2015-2021.aspx
IGME groundwater database	Data base (Access)	Water database from IGME https://info.igme.es/BDAguas/
MITECO groundwater database	Data base (Access)	Control network of the quantitative status of groundwater in the MITECO DB

Data	Type/Format	Origin
		https://www.miteco.gob.es/es/cartografia-y-sig/ide/descargas/agua/red-piezometrica.aspx
Spanish Land Movements database	Data base (Access)	Database MOVES from IGME https://info.igme.es/BD2DMoves/
Precipitation	Ascii (.txt)	Download service AEMET http://www.aemet.es/es/serviciosclimaticos/cambio_climat/datos_diarios
Spanish mining cadastre	Data base (Access)	https://geoportal.minetur.gob.es/CatastroMinero/BusquedaAlfanumerica
Industrial areas, Solid Urban Waste	Cartography (vector/shape)	Regional IDEE
Transport Network	Cartography (vector/shape)	http://centrodedescargas.cnig.es/CentroDescargas/index.jsp
New processed InSAR/PSI data	Vector (shape, csv, ascii), raster	New SARAI product
Classification of PS	Data base (Timescale / hyper table)	New SARAI product
Land movements susceptibility maps	Cartography (vector/shape	New SARAI product
Predicted land movements	Spatio temporal information (.txt/.csv)	New SARAI product
Scientific papers	.pdf	New SARAI product

1.4. **Data reuse** [Will you re-use any existing data and how?]

As mentioned in the previous sections, most of the data that will be used during this project comes from external sources to the SARAI project (see data origin in Table 1). All the data used correspond to public data or data resulting from previous projects of the participants in the SARAI project.

The reuse of all this data will lead to new products. In a simplified way, these new products are reflected in the last rows of Table 1. they are listed below:

1. New processed InSAR/PSI data (in vector/shape, csv, ascii, raster formats).
2. Database with temporal dimension that will collect for each PS point most of the information compiled in Table 1 (database format "nested" with time series).
3. Classification of PS performed automatically (database format "nested" with time series).
4. Land movements susceptibility maps (vector/shape FORMAT).
5. Predictions of ground movements in pilot zones (.txt/.csv format)
6. Scientific contributions (pdf format)

The data described in points 2 to 5 will be collected during the SARAI project and will be published at the end of it. The database referred to in point 1 is intended to be a database that will remain alive when the project ends.

1.5. **Expected data size** [What is the expected size of the data?]

We are currently evaluating the approximate volume that the data will occupy in MB / GB / TB. Two types of data volumes are expected: those derived from PS data that imply a large volume that can range from GB to TB (we are studying how to store them for efficient use of the productS), and the rest of the data that will not exceed 2GB. In the following updates of the DMP we will be able to give details about the proportions of raw data and processed data of the data that appear in Table 1, as well as other secondary outputs such as scientific articles. The implications of data volumes in terms of storage, access and preservation may also be included in future updates, in case additional costs need to be included.

1.6. **Impact and usefulness of the data** [To whom might it be useful (data utility)?]

The products of the SARAI project will detect areas susceptible to different geological hazards related to ground deformation. Consequently, the results of SARAI will support national, regional and local policies involved in risk mapping and management, allowing better preparation for climate change and stronger provision for the protection of society and the security of people against geological risks. Other end applications are geodesy and land management, urban and rural planning, cultural heritage protection, climate services, infrastructure development, mining and other natural resource extraction, dam and groundwater monitoring, insurance and litigation issues, structural and civil engineering, the real estate market and transportation. End users include geological surveys, road, rail and mining administrations, regulators and planners, authorities at European, national, regional and municipal levels, citizens of Copernicus participating states, SMEs and industry. The specific impacts expected and how they will be achieved are highlighted in Table 2. The innovation provided by the products of the SARAI project is aimed at multiple beneficiaries of science, the environment, society, technological progress and the economy/competitiveness (Table 3).

Table 2. Estimated Impacts (EI) of the SARAI project products and their relationship with the project objectives (O).

Estimated Impacts	¿How will the EI be achieved by SARAI products?

