ENVRI^{plus} DELIVERABLE

Deliverable 15.1: ENVRI RM training materials, suitable for classroom and e-learning use

WORK PACKAGE 15 – Training, e-Learning and courses

LEADING BENEFICIARY: CARDIFF UNIVERSITY

Author(s):	Beneficiary/Institution
Alex Hardisty	Cardiff University
Abraham Nieva de la Hidalga	Cardiff University
Aurora Constantin	University of Edinburgh
Malcolm Atkinson	University of Edinburgh

Accepted by: Jacco Konijn (WP 15 leader)

Deliverable type: REPORT

Dissemination level: PUBLIC

Deliverable due date: 31. October.2017/M30

Actual Date of Submission: 20. November.2017/M31

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 654182

ABSTRACT

The ENVRI Reference Model is an enabling tool the environmental research community can use to secure interoperability between infrastructures, to enable reuse of common components, to permit sharing of resources, and to provide a common language of communication between those responsible for the design and construction of Research Infrastructures.

This deliverable document describes work undertaken and results obtained to characterise the target training audience and to design and prepare training materials for that audience in a form suitable for both e-learning and classroom use.

The work reported has been undertaken as part of Task 15.1 in Work Package (WP) 15; also part of Theme 5 "Knowledge Transfer". This theme ensures cross-fertilisation and knowledge transfer of new technologies, best practices, approaches and policies among the Research Infrastructures. The present report and the ENVRI RM itself should help Theme 5 achieve this.

Project internal reviewer(s):

Project internal reviewer(s):	Beneficiary/Institution
Robert Huber	University of Bremen
Jacco Konijn	University of Amsterdam

Document history:

Date	Version
04.09.2017	Outline of document structure for comments
09.10.2017	First almost complete draft of text for comments
17.10.2017	Draft for internal review
15.11.2017	Corrected version
16.11.2017	Accepted by Jacco Konijn (WP15 leader)

DOCUMENT AMENDMENT PROCEDURE

Amendments, comments and suggestions should be sent to the editors (Alex Hardisty HardistyAR@cardiff.ac.uk, Abraham Nieva NievadelaHidalgaA@cardiff.ac.uk, or one of the authors listed above.)

TERMINOLOGY

Glossaries are provided online here:
https://confluence.egi.eu/pages/viewpage.action?pageId=14452608 and here:
https://wiki.envri.eu/display/EC/Appendix+B+Terminology+and+Glossary and as Appendix 7.

PROJECT SUMMARY

ENVRIplus is a Horizon 2020 project bringing together Environmental and Earth System Research Infrastructures, projects and networks together with technical specialist partners to create a coherent, interdisciplinary and interoperable cluster of Environmental Research Infrastructures across Europe. It is driven by three overarching goals: 1) promoting cross-fertilisation between RIs, 2) implementing innovative concepts and devices across RIs, and 3) facilitating research and innovation in the field of environmental understanding and decision-making for an increasing number of users outside the RIs.

ENVRIplus aligns its activities to a core strategic plan where sharing multi-disciplinary expertise will be most effective. The project aims to improve Earth observation monitoring systems and strategies, including actions to improve harmonisation and innovation, and generate common solutions to many shared information technology and data related challenges. It also seeks to harmonise policies for access and provide strategies for knowledge transfer amongst RIs. ENVRIplus develops guidelines to enhance trans-disciplinary use of data and data-products supported by applied use-cases involving RIs from different domains. The project coordinates actions to improve communication and cooperation, addressing Environmental RIs at all levels, from management to end-users, implementing RI-staff exchange programs, generating material for RI personnel, and proposing common strategic developments and actions for enhancing services to users and evaluating the socio-economic impacts.

ENVRIplus is expected to facilitate structuration and improve quality of services offered both within single RIs and at the inter-RI (European and Global) level. It promotes efficient and multi-disciplinary research offering new opportunities to users, new tools to RI managers and new communication strategies for environmental RI communities. The resulting solutions, services and other project outcomes are made available to all environmental RI initiatives, thus contributing to the development of a coherent European RI ecosystem.

Blank page

TABLE OF CONTENTS

TABLE OF FIGURES

TABLE OF TABLES

Executive summary

The ENVRI Reference Model (ENVRI RM) is an enabling tool – a framework, a 'context for participation' – that helps the Environmental Research Infrastructures (RI) to achieve their shared goals of cross-fertilisation, harmonisation and innovation between RIs; generating common solutions to many shared information and communications technology (ICT) and data related challenges. The ENVRI RM achieves this by offering a defined vocabulary for communication between participants, and a set of conceptual relationships between those terms as the basis for creating a common understanding and language for discourse. Addressing a wide range of RI needs, it supports data identification, cataloguing, generation of data products, publication, citation, semantic harmonisation, discovery and access, provenance, processing and use. However, ENVRI RM is complex, largely alien to the intended target audience and there is a need for training.

