RESEARCH PAPER

Fostering Data Sharing in Multidisciplinary Research Communities: A Case Study in the Geospatial Domain

Martina Zilioli, Simone Lanucara, Alessandro Oggioni, Cristiano Fugazza and Paola Carrara

National Research Council of Italy, Institute for Electromagnetic Sensing of the Environment, Milan, IT
Corresponding author: Martina Zilioli (zilioli.m@irea.cnr.it)

The sharing of research data allows for information reuse and knowledge advancement but its realization is often a challenge and seldom successful in practice. We propose a workflow for the design of a User Support System (USS) aimed at tutoring research groups in data sharing by considering their social and domain backgrounds. Our engagement approach focuses on multidisciplinary geospatial research, particularly when interoperable data sharing is required. Specifically, we first characterize the research community on the basis of the behavior and competences in data management by its groups and then target the needs of the latter with specific facilities. We address for the first time in literature the issue of modeling research groups as targets of the USS and provide a roadmap to standardize USS activities across different communities. We describe the implementation of the workflow in the context of an Italian research project and we assess the impact of the USS in terms of increase in the number of nodes and resources in the project's data infrastructure, and of fulfilment of the expectations by the research groups.

Keywords: geospatial data; user support systems; user profiles; communication facilities

Introduction

The digital revolution of the recent decades has created an explosion in the capacity to manage vast and complex data volumes, allowing research groups to share the groundwork of their publications. This is particularly important in the geospatial domain because sharing georeferenced data easily yield to (Lanucara et al. 2019): i) allows comparison between different areas of the globe; ii) enables new research, merging data from different domains; iii) provides easy readable maps to decision makers. However, even if stimulated by government mandates and journal policies, data sharing does not work as expected because many factors hamper this process. For example, although data infrastructures have been sufficiently established to ease open data practices, data protection and data quality are still open issues (Sayogo and Pardo 2013, Kervin et al. 2013), thus discouraging scientists to share their data. In addition, it seems that (inter)national, top-down mandates for data sharing do not necessarily commit researchers to planning data management at laboratory level (Haendel et al. 2012). Moreover, data sharing can be very difficult to achieve in multidisciplinary contexts due to providers' different practices (Lee et al. 2006). For instance, in the environmental sciences, different disciplines are required to interact: these are often characterized by gathering methods which vary individually (Karasti et al. 2006) or according to the community's common habits (Birney 2012).

In this paper, we present a case study in the geospatial domain, where the replacement of central production of geographic information with a network of producers creates a patchwork coverage of divergent practices. To normalize them, we propose a methodology for empowering multidisciplinary research community with data sharing skills and tools.

In modeling participation by research groups in data sharing, we distinguish between *involvement*, that is passive participation typically due to lack of appropriate supporting staff (Kervin et al. 2014), reward systems

(Arzberger et al. 2003), and data management training activities, and *engagement*, that is an active participation, supported by a data infrastructure where, beside technological solutions, also "soft interoperability" is achieved by coordinating social practices and organizational aspects of data sharing (Lee et al. 2017, Nedovic-Budic and Pinto 2001). To enable transition between these states, we propose a workflow to design a User Support System (USS) intended as "the assistance provided to the users of technology and other products" in the context of the e-Science infrastructure (Chunpir et al. 2014). This change in data sharing practices amounts to shifting from passive actions, which are not assumed to achieve the intended outputs, to active and motivated contributions that can effectively change scientific practices on the long-term as the difficulties of distinct research groups are considered separately by the USS.

In Section "Methodology", we describe the conceptual framework we used to deal with the challenges posed by multidisciplinary projects. Specifically, we pinpoint two key issues and cross-reference them with a selection of works in the state of the art which help to identify gaps and solutions which our approach has to develop (Section "Related work"). Then, in Section "Engagement approach application and impact assessment" we apply the workflow to project RITMARE, an Italian Flagship Project that aimed at bringing all contributions to Italian marine research under the same umbrella. With its broad and heterogeneous corpus of data providers and differing data management practices, RITMARE constituted the ideal use case for applying our methodology. We also provide the evaluation of the proposed USS by the community and describe its impact on the data infrastructure. Finally, in Section "Concluding remarks" we summarize the proposed approach and outline its limits and future outlook.

Methodology
Research context
Currently, governments, funding agencies, and other categories of stakeholders are enforcing data sharing practices (Borgman 2015); moreover, several journals are adopting policies to promote access to datasets (Savage & Vickers, 2009). The shifting nature of data sharing assumes a different inflection according to world region, age, work focus, and subject discipline of scientists (Tenopir et al. 2015). Consequently, the specific mechanisms of data sharing are difficult to be described as well as enforced in a unique way (Kim 2017). Particularly challenging are multidisciplinary projects where research groups belonging to different subject disciplines take part in a common process.

Since 2012, RITMARE–the Italian leading Marine Research project–involved this kind of community, composed of groups from the marine and maritime domains and belonging to eight disciplinary areas.[1] Enabling tools and a data policy were developed during the first three years of the project in order to foster data handling and distribution across domain boundaries. They were designed with the purpose of meeting the communities' sharing practices, on the basis of surveys about the needs for data preservation and reuse by the distinct groups.

These tools and guidelines contributed to creating the RITMARE Spatial Data Infrastructure (SDI) that allows for managing and sharing data, processes, and the information produced by the RITMARE subprojects. In particular, they comprise:

- A free and open source software suite for distribution of geospatial data on the Internet, named *Geoinformation Enabling Toolkit StarterKit ®* (GET-IT) (Fugazza et al. 2014). This suite allows participants to share their maps, observations, and documents through interoperable web services;
- a data policy (Basoni et al. 2015) described in a document that has been the basis for reaching an agreement among the participants on how to release data generated by RITMARE.

After three years, we found that the RITMARE SDI did not host the expected number of data resources. We identified the key issues that were curbing scientists' engagement in data sharing; specifically:

Key issue 1. The need for "*Motivating data sharing*" by those research groups that are neither technologically savvy nor enabled;
Key issue 2. the need for "*Normalizing heterogeneous practices*" of data sharing within the framework of the RITMARE SDI.

