Journal of Interdisciplinary Methodologies and Issues in Science

Managing and sharing multidisciplinary information in human-environment observatories: feedbacks and recommendations from the French DRIIHM network

Fanny ARNAUD[1,*], Emilie LERIGOLEUR[2], Arnaud JEAN-CHARLES[3], Iwan LE BERRE[4], Corinne PARDO[3], Jean-Claude RAYNAL[3], Jérôme FOZZANI[5], Kristell MICHEL[1], Marie-Laure APERS-TREMELO[6], Dad ROUX-MICHOLLET[7]

[1] CNRS UMR 5600 EVS, ENS de Lyon, France
[2] CNRS UMR 5602 GEODE, Toulouse, France
[3] CNRS FR3098 ECCOREV, Aix-en-Provence, France
[4] CNRS UMR 6554 LETG, Brest, France
[5] CNRS USR 3456 LEEISA, Cayenne, Guyane française
[6] CNRS UMR 7300 ESPACE, Avignon, France
[7] GRAIE, Villeurbanne, France

*Corresponding author: fanny.arnaud@ens-lyon.fr

DOI : 10.18713/JIMIS-120620-6-3
Soumis le 26 septembre 2019 – Accepté le 21 janvier 2020
Volume : **6** – Année : **2020**
Titre du numéro : **Observatoires scientifiques Milieux / Sociétés, nouveaux enjeux**
Éditeurs : *Maud Loireau, André Miralles*

Abstract

The LabEx DRIIHM is a research network that gathers together 13 Human-Environment Observatories (OHM) focused on anthropogenically modified socio-ecosystems in France and worldwide. Within the open science context, a Research Data Infrastructure (RDI) was implemented brick by brick to describe, visualize, and disseminate multidisciplinary long-tail datasets produced by the DRIIHM community. The RDI contains both tools and standards at the network scale, and specific tools at the individual observatory scale. This paper describes the building blocks of the RDI and analyses its strengths and challenges on the basis of engineer feedback and user surveys. Recommendations for improving the RDI, better measuring its effectiveness, and enhancing open science awareness, have been formalized for the SO-DRIIHM project that will start in 2020. Cross-disciplinary approaches using DataViz tools have emerged already, and these enrich the way in which scientific information is disseminated, and could raise new scientific issues.

Keywords

Research Data Infrastructure; Human-Environment Observatories; multidisciplinarity; collaborative tools; FAIR data

I INTRODUCTION

Over the past few decades, long-term monitoring programs and observatories have been set up to understand the evolutionary trajectories of Earth's environments, assess the sensitivity of systems to natural and anthropogenic pressures, and to quantify levels of change in space and time. Having first focused on biophysical systems (e.g. the international LTER network: Parr *et al.*, 2003; the German TERENO network: Bogena *et al.*, 2016), long-term observatories then investigated social dimensions of biophysical change (e.g. the American HERO network: MacEachren *et al.*, 2006; the international LTSER platforms: Dick *et al.*, 2018). The aim of these information devices is to observe (monitor, analyse, understand) on a territorial entity, the socio-environmental dynamics resulting from the dynamic interactions of socio-economic and biophysical systems (Libourel *et al.*, 2009).

Observatories are great places for producing, exchanging and sharing information and knowledge on a long-term basis (Libourel *et al.*, 2009). However, the multidisciplinary and complex nature of socio-ecosystems makes it challenging to develop tools to help research teams to work together effectively and to transfer scientific knowledge to the general public and decision makers. Environmental and social scientists typically produce heterogeneous "long-tail" datasets, which may have been collected by numerous methodologies and may vary in subject, nature, size, format, and complexity, because of the culture and organisation of particular disciplines (Genova, 2018). Research Data Infrastructures (RDIs) aim to provide tools and services for the storage, visualization, and dissemination of datasets, particularly long-tail ones (Latif *et al.*, 2019). Such RDIs (also called SDIs in the case of spatial datasets) have been developed in environment-society observatories (Cinnirella *et al.*, 2012; Loireau *et al.*, 2014; Zwirowicz-Rutkowska and Michalik, 2016; Gourmelon *et al.*, 2017). RDIs are powerful, as they provide environmental managers with a wide range of decision-support tools. However, their implementation and maintenance require considerable organizational resources, and even political support for public service infrastructures (Le Berre *et al.*, 2013) or "networks of networks" research infrastructures (e.g. the OZCAR network: Gaillardet *et al.*, 2018).

The implementation of RDIs was pushed in Europe by the Aarhus Convention (1998) for environmental data and the European INSPIRE directive (2007) for spatial environmental data. With the National Plan for Open Science (2018), France is adopting an open science policy that extends the French law "Loi pour une République numérique" (2016). Furthermore, the worldwide open science movement has fostered several transnational initiatives such as the Research Data Alliance (RDA), the Global Sustainability Coalition for Open Science Services (SCOSS), and GO FAIR, a bottom-up stakeholder-driven initiative that aims to implement the FAIR data principles of making data "Findable, Accessible, Interoperable, and Reusable" (Wilkinson *et al.*, 2016).

The development of RDIs in an open science context increases the resources allocated to data management and data visibility, which can foster more interdisciplinary works. This raises organizational challenges: with a top-down approach, funders and institutions formulate specific requirements for data management (e.g. Data Management Plans [DMPs] required by the French ANR funder), but most researchers still do not know how to proceed and/or are reluctant to make their data available. There is also surprisingly little agreement on the metrics for measuring the use of RDIs (Georis-Creuseveau *et al.*, 2018), although fulfilment of the user requirements for data access and analysis should be the primary goal of RDI implementation with a bottom-up approach (Zwirowicz-Rutkowska and Michalik, 2016; Latif *et al.*, 2019).

This paper focuses on an RDI developed within the French DRIIHM (Device for Interdisciplinary Research on Human-Environment Interactions) network, a network of environment-society observatories. The objectives of this paper are to (i) describe the building blocks of the DRIIHM RDI; (ii) analyse the strengths and challenges of the current RDI on the basis of engineer feedback and user surveys; and (iii) make recommendations to increase the involvement of researchers in data supply, enhance open science awareness, and promote cross-disciplinary practices within the DRIIHM network.

II PRESENTATION OF THE DRIIHM NETWORK

Research on (bio)physical systems was structured in France into observatories in the 1980s with the creation of OSUs (Observatories of the Sciences of the Universe), and then OREs (Environmental Research Observatories) in the early 2000s, supported by INSU-CNRS (Loireau *et al.*, 2014). The OHMs (Human-Environment Observatories) are another observation system launched in 2007 and structured in the Laboratory of Excellence (LabEx) DRIIHM in 2012 as part of the “Investments for the Future” French program. Supported by INEE-CNRS, the OHMs have a more integrated territorial entry on societal and environmental issues. Their specificity is to be 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 that changed the evolutionary trajectory of the socio-ecosystem (a stop in mining activity; major floods; implementation of the Great Green Wall) (Chenorkian, 2012). Currently, the LabEx DRIIHM gathers together 13 OHMs in France and worldwide (Figure 1).