To better understand users' needs related to about the RM, the concept of persona has been adopted. A 'persona' represents a specific target audience category, capturing their motivations, goals, attitude, frustrations; in other words, the "essence" of who they are and thus, what their training needs may be. There are 3 personas identified up until now among the potential RM users: i) RI Professional, committed to ensuring a reliable and effective RI meeting the needs of its users; ii) RI Systems Engineer, designing and developing sustainable and easily used systems; and iii) RI Strategist / Manager, managing/improving their RI and aligning it with other research infrastructures.

Creation of e-learning materials generally follows one of three main patterns: i) presenting physical classroom-based content over the internet (information dump); ii) delivery of content structured as small lessons (piecemeal) followed by reinforcing activity (quiz); or iii) providing content based around a specific problem or case-study, with practice and experience exercises for students (active learning). Training materials related to the ENVRI RM follow the latter approach.

Two specific training courses have been produced. These are each of 10 lessons and are suitable for both classroom and e-learning delivery. The first, addressing the RI Professional persona provides information on the value of modelling and an introduction to the basic concepts of the ENVRI Reference Model. The second, situated in a specific real-world case study teaches specific skills to use the ENVRI Reference Model and is aimed at RI Systems Engineers.

For mainly practical reasons, training materials have been produced using the widely-adopted Xerte toolkit for producing multi-media teaching and training content. Learning objects created with Xerte can be easily shared and repurposed or exported for off-line use (and preservation) in a variety of standard formats, including for delivery via the ENVRI Community training platform.

The materials have been used at different stages with different partners in the ENVRIplus RI site visits where some specific training events have been held. During these visits, participants have been invited to try out some of the lessons and asked to provide feedback about them. This feed has been generally positive and has helped to refine the training content.

1 Introduction

The ENVRI Reference Model (ENVRI RM) is an enabling tool – a framework, a 'context for participation' – that helps the Environmental Research Infrastructures (RI) to achieve their shared goals of cross-fertilisation, harmonisation and innovation between RIs; generating common solutions to many shared information and communications technology (ICT) and data related challenges. The ENVRI RM achieves this by offering a defined vocabulary for communication between participants, and a set of conceptual relationships between those terms as the basis for creating a common understanding and language for discourse.

Addressing a wide range of RI needs, it supports data identification, cataloguing, generation of data products, publication, citation, semantic harmonisation, discovery and access, provenance, processing and use.

However, it is complex, largely alien to the intended target audience and there is a need for training.

The present report describes work undertaken and results obtained to characterise the target training audience and to design and prepare training materials for that audience in a form suitable for both e-learning and classroom use.

2 Concordance with the task description

2.1 Work Package objective

The work reported in the present document is part of Task 15.1 in the Description of Activities of the ENVRIplus project. The main objective of the Work Package is to "*organize and develop concrete training programs, focussing on the development of training modules, e-learning environments and specific courses for the main stakeholders in the Environmental RI community".*

2.2 Task description

Work Package 15, Task 15.1 has the title "Teaching RI operators key skills". The task is split in two parts; respectively covering training materials and activities for the ENVRI Reference Model and for building e-Infrastructure environments. The task description is the following:

> *"This task will focus on two key technical topics to be taught to the technical operators of the RIs involved in ENVRIPLUS.*
>
> *1) The ENVRI Reference Model:*
>
> *The ENVRI Reference Model is used to establish both common grounds for communication, and to provide a framework for a common way to build / integrate environmental research infrastructures. In collaboration with Theme 2 (Reference Model development), a series of training sessions and virtual training packages will be developed for ESFRI RI and I3[1] staff members to help understanding the practical application of the Reference Model. The task aims to prepare combined training materials, suitable for both classroom and distance learning use (i.e., write once).*
>
> *2) Building e-Infrastructure environments:*
>
> *(text on this aspect omitted here)*

[1] In this context, "I3" refers to the European Commission's "Integrating Activities" style of projects i.e., those that are comprised of networking activities, trans-national and/or virtual access activities, and joint research activities; with the aim of consolidating emerging research infrastructure communities.

Training will be delivered in face-to-face events, remote content for Webinars, Online content for self-paced learners, or MOOCs, whatever is most appropriate in given circumstances."