---

[1] http://sp7.irea.cnr.it/wiki/index.php/Aree_di_ricerca.

We investigated how previous works tackled these issues for designing and developing the implementation of our engagement approach.

Related Works

To ease selection of relevant works in the state of the art and pinpoint their different and common aspects, we identified five characteristics that helped us to outline every paper and summarize its contribution; specifically:

1. Target: The categories of professionals that are considered.
2. Method: The methodology the authors used to investigate the issue that is under consideration.
3. Barrier: What the authors consider as the cause of the issue that is discussed.
4. Drivers: The motivating factors that stimulate data sharing in the community under consideration.
5. Enabling actions: The possible actions authors identified for addressing the issue at stake and overcoming barriers.

Two works are related to the first key issue, i.e. "Motivating data sharing". The method adopted in both cases is a case study analysis.

Haendel (2012) describes Eagle-i, a pilot project facilitating resource sharing in biomedical research by creating a network of suitable repositories. Its target community is composed of scientists and the barrier is constituted by researchers' difficulty in organizing their data, scientific process, and resources in a structured way: This hampered sharing of data in a reusable format. As for the identified driver, the authors noticed that the scientists took particular pride in intellectual autonomy and in mastery of the techniques necessary to answer their scientific queries. Therefore, the suggested solutions were a data collection tool to capture data directly from the laboratory and assistance in biocuration in order to provide scientists with simple tools and supporting know-how.

Foster and Gibbons (2005) focused on facilitating adoption of an institutional repository by University faculty members. Here, the barrier is constituted by the so called "adoption problem" (that is, changing the used practice for implementing a new one) and is related to the apparent misalignment between the benefits and services of an institutional repository with respect to the actual needs and desires of faculty members. The driver consists of leveraging faculty members' and researchers' will to do their research, write about it, and share it with others in an easy way. The proposed solution is a user support structure to reach out faculty members more easily and to support them in person and online. This structure helps identify specific problems of faculty members to ease submission of research products to the institutional repository. The user support staff (named *library liaison*) interacts with faculty members according to the terminology they use and the needs they put forward.

The remaining papers refer to the second key issue, i.e. *"Normalizing heterogeneous practices"*. Like in Haendel (2012), in Parsons et al. (2011) the target is constituted by scientists and the method is a case study analysis. The paper discusses the experience of data scientists who managed data from the International Polar Year (IPY) 2007–8, when they collected a plethora of data formats to address interdisciplinary science goals. The barrier is constituted by the means of structuring, representing, and describing data varying across disciplines. The proposed solution is to improve interpersonal communication and collaboration between data creators and data managers.

The method adopted in Dallmeier-Tiessen (2014) is to develop a conceptual model describing the data sharing process. The target is constituted by policy makers, researchers, data managers, infrastructure service providers, and publishers. In this study, a series of key requirements are identified for the development of a mature data sharing culture. The barrier is made apparent by observing that, although infrastructures and standards may be well-developed within individual disciplines, interdisciplinary data sharing is scarce because these do not interoperate effectively with infrastructures and standards outside of the specific domain. The solution that is identified comprises the creation of an appropriate infrastructure and the training of researchers and data managers in adopting common standards.

Star and Ruhleder (1996) studied the organizational challenges that hampered adoption of software designed to support biologists' work on data. This was done through semi-structured interviews and observations by 25 labs with more than 100 biologists over three years, targeting the software designers (computer scientists) and the users (biologists). While the barrier was the communication gap between them, the driver consists in enabling the users to understand both the formal, computational level (traditionally, the

domain of the computer programmer and system analyst) and the informal level of the workplace language. The proposed solution is to create multi-disciplinary development teams (comprising both developers and users) to mitigate the "transcontextual syndrome" that limits streamlined communication due to their distinct knowledge backgrounds.

Volk et al. (2014) proposed a 21-questions survey to 175 natural resources scientists to identify the issues affecting distributed research teams, that is, teams which need to work closely but are separated by space, institution, and/or training. The identified barrier is the lack of training in data management, the scarce adoption of software tools, and the heterogeneous terminology; the driver is constituted by formalizing communication and data sharing processes. Specifically, the authors propose the following solutions: 1) to use data dictionaries, read-me files, protocols, and other metadata tools to structure metadata; 2) to code null values in different datasets; 3) to define roles and responsibilities of team members using organizational charts; 4) to define the overall process using data flow diagrams; 5) to define the rules for data transfer; 6) to define timelines to accomplish the distinct phases of the data life cycles.

Work discussion and lessons learned from the literature

Leaving aside technological aspects, not considered by the literature presented so far, the weight of the human component in software adoption in data e-infrastructures and the relevance of social dynamics among distinct professional figures are stressed by these works. We see that these issues emerge also in small-scale projects, because roles heterogeneity (see below for their categorization) is independent of enterprise dimensions. Moreover, heterogeneities can also be found in seemingly homogeneous groups, whose participants can be ascribed to the same role, because of different careers and experiences. Also, these social aspects are exacerbated by lack of data delivery guidelines, such as when compliance with a data policy is not strictly mandated or when an institutional directive is not provided. Finally, the considered projects typically establish hybrid work communities where people with different knowledge backgrounds interact to accomplish a common goal. In such communities (e.g. a research group in RITMARE) these people hold distinct roles such as that of 1) researchers, 2) staff figures supporting the former in data management, e.g. data wranglers (Parsons et al. 2004), computer scientists, and 3) staff figures enforcing the project or community data policy, e.g. data stewards.

The selected studies provide hints on how to understand and deal with the cultural divide between the persons holding these roles. They provide an exhaustive view on the barriers of interest to us and suggest how to overcome them by identifying appropriate tools.

