Edition 2019

Promote Open Science within the DRIIHM Interdisciplinary Research Facility on Human-Environment Interactions: Co-design of an e-infrastructure implementing the FAIR Principles

Nom et prénom du responsable (coordinateur)	Emilie LERIGOLEUR - UMR 5602 Géographie de l'environnement (GEODE)
Organisme du coordinateur	Centre National de la Recherche Scientifique (CNRS) - Délégation Occitanie Ouest (N° SIRET : 180 089 013 00676) 16 avenue E. Belin - BP 24367 - 31055 Toulouse cedex 4 - France
Choix d'une des catégories	Outils - méthodes
Aide demandée	98,112.48 €
Coût complet	334,646.85 €
Durée	24 months
Mots clés de la proposition	e-infrastructure; co-design; interdisciplinarity; awareness; FAIR

Table of Contents

Project summary

As part of the international Open Science movement, this project deals with the structuration, sharing and opening of research data in the context of a network of thirteen CNRS-INEE observatories (OHMs): the Interdisciplinary Research Facility on Human-Environment Interactions supported by the ANR (LabEx DRIIHM 2012-2025). This network involves nearly 1000 scientists. Large amounts of heterogeneous data are produced or collected, covering research in Natural and Life Sciences, and Human and Social Sciences. Several initiatives have been carried out within the LabEx DRIIHM to increase visibility and data sharing to connect scientific teams that are not always linked, promote the re-use of data and potentially lead to the emergence of new research topics. A range of tools has been developed over the years: metadata geo-catalogs, web GIS platforms, photo libraries, HAL collections, etc. However, there are significant contrasts in the contribution to these tools and their use between OHMs, and two surveys conducted in 2017 and 2018 showed that researchers remain poorly informed about Open Science practices. The DRIIHM community is globally motivated by data sharing, but does not know how to proceed and identifies obstacles such as fear of hacking, misuse, security or loss of data ownership.

The objective of this project is to optimize the appropriation of Open Science by the DRIIHM community through: *i)* the organization of awareness campaigns showing the benefits of data sharing and openness; *ii)* the co-construction of a more ergonomic and interoperable e-infrastructure, integrating existing tools and accompanying researchers to find, share and (re-)use data through the concrete and gradual implementation of the FAIR principles. This project is part of international initiatives such as the *Research Data Alliance* and *GO FAIR*.

This Flash call offers the opportunity to strengthen the collaboration recently initiated with ergonomists and web development specialists to respond more closely to the community health needs and highly improve the existing infrastructure. The methodology is based on an iterative and incremental AGILE software development: the e-infrastructure will be enhanced with new features after each iteration. The originality of this project lies in the co-construction, researchers being involved at the early stages of the project. Challenges are to identify the current practices in data management and data access, and then to manage change with the integration of Open Science practices into the data lifecycle. Training workshops will enable the DRIIHM community to learn how to use and evaluate the new e-infrastructure efficiently. Indicators will be developed to measure the evolution of practices, the usability of the e-infrastructure and the level of data FAIRness.

The impacts of this project will be the directly quantifiable use of the e-infrastructure by the community of data producers and users. As a result, researchers will acquire a better knowledge of Open Science and their datasets will gain in visibility. They will be aware of data management using a Data Management Plan model and will have the opportunity to generate a data paper draft. The e-infrastructure source codes will be freely accessible and maintained on a dedicated repository platform. The scientific production will be open access and referenced in HAL. The methodological solutions developed and tested in this is project can be exploited beyond the LabEx DRIIHM. Finally, through the implementation of interoperability, research data will be visible in major national and international data infrastructures including the future *European Open Science Cloud* (EOSC) portal.

1. DESCRIPTION OF THE PROPOSAL

1.1. Proposal's context, state of the art and positioning within the International Open Science Dynamics

The emergence of Open Science since about fifteen years is changing the way in which scientific research is carried out, fostering much more collaborative works and new ways for the dissemination of knowledge. On one hand, funders and institutions formulate specific requirements for data management and sharing public data, but on the other hand the majority of researchers still do not know how or where to share their research data[1,2].

To help researchers, specific websites dedicated to Open Science emerged (*e.g.* DoRANum e-learning platform, "ouvrir la science" blog in France). The *Global Biodiversity Information Facility* has chosen to develop the *Integrated Publishing Toolkit* for generating a data paper manuscript. Another tactic to promote good practices is to tell believable horror stories based on real life examples: a lost backpack containing five years of research data, a fire in an office[3].