Improve capacities to assess and mitigate geological risks in urban areas, infrastructure and ecosystems (**EI1**).	- Updated displacement data with improved DInSAR and PSI processing and analysis tools: separating deformation signals from noise. (**O1**) - Automatic analysis tools to exploit DInSAR data and its relationship with the basic components of the spatial-temporal variability of ground motion. (**O2, O3, O5**) - New inventory of InSAR classified deformation based on all the processing results obtained by the IGME and CTTC teams in previous projects (**O4**)
Improve the prediction of geological hazards in urban areas and infrastructures. (**EI2**)	- Tools for prediction and early detection of ground deformation/geologic hazards in specific areas using artificial intelligence (**O7**) - Probability-based susceptibility maps for different geographic hazards at national scale (**O6**)
Improve civil infrastructure management strategies to support decision makers (**EI3**) Enhance resiliency and safety for geohazard related to ground deformation (**EI4**)	- Earth deformation maps that can show areas of potential geological risk. Classification of these signals according to the natural or anthropic causal phenomenon (**O5**) - Estimate the areas most likely to experience earth movements in the future (**O6**) - Short-term future time evolution of soil deformation (**O7**)
New modelling routines to automatically determine and predict ground motion from InSAR data (**EI5**)	- AI-based tools to classify active deformation (**O5**) - AI-based tools for ground motion prediction (**O7**)

Table 3 Intended beneficiaries in Science, Environment, Society, Technological progress, Economy/competitiveness. Observer promoting entities (EPO) that support the SARAI project.

Type	**Sector**	**EPO**	**Opportunities provided by SARAI products that are beyond the direct impact**
Science	University, research institutions (public) both nationally and internationally	- Uc3m-Santander Big Data Institute	Contribute to scientific knowledge of terrestrial movements characterization from satellite observations.
Society	Land planning (public)	- Spanish Geographic Institute (IGN)	Provide information on soil deformation that improves land use planning and lowers the cost for municipalities.

Type	Sector	EPO	Opportunities provided by SARAI products that are beyond the direct impact
Society	Geological risk management (public)	- Earth Observation and Geohazards Expert Group of the Geological Surveys of Europe	Improve plans dedicated to protection against natural hazards (for example, landslides) and anthropogenic hazards (for example, the impact of excavation work).
Environment	Environmental disaster risk management (public)	- Ministry for Ecological Transition and Demographic Challenge - WWF Adena	Locate ecosystems at risk due to overexploitation of water (which triggers land subsidence) or ground movements.
Technology and economic progress	Specialized companies on *remote sensing & InSAR* (private)	- Geokinesia (star-up) - Detektia (start-up) - Dares	Improve the tools used to analyse InSAR data.
Technology and economic progress	Digital technology companies (private)	- Singular People S.L.	Promote the development of digital technology companies (Big data, AI, Agile, Blockchain etc).
Technology and economic progress	Infrastructure managers (public/private)	- COMSA	Motivate entities in charge of civil infrastructure to use InSAR-based technology as a control tool
Technology and economic progress	Insurance consortiums and companies (public/private)	- Insurance Compensation Consortium	Activate the use of InSAR technology for insurance companies (private sector).

2. FAIR data

The data generated in the SARAI project follows the **FAIR principles (Findable, Accessible, Interoperable, Reusable),** which are a set of guiding principles to make research data easy to find, accessible, interoperable and reusable. (Wilkinson et al., 2016).

2.1. Findable Data

2.1.1. Ability to discover data generated by SARAI (provision of metadata and the standards for its creation) [Are the data produced and/or used in the project discoverable with metadata, identifiable and locatable by means of a standard identification mechanism (e.g. persistent and unique identifiers such as Digital Object Identifiers)? What metada will be created?]

The metadata will be produced trying to follow the INSPIRE TG2.0 format to the extent that the data allows, thus ensuring data discovery. Metadata will be provided for the following products listed in section 1.4:

- *Database with a temporal dimension that will collect for each PS point most of the information compiled in Table 1:* Description of each of the variables included in the aforementioned database, format and origin of the data collected. It will be necessary to study whether this database can form part of the institutional base of the IGME-CSIC.
- *PS classification performed automatically (database format "nested" with time series):* Description of the categories included in the classification and reference to the methodology.
- *Ground motion susceptibility maps:* description of the susceptibility values and reference of the methodology used**.**
- *Predictions of ground movements in pilot areas:* description of the variables included in the time series and its format.

The data stored in the Zenodo repository will include a readme file along with the different types of dataset, where the metadata will be included.

2.1.2. Persistent and unique identifiers such as digital object identifiers [What naming conventions do you follow?]

The main repository for the storage of open data will be ZENODO, so the digital object identifiers (DOI) will be used, which ZENODO generates for both journal articles, data sets or open source versions.