The present report explains the materials and activities associated with training for the ENVRI Reference Model[2]. Section 3 (page 11) describes the approach to designing and preparing training materials, including audience profiling, styles of training and the design of the courses, the authoring and delivery method and the case study used to situate some of the training content. Section 4 (page 16) explains the planned courses and their syllabuses, while section 5 (page 20) describes the materials themselves and initial testing and validation from organised training sessions. Section 6 (page 22) provides an outlook and next steps. Appendices 1 – 5 (page 25 onwards) provide supplementary information.

2.3 An introduction to the ENVRI Reference Model

All research infrastructures (RIs) for environmental sciences (the so-called 'ENV RIs'), although very diverse have many common characteristics. The ENVRI Reference Model (ENVRI RM, RM) exists to illustrate these common characteristics to provide a common language and understanding, to promote technology and solution sharing and to improve interoperability.

Like the confusion caused by multiple languages among the people in the biblical Tower of Babel story[3], when using different languages in specific ENV RI domains, it is difficult for people from different RIs to communicate with one another. This can also be a problem even between different sub-communities within a single RI.

Thus, ENVRI RM offers a common language, a common vision, uniform framework, and standard solutions. Therefore, it enables the achievement of a higher level of understanding, and interoperability. It allows reuse and sharing of resources. In addition, it helps identify duplicated and missing actions, as well as bottlenecks and areas for improvement. The ENVRI RM also bridges the world of users (scientists) and their work with the underlying technical services and their associated specifications.

The ENVRI RM is based on a typical data lifecycle model with five phases:

Data Acquisition: The RI collects raw data to be stored and made accessible within the infrastructure.

Data Curation: The RI stores, manages and ensures access to all data-sets produced within the infrastructure.

Data Publishing: The RI enables discovery and retrieval of scientific data by internal and external parties.

Data Processing: The RI provides a toolbox of services for a wide variety of data processing tasks.

Data Use: The RI supports users of an infrastructure to gaining access to and use data, and facilitates the preservation of derived data products (e.g., results of analysis).

Each ENV RI supports the data lifecycle to a different extent. According to the scope of a specific RI, its core activities align strongly with some of the phases while other phases are not so comprehensively supported.

Each lifecycle phase defines a set of functionalities, found to be common to multiple RIs. These functionalities, identified as a 'minimal model' defining the information and communication technologies (ICT) of a typical environmental RI are expressed from different perspectives or

[2] The second part of the task has been reported by partner 37 EGI.eu in deliverable D15.4.
[3] https://en.wikipedia.org/wiki/Tower_of_Babel

viewpoints[4]. Together they provide a basis for RI infrastructure providers to understand the critical elements, especially with respect to achieving interoperability between RIs.

Science Viewpoint (SV): The enterprise or Science Viewpoint of the ENVRI RM captures the functionalities of an environmental RI from the perspective of the people who perform their jobs and achieve their goals as mediated by the infrastructure. Modelling in this viewpoint derives the principles and properties of model objects through the analysis of the structure and functionality of communities, people interacting within and around those communities, and rules governing their interactions. This reveals the human behaviours that the RI's systems must support and potentially automate.

Information Viewpoint (IV): The goal of the IV is to provide a common abstract model for the shared research data handled by the infrastructure. The focus lies on the data itself, without considering any platform-specific or implementation details. It is independent from the computational interfaces and functions that manipulate the data or the nature of technology used to store it. The term 'data' is used in this viewpoint in a wide sense, meaning not only scientific data collected as part of the scientific process but also other kinds of artefact, such as encoded workflow representations, provenance trails, metadata, etc. that can all be managed as computer data.

Computational Viewpoint (CV): A research infrastructure (RI) provides a context in which investigators can interact with scientific data in a principled manner. The CV concerns the design of the data processing, analytical, modelling and simulation processes and applications provided by the system.

Engineering Viewpoint (EV): The EV tackles the problem of how to distribute functionality across a physically distributed research infrastructure. It gives prescriptions for supporting the computational interaction required within the system in a range of different concrete situations and answers how different middleware elements should work together.

Technology Viewpoint (TV): The TV concerns real world constraints, such as restrictions on the facilities and technologies available to implement the system and the existence of the relevant standards applied to existing computing platforms on which the computational processes must be executed.

2.4 Further reading

For further information about the ENVRI RM, see:

- The on-line training content reported herein (links in Appendix 1, page 29);
- The ENVRI Reference Model, on-line at http://envri.eu/rm. The sections on "Getting Started", "Introduction" and "Model Overview" provide a comprehensive introduction to the model and its concepts.