In their efforts to produce high quality data, researchers increasingly have to follow robust data management practices such as the machine-actionable FAIR principles in order to make their data Findable, Accessible, Interoperable, and Reusable[4]. These stipulate that digital objects should have unique identifiers, high-quality metadata, unambiguous licensing, adhere to data standards, etc.

In parallel, research funders are increasingly mandating Open Science practices. They require a *Data Management Plan* (DMP) as part of the proposal and evaluation process. All along the data lifecycle, a DMP helps researchers consider: when research is being designed and planned, how data will be managed during the project and shared afterwards with the wider research community. Online tools are available to help creating DMP (*e.g.* DMPonline, DMP OPIDoR). The encouragement by funders for researchers to provide open access data must be nonetheless tempered by the overriding legal requirements of the new General Data Protection Regulation compliance (*i.e.* for personal data).

[1] O'Caroll *et al.*, 2017. Providing researchers with the skills and competencies they need to practice Open Science. (10.2777/121253)

[2] Masuzzo *et al.*, 2017 (10.7287/peerj.preprints.2689v1)

[3] Todd & Morgan, 2017. Why horror stories don't lead to nightmares.

[4] Wilkinson *et al.*, 2016. The FAIR Guiding Principles for scientific data management and stewardship (10.1038/sdata.2016.18)

To overcome barriers to data access and reuse, many international and European top-down (*e.g. GO FAIR*, FAIRsharing[5]) and bottom-up initiatives (*e.g. Research Data Alliance* RDA) have appeared during the last decade, with the goal of producing recommendations and building the social and technical infrastructure to enable open sharing and re-use of data. *GO FAIR* particularly adopts the FAIRification Process scheme. The RDA-SHARC (SHAring Rewards & Credit) interest group gives recommendations to assess FAIRness[6]. The European Commission's H2020 program provides guidelines[7], supports the FAIRsFAIR project and proposes the establishment of a *European Open Science* Cloud (EOSC) to bring together existing and emerging data infrastructures. In this perspective, the *FAIR Data Expert Group* produced reports[8,9] with the broad range of changes required for the implementation of the FAIR data principles[10]. The ENVRI-FAIR project aims to enable data- and knowledge-based decisions connected to the EOSC-hub. Furthermore, it was recently developed a core set of FAIRness indicators measured by a semi-automated process, as well as a template to derive community-specific metrics evaluating FAIR aspects important to them[11].

With the National Plan for Open Science (2018), France is adopting an Open Science policy that extends and strengthens the French law "Loi pour une République numérique" (2016), encouraging the use of DMP models and FAIR principles among other measures. The French Government ensure also the sustainability of data produced by its research infrastructures. The French National Strategy on Research Infrastructures (nation roadmap 2018-2020) labels 99 data infrastructures: in the environmental domain, there are IR Système Terre, E-LTER-France RZA, E-LTER-France OZCAR. For social sciences, there are Huma-Num and Progedo.

1.2. The LabEx DRIIHM scientific community

The LabEx DRIIHM is an interdisciplinary research consortium focused on anthropogenically modified socio-ecosystems. Initiated in 2007 and structured in a LabEx consortium in 2012, the DRIIHM relies on a network of 13 Human-Environment Observatories (OHMs) in France and worldwide. Each of them are organized around a focal object or territory (*e.g.* a mining area in Tucson, USA; the highly anthropogenic Rhône river corridor in France; the desertification process in the Sahel) and a founding event which changed the evolutionary trajectory of the socio-ecosystem[12] (*e.g.* stop of the mining activity; major floods; implementation of the Great Green Wall in the Sahel).

The complexity of socio-ecosystems is understood through a broad disciplinary spectrum covering the Natural and Life Sciences and the Human and Social Sciences. 65 disciplinary fields are represented. In 6 years of activity (2012-2018), the DRIIHM has brought together more than 1000 scientists from 123 research units in 78 universities, 30 research organizations or french "Grandes Ecoles" from 13 different countries. The research carried out within the DRIIHM also has a transdisciplinary vocation, with close collaborations with territorial actors and knowledge transfer.