2.1.3. Nomenclature, keywords and version control [Will search keywords be provided that optimize possibilities for re-use? Do you provide clear version numbers?]

The nomenclature used and keywords to apply in search strategies will be developed in future updates of this DMP. The SARAI project has a duration of 4 years. As it is not a project with continuity, the version of the data will have a single version, this fact being able to be modified by possible location of errors or misprints, or by the existence of future projects that give continuity to some of its tasks.

2.2. Data accessibility [Accessible Data]

2.2.1. Data Availability [Which data produced and/or used in the project will be made openly available as the default? If certain datasets cannot be shared (or need to be shared under restrictions), explain why, clearly separating legal and contractual reasons from voluntary restrictions]

The products of the research promoted within the SARAI project and related to objectives O3, O4, O5, O6 and O7 will be open, as long as they have been validated and the scientific works related to the data and proposed as milestones have been published. The data produced by the SARAI project, which is related to objectives O1 and O2, will be hosted on the CTTC servers, due to the large size of the data.

2.2.2. Repositories where data will be stored [How will the data be made accessible (e.g. by deposition in a repository)?]

The data produced by the SARAI project related to the objectives of O3 to O7 will be stored in the ZENODO repository (https://zenodo.org/), while scientific publications will be stored in the Digital CSIC repository (https://open. igme.es/),

or the analog within the CSIC (https://digital.csic.es/). The data resulting from the O1 and O2 objectives will be stored in the CTTC services and will be accessible at their justified request.

Ideally, the generated database will be included in the institutional databases of the IGME. In this case, the database will also reside on the CSIC servers and the metadata in the IGME-CSIC metadata catalogue. The feasibility of publishing the database and the susceptibility map under a map service will also be evaluated.

2.2.3. Methods or software tools needed to access the data [What methods or software tools are needed to access the data?]

The software needed to access the data will be different depending on the format of the data:

1. *Database with temporal dimension that will collect for each PS point most of the information compiled in Table 1*: under study, an possible option is SQL
2. *Improved processing of InSAR/PSI data:* they will be defined in future versions of the DMP.
3. *PS classification performed automatically (database format "nested" with time series):* SQL
4. *Terrain movement susceptibility maps: Geographic information system (ArcGIS or Qgis)*
5. *Predictions of ground movements in pilot areas:* spreadsheet or text editor or SQL, format to be defined depending on the size of the data to be generated
6. *Scientific publications:* pdf reader

2.2.4. Location of deposit of data and associated metadata, documentation and codes. [Where will the data and associated metadata, documentation and code be deposited?]

In the aforementioned repositories, in the CTTC and in the CSIC. In addition, the feasibility of using IGME-CSIC servers and the IGME-CSIC metadata catalog will be studied.

2.2.5. Access to data with some type of restriction [If there are restrictions on use, how will access be provided? Is there a need for a data access committee? Are there well described conditions for access (i.e. a machine-readable license)? How will the identity of the person accessing the data be ascertained?]

Access to data not deposited in the repositories will be requested by email to those responsible for the project, who will evaluate and approve the access. The specific conditions of use of the data will be established and a commitment letter or contractual agreement will be signed.

2.3. Interoperable data:

2.3.1. Data and metadata vocabularies, standards or methodologies followed to facilitate interoperability [What data and metadata vocabularies, standards or methodologies will you follow to make your data interoperable?]

The concept of interoperability applies both at the data level and at the metadata level (see metadata format). The INSPIRE application rules on metadata will be

considered, which correspond to the ISO standards on metadata. Interoperability consists of the exchange and reuse of data between researchers, institutions, organizations, countries, etc. (use standard formats, and, whenever possible, access the data through open software). The data should be able to be combined and used with other data or tools. For this, the data format will be open and interpretable by different tools, including open software such as Qgis, R or Python (see section 2.2.3).

2.3.2. Standard vocabulary for all dataset data [Will you be using standard vocabularies for all data types present in your data set, to allow inter-disciplinary interoperability?]

In the next updates of the DMP, when progress has been made in the research activities, the need to define a standard vocabulary for all the data of the SARAI project can be evaluated. In any case, an attempt will be made to use the semantics of the data sets using the INSPIRE data specifications or those agreed internationally.

2.4. Data reuse

2.4.1. Type of license that will be used for the data generated in the SARAI project [How will the data be licensed to permit the widest re-use possible?]