For further contextual reading, see:

- An explanatory article on Reference Model guided system design [Zhao 2015];
- A book chapter on computational challenges in global environmental research infrastructures [Martin 2017].

2.5 Abbreviations used in the present document

CV Computational Viewpoint

[4] The ENVRI RM is structured according to the Open Distributed Processing (ODP) standard (ISO/IEC 10746), and is defined through five different perspectives or 'viewpoints'. The Science (or Enterprise), Information and Computational viewpoints have been comprehensively implemented and refined through several releases of the ENVRI Reference Model (versions 1, 1.1, 2.1, and 2.2). First versions of the Engineering and Technology Viewpoints have been released since version 2.2, October 2017.

EV[5]	Engineering Viewpoint
IV	Information Viewpoint
RI	Research Infrastructure (plural: RIs)
RM	Reference Model
SV[5]	Science Viewpoint
TV	Technology Viewpoint

3 The adopted approach to training materials

3.1 Criteria for creating engaging learning activities

The task description (section 2.2 above) states that training packages will be developed for ESFRI RI and project staff members to help them understand the practical application of the ENVRI Reference Model. We considered who are the principal and secondary audiences for the training materials. We carried out semi-structured interviews to identify possible different personas wishing to make use of the ENVRI Reference and we used these to establish an audience profile. Section 3.2 below discusses this in more detail.

We also considered styles of training and the delivery method, noting the task requirement to prepare combined training materials, suitable for both classroom and distance learning use (i.e., write once). This is further discussed in section 3.3 below.

Additionally, to really create engaging learning activities, we concluded that so far as is practical within constraints of time and resources, training materials should be based around a realistic real-world scenario that is contextually relevant. This should lead the student towards meaningful decision making and meaningful feedback based on the real-world results of their design decisions. This is further discussed in Appendix 3 (page 30).

3.2 Audience profiling

3.2.1 Principal audience and use of personas

The principal target audience for the RM has previously been identified as the ICT experts within RIs having responsibility for the architectural design and implementation of their RI and for its interoperations with other RIs. These persons are typically systems architects, designers, integrators and engineers. Secondary audiences include: i) research leaders steering the long-term balance of investment in each RI, who will appreciate the savings from shared solutions and the benefits of inter-operation; and ii) third-party solution or component providers wanting to understand how to shape their offerings to gain adoption.

As pointed out in deliverable D5.2 [Hardisty 2017] a range of instruments – patterns, guidelines for use, training materials, learning ramps, case studies, engineering tooling – are all needed to facilitate familiarisation, training, engagement and uptake. Several different steps are needed, to make the complexity of the RM accessible and to demonstrate its value at both theme-expert level within ENVRIplus project and at the technical-practice level in RIs. Preparation and delivery of e-training material is just one step.

To better understand users' needs related to the communication about and learning of the RM, the concept of persona from Human-Computer Interaction (HCI) studies has been adopted. A

[5] Note that formally in ODP, the abbreviations NV and EV are used to denote the Engineering Viewpoint and Enterprise Viewpoint respectively. However, in adaptation to the ENVRI context we use EV to denote the Engineering Viewpoint and SV to denote the Science Viewpoint (see previous footnote also). The Science Viewpoint is equivalent to ODPs Enterprise viewpoint, however, the decision to change the name addresses the need to reinforce the idea of the ENVRI RM supporting scientific research communities.

'persona' is a representation of a specific audience segment for a product or a service [Dix 2004]. It represents a specific category of users, and captures their motivations, goals, attitude, frustrations; in other words, the "essence" of who they are. The concept of persona is used in Human-Computer Interaction (HCI) and User Experience (UX) design as a tool to focus on users' needs, desires and skills when designing a product or service for them.

The benefits of using the persona concept in the design process include:

- Sharing a common understanding of various audience groups among the design team members;
- Guiding the proposed solution to meet the needs of individual user personas;
- Putting a human "face" on the description to create empathy for the personas;

There are several perspectives regarding personas. The one adopted here is the goal-directed perspective, which is a framework for discussion and an efficient psychological tool for looking at problems while considering users' unique characteristics. In this perspective, the focus is on the user's goals, including characteristics of their typical workflow, context and attitudes. A persona description in this perspective could capture any of or all the following attributes:

- A person's motivations;
- A person's goals;
- A person's current pain points or frustrations;
- A short biography, giving relevant information about their background;
- A person's technical ability along with which devices they use and how often;
- Some demographic data, such as age, location, sex, etc.;
- A quote that characterises their attitude; and,
- A picture that acts as a visual token for the persona to encourage empathy.