Two main methods are suggested by these works: 1) Surveying the community of interest with questionnaires or interviews and/or 2) developing a conceptual model that, by cross-referencing the issues at stake with those described in existing literature, allows to develop a solution strategy. In both cases, the outputs are guidelines and working frameworks centered on the use case in hand, which fall short of identifying reusable solutions that can be adapted to other contexts.

However, they identified some barriers recurring also in our project: Haendel (2012) discusses the lack of data management literacy among scientists that is an issue we encountered in RITMARE and suggested us to couple the software we developed with (i) a training team, that may facilitate its uptake and (ii) a training materials, that can be autonomously consulted by a research team.

Foster and Gibbons (2005) suggest that providing a user support staff may allow us to better shape both the training team and material by orienting it to the community needs. In fact, the training material would be made easier to understand by a clear glossary helping scientists to interact more easily with the software that are provided. Consequently, we focused on formalizing communication flows, for instance by collecting in web pages the terminology the heterogeneous community shall refer to. To improve communication, Star and Ruhleder (1996) suggested us how to extend our terminology so as to avoid ambiguity of terms used by different categories of workers, while Parsons (2011) suggested us to envision additional figures specifically devoted to ease interaction between data manager and researcher's roles.

In general, although researchers' education is strongly recommended by the different authors, these papers only helped us assess the aspects and the roles to be taken into consideration for planning our training team and materials; the lack of a coding system to represent the heterogeneous community of roles to assist as well as to model their distinct needs is an aspect not addressed in these papers. Also, human operators are required in support systems where users are not only researchers but also data managers, and services supporting the roles-related activities of both. For this reason, support system should provide not only software applications to ease collection, archiving, and publishing of data, but should be also in charge of:

- Understanding the motivations that can stimulate active data sharing and the barriers to be overcome to accomplish this;
- fostering dialogue with researchers, in order to agree upon a realistic overview of their needs and align these with institutional mandates;
- planning training activities according to the different data handling skills (to compensate for the data curation gaps by the different researchers);
- mediating the interactions between data providers and data managers in order to define a collaborative way for sharing data.

All these aspects lead us to consider a USS as a valid helping platform to set up the engagement approach for our case study, allowing us to implement our idea of *engagement*. We extend the notion of USS (see Introduction) by conjunctive introducing for the first time in literature two new perspectives:

1. The USS will be addressed not only to researchers, but also to staff such as system administrators, data stewards, data managers, and developers. In fact, multidisciplinary science requires all these roles continuously to update and extend their practices. All these categories of users will coalesce in the notion of research group;
2. the USS design is based on a workflow modeling the research groups (abridged as "groups") of the geospatial research community under consideration (abridged as "community").

Developing the engagement approach: the workflow

The workflow allows planning the USS in a structured way. We assume that the community under consideration is a heterogeneous research environment characterized by the coexistence of distinct units, the groups, that in turn consist of a variable number of data providers and data management staff. We encountered this arrangement also in other projects (e.g., LTER) where we were previously involved, and we based on the observed geospatial management practices the following modeling. Also, we referred to the RITMARE gap analysis to enrich the groundwork documentation. Even if the groups belong to the same community where they share the common goal of enabling access to geographic information through an SDI, categorizing them is needed to tailor the *facilities*, i.e. communication tools and services that are provided through e-learning platform (educational) or team assistance (supporting), to address their specific needs.

For each group, we enact the two phases in **Figure 1** in order to obtain:

1. A profile for the group that is investigated, modelling its geospatial data management activities;
2. The characterization of the facilities matching the needs of the group.

Phase 1 – Profiling the groups

Phase 1 aims to obtain a concise profile for each specific group by modeling it with respect to two aspects we name *behavior* and *competence* in the domain of geospatial data management.

The *behavior* of the group is formalized on the basis of the features summarized in **Table 1**.

Each feature can either be present or absent inside a group, thus it can be represented as a binary value (1 or 0); as a consequence, each group can be associated with a triple representing its behavior. Only five among the eight possible triples are considered. In the community, F3 implies F2 (i.e., data need to be web-distributed in order to be accessed through standard web geoservices) and F3 implies F1 (i.e., data need to be georeferenced in order to be accessed through standard web geoservices). Thus, the combinations corresponding to triples (0,1,1), (1,0,1), and (0,0,1) are not possible.

We associate each of the five remaining triples to a *behavior category* (BC) representing the data management behavior of the group:

- *Committed*: Identifies the groups that produce spatial data during their research activities and distribute them using interoperable standard web geoservices (triple 1,1,1).
- *Competent*: Identifies the groups that produce spatial data and share them over the Web but not via interoperable standards (triple 1,1,0).
- *Web-ready*: Identifies the groups that produce spatial data but do not have an appropriate public infrastructure or repository for distributing them over the Web (triple 1,0,0).
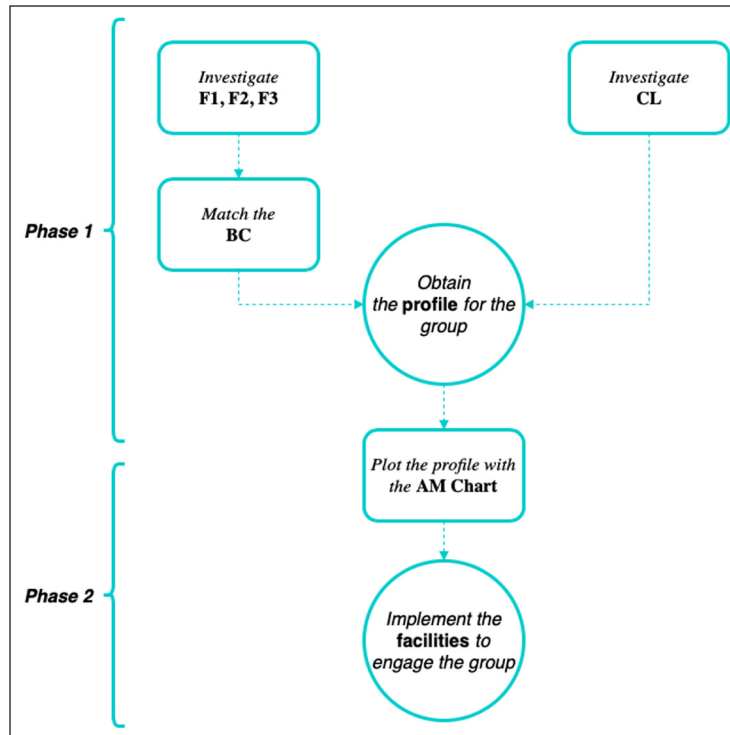
Figure 1: The workflow for the engagement of a group.