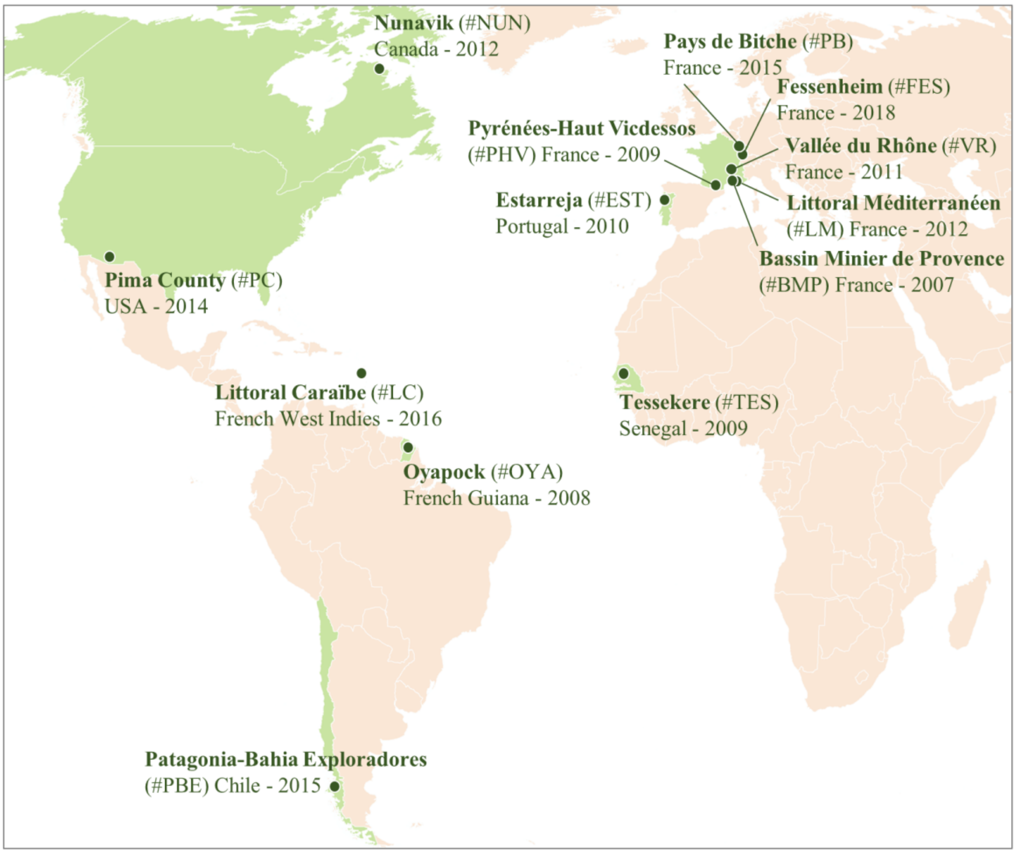

Figure 1: Locations and years of creation of the 13 OHMs. #xx is the OHM acronym.

The complexities of anthropogenically modified socio-ecosystems are investigated through a broad disciplinary spectrum covering natural and life sciences and human and social sciences. Between 2012 and 2018, the DRIIHM community involved more than 1000 scientists from 123 research labs in 78 universities, 30 French "Grandes Ecoles" and other research organizations, from a total of 13 different countries. The DRIIHM also has a transdisciplinary vocation, with close collaborations with territorial actors and knowledge transfer to the general public.

Research is based on annual calls for projects (bi-annual calls since January 2020) which are open to the DRIIHM scientific community and any external team interested in the research themes. A total of 748 projects have been conducted since 2012, including 492 projects funded by the French ANR and 256 foreign projects supported by the CNRS. Thirty-one projects were conducted with the involvement of two or more OHMs. The ANR also funded 12 PhD thesis and 15 post-doctoral fellowships.

III IMPLEMENTATION OF THE DRIIHM RDI

3.1 Overview

The Data-DRIIHM engineer group was created in 2012. Initially composed of 6 engineers in geomatics and data science from the three oldest OHMs, the group was founded to pool the existing tools, skills and increase the visibility of the DRIIHM network (Fantino *et al.*, 2013).

Journal of Interdisciplinary Methodologies and Issues in Science

The group started three technical tasks (metadata cataloguing, DataViz, photo gallery) and one dissemination task (knowledge diffusion and workshops), which have progressed more or less quickly over the years, given the engineers involved (Figure 2). The Data-DRIIHM group currently has 9 engineers located in different OHMs, 7 scientific coordinators, and a pilot web engineer who provides technical support for the OHMs that have fewer human resources available.

		2008 ….	2012	2013	2014	2015	2016	2017	2018	2019	2020
	Data-DRIIHM group		**Group creation 6 engineers**	**OGIIS study**	**Pilot web engineer**					**9 engineers 7 scientific coordinators**	**SO-DRIIHM project**
Technical tasks	Metadata cataloguing	Meta#VR	Meta#BMP	Meta#PHV		**Meta-DRIIHM** Meta#EST, #NUN, #OYA, #PB, #TES	Meta#PIM				
Technical tasks	Web GIS / DataViz		GeoOSR#VR	EXPRIHME #BMP		WebGIS#PHV		Cartoyamar #OYA	Timeline#VR	StoryMap#LC	
Technical tasks	Photo gallery		Photo#BMP Photo#PHV			Early interop. #PHV			**Photo-DRIIHM**		**Interop. Photo-DRIIHM**
Dissemination task	Knowledge diffusion & workshops (W)				**Website Submission platform**	**W-metadata**	**W-photo gallery**	**W-webGIS** W-metadata #TES	**W-websites W-photo gallery HAL collection**	**W-timeline**	**W-open science**

Figure 2: Temporal view of the Data-DRIIHM group's tasks over 2008-2020. Tools/actions in dark blue were developed at the DRIIHM scale. Tools/actions in light blue were developed at the OHM scale (#xx is the OHM name; see Figure 1 for acronym details).

The RDI was implemented brick by brick to support research workflow (Figure 3). The DRIIHM web site launches a project submission platform whose data model has been adapted from Fantino *et al.* (2013). Once the project has been accepted (steps 1-2; Figure 3), each scientific partner signs the DRIIHM charter and is committed to describe and provide all or part of the project's research data (step 3). The long-tail datasets can be very diverse. They include spatial data (airborne photographs, field surveys, GIS layers), tabular data, textual resources (interview transcripts, questionnaires), and iconographies (old postcards). The Data-DRIIHM group makes tools and services available to researchers at every stage of the research data lifecycle (step 4), which is the way to conceptualize the stages through which scientific data generally passes in a research process (Latif *et al.*, 2019). Tools at the DRIIHM scale aim to "glue" long-tail datasets and connect services (e.g. central metadata catalogue, photo gallery), and tools at the OHM scale were developed for specific needs (e.g. web GIS platforms, story map). Harvesting protocols make tools and services visible on national and European data portals (see the "(Inter)national scale" band in Figure 3).