1.3. The DRIIHM data infrastructure: early achievements towards Open Science

To promote research works within and between OHMs, the LabEx DRIIHM launches an annual call for research proposals, using its online submit website. Each scientific partner signs the DRIIHM Charter and is committed to describe and provide all or part of its research data. These data are

[5] Sansone *et al.*, 2019. FAIRsharing as a community approach to standards, repositories and policies (10.1038/s41587-019-0080-8DO)

[6] David *et al.*, 2019. How to assess FAIRness to improve crediting and rewarding processes for data sharing?... (10.5281/zenodo.2625721)

[7] H2020 Programme Guidelines on FAIR Data Management in Horizon 2020 (2016)

[8] Turning FAIR into reality (Final report and action plan from the European Commission expert group on FAIR data - 2018)

[9] Simon *et al.* 2018. FAIR Data Action Plan: Interim recommendations and actions from the EC Expert Group on FAIR data

[10] Monst *et al.*, 2017. Cloudy, increasingly FAIR; Revisiting the FAIR Data guiding principles for the EOSC (10.3233/ISU-170824)

[11] Wilkinson *et al.*, 2018. A design framework and exemplar metrics for FAIRness. Sci. Data 5:180118 (doi: 10.1038/sdata.2018.118)

[12] Chenorkian, 2012. Les Observatoires Hommes-Milieux : un nouveau dispositif pour une approche intégrante des interactions environnements-sociétés et de leurs dynamiques. (doi.org/10.4000/soe.159)

essentially geospatial data (*e.g.* serial data, raster and vector datasets, sensor data), tabular data (*e.g.* observation data), textual resources (*e.g.* social surveys, transcripts), and iconographies/images.

In 2012, the Spatial Data Infrastructure (SDI) group was created to provide the necessary technical and human assistance for data stewardship in the OHMs. The group gathers engineers in geomatics, informatics and OHMs' scientific coordinators who propose and set up online tools for the census and dissemination of data[13,14]: metadata geo-catalogs, web Geographic Information Systems (GIS) platforms, photo libraries, an on-going DOI (*Digital Object Identifier*) attribution procedure for research data, OHMs web sites, etc. The use of standards and harvesting protocols allows interoperability, particularly for geospatial data in compliance with the INSPIRE European Directive.

Another initiative towards Open Science was recently implemented: now all publications (at least bibliographic references) related to research work (co)funded by the LabEx are deposited in the DRIIHM collection of the HAL open archive. This major commitment of LabEx researchers contributes to a better international visibility thanks to the many general search engines, notably *Google Scholar*, the *Bielefeld Academic search engine*, *DART-Europe* and other more disciplinary ones such as *Research Papers in Economics*, *PubMedCentral* or *Isidore*.

1.4. The DRIIHM data infrastructure status: strengths and gaps

Heterogeneous human resources between OHMs cause important contrasts in the contents and the maintenance of data sharing tools (*e.g.* geo-catalogs metatada completed from 5 to 90% depending on OHMs). Among the 445 metadata records of the Geo-DRIIHM mutual geo-catalog, only 5% provide a hyperlink for data download. In most OHMs, the online applications remain poorly used.

Table 1: SWOT matrix of the current DRIIHM data infrastructure

Strengths	Weaknesses
. SDI engineer group . Online platform for the DRIIHM annual research call . Online tools for heterogeneous data description and visualization . Use of metadata standards . HAL deposit procedure for research papers . DOI attribution procedure for research data	. Contrasted human resources between OHMs . Insufficient communication on the interest of data sharing . Researcher fears of data sharing: hacking, misuse, low security, loss of property, etc. . Non-ergonomic metadata capture tool . Very few open access data . Lack of quantitative indicators
Opportunities	**Threats**
. Increasing awareness with Open Science movement and laws . Increasing interdisciplinary and inter-OHMs research . RDA, GO FAIR and other large-scale initiatives . National Research Data Infrastructures . Future EOSC portal	. loss of data and risk of no research continuity . Lack of connections between research teams and lack of interdisciplinary research . Lack of international visibility . Increase the rate of not-shared data in the increasing data production context . Increase the misunderstanding between researchers and data managers who use different vocabularies

The SDI group conducted two surveys in order to better understand the uses and users of the data infrastructure (60 people surveyed in 2017) and, secondly, to verify the degree of acceptability of a future data deposit and sharing platform (80 people surveyed in 2018). Results showed that the community seems aware of the Open Science shift. However, several obstacles persist for a practice:

[13] Fantino *et al.*, 2013 - OGIIS : Etude de faisabilité d'un outil de gestion intégrée de l'information scientifique, 74 p.

[14] Trémélo *et al.*, 2018. Offre d'outils et de services communs pour la gestion et la valorisation de l'information scientifique des OHM.

fears of hacking, misuse of data, security, loss of property, etc. The complexity of some user interfaces, the vocabulary too technical, the lack of information and training also explain this lack of enthusiasm. These preliminary results substantiate the conclusions of a report produced by the *Education and Skills under Open Science*[1] working group (WG).

The SWOT analysis summarizes the strengths explained in 1.3 and the weaknesses cited above (**Table 1**). It also points out the opportunities in the Open Science recent context, but also the threats particularly important, delaying the interdisciplinary research integrated phase.

