A Help Desk to support Data Sharing in Environmental Research Communities

Zilioli Martina[1], Lanucara Simone[2], Oggioni Alessandro[3] and Carrara Paola[4]

1. Introduction

Environmental research is more and more involving multidisciplinary communities, where different research groups are providing different data. To foster interdisciplinary research, scientists need to exchange and integrate data and metadata, and informatics offers a growing number of means to address data interoperability across environmental subjects. We propose a Help Desk (HD) that provides facilities and services to support data interoperability among research groups of the RITMARE project. The HD aims at enabling research groups to share their data through the RITMARE Spatial Data Infrastructure (SDI), an interoperable infrastructure built to enhance the (open) access and reuse of data inside the project community. In Section 2 (Methods) we briefly introduce the issues at stake and describe the HD design. We report in section 3 (Results) the HD implementation in our research framework and the HD impact on RITMARE community. We finally discuss in Section 4 if the HD could be exported in other environmental research communities.

2. Methods

2.1 Approaches to manage heterogeneity of data and data providers

Interdisciplinary research is challenging different fields like climate science and ecology [1, 2], where means to share and integrate heterogeneous data are necessary to extract new knowledge. Informatics aids to bring data at different levels of interoperability, allowing to make them comparable and usable across multiple contexts and disciplines. For instance, in the last decades interoperability of geographic information has been strongly influenced by Spatial Data Infrastructure (SDI) initiatives [3]. These cyber-infrastructures are developed to increase and coordinate the distribution of several types of geospatial data and metadata. SDIs allow data to become Findable, Accessible, Interoperable and Reusable (according to FAIR

[1] National Research Council of Italy, Institute for Electromagnetic Sensing of Environment, zilioli.m@irea.cnr.it
[2] National Research Council of Italy, Institute for Electromagnetic Sensing of Environment, lanucara.s@irea.cnr.it
[3] National Research Council of Italy, Institute for Electromagnetic Sensing of Environment, oggioni.a@irea.cnr.it
[4] National Research Council of Italy, Institute for Electromagnetic Sensing of Environment, carrara.p@irea.cnr.it

principles [4]) among a wide range of producers and users by addressing technical, policy, economic and social issues. In fact, to be better exploited by their users, SDIs need to be populated with data according to technical and policy requirements, mostly in order to assure interoperability between different domains. However, data management inside a research group depends on several variables such as the subject discipline it belongs to [5], its community practice [6], and members' skills and competences to use informatics tools. This means that data providers are not always able to produce interoperable and heterogeneous data, so that they need to be suitably enabled to exploit the data infrastructure and other informatics tools to foster data sharing and reuse as most as possible.

2.2 The RITMARE Case Study

A heterogeneous community of data providers shapes the RITMARE project (http://www.ritmare.it), a flagship project funded by the Italian Ministry of University and Research, which involves several research groups working on different marine disciplinary domains (physical oceanography, chemical oceanography, geophysics, geology, ecosystem, molecular life sciences, fishing and agriculture, coastal system). After three years of project, we have found that RITMARE SDI does not host the expected number of data produced by research groups nor was the RITMARE Data Policy (RDP) fully addressed, yet. To maximize the SDI sustainability by data enrichment, we have designed and implemented a Help Desk (HD) to specifically support heterogeneous research groups in data sharing, by considering their differences in terms of data produced and data management competences. The HD is conceived as a series of web facilities and services which are created, managed and delivered by a staff of people engaged in enabling research groups to make data interoperable with those produced by other groups.

2.3 The RITMARE HD design

To provide the multidisciplinary community with an appropriate HD, we first defined what an HD needs to provide. The first aim of the HD is to allow users (members of the research groups) to contact a support staff [7] and receive useful information to address their own data management problems. In order to identify what contents the staff have to produce and how it should interact with users, we designed the HD by (I) first mapping how many data management competence levels were present in the research community and (II) which HD support activities could better overcome the data management gaps of research groups. This was done through surveys, enriched also with information provided by the RITMARE Data Team, i.e. the working group in charge of develop the RITMARE SDI. Finally, we identified different facilities and services by crossing the data management competence levels and the HD support activities.