The IN2P3-CNRS Computing Centre hosts the submission platform database as well as the DRIIHM tools. On the other hand, the storage of research data is the responsibility of each OHM based on institutional resources, e.g. the data centre of the Ecole Normale Supérieure de Lyon for the OHM Rhône Valley, the Toulouse Jean Jaurès University data centre for the OHM Pyrenees-Haut Vicdessos. The building blocks of the RDI are described below.

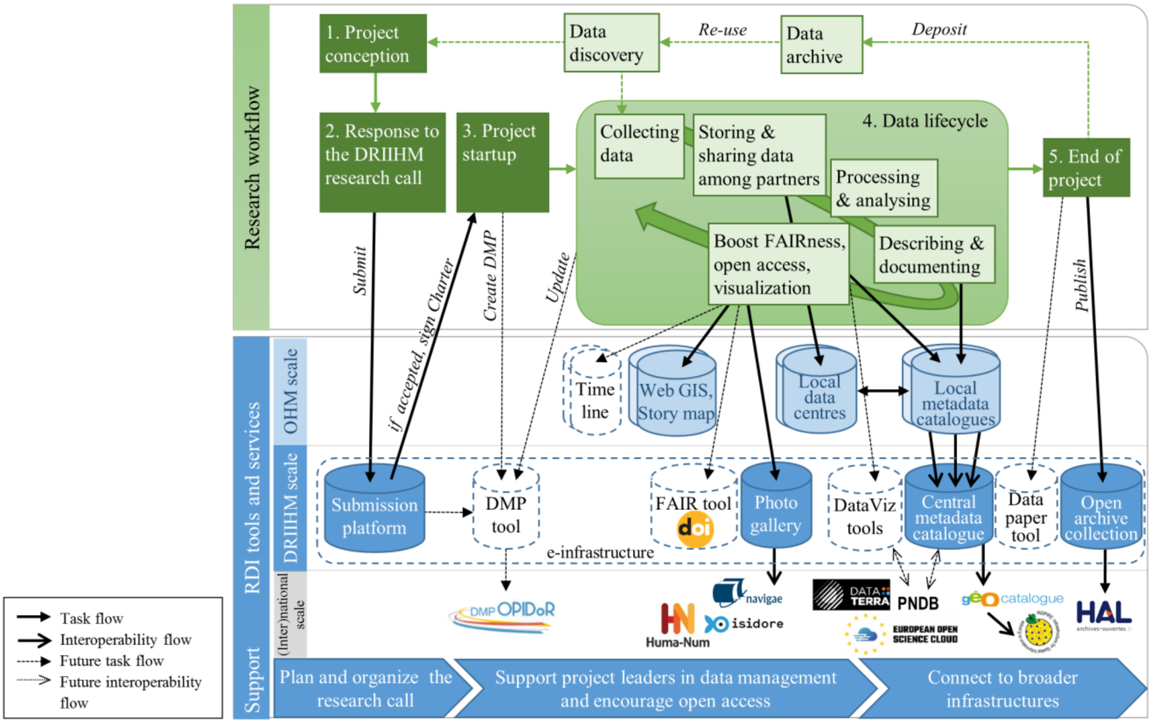

Figure 3: Integrative view of the RDI supporting the DRIIHM research workflow. Steps 1 to 4 extend over the duration of a research project, i.e. 1 to 2 years. Dashed arrows and boxes indicate tools under development (see Section V).
Adapted from: https://guides.library.oregonstate.edu/research-data-services/lifecycle

3.2 Data cataloguing

The Data-DRIIHM group has implemented a public central metadata catalogue to explore and access the DRIIHM data products. Metadata provide information on the data attributes, their units, precision, quality, spatio-temporal extent, the data authors, and the lineage describing how the data was measured, acquired, or computed. These are essential elements to ensure that the observatory data are compatible with the FAIR data principles and make them available as community resources (Horsburgh *et al.*, 2011). The DRIIHM metadata catalogue is based on the GeoNetwork application. It harvests each OHM catalogue and is then harvested in turn by the French geocatalogue and the European INSPIRE Geoportal, so that data discovery is possible at several hierarchical levels. Metadata entry is performed using either a spreadsheet file or online entry, based on the ISO 19115-19139 standard for spatial data in compliance with INSPIRE, and the ISO 15836-1:2017 standard (the Dublin Core metadata element set) for other kinds of data. The DRIIHM catalogue currently contains 495 open metadata.

3.3 Data visualization

Web GIS platforms make it easier to discover spatial data products in an interactive way for a given territory. Some platforms enable simple visualization and navigation, while others provide data analysis and downloading tools with either free access or access through login and password (Horsburgh *et al.*, 2011; Manley *et al.*, 2015; Gaillardet *et al.*, 2018). Some OHMs have set up web-based GIS applications to share data among scientific partners and/or present data products to territorial actors, developing their own interfaces based on available skills and resources. For example, the web GIS of the OHM Rhône Valley GeoOSR is based on Portal for ArcGIS (ESRI). It allows the user to display historical and contemporary data on the alluvial dynamics of the Rhône via four thematic maps, and to perform queries with analysis widgets. Cartoyamar is the GIS platform of the OHM Oyapock. It was developed with open source bricks (GeoServer, Openlayers, Cesium) including photogrammetry, but has not yet been launched online. The web GIS of the OHM Pyrenees-Haut Vicdessos, based on ArcGIS Server, provides to logged in researchers on the geolocation of monitoring instruments and sampling sites, with information on who is doing what, where, and how in the field (Galop and Lerigoleur, 2015). This OHM also implements a DataViz interface facilitating data import and export to and from a hydrological database on peatlands, including a search engine and preview tools (Rosset *et al.*, 2019).

The implementation of multimedia links considerably increases the richness of webmapping applications. Going beyond the simple geolocation of data, story map applications help to connect places and events and build a narrative of their evolution. They integrate various supporting items such as animations, videos, and cyclic or linear timelines at various time steps (Caquard and Dimitrovas, 2015; Cunty *et al.* 2015). Such a spatio-temporal representation is implemented by the OHM Littoral Caraïbe (Le Berre *et al.*, 2019; Figure 4). The story map is based on the Knightlab open source software, which is linked to a timeline and the DRIIHM photo gallery. The aim is to reconstitute the trajectory of the port complex of Guadeloupe in its socio-political, economic, and environmental context, and to identify the events that figured in its evolution.

Figure 4: Screenshot of the Story map of Pointe-à-Pitre port development in the 20th century.

A few years earlier, a desktop application called EXPRIHME was also developed by the OHM Bassin Minier de Provence to display maps, photos, videos, and text to disseminate interdisciplinary research results on human-environment interactions (Raynal *et al.*, 2013).