Table 1: Data management features.

		Features
F1	*Spatial data production*	Feature 1 describes if the group associates its data to a geospatial coordinate system. This is necessary to refer data to a location or a geographical area.
F2	*Web-based data distribution*	Feature 2 describes if the group provides access to data through web-based tools such as online databases, repositories, geoportals, or desktop applications.
F3	*Interoperable web-based data distribution*	Feature 3 describes if the group provides interoperable access to data through standard web geoservices.

- *Receptive*: Identifies the groups that produce data that are not necessarily georeferenced or geotagged and that distribute them over the Web, e.g. via online databases or public archives (triple 0,1,0).
- *Locked*: Identifies the groups that produce data that are not necessarily georeferenced and that are not shared over the Web. It is the case of groups collecting data in physical or digital archives and that do not publish them via a web-based system (triple 0,0,0).

The second aspect we modeled is the *competence level* (CL) of the group, which describes its knowledge background: it depends on the thematic area the group belongs to and the expertise that are shared among its members. The latter is measured by the competence in terms of the skills, education, and domain practices that are necessary to effective data sharing.

The competence is not conceived to be static: it is a descriptor of the technological capacity of a group that can be improved in time. Informative actions can both upgrade the group from one level to another by delivering contents for training through the facilities, and overcome the gaps of the group expertise. The levels and informative actions are described in **Table 2**.

We suggest that similar competences regarding maturity of data policy practices for a group should be considered in order to assess the global competence of a group, w.r.t. data management. In fact, maturity of the data policy practices in place may slow down or accelerate the data sharing through the SDI. Moreover, this maturity is independent of the technological maturity of a group, whereas a correlation may exist (e.g. when complying with a mandatory data policy stimulates consolidation of the data management skill set).

Table 2: Data management competence levels and informative actions.

Level	Competence	Description
0	*Capturing competence*	This level describes groups that have basic expertise in describing the phenomena that are their research objects and in acquiring and preparing data according to the practices of the specific domain or lab.
1	*Digital competence*	This level describes groups that have the necessary know-how and tools to digitally manage geospatial data. The informative actions to reach this level provides contents on how to collect georeferenced data (layer, observations or measurements) in digital formats, organize them in digital systems (spatial databases, relational databases, file systems), and store them in different types of repositories (institutional servers, projects repositories, desktop PCs, cloud systems).
2	*Interoperability competence*	This level describes groups that have the know-how and tools to share geospatial data by using standard web geoservices. The informative actions to reach this level provides contents on how to make datasets distribution interoperable with the datasets produced by other groups (e.g., by using an SDI).

As a consequence, we postpone identification and description of the competences associated with the data policy to further studies in order to address this aspect with specific facilities.

We also propose to assess behavior and competences independently by using focus groups, questionnaires, or interviews to investigate with different questions or in separate surveys the staff composition, skills, and their disciplinary domains. These two aspects are considered separately since the presence of competences inside the group doesn't automatically imply a congruent behavior. For example, as will be shown later, a group could be competent in digitally preparing datasets but there could be no apparent activity for data sharing over the Web.

The evaluation of both behavior and competence allows us to create the group profile, as the competence depends on the background of individual staff members and can vary independently of the behavior expressed by the group as a whole.

Phase 2 – Planning the User Support System

Once a profile is assigned to every group, our approach identifies how to support data management activities by selecting the set of facilities that will constitute the USS.

The selected facilities can be divided in two classes. The first is the *educational* class, which encompasses tools that are also used by other authors and initiatives, such as glossaries (Volk et al. 2014), FAQs, video tutorials, guidance materials and how-to documents (Parsons et al. 2005), that allow the groups to receive directions without interacting with the external mediators of the USS. The second *supporting* class encompasses webinars (DataOne, Digital Curation Centre), on-demand activities, mail contact and consulting (Chunpir et al. 2014), on-site meetings, which allow the group to receive direct support by external mediators. These facilities, which require professional skills in order to be effective, are:

1. Glossary, which provides fundamentals and basic knowledge about a particular topic; it uniquely describes terms that could not be understood by groups that, albeit in the same project community, belong to different thematic domains. It provides a shared vocabulary the research community can refer to, and research members can autonomously use this tool without the need for direct assistance by mediators.
2. Frequently Asked Questions (FAQ), which provide answers to recurring questions on the project's topics of interest or on how to accomplish the data management goals.
3. Video tutorials and how-to documents, which provide assistance to groups members for accomplishing the technical activities that are required for sharing data;
4. Webinars, which are conceived for users that require specific or deeper formation about the project's topics of interest, fostering discussion among different behavior categories;
5. On-demand assistance through online services (the helpdesk unit), which is a team collecting help requests through online tools (such as e-mails) and providing personal assistance to groups for dealing with specific issues.

6. On-demand assistance through meetings and training courses, which are training sessions for groups. They can be planned on demand once the researchers or the system administrators have made explicit their training needs.

In order to assign the suitable facilities to the different profiles in the community, we prepared a chart (**Table 3**, described below) listing those which better fit the needs of a particular profile.