1.5. Main objectives

In order to overcome disciplinary, technical, methodological, and organisational barriers to move forwards Open Science in the DRIIHM consortium, this project is built on two main issues:
1- Promote Open Science and encourage Open Science practices, showing the challenges and benefits of structuring, accessing, sharing and opening research data
2- Enhance the existing infrastructure to a more user-friendly and interoperable e-infrastructure with a single access portal. This e-infrastructure will offer new functionalities that will be prioritized according to users' expectations and motivations:

- manage the datasets of each project using a DMP model by using/integrating DMP OPIDoR
- make these datasets FAIR (towards 5 Star Linked data[15])
- manage the datasets using a dashboard to view/update FAIR status, license, metadata, DOI, etc.
- generate a draft data paper from the enriched metadata
- more efficiently search for datasets in the e-infrastructure
- visualize the datasets in the form of interactive graphs and/or in a web GIS
- download the data

As research activities go increasingly online and produce vast amounts of data, e-infrastructures are keys to support Open Science. An e-infrastructure refers to a combination of digital technologies (hardware and software), resources (data, services, digital libraries), communications (protocols, access rights and networks), and the people and organisational structures needed to manage them[16].

According to this Flash call, the project deals with research data on a broader heterogenous scope: geographical, tabular, textual and iconographic data, and their associated metadata.

On the technical level, the interoperability of tools and data through the use of appropriate norms, standards, ontologies and vocabularies will be ensured. The ultimate objective is to be visible in EOSC portal, national and European infrastructures, reference disciplinary data warehouses, etc.

1.6. Working program, action plan and tasks

The definition and implementation of Open Science as well as its adoption and change management among the research and data stewarding community are constantly evolving. Thus, this proposal is based on the co-design[17] software-engineering process. The aim of this process is to manage dynamic and integrative view on tools and practices to elaborate expectations, requirements and project management. The design approach is the idea of co-evolution of tools, practices and actors. The development of a software system is not seen as an end in itself but as a means to foster novel practices which then pose new challenges and requirements for the tools used. As a consequence, it is necessary that practitioners and developers closely collaborate during all stages of the co-design process. **Figure 1** describes the SO-DRIIHM general architecture and mentions actors, practices, tools, data management and resources for improving the current solution and promote Open Science among the DRIIHM community.

[15] Hasnain & Rebholz-Schuhman, 2018. Assessing FAIR Data Principles Against the 5-Star Open Data Principles.
[16] Hologne, 2018. TDM, e-infrastructure et Open Science. DIST INRA, journée OMTD, 4 juin 2018
[17] Allert *et al.*, 2007. KP-LAB Knowledge Practices Laboratory... (hal-00593146)

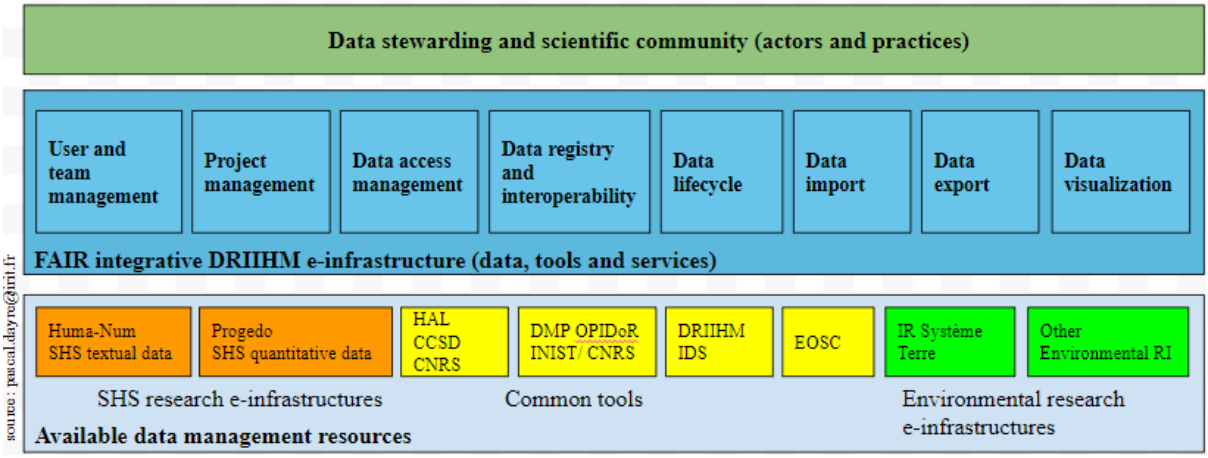

Figure 1: The SO-DRIIHM general architecture

The prime contractor itself will be provided by the *Rational Unified Process* (RUP) for building an e-infrastructure solution from requirements established by the co-design process. The RUP is piloted by use cases (risk management and usefulness), is architecture centric and iterative, and is an incremental method (roadmap of prototype). The RUP ensures the production of high-quality software that meets the needs of end-users, within a predictable schedule and budget[18] with tools like UML (*Unified Modeling Language*). It is well fitted for Object-Oriented and Web-Enabled systems.