At the DRIIHM scale, engineers implemented a photo gallery to display field images and archive resources using the open source Piwigo software, with images being geolocated on an interactive map. Thanks to the Dublin Core metadata standard and the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH), the DRIIHM photo gallery is harvested by the CNRS HumaNum Isidore search engine for human and social sciences digital data, and the CNRS Navigae platform for photographic and cartographic data. The DRIIHM photo gallery currently contains 9303 public images from 6 OHMs, consisting mainly of old postcards, old maps, and photos taken regularly at the same location over different decades and showing landscape changes.

3.4 Knowledge diffusion

All scientific publications related to projects (co)funded by the LabEx DRIIHM are referenced in the HAL open archive, and some are also downloadable. A specific procedure was described to help authors deposit their publications in the HAL DRIIHM collection, which currently contains 909 documents. This major commitment enhances the international visibility of the DRIIHM network through Google Scholar, the Bielefeld Academic Search Engine, DART-Europe, and discipline-specific search engines such as Research Papers in Economics and PubMed Central.

IV STRENGTHS AND CHALLENGES OF THE CURRENT RDI

A SWOT analysis was conducted to evaluate the effectiveness of the current DRIIHM RDI and identify areas for improvement (Table 1).

a. Strengths (see 4.1)	b. Weaknesses (see 4.2)
• Data-DRIIHM engineer group • Submission platform for the DRIIHM annual research call • General and specific tools for long-tail dataset management • Use of metadata standards, interoperability with generic infrastructures at the national scale • HAL collection for scientific publications	• Different human resources cause heterogeneity in the RDI tools and contents • Insufficient communication on the benefits of sharing and opening data • Very little open access data • Lack of RDI evaluation indicators • Insufficient interoperability of some tools
c. Opportunities (see 5)	**d. Threats (see 4.2, 4.3)**
• Increasing open science awareness in compliance with RDA, EOSC, GO FAIR initiatives • Promoting open-access data papers • More user-friendly e-infrastructure,interoperable with (inter)national infrastructures and certified repositories • Procedure for assigning DOIs to datasets • Knowledge dissemination tools	• Isolation of some OHMs with fewer resources • Increasing rate of unshared data in the context of increasing data production • Lack of tools allowing cross-disciplinary works • Lack of international visibility of the RDI • Misunderstandings between researchers and data managers who use different vocabularies

Table 1: SWOT analysis of the current DRIIHM RDI.

4.1 Reliable technical support

The DRIIHM RDI is the result of collective efforts over several years. The challenge was to build upon the huge heterogeneity in data production and management practices that is commonly found in multidisciplinary observatories (Gaillardet *et al.*, 2018). At this stage, the

Data-DRIIHM group achieved several successes by developing both general and specific tools to assist researchers in their research workflow: Figure 2 shows the addition of software bricks of the individual OHMs through time, and the mutualisation efforts with the development of central tools. The content of the HAL collection corresponds to 90% of the DRIIHM publications over the period 2010-2019, which is another indicator of the RDI's success (Table 1a).

The interoperability of data, metadata, and services is a key factor for RDI effectiveness, particularly for spatial data in compliance with the INSPIRE directive. It relies on standards and harvesting protocols shared by the community, which can be challenging to implement in multidisciplinary networks (Manley *et al.*, 2015; Gaillardet *et al.*, 2018). The Data-DRIIHM group achieved early interoperability with its metadata catalogues and is completing interoperability with its photo gallery. If catalogues are fed regularly, they can become powerful communication tools. For example, the OHM Rhône Valley produces annual synthesis figures to illustrate the state of progress of metadata cataloguing (Figure 5). These syntheses are an integrated and playful view of the production of scientific data on the Rhône, speaking for both researchers and practitioners.

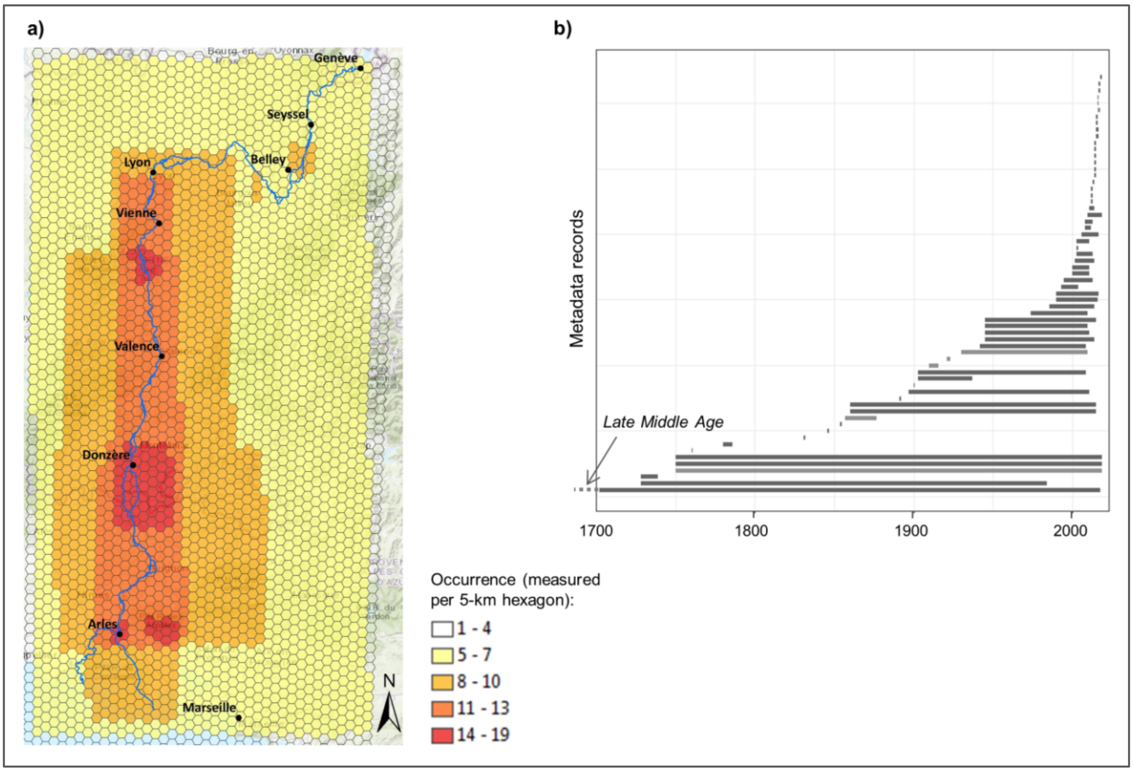

Figure 5: Examples of syntheses extracted from the Rhône metadata records for 2011 to 2018 (n=66): a) heat map showing the spatial extent of datasets; b) temporal extent of datasets.