The chart contains three types of cells:

1. cells listing the facilities suitable for the specific profile.
2. cells marked with[a]: This marker is applied when the feature underlying a competence is present. Hence, the latter can be taken for granted and the facilities envisaged for this profile are either not necessary or can be chosen freely depending on the specific case. For example, if F2 is 1 (i.e., the group is active in *web-based data distribution*), then CL1 = 1 and CL0 = 1, that is these two competences can be taken for granted since the group shares its data over the Web. However, it could also be the case that the group needs to master its competence and then there is the need for choosing facility contents that can specifically improve this activity.
3. cells marked with[b]: This marker is applied when a particular feature is absent in the profile under consideration. Hence, a priority for the competence to master has to be established and it is reported in the cell. For example, if F1 for the group is 0 (i.e., it isn't active in *spatial data production*), it is necessary to prioritize which competence the group has to gain according to the specific case (CL0 or CL1). It reasonable to suppose that factors which are confining it to a particular behavior are other than capacity gaps.

In Phase 2, we match every profile to the cells to identify which facilities the USS ensures for the appropriate technological enablement of the group. We focused on a selection of the facilities that are useful to achieve the technological maturity for every possible profile, but other tools such as short courses (e.g., Data Carpentry) can be considered in further implementation.

Application of the engagement approach and impact assessment

This section describes application of the workflow to the RITMARE case study by describing (I) the USS planning and implementation following the workflow, (II) the expectations of the surveyed RITMARE groups with respect to the facilities proposed, and (III) the impact of the USS on the technological development of the RITMARE SDI.

Enacting the engagement approach in RITMARE

We applied the workflow to (I) profile the groups of the RITMARE community according to Phase 1 and to (II) plan and implement the suitable facilities for the RITMARE community according to Phase 2.

Phase 1 – Profiling the RITMARE groups

The RITMARE behavior categories. We derived the three feature combinations for every group of the community through a series of questionnaires submitted to the representatives of the RITMARE Research Work Packages and Research Actions.

We found that the community is characterized by all five categories for which we provide the following descriptions:

Table 3: Facilities chart.

		Behavior categories				
		Committed (1,1,1,)	Competent (1,1,0)	Web-ready (1,0,0)	Receptive (0,1,0)	Locked (0,0,0)
Competence levels	0	F2 = 1 → CL0 = 1[a]	F2 = 1 → CL0 = 1[a]	F1 = 1 → CL0 = 1[a]	Glossary, FAQ	Glossary, FAQ
	1	F2 = 1 → CL1 = 1[a]	F2 = 1 → CL1 = 1[a]	Glossary, FAQ, Webinar	F1 = 0 → CL0 has priority[b]	F1 = 0 → CL0 has priority[b]
	2	On-demand (online), On-demand, Webinars	On-demand (online), Webinars	F2 = 0 → CL1 has priority[b]	F1 = 0 → CL1 has priority[b]	F1 = 0 → CL1 has priority[b]

1. *Committed*: In RITMARE, this category comprises groups which already distribute their resources through the RITMARE SDI (F3 = 1). They hold within their research team the skills and figures (such as system administrator, data managers or researchers with technological literacy) to process data through GIS and use web-GIS software to deliver data through the services supported by the SDI (WMS, WFS, WCS, SOS, CSW). The groups which are distributing their resources in an interoperable way constitute active data nodes[2] in the RITMARE Data Portal (v0.0) (they are accessible in the directory on the right column at: http://portale0-sp7.ismar.cnr.it/#/nodes/list), a prototype of the project geoportal and single access point to the RITMARE SDI.

2. *Competent*: In RITMARE, this category comprises the groups that produce georeferenced data belonging to their disciplinary area such as physical oceanography, geology, or ecology, where standard web-distribution of geospatial data is already an adopted practice. These groups have previously experienced web distribution in previous projects (LTER, SeaDataNet, MyOcean, EMODnet) and their data are already collected and structured according to community practices, supported by existing software tools or by those developed during RITMARE.

3. *Web-ready*: In RITMARE, this category comprises groups that record their data through spreadsheets or even paper notes. Their data need to be either digitized or managed with GIS or WebGIS software, also because they have long-term time series or historical collection related to the investigated area (Minelli et al. 2018).

4. *Receptive*: In RITMARE, this category comprises groups that rely on external data infrastructures (such as GenBank) or those provided by research institutes (e.g. the European Bioinformatic Institute), thus being predisposed to Web data sharing in RITMARE. These groups include researchers from molecular life sciences where data archiving in public repositories is formally established. They are compliant with data sharing practices proper of their community but they are less aware of spatial representation of data and standard services, even if some domains, such as metagenomics, are more prone to adopting the suggested practices (Parks et al. 2009).

5. *Locked*: This category comprises ecology teams that manually collect data through paper notes and without using geotagging or georeferencing devices (Michener et al. 2006);

The RITMARE competence levels. To complete groups profiles, we assessed the competence for every group featured in the community thanks to the contacts between them and the RITMARE Data Team. We observed that the groups belonging to the *committed* and *competent* behavior only needed to acquire or deepen their *interoperable competence* (for instance, by adopting a specific terminology in order to enable semantic interoperability). The groups representing the *receptive, locked,* and *web-ready* behaviors needed to fully achieve or complement the necessary *digital competence,* as they were not always aware of geospatial practices.

　The RITMARE profiles. The profiles that have been identified are displayed in **Table 4**, where we show a plus sign "+" for combinations that were found in the research community (positive profiles) and a minus sign "−" when the particular combination is absent (negative profile). It represents a synthetic and qualitative summary of the RITMARE community based on the profiles of its groups. This summary allowed us to direct planning and implementation of the USS.

　We report only the profiles obtained for the groups which interacted with the RITMARE Data Team, since for these groups it was possible to assess the competence.

Phase 2 – Planning and implementing the RITMARE USS

We identified the facilities the USS has to set up to engage the distinct groups in data sharing by matching the profiles against the facilities chart and selecting the cells corresponding to positive profiles (**Table 5**). In RITMARE, we decided to implement all the facilities suggested by the chart as the project had the necessary resources. In case the profiles overlapped with cells marked as [a]*or*[b], we decided to provide only on-demand assistance to meet the needs of the related groups.

