MS23: First face-to-face FAIR data management training event for ENVRI data centre staff completed

Author(s)	Maggie Hellström, Maria Johnsson, Nicola Fiore, Lucia Vaira, Jacco Konijn
Lead partner	LifeWatch ERIC
Work Package	6
Status	Final
Date	2020-03-18

ENVRI-FAIR (www.envri-fair.eu) has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824068

TABLE OF CONTENTS

MS23: First face-to-face FAIR data management training event for ENVRI data centre staff completed

Background

The objective of Work Package 6 is to provide training to ENVRIs and key ENVRI stakeholder groups about the FAIR principles, how to implement them in RI services and data management activities at data centre level, how to evaluate the degree of implementation using FAIR metrics, as well as relevant legal and policy requirements.

While Task 6.1 is concerned with preparing and disseminating training resources, the main responsibility of Task 6.2 is to organise and provide training for data centre staff from RIs (both inside and outside of ENVRI-FAIR) on FAIR implementation.

The primary target group for the training is staff at the ENVRI data centres, especially those concerned with data management and service architecture. This group will benefit both from self-paced study activities and from tutorial sessions organized e.g. in connection to project collaboration meetings and Webinars. A second important target group for the training is the staff of data centres of key local, regional and national institutions dealing with environmental data. In fact, many RIs are concerned with coordinating and disseminating data products (and services) produced by external contributors. These data provider communities will also benefit from training in FAIR practices. To promote this, tutorial sessions will be organized in connection to domain-specific conferences.

First face-to-face training event

Here we describe the first face-to-face training event targeting data centre staff, held on February 5, 2020 in connection to the ENVRI Week meeting in Dresden, Germany.

Planning and preparations

Target groups

Data centre staff (IT people with good grasp of metadata as such, but much less knowledge of ontologies, linked open data, vocabularies etc.)

Choice of theme

The gap analyses performed by WP6 Task 6.1, and reported on in deliverable D6.1 [Hellström 2019], indicated that ontologies, terminologies and semantic linking are topics judged to be of high importance and priority by the RIs in ENVRI-FAIR. As some project partners (Environmental Agency Austria/eLTER and the TIB Leibniz Information Centre for Science and Technology among others) have considerable knowledge and competence in these fields, we choose to build up the program of the first face-to-face ENVRI-FAIR training event around the theme “Terminologies for ENVRIs: why, what and how”.

Goals

The aim of the course was to give a good introduction into ontologies and how these can help “putting the I into FAIR”. After the course, participants should have a good understanding of key concepts of the Semantic Web and knowledge representation techniques, and be familiar with different examples of ontology & vocabulary portals.

Importantly, the course was intended to be relevant to all ENVRI-FAIR sub-domains (atmosphere, marine, solid earth, and terrestrial ecosystem/biodiversity).

Selecting instructors

Selecting the instructors for the training event, we wanted to have contributions both from experts that are familiar with the ENVRI community organisations and their terminology-related requirements, as well as from specialists from other disciplines that can contribute their perspectives. Based on a combination of their credentials and positive experiences from the 2018 and 2019 summer schools organised by LifeWatch ERIC and ENVRIplus, we decided to invite Clement Jonquet, Barbara Magagna and Markus Stocker as teachers:

Dr. Clement Jonquet (LIRMM, Univ. of Montpellier), PhD in Informatics, Associate Professor and former postdoc & visiting scholar at Stanford University, has 12-year experience in ontologies and semantic Web research applied to biomedicine and agronomy. He works on the design and development of ontology repositories, ontology-based services especially semantic annotation as the PI of ANR-JCJC & H2020-MSCA supported SIFR project (http://www.lirmm.fr/sifr), co-PI of ANR PractiK-Pharma (http://practikpharma.loria.fr) and PI of ANR D2KAB (http://ww.d2kab.org). Since 2015, he gathers the national and international agronomy community around the AgroPortal initiative to build a reference vocabulary and ontology repository for agronomy, food and plant sciences.

Barbara Magagna: holding a master degree in landscape planning and in geoinformatics has 14 years of experience in ontology engineering and process facilitation, abilities she could apply in several national and European semantic projects. She had been working for the University of Natural Resources and Applied Life Sciences, the University of Vienna and since December 2007 for the Environment Agency Austria (Umweltbundesamt GmbH) where she undertakes the function of a semantic analyst and database designer. She was involved in the development process of terminologies like SERONTO and EnvThes. She has experience as work package lead related to data management, in the design of UML models and XML schemas in the air quality data reporting area, in the design of semantic models in projects related to Environmental Research Infrastructures (ENVRI). She is now co-chairing the RDA Working Group I-ADOPT which aims at producing an interoperability framework for the semantic representation of observable properties.

Markus Stocker: Markus leads the Knowledge Infrastructures research group at the TIB Leibniz Information Centre for Science and Technology. He holds a PhD in Environmental Informatics from the University of Eastern Finland; a MSc in Environmental Science from the University of Eastern Finland; and a Diploma (MSc) in Informatics from the University of Zurich, Switzerland. His research interests lie at the intersection between research infrastructures and research communities, and how such infrastructures acquire, maintain, and share scientific knowledge about human and natural worlds. Prior to TIB, he held a postdoctoral research associate position at PANGAEA, the Data Publisher for Earth & Environmental Science, at the MARUM Center for Marine Environmental Sciences, University of Bremen, Germany. As a member of the Research Data Alliance (RDA), he is involved in various groups, in particular the WG Persistent Identification of Instruments.