4.2 Heterogeneities in the use of the RDI and difficulties in measuring its effectiveness

Continuous technical support is required to assist researchers with inputting data into the databases. However, most engineers work only part-time on OHMs, and some OHMs do not have any dedicated engineers, with the result that considerable variability exists in the contents and maintenance of tools (Table 1b). For example, metadata catalogues show completion rates from 5% to 90%, depending on the OHM, which makes the 'Findability' of datasets (from the 'F'AIR point of view) very heterogeneous. Seventy-five percent of metadata records have at least one hyperlink of any kind (e.g. the project website, a web GIS platform), however only 9% of metadata records provide a hyperlink for data download. Researchers may use other channels to share their data, such as data repositories (e.g. PANGAEA, Zenodo), supplementary data in journal articles, or provision of their contact details in metadata to provide data on request. These alternative data-availability approaches may lead to underestimation of the F'A'IR 'Accessibility' rate. With regard to the 'Interoperability' of FA'I'R, some web services of the DRIIHM RDI are not yet connected to each other and are thus under-used. However, web service standards provide the glue by which both machines and people can interact to form a greater networking system, accessible from any other machine in the DRIIHM network. The FAI'R' 'Re-use' rate is also difficult to estimate, which highlights the necessity to develop RDI evaluation indicators. This rate could be low, especially because the web applications are currently in French, which restricts the re-use of data by regional partners (Table 1d). One of the probable reasons for the low FAIR rates is insufficient communication on sharing and opening data within the DRIIHM community, which is corroborated by some of the results of the user surveys.

4.3 A need to increase the impact of the RDI on users

The Data-DRIIHM group conducted surveys to better understand the RDI uses and users (60 people surveyed in 2017), and to gather opinion on sharing and opening data through a potential new platform (80 people surveyed in 2018). Most respondents were senior researchers or engineers. Older OHMs and metropolitan French OHMs were more represented than others. The first survey revealed that a large proportion of respondents did not know what an RDI or SDI was, even though they knew at least one metadata catalogue or web GIS tool in their OHM. This vocabulary problem is indicative of the "valley of death" between e-infrastructure providers and domain specialists (Mons, 2016) (Table 1d). However, the second survey showed that the community seems to be aware of the open science shift. Researchers want to share their data (76%; Figure 6a), preferably after publishing related research articles or on request, but many do not know how to proceed (69%; Figure 6a). The main motivations for data sharing are ensuring the continuity of the scientific observatory (65% of correspondents), promoting new research questions and cross-disciplinary approaches (50% and 44% of correspondents respectively), and long-term data safeguarding (49% of correspondents; Figure 6b).

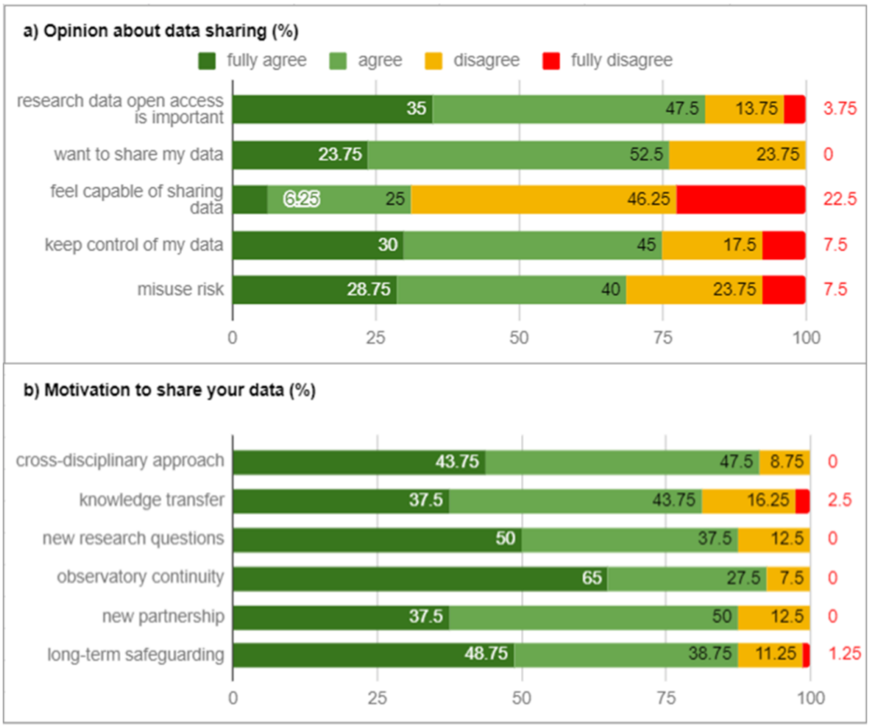

Figure 6: Main results of the 2018 survey (80 respondents).

However, several obstacles persist, including fears of hacking, misuse of data, security, and loss of property. The DRIIHM network is rather well advanced on metadata storage, but data storage in OHM data centres has difficulties to be done. There are still disciplinary disincentives to provide data to the common pot, before even opening data outwards. Most respondents ask for more skills and require support to begin and continue to practise open science. These results corroborate the conclusions of the Open Science Skills Working Group mandated by the European Commission (O'Caroll *et al.*, 2017).

The users questioned in the first survey proposed some action levers: learn how to specify data access and use conditions, be aware of the legal and regulatory environment, complete metadata to avoid data misuse, automate the metadata writing process, make the handling of RDI tools easier, ensure RDI sustainability, use different web tools for dissemination of research results, and increase the number/availability of RDI engineers. For the last question of the second survey "What would be the functionalities of an ideal RDI platform?", half of the respondents suggested "an ergonomic and easier to use platform". Other frequent keywords mentioned were "security", "interoperability", and "visualization tools". Six percent of people have a more accurate idea of the issue, whereas 18% of people did not respond to the question.

V RECOMMENDATIONS FOR IMPROVING THE RDI TOWARDS MORE OPEN SCIENCE AND CROSS-DISCIPLINARY PRACTICES

Opportunities identified in the SWOT analysis (Table 1c) have been formalized in the SO-DRIIHM project, which results from the ANR 2019 Flash call "Open science: research practices and open research data". This 2-year project will involve engineers of the Data-DRIIHM group, as well as volunteer researchers, ergonomists, and web development experts

(Toulouse Institute of Computer Science Research, Makina corpus). The objectives are to (Lerigoleur *et al.*, 2019):

- promote open science within the DRIIHM community, showing the benefits of structuring, accessing, sharing, and opening data (see 5.1);
- enhance the RDI to make it a more user-friendly and interoperable e-infrastructure (see 5.2).

5.1 Enhancing open science awareness and researcher involvement

When promoting open science, sociological aspects are more challenging than technical ones, which is why incentives for data sharing are a key issue (Genova, 2018). Workshops and webinars will be organized during the SO-DRIIHM project, with invited experts and presentations of success stories. Another tactic to promote good practices is to tell believable horror stories based on real life examples, e.g., a lost backpack containing five years of research data or a fire in an office (Todd and Morgan, 2017).