　We decided to provide both the educational and the supporting facilities. The educational ones allowed us to address both key issues of the research context.

　Table 6 outlines the facilities provided. Column "items" contains either the number of items in a given facility (e.g., the glossary) or the number of requests issued to the RITMARE Data Team for a specific need. We

---

[2] A data node is a logic unit in the SDI. In RITMARE, it corresponds to the standard web services that are needed to distribute data through the RITMARE geoportal. It also corresponds to the group which creates the services and give access to their own data through standard geoservices.

Table 4: The RITMARE group profiles.

		Behavior categories				
		Committed (1,1,1,)	Competent (1,1,0)	Web-ready (1,0,0)	Receptive (0,1,0)	Locked (0,0,0)
Competence levels	0	–	–	–	+	+
	1	+[a]	+[a]	+	+[b]	+[b]
	2	+	+	–	–	–

Table 5: RITMARE facilities.

		Behavior categories				
		Committed (1,1,1,)	Competent (1,1,0)	Web-ready (1,0,0)	Receptive (0,1,0)	Locked (0,0,0)
Competence levels	0	NO	NO	NO	Glossary, FAQ	Glossary, FAQ
	1	To master	To master	Glossary, FAQ, Webinar	NO	NO
	2	On-demand (online), On-demand (F2F), Webinar	On-demand (online)	NO	NO	NO

set up a helpdesk unit made up of two persons dedicated to provide on-demand assistance through e-mails and messaging software. They assisted the groups in the creation and management of new RITMARE data nodes, in the enablement of external data nodes, and in the transfer of two preexisting RITMARE data nodes.

USS impact in the RITMARE community

We surveyed how the educational and supporting facilities proposed by the USS met the groups' needs. We submitted 24 questions through an online form and elaborated the anonymous responses using infographics.[3] The implemented facilities were considered strongly formative (60%) for accomplishing the data management tasks and useful (30%) for overcoming the more general obstacles related to data management. In particular, almost all respondents (90%) considered facilities such as the Glossary, the FAQ, and the on-demand support very useful for the purpose of facilitating data management and sharing. The researchers considered the Glossary and the FAQ very helpful to improve their competences, while the other facilities were deemed as useful by researchers and system administrators indifferently.

The respondents cover primarily the *committed*, *competent*, and *web-ready* categories but, considering also the groups publishing their datasets in third-party infrastructures, 80% of the respondents are compliant with the FAIR principles. It is interesting to note that, albeit 60% of the respondents are strongly devoted to data collection and analysis, only 30% of them are committed to data stewardship, as the others consider it as an optional activity. However, whatever their commitment to data stewardship, 60% of the respondents think it deserves more support through the USS.

As for the more general comments by the interviewed users, together with rewarding mechanisms, the lack of guidelines in data management is considered as a limiting factor to data sharing (see also Kervin et al. 2014; Arzberger et al. 2003) by the 30% of groups and it is considered as crucial for 70% of them. Also, receiving indications on how to handle data is considered by groups as a motivational factor and more all-round data management education is strongly indicated (60%) as a driving force.

USS impact assessment on the RITMARE SDI

We verified how the USS activities influenced the RITMARE SDI in terms of new data nodes and number of resources distributed through them. Before the establishment of the USS, the SDI included 12 data nodes; one year after enacting the USS, three new groups set up their data nodes and the enablement of three existing data nodes was completed. The three new data nodes distributed 1562 new resources through interoperable standard services by raising to 18200 the resources available in the SDI. The data categories which were more enriched are oceanography, geology, geophysics, ecosystems, fisheries, and aquaculture, thus providing a multidisciplinary platform to a wide range of users.

---

[3] Martina Zilioli, Alessandro Oggioni, Simone Lanucara, & Paola Carrara. (2018, December 30). USS impact in the RITMARE community: survey results and questionnaire. Zenodo. http://doi.org/10.5281/zenodo.2529049.

Table 6: Facilities in the RITMARE USS.

		Facilities			
Class	**Name**	**Contents**	**Items**	**Container**	**Identifier**
Educational	Glossary	Definition of a common terminology (e.g. describing file formats, access protocols, and mandatory regulations)	45	Wiki	http://sp7.irea.cnr.it/wiki/index.php/
Educational	FAQ	The RITMARE Data Portal, Data Nodes (DNs), RITMARE Data Policy	24	Wiki	http://sp7.irea.cnr.it/wiki/index.php/
Educational	Webinar	Policy and Technological Components of the RITMARE SDI, How to create autonomous RITMARE DNs	2	Wiki	http://sp7.irea.cnr.it/wiki/index.php/
Educational	Videotutorials,	How to create autonomous RITMARE DNs	1	GET-IT website	http://www.get-it.it/
Supporting	Service	Installing and mantaining new RITMARE DNs	4	Help desk	help.skritmare@irea.cnr.it
Supporting	Service	External DNs Enablement	2	Help desk	help.skritmare@irea.cnr.it
Supporting	Service	Data Transfer to RITMARE DNs	1	Help desk	help.skritmare@irea.cnr.it
Supporting	Service	Data Connection to a RITMARE DNs	1	Help desk	help.skritmare@irea.cnr.it

Concluding remarks

This paper describes an approach to engage multidisciplinary research groups in data sharing by implementing a workflow-based USS. Its application to the RITMARE case study suggests that it represents an essential component of any engagement activity, also addressing organizational and coordination obstacles encountered in other community (Nedovic-Budic and Pinto, 2001, Hartter et al. 2013). The workflow constitutes a roadmap to deal with the social component of SDIs, which is often disregarded in favor of the technological, political, and economical ones (Feeney et al. 2001; Rajabifard and Williamson, 2002). However, the social aspects of data sharing have not been fully explained (Omran et al. 2007) as the study of organizational behavior in spatial data sharing across different cultural backgrounds is still an emerging issue (Yu, J.X. 2010). The lack of a clear categorization of roles in the groups makes it difficult to execute the user profiling that is essential to every engagement and communication strategy.