We classified the RITMARE research groups in four data management competence levels in order to obtain a summary on how research groups differentially provide data to SDI. They are:

- *Level 0 - Not GIS enabled*. This level identifies research groups which collect and store their geospatial data via images, spreadsheets or texts without georeferenced information;
- *Level 1 - GIS enabled*. This level identifies research groups which store their data by using geospatial file format (shapefile, geoJSON, geospatial database, GML) without choosing distribution services to share them on the Web;
- *Level 2 - GIS enabled*. This level identifies research groups which store and share their data by WebGIS applications without interoperable standard services;
- *Level 3 - GIS enabled*. This level identifies research groups which store and share their data by using WebGIS applications and interoperable OGC standard services. They are also metadata aware.

To convey instructions and assistance to research groups according to their data management competence levels, we identified two types of support activities. They were both educational facilities to stimulate data providers in gaining upper competence levels and technological services to overcome technical limitations to data inclusion in the RITMARE SDI. We then associated educational facilities to all the data management competence levels by tailoring contents to the user level and by using different delivering tools; instead, we associated the technological services only to Level 2 and 3, as only partially enabled research groups could be supported in this sense. By associating data management competence levels and support activities we have then chosen communication tools and contents to shape the RITMARE HD. The selected tools are reported in Fig. 1 and are divided in three main activities modules:

- the learning (1) and training (2) modules consist of educational web-facilities for research groups which needs to initially organize and store their geospatial data according to the RITMARE SDI requirements. These educational facilities provide contents through different communication channels which are glossary, FAQs, webinars and webcasts, and a software documentation. These tools are accessible through two online platforms and curated by the HD staff;

- the supporting (3) module consists of a technological service for more advanced research groups (Level 2 and 3), which already share their research data by WebGIS or (standard) web services, but still need to be supported according to some specific issues of RITMARE. The technological services correspond to an online assistance provided by the technical staff, available through a unique contact point (by email, calls and conference calls). The HD staff receives, tracks and manages help requests, and provides technical solutions to reach the technological enablement of research groups. Four types of technological support are described in the following chapter.

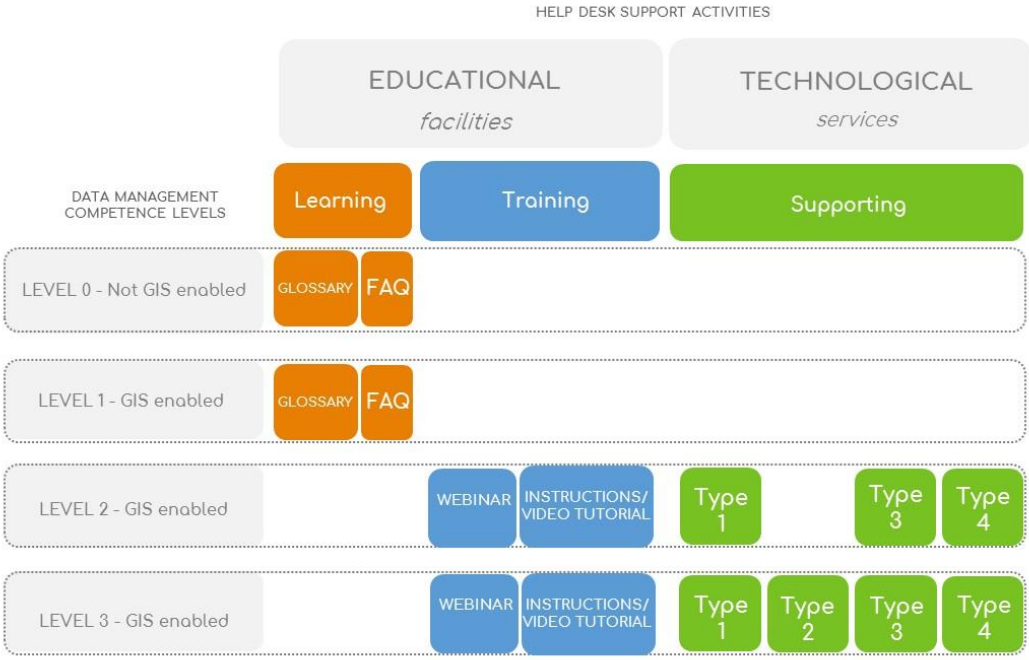

Fig. 1. The RITMARE HD design

The RITMARE HD design identifies four data management competence levels (reported on left) and two types of support activities (reported above). Facilities and services tools which shape the HD are listed with different colors according to the activity module (learning, training and supporting) they belong to. Technological support services are divided in four types: installing and assisting new RITMARE SDI nodes (Type 1); external nodes enablement (Type 2); data transfer to RITMARE SDI nodes (Type 3); data connection to RITMARE SDI nodes (Type 4).