3.2.2 Identifying useful personas

We carried out an extensive series of semi-structured interviews and questionnaire analysis in conjunction with representatives of RIs during site visits to several RIs (Appendix 1).

In general, the ENVRI RM should be of interest to different research infrastructure (RI) personnel and to computer scientists (CS). Up until now, 3 personas have been identified among the first of these target audiences:

RI Professional: Philip is committed to making his work for his RI accurate, reliable and effective. When changes are needed he must balance helping with their implementation against meeting his current commitments.

RI Systems Engineer: Edward needs to communicate well with his users and other stakeholders as he tries to meet their priorities, comply with regulations and negotiate for resources. He wants help finding the best ways of developing agreements that lead to a sustainable provision and use of ICT for his RI.

RI Strategist / Manager: Sabina is interested in building/improving their Research Infrastructure and aligning it with other EU research infrastructures.

Besides these three personas covering the users who work within the RI, we recognise another category (or categories) that includes people outside the RI (e.g., Computer Science researchers or IT providers who wish to shape offerings for RIs). The persona(s) describing such people is not within the ENVRIplus project remit as the available resources are fully committed to meeting RIs' direct needs. Future programmes of work developing and supporting the ENVRI RM may address training material for such users to build an open RM sustainable community.

3.2.3 Targeting training to different personas

3.2.3.1 Training targeted towards RI Professionals

Specific individuals in the role of the RI Professional – one of the two primary target personas for training – have a wide range of ability in computing and varied experience of ICT systems. For some, building computing systems is their principal skill. However, many are disciplinary scientists that have come to IT systems via indirect routes, as their scientific and professional careers forced engagement with such systems. For persons at this end of the spectrum, the technicalities of the ENVRI RM, its benefits in systems views of infrastructure and the systems-oriented thought processes needed to effectively exploit the RM are particularly challenging. This must be addressed by the training. Such training must carefully avoid the use of technical jargon.

3.2.3.1 Training targeted towards RI Systems Engineers

Individuals in the role of RI Systems Engineer persona have much more in-depth knowledge about computing and distributed ICT systems, and broader understanding of the technical challenges and trade-offs implicit in the architectural and technical design of such systems. For individuals identifying with this persona, deeper training is needed on each of the specific modelling concepts in the ENVRI RM, with examples of how that can be applied in real-world situations. Thus, we situate the lessons in a real-world case study context (see Appendix 3).

3.2.3.1 Training targeted towards RI strategist/manager and other personas

Training targeted towards the RI strategist/manager persona is deemed of lower priority, as the need can be initially satisfied with the training material aimed at the RI Professional persona.

Future work may design and create specific training materials for the RI strategist/manager and other personas as and when the need arises, and resources become available.

3.3 Styles of training and design of the course modules

3.3.1 Styles of training/learning design

Creation of e-learning materials generally follows one of three main patterns: i) presenting physical classroom-based content over the internet (information dump); ii) delivery of content structured as small lessons (piecemeal) followed by reinforcing activity (quiz); or iii) providing content based around a specific problem or case-study, with practice and experience exercises for students (active learning) [Cook 2015] [Moore 2016].

These three design strategies have been further explored by designing a brief introductory ENVRI RM course and exposing it using the three styles.

The information dump style is a very common structure for on-line courses. In this style, information is delivered to the student as a continuous stream of text (or talking) and illustrations. This approach relies on the student working their way through the content by reading (listening to) and absorbing it. For complex content, this is difficult, taking a long time to achieve learning outcomes[6].

In the piecemeal approach style[7], content is organised in 'bite-sized' pieces. This makes the material easier to comprehend than the simple 'information dump' style. However, students are bombarded with new content with little immediate opportunity to apply the acquired knowledge and reinforce the learning. This often leads to information overload and disconnect between learning and practice.

[6] Example of information dump style - https://xerte.cardiff.ac.uk/play_3821
[7] Example of piecemeal style - https://xerte.cardiff.ac.uk/play_3850

In general, the information dump and piece meal styles are common structures of typical on-line courses (as illustrated in the top part of Figure 1). The main characteristic is that they rely on the upfront delivery of content, and reinforcement at the end or after the teaching. Active learning is proposed as an alternative style (illustrated in the bottom part of Figure 1).