We included as targets of the USS activities not only researchers but also data managers to facilitate their interaction. We demonstrated that a USS can more effectively help shifting groups from the involvement to the engagement state if they are modeled with a fine-grained description. This is particularly important to streamline communication among heterogeneous cultural groups and to pinpoint the suitable enablers to fit the diverse needs.

The workflow we proposed profiles the groups in an unprecedented way by modeling them separately addressing two distinct aspects of data management: *behavior* takes into account the practices enacted by a group, while *competence* expresses the skill sets and backgrounds shared among its members.

To demonstrate impact of the USS on the RITMARE SDI, we applied the workflow to its multidisciplinary community and verified that it helped enriching the project's SDI with more data nodes. Specifically, we observed that the RITMARE groups featured all five behaviors, spanning from groups that were actively involved in the project's infrastructure and groups that were reluctant to share data. For each of these, on the basis of surveys, we applied the workflow and implemented the appropriate USS facilities. Thanks to this approach, we engaged new groups in managing their data according to the project practices. In particular, we obtained a gain of 20% in the number of data nodes in the SDI and of 8% in the number of datasets made available by the RITMARE Data Portal. The profiles that were more effectively engaged were the *committed*-competence 2, the *competent*-competence 1, 2, and the *web-ready*-competence 1. While engagement of the first two profiles was expected, the positive feedback of the latter suggested us to improve the USS with services to check file structures and the related metadata, to flag missing information, and to make file formats apt to GIS/Web-GIS distribution.

In general, we obtained a positive evaluation on the facilities by the whole RITMARE community and the survey based on questionnaire helped us identify whether its expectations were met and which obstacles were addressed. Particularly, these are the lack of (I) an organizational pipeline to distribute data, (II) full-time professionals and education to manage data, and (III) competence in applying licenses to data. While the

community recognized the helpfulness of the USS in group education, facilities for improving competence in data protection (comprising unrewarded or irregular reuse) are requested. Even if we did not address this topic in this paper, we consider it crucial to effective enablement of the groups.

Although we have used indirect indicators to assess the uptake of the facilities (such as the questionnaire results and the SDI population), we intend to quantify it more precisely with access counters to the educational platform or by ranking systems. We consider it helpful to address researches focusing on infrastructure performance indicators (Fledderus, 2017) to relate our services to data infrastructure sustainability.

Finally, we believe that tests of the workflow in other case studies are needed. We argue that our workflow-based methodology constitutes the first example of structured composition of a USS. We look forward to further case studies to be addressed with the same methodology to add accuracy and rigor to our proposal.

Acknowledgements

Competing Interests

The authors have no competing interests to declare.

References

Arzberger, P, Schroeder, P, Beaulieu, A, Bowker, G, Casey, K, Laaksonen, L, David Moorman, D, Uhlir, P and **Wouters, P.** 2003. An International Framework to Promote Access to Data. *Science*, 303 (5665): 1777–1778. DOI: https://doi.org/10.1126/science.1095958

Basoni, A, Menegon, S and **Sarretta, A.** 2015. Sailing towards Open Marine Data: the RITMARE Data Policy. Available at: https://ercim-news.ercim.eu/en100/special/sailing-towards-open-marine-data-the-ritmare-data-policy. ERCIM News.

Birney, E. 2012. Lessons from big data project. *Nature*, 489: 49–51. DOI: https://doi.org/10.1038/489049a

Borgman, CL. 2015. If the Data Sharing is the Answer, what is the Question? Available at: https://ercim-news.ercim.eu/en100/special/if-data-sharing-is-the-answer-what-is-the-question. ERCIM News.

Chunpir, HI, Badewi, AA and **Ludwig, T.** 2014. User Support System in the Complex Environment. In: Design, User Experience, and Usability. User Experience Design Practice. DUXU 2014. *Lecture Notes in Computer Science*, 8520. Cham: Springer. DOI: https://doi.org/10.1007/978-3-319-07638-6_38

Dallmeier-Tiessen, S, Darby, R, Gitmans, K, Lambert, S, Matthews, B, Mele, S, Suhonen, I and **Wilson, M.** 2014. Enabling Sharing and Reuse of Scientific Data. *New Review of Information Networking*, 19(1): 16–43. DOI: https://doi.org/10.1080/13614576.2014.883936

Feeney, M, Rajabifard, A and **Williamson, I.** 2001. Spatial Data Infrastructure Frameworks to Support Decision-Making for Sustainable Development. In: *Proceedings of the 5th Global Spatial Data Infrastructures*, Cartagena, 1–15.

Fledderus, E and **von Voigt Wiebelitz Lawenda, G.** 2017. Evaluation of e-Infrastructures and the development of related Key Performance Indicators. Available at: https://www.rd-alliance.org/sites/default/files/attachment/Evaluation%20of%20e-Infrastructures%20and%20the%20development%20of%20related%20Key%20Performance%20Indicators%20%281%29.pdf.

Fugazza, C, Menegon, S, Pepe, M, Oggioni, A and **Carrara, P.** 2014. The RITMARE Starter Kit–Bottom-up Capacity Building for Geospatial Data Providers. In: *Proceedings of 9th International Conference on Software Paradigm Trends (ICSOFT-PT)*, Vienna, 169–176. DOI: https://doi.org/10.5220/0004999801690176

Haendel, MA, Vasilevsky, NA and **Wirz, JA.** 2012. Dealing with Data: A Case Study on Information and Data Management Literacy. *PLoS Biol*, 10(5). DOI: https://doi.org/10.1371/journal.pbio.1001339

Karasti, H, Baker, KS and **Halkola, E.** 2006. Enriching the Notion of Data Curation in E-Science: Data Managing and Information Infrastructuring in the Long Term Ecological Research (LTER) Network. *Computer Supported Cooperative Work*, 15: 321. DOI: https://doi.org/10.1007/s10606-006-9023-2