Opening data does not mean distributing everything to the entire world without delay. It is more about making data available to advance science while respecting the legal, ethical, and contractual frameworks of research projects (Boistel *et al.*, 2017). This principle is summarized by the European Commission's (2018) recommendation "as open as possible, as closed as necessary". Data should also be made as FAIR as possible by increasing the degree of "FAIRness" (Hodson *et al.*, 2018). Training and support on research data management, research integrity, ethics, and legal constraints are needed. It is also important to show how researchers can interact with the general public to enhance research impact. Last but not least, the members of the Data-DRIIHM group will have to train/upgrade themselves on open science practices in order to be exemplary, especially the scientific coordinators which are not data scientists.

The challenge of the SO-DRIIHM project is to promote the use of:

- a DMP model for each research project;
- digital unique identifiers for datasets (DOI, Handle) and data producers (ORCID);
- institutional and generic/thematic certified repositories for data storage;
- generic standards to describe core metadata, and as far as possible, controlled vocabularies and reference ontologies of the scientific fields, to optimize data re-use. The semantic web approach will particularly help making data machine-readable.

The writing of "data papers" will also be encouraged to reward metadata entry and data curation tasks. The popularity of open-access data journals (e.g. Data in Brief, Scientific Data) and citations to these journals has grown rapidly in recent years (Berghmans *et al.*, 2017), and studies have shown that the publishing of datasets has a positive effect on the bibliometric impact of published papers (Belter, 2014; SPARC, 2017).

5.2 Technical prospects for improving the RDI

A research-centric approach based on a new agile co-design method involving data engineers and researchers is planned to improve the existing RDI (Lerigoleur *et al.*, 2019). The future DRIIHM e-infrastructure should provide new functionalities under a single access portal, which will be prioritized according to user expectations and motivations (see dashed arrows and boxes in Figure 3):

- manage the datasets using a DMP model based on the DMP OPIDoR platform;

- automate as far as possible the metadata generation and publication, e.g. using the R script "geoflow" (DOI: 10.5281/zenodo.3138921);
- make the datasets more findable and accessible thanks to the interoperability of the future RDI with (inter)national infrastructures (e.g. Data Terra, National Biodiversity Data Pole, EOSC);
- use a dashboard to view/update the FAIRness status, metadata, DOI, license;
- more efficiently search for datasets produced by the DRIIHM community;
- visualize the datasets via interactive graphs and web GIS;
- download the datasets;
- generate a draft data paper.

The data storage strategy is not to develop a data centre at the DRIIHM scale, but rather to encourage the use of certified external repositories, for the main following reasons:

- avoid data duplication in cases where researchers already have to deposit their data into an institutional repository;
- minimize the volumetry of stored data;
- avoid the cumbersome repository certification process such as the CoreTrustSeal certification;
- be more realistic about resources allocated for research and the potentially unsustainable LabEx DRIIHM.

However, a specific data repository can be temporarily implemented to store orphan datasets, if no other solution is found.

By following a continuous improvement approach, key performance indicators (KPIs) will be proposed to evaluate the effectiveness of the awareness process and the use of the new e-infrastructure. For data producers, these KPIs would consist of the number of metadata records, the proportion of researchers who referenced and opened their datasets, and the level of FAIRness of the datasets, while for data users, they would consist of the number of visits on the e-infrastructure, connection duration, number of consulted metadata, number of data download, and number of data visualization. Finally, evaluation questionnaires will be filled and analysed after each training/seminar event.

5.3 Increasing cross-disciplinary approaches: chrono-systemic timelines

Multi-dimensional and comparative approaches are needed to investigate the complex interactions between present and past conditions of socio-ecosystems and their spatial dynamics (Le Tourneau *et al.*, 2013). The Data-DRIIHM group is working on such an approach based on a chrono-systemic timeline tool (Bergeret *et al.*, 2015). Participative workshops were held in 2018 and 2019 by the OHMs Rhône Valley and Pays de Bitche, gathering researchers and territorial actors to feed the timelines and specify the users' expectations. The Rhône timeline currently describes 281 significant events in the socio-ecological trajectory of the fluvial corridor, categorized into 26 classes (e.g. socio-economic context, hydro-climatic events, river restoration). Furthermore, researchers summarized the evolutionary trends of 8 research topics (e.g. biological patterns, morphological patterns, media discourse) transversal to several research projects conducted since the creation of the OHM. This timeline is considered a communication tool for river managers and a working tool for researchers, which will foster cross-disciplinary approaches and possibly raise new scientific issues in terms of causal relationships and drivers of change.

The challenge now lies in developing an interactive tool that allows the user to respond to a specific question by exploring the causal links between events. Promising solutions based on multiple coordinated-view geo-visualization interfaces have been identified (Arnaud *et al.*, 2018). In October 2019, a workshop brought together forty researchers from 9 OHMs, which confirmed the interest of the DRIIHM community in implementing a timeline application at the network scale.

VI CONCLUSION

In recent years, many open science initiatives have fostered dissemination of scientific knowledge worldwide, based on digital technologies and new collaborative tools. Science is becoming increasingly cooperative and transparent, with improved access to and re-usability of research products for laboratories, decision makers, innovative enterprises, and the general public. However, discipline-specific barriers still exist, and these cause challenges for the management of heterogeneous long-tail datasets across research communities.

In this paper, we have presented the architecture of a research data infrastructure that enables collection, description, visualization, and dissemination of long-tail datasets produced by a network of 13 environment-society observatories. Tools and standards have been developed at the network scale, as well as specific tools at the individual observatory scale. Permanent human assistance is critical for achieving a gradual cultural change in respect to researchers sharing their research data. Maintaining technical skills over the long term is thus needed in such scientific observatories. On the other hand, developing central tools and services partly counteracts the heterogeneity in human resources and moves towards greater data sharing.

The under-used RDI still lacks quantitative evaluation indicators, and interoperability with (inter)national infrastructures must be enhanced. This will be developed in the SO-DRIIHM project that will start in 2020 as well as with more education on open science practices. Feedbacks from the DRIIHM-RDI showed that the technical device increased its connectivity according to the FAIR principles and is now becoming a scientific stimulator on interdisciplinary issues: spatio-temporal approaches have already emerged (story maps, chrono-systemic timelines). These have enriched the way in which scientific information is spread out, and could raise new scientific issues on socio-ecosystem functioning.

References

Arnaud F., Roux-Michollet D., Antonio A., Barthélémy C., Carrel G., Comby E., Durey D., Franquet E., Graillot D., Grelot F., Honegger A., Lamouroux N., Lepage H., Marmonier P., Morardet S., Olivier J. M., Piégay H., Poirier C., Radakovitch O., Sivade E., Wichroff E. (2018, November). Croiser les disciplines et partager la connaissance produite dans un observatoire : élaboration d'une frise chrono-systémique pour l'OHM Vallée du Rhône. In *Spatial Analysis and GEOmatics conference*, Montpellier, France.

Belter C. W. (2014). Measuring the value of research data: a citation analysis of oceanographic data sets. *PLoS ONE* 9(3), e92590.