The assigned license will try to be the least restrictive, but **will be defined in future DMP updates**. The type of licenses could be:

- In Creative Commons Licenses, CC BY (allow commercial use and allow modifications) will be used.
- In Open Data Commons Licenses (specific licenses for data), Public Domain Dedication and License (PDDL — "Public Domain for data/databases") will be used, which allows free sharing, modification and use of the work without restrictions.
- Licenses for computer code or software.

The data housed in the institutional databases or the IGME-CSIC servers will be accessible and reusable under the IGME-CSIC data license:: https://www.igme.es/Ayuda/CondiUso.htm.

2.4.2. Data embargo period

As mentioned in section 2.2.1. the data will be open once they have been validated and the scientific publications committed to in the different milestones of the project have been made. Except for the data of the objectives O1 and O2.

Scientific publications will be subject to the license conditions of each journal. Some of them have an embargo period, which depends on the publisher and the magazine. In any case, the versions allowed by these licenses will be stored in the repository mentioned in point 2.2.2.

2.4.3. Use by third parties [Are the data produced and/or used in the project useable by third parties, in particular after the end of the project?]

Once the SARAI project has ended, the use by third parties of the data created or used in the project is allowed.

The use of data and information from the SARAI project will be carried out by the users or reuse agents at their own risk and responsibility, and they will be exclusively liable to third parties for damages that may arise from it.

The CTTC, the IGME-CSIC, and those responsible for the SARAI project will not be responsible for any loss, claim or damage derived from the use or misuse of their information by the reusing agents, nor for the damages suffered or economic losses that, directly or indirectly, produce or may produce economic, material or data damage, caused by the use of the reused information or the decisions derived from its use or manipulation.

The CTTC, the IGME-CSIC, and those responsible for the SARAI project do not assume responsibility for any error or omission contained therein.

2.4.4. Data quality and organization [Are data quality assurance processes described?]

To ensure the quality of the data, they will be structured in different folders, whose names identify their content and version. The version will be assigned by a number, decimal or integer, preceded by the letter "v". This number will increase as the folders and data files are updated. The collected data will have to be stored in folders that are not modified to ensure its quality. The data generated will be validated in the pilot areas with on-site work, and in the areas where results from previous projects are available, comparing these results with those obtained in the SARAI project. Likewise, the mathematical tools referred to in section 1.2., where the relationship of the data with the project objectives are described, will allow estimations of uncertainty and errors of most of the products generated in the SARAI project.

2.4.5. Data Reuse Period [How long is it intended that the data remains re-usable?]

The period of time during which the data stored in the Zenodo repository or in the institutional databases or map services of the IGME-CSIC will remain reusable is indefinite. The data with restricted access will be subject to the conditions of the commitment agreements acquired.

2.4.6. Intellectual property

Ownership of the protectable or non-protectable results generated as a result of the execution and development of the SARAI project will correspond to the CTTC and/or the IGME-CSIC, depending on which part or in what proportion they have been obtained. To the extent that these results are subject to legal protection, the ownership of the Industrial Property Rights related to the inventions or other property titles that may derive from the aforementioned results will correspond to the institution(s) that have obtained them, having preference to request the corresponding protection titles, appearing as inventors or authors those researchers who have contributed intellectually to obtaining these results. In the event that both institutions have generated results jointly, both will be co-owners of those results.

The data is protected by the corresponding intellectual and industrial property rights and belongs to the CTTC and/or IGME-CSIC, and the corresponding citation indications must be respected. This citation must be made in one of the following ways: “Data origin: “©Geological and Mining Institute of Spain (IGME-CSIC)”, “Data origin: “©Centre Tecnològic Telecomunicacions Catalunya (CTTC)”, or both organisms. However, and in accordance with the general conditions for reuse in accordance with the ANNEX of RD 1495/2011, of October 24, implementing Law 37/2007 on the reuse of public sector information, the use, reproduction total or partial, distribution, public communication, reuse or any other similar or analogous activity of the data. The transformation and modification of the information in its

reuse may not imply any kind of alteration or distortion of its content. Certain databases or data sets may have specific conditions of use, which limit access and download of information in some cases and which will be included in the specific conditions of your reuse license.

3. Allocation of resources

3.1. **Human and financial resources to ensure that research data complies with FAIR principles** [What are the costs for making data FAIR in your project?]