Kervin, KE, Cook, RB and **Michener, WK.** 2014. The Backstage Work of Data Sharing. In: *Proceedings of the 18th International Conference on Supporting Group Work (GROUP)*. Sanibel Island, FL, New York. DOI: https://doi.org/10.1145/2660398.2660406

Kervin, KE, Michener, WK and **Cook, RB.** 2013. Common Errors in Ecological Data Sharing. *Journal of eScience Librarianship*, 2(2): e1024. DOI: https://doi.org/10.7191/jeslib.2013.1024

Kim, Y. 2017. Fostering scientists' data sharing behaviours via data repositories, journal supplements, and personal communication methods. *Information Processing & Management*, 30(4): 871–885. DOI: https://doi.org/10.1016/j.ipm.2017.03.003

Lanucara, S, Praticò, S and **Modica, G.** 2019. Harmonization and Interoperable Sharing of Multi-temporal Geospatial Data of Rural Landscapes. In: Calabrò, F, Della Spina, L and Bevilacqua, C (eds.), New Metropolitan Perspectives. ISHT 2018. Smart Innovation. *Systems and Technologies*, 100. Cham: Springer. DOI: https://doi.org/10.1007/978-3-319-92099-3_7

Lee, CP. 2006. The Human Infrastructure of Cyberinfrastructure. In: *Proceedings of ACM Conference on CSCW*, 483–492. New York. DOI: https://doi.org/10.1145/1180875.1180950

Lee, DJ and **Stvilia, B.** 2017. Practices of research data curation in institutional repositories: A qualitative view from repository staff. *PLOS ONE*, 12(3). DOI: https://doi.org/10.1371/journal.pone.0173987

Michener, WK. 2006. Meta-information concepts for ecological data management. *Ecological Informatics*, 1: 1. DOI: https://doi.org/10.1016/j.ecoinf.2005.08.004

Minelli, A, Oggioni, A, Pugnetti, A, Sarretta, A, Bastianini, M, Bergami, C, Bernardi, AF, Camatti, E, Scovacricchi, T and **Socal, G.** 2018. The project EcoNAOS: vision and practice towards an open approach in the Northern Adriatic Sea ecological observatory. *Research Ideas and Outcomes*, 4: e24224. DOI: https://doi.org/10.3897/rio.4.e24224

Nedovic-Budic, Z and **Pinto, JK.** 2001. Organizational (soft) GIS interoperability: lessons from the U.S. *International Journal of Applied Earth Observation and Geoinformation*, 3(3): 290–298. DOI: https://doi.org/10.1016/S0303-2434(01)85035-2

Omran, EE. 2007. Spatial Data Sharing: A Cross-Cultural Conceptual Model. Research and Theory in Advancing Spatial Data Infrastructure Concepts. Redlands, CA: ESRI Press. Available at: https://www.semanticscholar.org/paper/Spatial-Data-Sharing-%3A-A-Cross-Cultural-Conceptual-Omran/433067af85b091669f955f0951b04edde0b63ba3.

Parks, DH, Porter, M, Churcher, S, Wang, S, Blouin, C, Whalley, J, Brooks, S and **Beiko, RG.** 2009. GenGIS: A geospatial information system for genomic data. *Genome Research*, 19: 1896–1904. DOI https://doi.org/10.1101/gr.095612.109

Parsons, MA, Brodzik, MJ and **Rutter, NJ.** 2004. Data management for the Cold Land Processes Experiment: improving hydrological science. *Hydrological Processes*, 18: 3637–3653. DOI: https://doi.org/10.1002/hyp.5801

Parsons, MA and **Duerr, R.** 2005. Designating user communities for scientific data: challenges and solutions. *Data Science Journal*, 4. DOI: https://doi.org/10.2481/dsj.4.31

Parsons, MA, Godøy, Ø, LeDrew, E, de Bruin, FT, Danis, B, Tomlinson, S and **Carlson, D.** 2011. A conceptual framework for managing very diverse data for complex, interdisciplinary science. *Journal of Information Science*, 37(6): 555–569. DOI: https://doi.org/10.1177/0165551511412705

Rajabifard, A and **Williamson, IP.** 2002. Spatial Data Infrastructures: An Initiative to Facilitate Data Sharing, Global Environmental DBs–Present Situation and Future Directions 2. Hong Kong.

Savage, CJ and **Vickers, AJ.** 2009. Empirical Study of Data Sharing by Authors Publishing in PLoS Journals. *PLoS ONE*, 4(9). DOI: https://doi.org/10.1371/journal.pone.0007078

Sayogo, DS and **Pardo, TA.** 2013. Exploring the determinants of scientific data sharing: Understanding the motivation to publish research data. *Government Information Quarterly*, 30: 19–31. DOI: https://doi.org/10.1016/j.giq.2012.06.011

Star, SL and **Ruhleder, K.** 1996. Steps Toward an Ecology of Infrastructure: Design and Access for Large Information Spaces. *Information Systems Research*, 7(1): 111–134. DOI: https://doi.org/10.1287/isre.7.1.111

Tenopir, C, Dalton, ED, Allard, S, Frame, M, Pjesivac, I, Birch, B, Pollok, D and **Dorsett, K.** 2015. Changes in Data Sharing and Data Reuse Practices and Perceptions among Scientists Worldwide. *PLoS ONE*, 10(8). DOI: https://doi.org/10.1371/journal.pone.0134826

Volk, CJ, Lucero, Y and **Barnas, K.** 2014. Why is Data Sharing in Collaborative Natural Resource Efforts so Hard and What can We Do to Improve it? *Environmental Management*, 53: 883. DOI: https://doi.org/10.1007/s00267-014-0258-2

Yu, JX, Chen, N, He, J, Cao, Y, Ma, L and **Yang, H.** 2010. Research on geographic information sharing in different cultural backgrounds. In: *Research and Theory in Advancing Spatial Data Infrastructure Concepts*, Redlands, CA, 2007: ESRI Press. DOI: https://doi.org/10.1109/GEOINFORMATICS.2010.5567613