According to the RUP, a preliminary phase then three iterations of six months will be organized. The first iteration will be the inception phase of the project. The second iteration will be an elaborate phase for building prototype 1 and general architecture. The third and fourth iterations will be construction phases for building prototypes 2 and 3, respectively.

The SO-DRIIHM working plan is divided into the simultaneous four tasks that fit the classical steps of an IT development process with, in addition, the integration of awareness campaigns and co-design workshops (see Gantt diagram **Figure 2**):

• **Task 0 "Project management and change management"** is a transversal task dedicated to project organisation, management and financial activities. The SO-DRIIHM Steering Committee (SC) will be in charge of monitoring the RUP, creating and updating a DMP (D0.1*n* deliverables), defining a Vision document (*i.e.* the core project's requirements, key features, and main constraints) and a project plan, showing phases and iterations (D0.2). A SC member will participate to targeted *RDA* or *GO FAIR* events. Meetings will be regularly organized. A consortium agreement concerning the private partner will be sent to the ANR agency as well as scientific and financial reports (D0.3*n*).

• **Task 1 "Promoting Open Science and change in research activities"** corresponds to the organization of an Open Science seminar with intervention of specialists (*e.g.* CNRS-INIST) and four co-design workshops. After making a deep analysis of the existing infrastructure (D1.0), the ergonomist approach will allow to scaffold end-users (*i.e.* project leaders, data managers, data producers and data users) and developers to co-design the transformed activity[19]. As a result, the state of requirements will be refined after each iteration (D1.1*n*).

• **Task 2 "Co-specification of the DRIIHM e-infrastructure and its intermediation"** is divided in a functional co-specification subtask (*i.e.* the reformulation of the needs expressed by the users in task 1 by the IT team with use cases and mockups) and a technical co-specification subtask (D2.1*n*). The latter includes an inventory of the open source technologies and resources like research infrastructures that will actually be used (see **Figure 1**), according to web standards and technological recommendations

[18] Gornik & Davor, 2001. Rational Unified Process - Best Practices for Software Development Teams.

[19] Dayre, 2012. Un outil d'aide à la transformation du travail. In: Congrès international de la Société d'Ergonomie de Langue Française

of international consortia (*W3C*, *Open Geospatial Consortium*,…) and laws (INSPIRE,…). The software architecture will thus be performed.

• **Task 3 "Implementation, test and deployment of the DRIIHM e-infrastructure and its web interfaces"** includes the classical steps of an IT development project: unit development followed by unit tests, then integration tests to progressively lead to the finished product (D3.1*n* and D3.2). At the end of the user testing phase, a questionnaire will be filled by testers to collect comments and suggestions for validating the usefulness and usability of the solutions developed.