3. Results

3.1 The RITMARE HD implementation

We implemented the RITMARE HD by activating the HD staff employed to produce contents for educational facilities and to provide technological services. The results of the RITMARE HD are:

- GLOSSARY: a 45 items section, accessible through a MediaWiki platform (http://sp7.irea.cnr.it/wiki/index.php/MediaWiki:SP7_Wiki), which provided educational contents related to collection, format, organization, storage, web distribution, access of geospatial data and their metadata;
- FAQs: a 24 items section, accessible through a MediaWiki platform (http://sp7.irea.cnr.it/wiki/index.php/MediaWiki:SP7_Wiki), which provided contents about RITMARE Data Portal, RITMARE Data Nodes, RITMARE Data Policy;
- WEBCAST: a 2 items multimedia program, accessible through a MediaWiki platform (http://sp7.irea.cnr.it/wiki/index.php/MediaWiki:SP7_Wiki), dedicated to (i) The RITMARE SDI: what is it? Policy and Technological Components and (ii) How to create a RITMARE Data Node: the GET-IT software suite[5]. The items were prior distributed to the RITMARE community through webinar channels;
- INSTRUCTIONS / VIDEO TUTORIALS: a detailed documentation, accessible through a dedicated website (http://www.get-it.it/), describes how to create a RITMARE Data Node by using the GET-IT software suite (this is mainly devoted to groups that do not already distribute data by web services);
- DIRECT SUPPORT ACTIVITIES: a technical staff was employed to provide four types of technological enablement by collecting requests through an email address (help.skritmare@irea.cnr.it), followed, if needed, by calls and conference calls. We report here the percentage of support for every type of support provided:
    - Type 1 technological support (Installing and assisting new RITMARE SDI nodes) – 50%
    - Type 2 technological support (External nodes enablement) - 25%
    - Type 3 technological support (Data transfer to RITMARE SDI nodes) – 12%
    - Type 4 technological support (Data connection to RITMARE SDI nodes) – 13%

[5] The GET-IT Geoinformation Enabling ToolkIT starterkit®, is the open-source software suite developed by the RITMARE Data Team to enable RITMARE researchers to share their geospatial data and metadata through OCG Web standard services [8], creating new nodes of the RITMARE SDI

3.2 Help Desk activities impact on the RITMARE community

The RITMARE HD design and implementation aimed at enabling a growing number of research groups to exploit informatics means offered in the project and make their data interoperable by populating the RITMARE SDI. To estimate HD efficacy, we try to assess the RITMARE HD impact on the RITMARE community. We choose to state HD efficacy by using three parameters reported in Table 1 and listed here following:

- new data nodes created in the RITMARE SDI;
- new data resources provided through the RITMARE SDI;
- enablement of new data providers.

We considered the SDI population in terms of nodes and data as a proxy of the HD success. Results were reported in Table 1, where we indicated the number of data nodes created, the number of data resources shared and the number of research groups enabled to use and connect to the RITMARE SDI. We report figures referred to both before and after the RITMARE HD implementation.

HD Efficacy Parameters	Before HD activation	After HD activation
RITMARE SDI Data Nodes (number of registered nodes)	12	15
RITMARE SDI Data Resources (number of accessible dataset through the RITMARE SDI)	16628	18200
RITMARE Data Providers (number of enabled research groups)	ND	3

Tab. 1. RITMARE HD impact on RITMARE SDI

The data domains which are more involved are oceanography, geology, geophysics, ecosystems, fisheries and aquaculture, thus multidisciplinary enriching the RITMARE SDI [9]. The HD is still currently assisting three research groups belonging to two data management competence levels: one to Level 4 competence, two to Level 3.