Training contents

The details of the training contents were decided by the instructors in dialogue with the WP6 core team. Because of the time limitation – a maximum of 3 hours – it was only

possible to scratch the surface of what is a very broad and complex topic. The instructor team therefore settled on a combination of introductory lectures and teacher-led demonstrations.

The lectures introduced the basics about terminologies and the Semantic Web, covering:

- The semantic gradient (taxonomies, thesauri, ontologies)
- Knowledge representation languages (RDFS, OWL)
- Basic features of terminologies (classes, properties, assertions, etc.)
- Lightweight exercise in Protégé (building a small environmental terminology)

This introduction was followed by presentations and demonstrations of how to use ontologies (and other semantic resources) through domain-specific ontology repositories such as the BioPortal, AgroPortal and EcoPortal, including:

- Ontology selection and recommendation
- How to use an ontology in the repository
- Semantic annotation of text data
- Ontology alignments management
- Automatic access to ontologies within the repositories (SPARQL & REST)

Implementation

Advertising the event

With the help of the ENVRI-FAIR communications team, the event was advertised on the ENVRI-FAIR web site (https://envri.eu/event/envri-fair-first-training-event/), and was also communicated via relevant social media accounts. In addition, information about the event was spread outside of the project to both ENVRI Community organisations and external networks, including research data management-related mailing lists. Please see the Appendix for examples of images used to advertise the training event.

Selecting the date and venue

The training took place on February 5, 2020 as part of the ENVRI Week 2020 collaboration meeting (https://envri.eu/event/next-envri-week/) held at Hotel Elbflorenz, Dresden. The room was the largest one available to the meeting, with a total seating capacity of about 150 persons. The on-site equipment included dual projectors, as well as a professional-quality audio system with loudspeakers and microphones.

In addition to this in situ facility, it was decided to also provide real-time streaming of the entire event to remote participants. This was accomplished by means of a Zoom video meeting, coordinated and hosted by Lund University. The Zoom software also supported the recording of the meeting, which was stored on two separate computers in Lund.

Participants

In order to estimate the potential number of participants, we asked people to register beforehand by sending an e-mail to the ENVRI-FAIR project office. Registrants were asked to provide contact details, as well as to indicate their affiliation and whether they planned to attend in person or via video link. A total of 54 registrations were received, indicating a strong interest in the theme of the training. Around 45 persons in total attended the training session: about 35 persons physically present and about 10 persons connecting remotely via Zoom. The majority of the participants came from research infrastructures

participating in ENVRI FAIR, with the rest and the others were from universities and other research related organisations.

Post-event activities

Disseminating course materials

All the PowerPoint slide decks shown at the event were converted to PDF format. To simplify access to the individual teachers' presentations, the recording made from the Zoom session was split into parts and saved as MP4-format video files. All these materials were uploaded to the open-access folder at https://fileshare.icos-cp.eu/s/oq8GDMMqgP7CRm2 (read access to anyone with the link).

A corresponding entry was made in the ENVRI Community training catalogue (see Milestone report MS22 for details), see https://trainingcatalogue.envri.eu/course/40.

Event evaluation

After the event, a course evaluation form of 8 questions were sent out to all participants. The evaluation also included a reminder of the availability to the recording of the training. The feedback from those who have answered the evaluation was in general very positive. They found the theme of the training relevant, and the presentations and the lecturers both received high scores.

Lessons learned

The Terminologies training event was the first one in a series that will be organised in the coming years under the ENVRI-FAIR label. In order to serve the ENVRI community in the best possible way, a high priority for WP6 will be to keep a high level of quality, both when it comes to the topics covered and the methods in which these are presented. Thus, it is very important to appraise all training events, looking carefully at what worked well and where improvements may be made.

Some of the most important "lessons learned", that we will take into consideration when going forward, are:

- Combining face-to-face and virtual meetings is a big challenge, especially regarding ensuring adequate sound quality. We recommend that careful attention is paid to the technical equipment that is already present at a considered venue, and that provisions are made to install loudspeakers and microphones that can be used by the in situ audience.
- Coordination among instructors is essential to avoid gaps and/or overlaps. In our case, this was achieved by video meeting and exchange of draft materials via e-mail in the weeks leading up to the event.
- Zoom is a good platform for virtual meetings, but creating recordings that can be distributed for later viewing requires careful setup and preparation. In addition, the learning curve for how to best use video editing software packages can be quite steep, especially for those programs that offer rich functionality.

Appendix

Images produced by the ENVRI-FAIR communications team and used for promoting the "Terminologies for ENVRIs: why, what & how" training event.

Figure 1: Advertisement banner used on the ENVRI Community web site and in newsletters and social media.

Figure 2: Infographics describing the training event, published on the ENVRI FAIR web site.