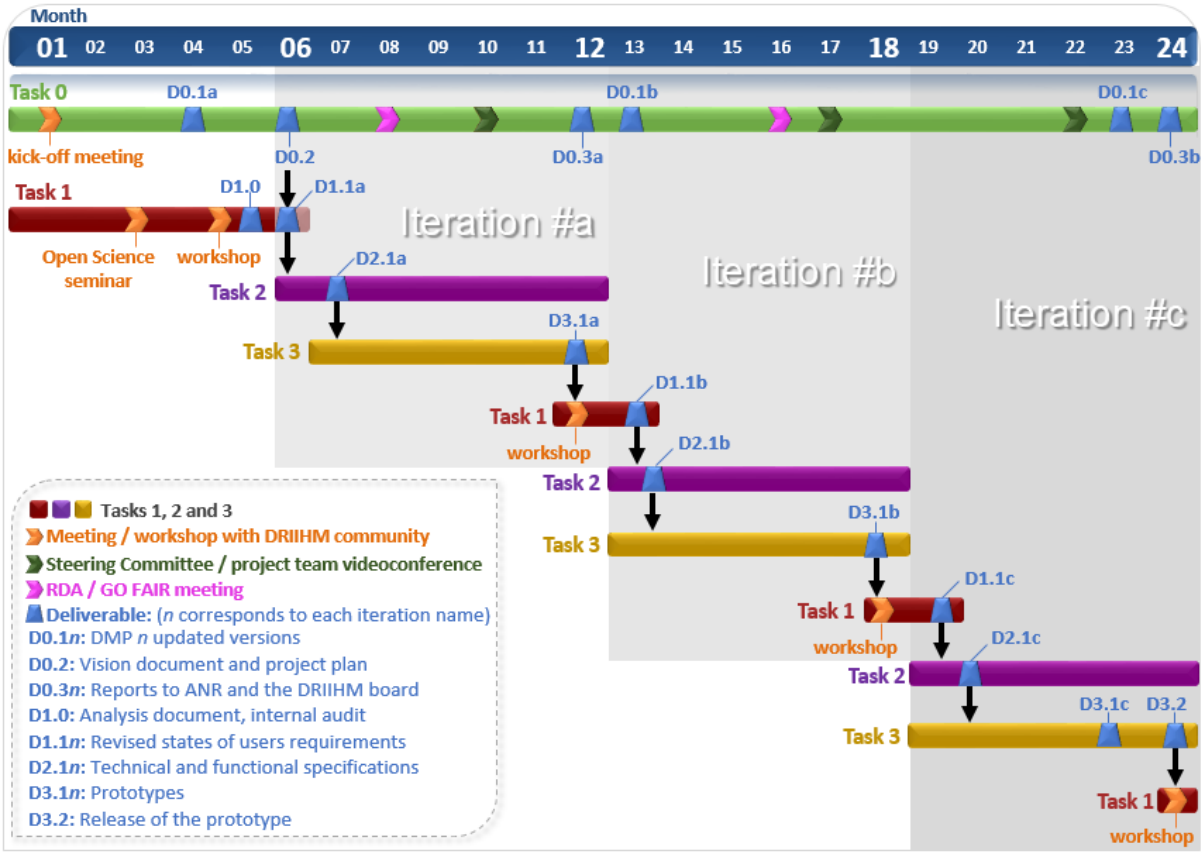

Figure 2: The SO-DRIIHM Gantt diagram, including the main deliverables per task

2. CONSORTIUM

This Flash call offers the opportunity to strengthen the collaboration between members of the SDI group (Partners 1-6) and ergonomists and web development specialists (Partners 7 and 8) to respond more closely to the community health needs and highly improve the existing DRIIHM infrastructure. Partners 1-6 are project owners, playing the role of data manager/producer/user. Partners 2 and 8 are involved in the e-infrastructure technical implementation. Partners 1 and 7 play a transversal role.

Partner 1 UMR 5602 Géographie de l'environnement (GEODE) - 24 pers.month / tasks 0, 1, 2, 3: GEODE laboratory focus on studies of environmental and landscape processes from the reconstruction of natural and socio-ecological processes in a long term duration to spatial modelling.

E. Lerigoleur is a GIS specialist from CNRS, in charge of the data stewardship of the "OHM Pyrénées". In April 2018 she submitted first to the SDI group then to the DRIIHM steering committee a proposal called "data2driihm" which aims to build a user-friendly infrastructure helping researchers to deposit and share their data. This Flash call is now the opportunity to achieve the "data2driihm" aim so far. She is the SO-DRIIHM coordinator, co-leader of task 0, leader of task 1 and member of task 2. **A non-**

permanent web developer will work closely with her and also with Partners 2 and 8 to develop technically the e-infrastructure.

Partner 2 FR 3098 Fédération de Recherche sur les Écosystèmes COntinentaux et Risques Environnementaux (ECCOREV) - 16.8 pers.month / tasks 0, 1, 2, 3: ECCOREV presents more than 30 laboratories with the aim of fostering interdisciplinary research in the environmental sciences. It is in charge of governance of the LabEx DRIIHM and one of its 13 observatories, the OHM "Bassin Minier de Provence" with two dedicated CNRS permanent staff (*i.e.* **C. Pardo** in tasks 0, 1, 3 and **JC. Raynal** in tasks 1, 3 respectively) and the **non-permanent computer engineer** (member of tasks 0, 1 and 3; co-leader of task 2) for the DRIIHM infrastructure.

Partner 3 UMR 5600 Environnement Ville Société (EVS) - 2.4 persons.month / tasks 1, 3: EVS brings together research communities from a wide range of disciplines in social sciences and humanities. Researchers belong to universities and colleges of the Lyon Saint-Etienne Pole. EVS provides scientific leadership in the OHM "Rhône Valley". **F. Arnaud** (CNRS GIS engineer) manages and enriches data sharing and visualization tools in this OHM. She will be involved in tasks 1 and 3.