Bergeret A., George-Marcelpoil E., Delannoy J. J., Piazza-Morel D., Berthier-Foglar S., Bonnemains A., Bourdeau P., Duval M., François H., Girard S., Laforgue D., Lamarque P., Madelrieux S., Tolazzi S. (2015). *L'outil-frise : une expérimentation interdisciplinaire : Comment représenter des processus de changements en territoires de montagne ?* Grenoble : Les Carnets du LabEx ITEM. https://www.espacestemps.net/articles/loutil-frise-dispositif-detude-interdisciplinaire-du-changement-territorial/. Access: 2020-07-03.

Berghmans S., Cousijn H., Deakin G., Meijer I., Mulligan A., Plume A., de Rijcke S., Rushforth A., Tatum C., van Leeuwen T., Waltman L. (2017). *Open Data: The Researcher Perspective*. CWTS, Elsevier, Universiteit Leiden. https://www.elsevier.com/about/open-science/research-data/open-data-report. Access: 2020-07-03.

Bogena H., Borg E., Brauer A., Dietrich P., Hajnsek I., Heinrich I., Kiese R., Kunkel R., Kunstmann H., Merz B., Priesack E., Pütz T., Schmidt H. P., Wollschlaeger U., Vereecken H., Zacharias S. (2016). TERENO: German network of terrestrial environmental observatories. *Journal of Large-Scale Research Facilities* 2, 1–8.

Boistel R., Bordignon F., Du-Pasquier D., Masteau C., Saquet G., Renouf B. (2017). *As open as possible, as closed as necessary*. Paris : Pôle IST École des Ponts ParisTech. https://patrimoine.enpc.fr/exhibits/show/dataincognita/. Access: 2020-07-03.

Caquard S., Dimitrovas S. (2017). Story Maps & Co. Un état de l'art de la cartographie des récits sur Internet. *Mappemonde* 121, 1–31.

Chenorkian R. (2012). Les Observatoires Hommes-Milieux : un nouveau dispositif pour une approche intégrante des interactions environnements-sociétés et de leurs dynamiques. Sud-Ouest européen. *Revue géographique des Pyrénées et du Sud-Ouest* 33, 3–10.

Cinnirella S., March D., O'Higgins T., Murciano C., Sardà R., Albaigés J., Pirrone N. (2012). A multidisciplinary spatial data infrastructure for the Mediterranean to support implementation of the Marine Strategy Framework Directive. *International Journal of Spatial Data Infrastructures Research* 7, 323–351.

Cunty C., Mathian H., Groupe Cartomouv (2017). Les pratiques de cartographie animée pour représenter le changement. *Mappemonde* 120, 1–16.

Dick J., Orenstein D. E., Holzer J. M., Wohner C., Achard A. L., Andrews C., Avriel-Avni N., Beja P., Blond N.,Cabello J., Chen C., Díaz-Delgado R., Giannakis G. V., Giannakis G. V., Gingrich S., Izakovicova Z., Krauze K., Lamouroux N., Leca S., Melecis V., Miklós K., Mimikou M., Niedrist G., Piscart C., Postolache C., Psomas A., Santos-Reis M., Tappeiner U., Vanderbilt K., Van Ryckegem G. (2018). What is socio-ecological research delivering? A literature survey across 25 international LTSER platforms. *Science of the Total Environment* 622, 1225–1240.

European Commission (2018). *Commission recommendation (EU) 2018/790 of 25 April 2018 on access to and preservation of scientific information*. http://data.europa.eu/eli/reco/2018/790/oj. Access : 2020-07-03.

Fantino G., Michel K., Maignan M., Lerigoleur E. (2013). *OGIIS : Etude de faisabilité d'un outil de gestion intégrée de l'information scientifique*. Aix-en-Provence: LabEx DRIIHM.

Genova F. (2018). Data as a research infrastructure. CDS, the Virtual Observatory, astronomy, and beyond. *EPJ Web of Conferences* 186, 1–10.

Gaillardet J., Braud I., Hankard, F., Anquetin S., Bour O., Dorfliger N., de Dreuzy J., Galle S., Galy C., Gogo S., Gourcy L., Habets F., Laggoun F., Longuevergne L., Le Borgne T., Naaim-Bouvet F., Nord G., Simonneaux V., Six D., Tallec T., Valentin C., Abril G., Allemand P., Arènes A., Arfib B., Arnaud L., Arnaud N., Arnaud P., Audry S., Comte V. B., Batiot C., Battais A., Bellot H., Bernard E., Bertrand, C., Bessière H., Binet S., Bodin J., Bodin X., Boithias L., Bouchez J., Boudevillain B., Moussa I. B., Branger F., Braun J. J., Brunet P., Caceres B., Calmels D., Cappelaere B., Celle-Jeanton H., Chabaux F., Chalikakis K., Champollion C., Copard Y., Cotel C., Davy P., Deline P., Delrieu G., Demarty J., Dessert C., Dumont M., Emblanch C., Ezzahar J., Estèves M., Favier V., Faucheux M., Filizola N., Flammarion P., Floury P., Fovet O., Fournier M., Francez A. J., Gandois L., Gascuel C., Gayer E., Genthon C., Gérard M. F., Gilbert D., Gouttevin I., Grippa M., Gruau G., Jardani A., Jeanneau L., Join J. L., Jourde H., Karbou F., Labat D., Lagadeuc Y., Lajeunesse E., Lastennet R., Lavado W., Lawin E., Lebel T., Le Bouteiller C., Legout C., Lejeune Y., Le Meur E., Le Moigne N., Lions J., Lucas A., Malet J. P., Marais-Sicre C., Maréchal J. C., Marlin, C., Martin P., Martins J., Martinez J. M., Massei N., Mauclerc A., Mazzilli N., Molénat, J., Moreira-Turcq P., Mougin E., Morin S., Ngoupayou J. N., Panthou G., Peugeot C., Picard G., Pierret M. C., Porel G., Probst A., Probst J. L., Rabatel A., Raclot D., Ravanel L., Rejiba F., René P., Ribolzi O., Riotte J., Rivière A., Robain H., Ruiz L., Sanchez-Perez J. M., Santini W., Sauvage S., Schoeneich P., Seidel J. L., Sekhar M., Sengtaheuanghoung O., Silvera, N., Steinmann M., Soruco A., Tallec G., Thibert E., Lao D. V., Vincent C., Viville D., Wagnon P., Zitouna R. (2018). OZCAR: the French network of critical zone observatories. *Vadose Zone Journal* 17(1), 1–24.

Galop D., Lerigoleur E. (2015). *Rapport d'évaluation de l'Observatoire Homme-Milieu Pyrénées Haut Vicdessos pour le Conseil Scientifique du Labex DRIIHM*. Toulouse : GEODE.

Georis-Creuseveau J., Claramunt C., Gourmelon F., Pinaud B., David L. (2018). A diachronic perspective on the use of French Spatial Data Infrastructures. *Journal of Geographic Information System* 10, 344–361.