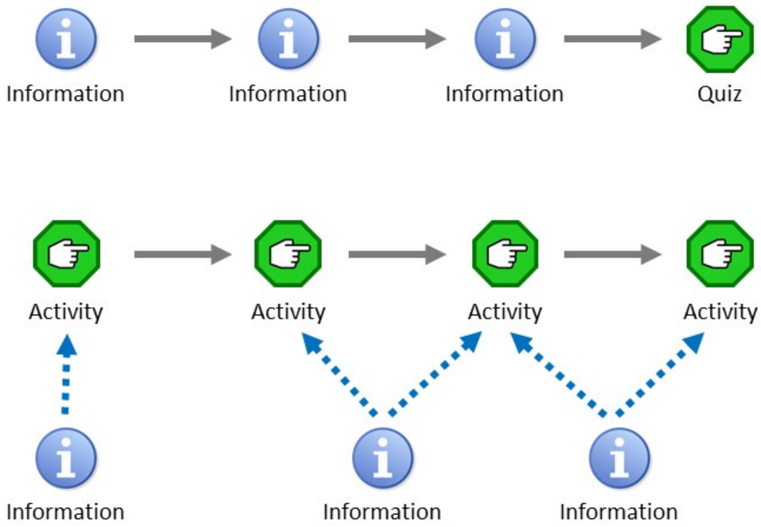


FIGURE 1 GENERAL STRUCTURE OF ON-LINE E-LEARNING COURSES.

ONLINE COURSES CAN PRESENT A STUCTURE THAT PUTS CONTENT AT THE FRONT AND LEAVES REINFORCING ACTIVITIES TO THE END (TOP PART OF THE DIAGRAM). ALTERNATIVELY, ACTIVE LEARNING EMPHASISES APPLICATION AND EXPLORATION REINFORCED BY DELIVERING INFORMATION AS NEEDED TO SOLVE THE PROBLEM AT HAND. ACTIVE LEARNING IS APPROPRIATE TO THE NEEDS OF THE ENVRI RM TARGET AUDIENCE.

In active learning courses[8], content is often organised around a specific problem or case-study. Delivery of information is interleaved with practice and experience exercises to reinforce understanding and learning points. When the case-study or problem addressed is highly relevant to the student's situation, learning occurs more quickly. Active learning styles often present opportunities for the student to refresh and re-prise their understanding by going back over and repeating parts of the material easily. Ease of navigation through the material becomes important.

The active learning approach offers the most advantages for designing online in-depth training materials suitable for an RI Systems Engineer persona wanting to learn in-depth about the ENVRI RM. Moreover, the material designed in this way can support blended learning. In blended learning, students both try the material for themselves and can interact face-to-face with instructors to solve real-world problems by applying the acquired knowledge [Singh 2003]. This can be supported because the relevant taught information is delivered in relation to the case-study or problem being addressed. The online materials provide opportunities for the student to refresh and reprise their understanding while solving problems[9].

The other styles are more appropriate for overview and introductory levels of material, such as those high-level ones aimed at the RI Professional persona.

3.3.2 Design to focus on target audience

The key to understanding how the ENVRI RM supports the development of an RI requires understanding how the five model viewpoints are integrated to describe RI systems as complex

[8] Examples of active learning style - https://xerte.cardiff.ac.uk/play_3877 and https://xerte.cardiff.ac.uk/play_3863

[9] In the present project and due to resource constraints, blended learning support has been provided to a limited extent, for example in post-training support to further the DASSH case study, EUFAR and EPOS work.

composite systems. The components of RI systems integrate a set of (sub)systems operating together to support the research data lifecycle. The main idea behind the viewpoints structure is to allow the description of a distributed system as a complex entity that needs to be understood by different groups of stakeholders [ISO/IEC 10746-n], [Linington 2012]. The use of the reference model allows practitioners to keep the design discussion centred at the right level. This allows moving from a high-level description of systems, understandable for researchers and sponsors to a lower design level for IT systems engineers and developers. Table 1 shows some mappings between some stakeholder types and different viewpoints of the RM. Note though, the level of knowledge needed about each viewpoint by different stakeholder types can vary.