Partner 4 UMR 7300 Etude des Structures, des Processus d'Adaptation et des Changements de l'Espace (ESPACE) - 2.4 pers.month / tasks 1, 3: ESPACE is a multi-thematic unit whose major fields of research are on one hand the space-nature-society interactions around environmental risks and urban systems, and on the other hand the theorization and formalization of spatial analysis. ESPACE leads the OHM "Mediterranean coastal". **ML. Trémélo** (CNRS engineer) will participate in tasks 1, 3.

Partner 5 UMR 6554 Littoral, Environnement, Géomatique, Télédétection (LETG) - 2.4 pers.month / tasks 1, 3: LETG focus on geography of the environment and the development of expertise in human geography, physical geography and geomatics. LETG co-leads the OHM "Port Caraïbe" and the development of its infrastructure, involved in several research programs dedicated to the geohistory of the Great Marine Port of Guadeloupe. As a GIS specialist, **I. Le Berre** (university lecturer of Université Bretagne Occidentale) will be also involved in tasks 1 and 3.

Partner 6 UMI 3189 Environnement, Santé, Sociétés (ESS) - 2.4 pers.month / tasks 1, 3: UMI 3189 ESS is an international laboratory, promoting shared scientific relations between North and South countries. **P. Duboz** (CNRS engineer) coordinates researches of the International OHM "Téssékére" leaned to ESS. She will be involved in tasks 1 and 3.

Partner 7 UMR 5505 Institut de Recherche en Informatique de Toulouse (IRIT) - 2.4 pers.month / tasks 0, 1, 2: IRIT is one of the main French research centers in Computer Science. A strategic research line deals with big data and its relation with artificial intelligence. **P. Dayre** (CNRS research engineer) is involved in various topics related to e-infrastructures for data management, data science and knowledge practices. He will co-lead task 0 and participate in tasks 1 and 2 for helping in the modelisation of research (changing) activities and in engineering the e-infrastructure.

Partner 8 Company Makina corpus - 13 pers.month / tasks 0, 1, 2, 3: This free software expert private company designs innovative business applications, mainly in the field of cartography and data analysis, ensuring independence, durability and technical quality. **A. Defays**, a UX and ergonomics researcher, has collaborated with E. Lerigoleur since 2018 to conduct the survey verifying the degree of acceptability of a future data deposit and sharing platform. She will co-lead task 2 and lead task 3. **S. Boureliou** is a web developer involved in tasks 2 and 3.

3. IMPACT OF THE PROPOSAL

3.1. Potential for using the results of the proposal

The target audience is the DRIIHM scientific community gathering more than 1000 scientists from 65 different disciplinary fields. In constant evolution, this community increases by an average of 135% per year due to the creation of 7 new observatories since 2012 and the intrinsic functioning of the Labex, which opens its annual calls for projects to the entire French and international scientific community (100 institutions in 13 countries). More than 200 projects (*i.e.* more than a third) have been

carried out by 102 foreign researchers with nearly 10 full-time equivalents worked. 43 post-doctoral fellows and 50 foreign doctoral students are also occasionally involved in these projects.

As any innovative, complex and interdisciplinary device, a maturation period is necessary to see results, and therefore data, emerge. The three most recent OHMs are currently in the initial phase. The 13 OHMs would be fully operational by 2021-2022.

With its new functionalities, the e-infrastructure will offer a directly quantifiable potential use by the community of data producers and users (see 5.). Through awareness of Open Science practices, researchers will gain skills in this field, usable beyond the DRIIHM sphere, and potentially exploitable within their disciplinary community. Particular attention will also be paid to respecting the transdisciplinary vocation of OHMs. The SDI group will therefore ensure that the user interfaces are adapted for consultation by the general public, territorial actors and donors.

3.2. Duplicability or replicability of results

The main interest in designing an open source e-infrastructure, directly accessible by the scientific community, is its possible re-use.

The e-infrastructure architecture will integrate the most generic software components, easily configurable for an application manager. It must also help to add, modify and delete web services provided by other organizations specialized in FAIR data management. This will be made possible by using standardized protocols and domain-specific ontologies known from the scientific community. The source codes produced during SO-DRIIHM will be free and maintained on a dedicated repository platform to encourage their re-use.

This approach, which is resolutely focused on redistribution, is perfectly in line with our strategy for the dissemination and sustainability of results (see 4.).

Finally, with the participation in the *RDA* and/or *GO FAIR WG*, our experience will be shared and we hope to come up with new recommendations that will be useful to all.