Gourmelon F., Rouan M., Nabucet J. (2017). Infrastructures de données géographiques et observatoires de recherche en environnement. Revue Internationale de Géomatique 27(3), 355–373.

Hodson S., Jones S., Collins S., Genova F., Harrower N., Laaksonen L., Mietchen D., Petrauskaité R., Wittenburg P. (2018). *Turning FAIR data into reality*. Interim report from the European Commission Expert Group on FAIR data. https://ec.europa.eu/info/sites/info/files/turning_fair_into_reality_1.pdf. Access: 2020-07-03.

Horsburgh J. S., Tarboton D. G., Maidment D. R., Zaslavsky I. (2011). Components of an environmental observatory information system. *Computers & Geosciences* 37(2), 207–218.

Latif A., Limani F., Tochtermann K. (2019). A generic Research Data Infrastructure for long tail research data management. *Data Science Journal* 18(17), 1–11.

Le Berre I., David L., Le Tixerant M., Defenouillère J., Nogues L. (2013). Infrastructure de données géographiques et gestion intégrée de la zone côtière. *Cybergeo: Revue européenne de géographie*, document-652.

Le Berre I., Ranély Vergé-Dépré C., Terral R., Rabévolo C. (2019). L'emprise portuaire du Grand Port Maritime de Guadeloupe, d'hier à aujourd'hui : apports de la cartographie dynamique géohistorique. *Norois* 250, 49–64.

Lerigoleur E., Arnaud F., Dayre P., Defays A., Duboz P., Jean-Charles A., Le Berre I., Pardo C., Raynal J. C., Trémélo M. L. (2019, October). SO-DRIIHM: promoting Open Science within the LabEx DRIIHM and improving its Research Data Infrastructure. In *APSEM Conference*, Toulouse, France.

Le Tourneau F. M., Marchand G., Greissing A., Nasuti S., Droulers M., Bursztyn M., Léna P., Dubreuil V. (2013). The DURAMAZ indicator system: a cross-disciplinary comparative tool for assessing ecological and social changes in the Amazon. *Philosophical Transactions of the Royal Society B: Biological Sciences* 368, 20120475.

Libourel T., Passouant M., Loireau M. (2009). Introduction du chapitre « Systèmes d'Information et Observatoires ». *Dossier Agropolis International* 9, 26–27.

Loireau M., Fargette M., Desconnets J. C., Mougenot I., Libourel T. (2014, November). Observatoire Scientifique en Appui à la Gestion du territoire (OSAGE) : entre espaces, temps, milieux, sociétés et informatique. In *Spatial Analysis and GEOmatics conference*, Grenoble, France.

MacEachren A. M., Pike W., Yu C., Brewer I., Gahegan M., Weave S. D., Yarnal B. (2006). Building a geocollaboratory: Supporting Human–Environment Regional Observatory (HERO) collaborative science activities. *Computers, Environment and Urban Systems* 30(2), 201–225.

Manley W. F., Gaylord A. G., Kassin A., Cody R., Copenhaver W. A., Dover M., Escarzaga S. M., Font R., García A. E., Haberman T., Lin D.D., Score R., Villarreal S., Tweedie C. E. (2015). The US Arctic Observing Viewer: A web-mapping application for enhancing environmental observation of the changing Arctic. *Arctic* 68(1), 100–110.

Mons B. (2016, February). Open Science as a social machine. In *European Open Science Cloud for Research Workshop*, Rome, Italy.

O'Caroll C., Kamerlin C. L., Brennan N., Hyllseth B., Kohl U., O'Neill G., Van Den Berg R. (2017). *Providing researchers with the skills and competencies they need to practise Open Science*. Open Science Skills Working Group Report, European Commission, Luxembourg. https://ec.europa.eu/research/openscience/pdf/os_skills_wgreport_final.pdf. Access: 2020-07-03.

Parr T. W., Sier A. R., Battarbee R. W., Mackay A., Burgess J. (2003). Detecting environmental change: science and society—perspectives on long-term research and monitoring in the 21st century. *Science of the Total Environment* 310(1-3), 1–8.

Raynal J. C., Rotereau A., Batteau P., Gachet S., Boutin N. (2013, October). Facteurs déterminants de la représentation de l'environnement chez les enfants avec approche spatialisée sur le territoire de l'Observatoire Hommes-Milieux Bassin Minier de Provence vu à travers l'application EXPRIHME. In *ESRI SIG Conference*, Versailles, France.

Rosset T., Binet S., Antoine J. M., Lerigoleur E., Rigal F., Gandois L. (2019). Drivers of seasonal and event scale DOC dynamics at the outlet of mountainous peatlands revealed by high frequency monitoring. *Biogeosciences Discussions*, 1–33.

SPARC (2017). *The Open Data citation advantage. Briefing paper*. https://sparceurope.org/open-data-citation-advantage/. Access: 2020-07-03.

Todd H., Morgan H. (2017, August). Why horror stories don't lead to nightmares. In *IFLA World Library and Information Congress*, Wrocław, Poland.

Wilkinson M. D., Dumontier M., Aalbersberg I. J., Appleton G., Axton M., Baak A., Blomberg N., Boiten J. W., da Silva Santos L. B., Bourne P. E., Bouwman J., Brookes A. J., Clark T., Crosas M., Dillo I., Dumon O., Edmunds S., Evelo C. T., Finkers R., Gonzalez-Beltran A., Gray A. J., Groth P., Goble C., Grethe J. S., Heringa J., 't Hoen P. A., Hooft R., Kuhn T., Kok R., Kok J., Lusher S. J., Martone M. E., Mons A., Packer A. L., Persson B., Rocca-Serra P., Roos M., van Schaik R., Sansone S. A., Schultes E., Sengstag T., Slater T., Strawn G., Swertz M. A., Thompson M., van der Lei J., van Mulligen E., Velterop J., Waagmeester A., Wittenburg P., Wolstencroft K., Zhao J., Mons B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* 3, 1–9.

Zwirowicz-Rutkowska A., Michalik A. (2016). The use of spatial data infrastructure in environmental management: an example from the spatial planning practice in Poland. *Environmental Management* 58(4), 619–635.

Acknowledgements
This research was funded by the LabEx DRIIHM, French program "Investissements d'Avenir" (ANR-11-LABX-0010) managed by the French National Research Agency (ANR). The first author works within the Zone Atelier Bassin du Rhône and the EUR H2O'Lyon (ANR-17-EURE-0018). We thank colleagues involved in the construction of the DRIIHM RDI since 2007: A. Antonio, H. Barcet, L. Chirol, P. Duboz, G. Fantino, G. Hinot, M. Maignan, C. Mouquet-Noppe, H. Parmentier, H. Piégay, S. Robert, A. Rotereau. We also thank P. Dayre and A. Defays for their involvement in the SO-DRIIHM project. Three anonymous reviewers made fruitful comments which substantially improved the quality of this manuscript.