The personnel designated to ensure the correct management of the data will be responsible for each one of the objectives of the SARAI project, the four IPs that may delegate the construction and storage of the data to other personnel of the research team and the work team, according to the schedule designated in the project report. In addition, it will be necessary to train personnel involved in the objective O4 Generation of an InSAR deformation database, based on TimescaleDB-type time series data, SQL type that is scalable for multidimensional time series data.

The personnel assigned to control the structure of the data and metadata are expert personnel in data management with the INSPIRE standard format (Information Systems area of the IGME-CSIC).

There is no direct budget allocated for data management, but there is an indirect budget, as technical experts are dedicated exclusively to ensuring that the research data complies with the FAIR principles. €15,000.0 has also been allocated for the development of a website that can be the access portal to all the products of the SARAI project, as well as a web viewer of the data. There is a hardware budget for data storage, consisting of:

- 1 desktop workstation (6.000,0€)
- 2 laptops (3.000,00 €)
- External hard drives (1.000 €)
- Data server (4.000 €)
- External hard drives (8.000 €)
- 3 high capacity data servers (12.000 €).

3.2. **Responsible for data management in the project** [Who will be responsible for data management in your project?]

The people designated for data management depending on the project objectives are:

- O1 y O2: Oriol Monserrat and Michele Crosetto.
- O4 y O6: Marta Béjar
- O5 y O7: Carolina Guardiola

These managers will ensure the correct collection of metadata, the reliability of the data, its security, as well as correct collaboration between the activities related to the different objectives.

3.3. **Long term preservation** [Are the resources for long term preservation discussed (costs and potential value, who decides and how what data will be kept and for how long)?]

There is no explicit budget in the SARAI project for the long-term preservation of the data generated, beyond the storage hardware described in section 3.1. Within the tasks

and budgets of the IGME-CSIC is the preservation of institutional databases. During this project, the feasibility of these data being part of aforementioned databases to guarantee their long-term preservation will be evaluated. If the result of this evaluation is positive, it will be requested through the established procedure that the data be part of an institutional database.

4. Storage and safety safeguard [Data security]

4.1. **Data recovery, safe storage** [What provisions are in place for data security (including data recovery as well as secure storage and transfer of sensitive data)?]

Data recovery will be guaranteed by making regular backup copies, preferably daily, on the external hard drives and the data storage server located at the IGME and the CTTC. The secure storage and long-term preservation of the data will be guaranteed by depositing the data in the repositories mentioned in section 2.2.2. The person responsible for the safeguarding and restoration of the data will be the person in charge of its management, or a person delegated by this person in charge.

The IGME-CSIC database and metadata catalogue are stored on servers that have an established maintenance and security plan, therefore they will adhere to that plan.

The management and control of access to collaborators to work with the data or data transfers while maintaining security, integrity and confidentiality (if applicable) will be managed and agreed upon by those responsible for the project.

5. Ethical aspects

5.1. **Ethical aspects and data protection** [Are there any ethical or legal issues that can have an impact on data sharing? These can also be discussed in the context of the ethics review. If relevant, include references to ethics deliverables and ethics chapter in the Description of the Action (DoA). Is informed consent for data sharing and long-term preservation included in questionnaires dealing with personal data?]

The ethical aspects of the use of artificial intelligence and Earth observation satellite data are today an open issue. In general, the SARAI project will guarantee that the use of data within the proposed activities is subject to international scientific standards of ethical conduct (see references at https://www.csic.es/es/el-csic/etica/etica-en-la-investigacion). In addition, the SARAI project will ensure that the application of artificial intelligence is in accordance with the Artificial Intelligence Ethics Guide published by the European Union in April 2019 (https://data.europa.eu/doi/10.2759/177365). More information on the ethics of applying artificial intelligence to satellite data can be found on the website of the "Artificial Intelligence for Earth Observation" project https://ai4eo.de/

In accordance with Organic Law 15/1999, of December 13, on the Protection of Personal Data (LOPD), the personal data of the members of the project, their collaborators, users of the data or the observer promoting entities will be incorporated to the contact lists of those responsible for the project, for the correct management of the project, the dissemination of the results and future scientific collaborations. The right of access, rectification, cancellation and opposition provided for the aforementioned Law may be exercised by sending an email to the email address of

those responsible for the project. The storage of the data must guarantee the security of the information and in no case will these data be transferred to third parties.

6. Others

6.1. Procedures, policies or references

This DMP has been developed from the template prepared by the University of Malaga, which in turn used the "DMPonline" tool (https://dmponline.dcc.ac.uk/plans Horizon 2020 DMP) as template.