4. RESULTS DISSEMINATION STRATEGY AND PROCESS FOR SUSTAINING THE PROPOSAL AND/OR RESULTS

As requested by the ANR, "*all results (codes, documentation, data, and publications) will be published under a Creative Commons3 CC-BY license or equivalent and made available on a permanent storage infrastructure no later than the end of the project*". No exception to the principle of openness of our results has been identified.

Publications and communication media will be open access and referenced in HAL. Communications during national or international conferences or symposia will be scheduled. At least one publication in a "research data management" engineering journal will valorize the results obtained. A special issue in a scientific journal would compile the contributions of specialists invited to the seminar on Open Science. The webinars will be broadly distributed in the disciplinary mailing lists. A publication will also evaluate the Open Science awareness in the scientific community, providing feedback and enhancing the analysis of preliminary investigations.

Regarding the source codes, Partner 8 commits to make the code available in the long run by maintaining a version on a collaborative forge and migrating it to a new forge if changes in the conditions of use do not meet the requirements of openness. They will ensure that the published code is not tied up to a specific forge that could become a single point of failure. By using standard versioning tools, they will guarantee anybody can make a copy of the source code and publish it in different locations. This way, the code will always remain available for anybody interested in using it.

The e-infrastructure is itself a vector of dissemination of good practices in Open Science. The computer engineer funded by LabEx DRIIHM will ensure its maintenance until at least December 2024. The IDS group will continue its mission of raising awareness and training in Open Science practices beyond the project, while participating in *RDA* and/or *GO FAIR WG*.

5. SELF-EVALUATION PROCESS

Updating the SO-DRIIHM DMP and the Gantt diagram will help us to better organize and plan our working tasks. Moreover SWOT analyses will be regularly planned.

By following a continuous improvement approach, some Key Performance Indicators (KPIs) are defined to evaluate the effectiveness of the awareness process and the use of the new e-infrastructure. To compare and measure successful open data initiatives, the European Data Portal defines 3 categories of KPIs: data readiness, implementation and impact.

To evaluate the success of the project, it will be measure:

- For data producers:
  - the number of metadata records (total and per project)
  - the proportion of researchers who referenced / indexed their datasets
- For data users:
  - the number of visits (distinguishing new visitors from those who visit it again)
  - the connexion duration
  - the bounce rate (ratio between the total number of visits and the number of visits to a page)
  - the number of consulted metadata
  - the number of requests access to the data not the metadata (download or visualisation)

Finally, evaluation questionnaires will be filled after each training/seminar event and analysed.

6. REQUESTED BUDGET

Research within the OHMs network is funded by the LabEx DRIIHM ANR-11-LABX-0010 (2012-2025). The SO-DRIIHM project is approved by the LabEx governance and is financed entirely by this Flash call. However, it should be noted that LabEx DRIIHM finances a computer engineer who will be working 50% of his time on the SO-DRIIHM project. In addition, awareness campaigns and workshops will be held in conjunction with the annual LabEx-funded seminars, to reduce mission costs. The LabEx DRIIHM will also support the publication costs in an Open Science journal.

The full cost of the SO-DRIIHM proposal is **334,646.85 €** which can be divided in four parts:

- Self-funding from public organisms of **153,384.43 €**, corresponding to the 27.6 person.month permanent staff expenses (Partners 1 to 7)
- Self-funding from private company (Partner 8) of **38,848.77 €** (*i.e.* 55% of global cost)
- Co-funding from the LabEx DRIIHM of **44,301.17 €** (*i.e.* 12 person.month of Partner 2)
- Requested budget to the ANR of **98,112.48 €** detailed in **Table 2**. Partner 1 will defray the mission expenses of Partners 2 to 7. Partner 8 will receive funds from ANR to complete self-funding for co-financing 13.09 person.month and mission costs.

Table 2 : Requested budget to ANR (eligible costs) allocated to Partners 1 and 8

ANR eligible costs		Partner 1	Partner 8
Staff expenses	Non-permanent staff	12 pers.month 42,914.01 €	13.09 pers.month 27,940.36 €
Outsourcing	English proofreading costs for publications	1,500.00 €	0.00 €
General and admin. costs	Mission expenses and travel costs: - kick-off meeting - RDA/GOFAIR meetings - Open Science seminar - Complementary funding for workshops Environment costs	 3,000.00 € 3,000.00 € 5,000.00 € 6,000.00 € 4,913.12 €	 0.00 € 0.00 € 0.00 € 105.00 € 0.00 €
Total requested budget per partner		**66,327.13 €**	**31,785.35 €**

7. WORKING PLAN (see Gantt diagram **Figure 2** in part 1.6)