TABLE 1 RELATIONSHIP BETWEEN STAKEHOLDER TYPES AND VIEWPOINTS

Stakeholder	Viewpoints	Stakeholder	Viewpoints
RI System Administrators	Computational Viewpoint Technology Viewpoint Engineering Viewpoint	Environmental Researchers	Science Viewpoint Information Viewpoint
RI Database Administrators	Science Viewpoint Information Viewpoint Computational Viewpoint Technology Viewpoint	IT/Tech Suppliers	Technology Viewpoint Engineering Viewpoint
RI System engineers, Programmers, Developers	Science Viewpoint Information Viewpoint Computational Viewpoint Technology Viewpoint Engineering Viewpoint	RI Professionals, RI Strategists / Managers	Science Viewpoint Information Viewpoint Computational Viewpoint

It is expected that, by using the ENVRI RM, RIs will be able to create a set of models that separate concerns neatly but at the same time keep the consistency of the RI systems, in ways that are accessible to all RI stakeholders. The design of the practical e-learning materials follows this general structure, providing materials that are accessible to a target audience composed of the different RI stakeholders.

3.4 Authoring and delivery method

3.4.1 Authoring content

For mainly practical reasons, we selected the widely-used Xerte toolkit[10] for producing multi-media teaching and training content for the ENVRI training platform [Ball 2008], [Hetrick 2011]. Xerte is a University of Nottingham initiative with goals to "*provide free, high-quality content authoring software that is easy to use for non-technical content authors; that provides best of breed accessibility, and which nurtures a positive and friendly community of users and developers*". The software has already been adopted for general use by Cardiff University, the main beneficiary authoring content in the present project. This gives access to resources and expertise that could be exploited in the present task.

In addition to the convenience of the existing support, using Xerte has the following benefits:

- **Uniform presentation**: Keeping a consistent presentation of content reduces distraction and facilitates easier understanding. Xerte allows reuse of contents and templates for presentations, including bespoke colour schemes, custom logos, and custom interactions.

[10] https://www.nottingham.ac.uk/xerte/

- **Fast construction**: Xerte provides tried and tested templates for learning materials. Xerte includes templates for presenting text, pictures, audio and presentations. In addition to this, it also includes pro-forma evaluation activities and quizzes that can be included into learning materials.
- **Collaborative development**: Xerte learning objects can be shared and edited by more than one author, facilitating review, maintenance and transference.
- **SCORM Ready**: Xerte supports exporting learning objects in SCORM format. SCORM, the Sharable Content Object Reference Model[11] is a collection of standards and specifications for web-based e-learning materials that is widely supported by many of the mainstream e-learning content authoring and delivery systems. This makes it an ideal way for integrating content developed in Xerte into the ENVRI Community training platform.
- **Durability**: Originally started by University of Nottingham in 2004, the Xerte Toolkit is a mature open-source community project of the Apereo Foundation[12], used by more than 100 educational institutions around the World. Learning objects created with Xerte can be easily shared and repurposed or exported for off-line use in a variety of standard formats. This latter point makes it easier to migrate content to other platforms, should that become necessary.

3.4.2 Delivering content

The ENVRI training platform[13] is supported by a deployment of the Moodle learning management system[14], deployed and maintained by LifeWatch Italy / University of Salento on behalf of the ENVRI Community. Using Moodle as the learning management system allows independent creation of learning objects as SCORM objects [ADLI 2004], which can then be integrated into lessons and courses structures in Moodle.

Using Xerte allows the development and delivery of content following a process in which the content can be developed and reviewed by partners, exported as a SCORM learning object, and then published in Moodle as part of a course. In this way, content can be made directly available for evaluation directly on the Xerte server. After going through a review and refinement process, the content can be exported as SCORM objects and published in Moodle, as part of a greater course or as independent learning materials.

At the time of writing the present report, beneficiaries are working together to design and build a landing page to serve as a common entry point to the ENVRI RM training materials within the Moodle platform. This landing page will guide learners to the appropriate content by allowing them to select topics, either through self-identifying with specific personas and being presented with relevant options for that persona, or more directly to specific training topics of interest to them. Appendix 2 provides an illustrative mock-up of what such a landing page could look like.

4 Planned courses and lessons

Starting from the three personas description explained above – specifically their goals, challenges and background (see Appendix 1) – we have developed suites of lessons (i.e., short courses), focusing first on the RI Professional and RI Systems Engineer personas.

We have identified that several levels of course material (Table 2) could be helpful over the longer-term, although resource limitations in the present project dictate we worked only on preparing material for the first two levels listed and described:

[11] https://en.wikipedia.org/wiki/Sharable_Content_Object_Reference_Model
[12] The mission of the Apereo Foundation (www.apereo.org) is to help educational organizations to collaborate to foster, develop, and sustain open technologies and innovation to support learning, teaching, and research. It has 85 members world-wide.
[13] https://training.envri.eu/
[14] https://moodle.org/

- The value of modelling and the Reference Model approach;
- Specific skills to make use of the ENVRI Reference Model;
- Agile modelling for resource limited teams; and,
- In-depth modules on specific RM aspects.