4. Discussion

Open science is a technological and social challenge faced by addressing informatics tasks and users needs, both becoming more critical in multidisciplinary research communities. In fact, the development of cyber-infrastructures to reduce barriers to data sharing posed some obstacles, particularly due to different data management competence of research groups. RITMARE is a paradigmatic framework as well as a multidisciplinary environmental project where we have tried to improve the RITMARE SDI sustainability with a Help Desk taskforce. By designing the RITMARE HD, we have implemented facilities and services able to increase the number of SDI nodes, the number of SDI interoperable data resources and the number of data providers sharing interoperable data. The HD provides education facilities and technological services to support research groups, which exhibit different data management backgrounds, by addressing their specific needs. The supporting facilities tools can be easily accessed by researchers of different domains with different technological abilities (a heterogeneous user target) through two online web platforms curated by the HD staff. They are maintained and could be steadily enriched of contents according to new environmental communities requirements. Also technological services could operatively address and overcome different data provider needs since four types of technological support were requested and met.

We think that our multi-level approach to design a HD for supporting interoperable data management in complex, multi-disciplinary projects, can be exported in other initiatives in three different ways:

- by applying the HD design and implementation pipeline in other research communities;
- by allowing access to RITMARE facilities to other research communities and by making up contents suitable to them;
- by offering technical staff support to other research communities and improving the technological enablement of research groups.

In particular, we propose to test in future if the HD taskforce could be created (i) ex-novo by following the pipeline or (ii) simply by re-organizing and extend the RITMARE HD platforms for new users, for example when an SDI development or a long-term management of data are required by interdisciplinary environmental studies. Our HD design and implementation would be a new roadmap to offer a standardized system to enable a growing number of data providers in data sharing, particularly with a glance to Data Management Plans required in Horizon 2020 framework [10] and to the application of FAIR principles.

References

[1] Beattie, A. J., Hay, M., Magnusson, B., De Nys, R., Smeathers, J., & Vincent, J. F. V. (2011). Ecology and bioprospecting. Austral Ecology, 36(3), 341–356. doi: 10.1111/j.1442-9993.2010.02170.

[2] Parsons, M.A., Godøy, Ø., LeDrew, E., F. de Bruin, T., Danis, B., Tomlinson, S., Carlson, D. (2011). A conceptual framework for managing very diverse data for complex, interdisciplinary science, Journal of Information Science, 37(6), 555–569. doi: 10.1177/0165551511412705

[3] Aalders, H. J. G. L., Moellering, H. Spatial Data Infrastrucutre. Retrieved from https://repository.tudelft.nl/islandora/object/uuid:2fe63c14-f620-4208-b11d-584991d48d3d/datastream/OBJ

[4] Wilkinson, M. D. et al. (2016) The FAIR Guiding Principles for scientific data management and stewardship. Sci. Data 3:160018. doi: 10.1038/sdata.2016.18.

[5] Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A.U., Wu, L. (2011) Data Sharing by Scientists: Practices and Perceptions. PLoS ONE, 6(6). doi:10.1371/journal.pone.0021101

[6] Birney, E. (2012) The making of ENCODE: Lessons for big-data projects. Nature. 2012 Sep 6;489(7414):49-51. doi: 10.1038/489049a.

[7] Jäntti, M.: Examining Challenges in IT Service Desk System and Processes: A Case Study (c), 105–108 (2012)

[8] Fugazza, C., Menegon S., Pepe, M., Oggioni, A., Carrara, P. (2014). The RITMARE Starter Kit - Bottom-up Capacity Building for Geospatial Data Providers. In Proceedings of the 9th International Conference on Software Paradigm Trends (ICSOFT 2014). ISBN 978-989-758-037-6, pages 169-176. doi:10.5220/0004999801690176

[9] Zilioli, M, Oggioni, A., Carrara, P. (2016) Censimento delle risorse erogate da RITMARE Data Portal (v0.0), prima implementazione del geoportale RITMARE. Retrieved from http://sp7.irea.cnr.it/wiki/index.php/Censimento_delle_risorse_di_RITMARE_Data_Portal_(v.0.0)

[10] European Commission. Directorate-General for Research and Innovation. Guidelines on FAIR Data Management in Horizon 2020. Retrieved from http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

The activities described in this paper have been funded by the Italian Flagship Project RITMARE