TABLE 2: FOUR LEVELS OF COURSE MATERIAL ABOUT THE ENVRI RM

The value of modelling and the Reference Model approach
An overview on why it is worth even thinking about becoming involved with the ENVRI RM. This is an introductory level course that would benefit those thinking of becoming familiar with the RM, and those supervising staff in that position.

Specific skills to use the ENVRI Reference Model
An introductory level course teaching the basics of the ENVRI RM (i.e., the core concepts), situated in a typical use case context. There are no pre-requisites for starting this course other than the student having a general understanding of the purpose and goals of research infrastructures, and a desire to learn about the ENVRI RM. In American parlance, this would be known as the "RM 101".

Agile modelling for resource limited teams
Focusses on how to use specific parts of the RM in an agile way to suit own circumstances and address specific problems. It's less about the RM details and much more about how to use it. Prior knowledge of the ENVRI RM is a pre-requisite.

In-depth on specific aspects of the RM
A suite of modules exploring specific aspects of the ENVRI RM in more depth. For example, on the use of correspondences, or on the incorporation of specific policies of the RI that must be reinforced through design.

Table 3 below shows the relevance of these different levels of training material to the defined personas, and indicates present availability.

TABLE 3: RELEVANCE OF DIFFERENT LEVELS OF TRAINING MATERIAL TO THE PERSONA

Training level	Personas		
	RI Professional	RI Systems engineer	RI Strategist/manager
The value of modelling and the Reference Model approach	Applicable (Available by end 2017)	Applicable (Available by end 2017)	Applicable (Available by end 2017)
Specific skills to use the ENVRI Reference Model	Not applicable	Applicable (Available now)	Partially applicable
Agile modelling for resource limited teams	Not applicable	Applicable (future work)	Partially applicable (future work)
In-depth on specific aspects of the RM	Not applicable	Applicable (future work)	Applicable (future work)

The thinking behind this multi-level proposal is that it is first necessary to help students (i.e., RI trainees) develop an understanding of the value and power of the RM approach before giving them basic training in skills they could use to exploit some of that power. Recognising that resource limitations are often a key concern for engineering teams, we then want to be able to help students move beyond basic modelling and into a more flexible and agile way of exploiting the RM for their own specific problems. Finally, there are in-depth specific aspects of the RM for which specialised or advanced training could be needed.

The first two training levels are elaborated further below.

- The value of modelling and the Reference Model approach (4.1);
- Specific skills to make use of the ENVRI Reference Model (4.2).

In each case, the scope of the training level is explained, together with the criteria to be met by the course / module design for that level. For the first two levels, worked on in the present project a specific, typical lesson syllabus and structure is given.

The third and fourth training levels are mentioned again as part of further course development (6.1.3, page 22 below).

4.1 The Reference Model approach and the value of modelling

A short introduction to the ENVRI RM is included in section 2.3 (page 9) above. This is like the starting point of the training module giving an overview of the Reference Model approach and the value of modelling. The module covers why it is worth thinking about becoming involved with the ENVRI RM and is an introductory level course that would benefit those thinking of becoming familiar with the RM, and those supervising staff in that position.

This introduction must be clear, concise and understandable, using as little technical jargon as possible. It must emphasise the concrete benefits and value of systems-oriented modelling and the role an RM plays in that. It must give simple but concrete examples, as well as forward references into the next level of training and tools available.

The lesson syllabus covers the following:

- **Lesson 1:** ENVRI RM Overview. The overview introduces the value of a Reference Model approach. It explains the alignment of the ENVRI RM with a typical research data life cycle, introducing the key phases of data acquisition, data curation, data publishing, data processing and data use. It introduces the key concept of 'viewpoints' and explains the 5 viewpoints that can be taken when modelling a system i.e., science (or enterprise) view, information view, computational view, engineering view and technology view.
- **Lesson 2:** Science Viewpoint – SV Community Roles. A role in a community identifies a behaviour performed any number of times concurrently or successively. After defining the community concept, this lesson introduces the SV community roles associated with each data lifecycle phase.
- **Lesson 3:** Science Viewpoint – SV Community Behaviours. A behaviour of a community is a composition of actions performed by roles normally addressing specific science requirements. In the ENVRI RM, the modelling of behaviours is based on analysis of the common operations of RIs, which has led to a list of common functions. This lesson introduces the SV community behaviours at each stage of the data lifecycle.
- **Lesson 4:** Information Viewpoint – IV Components. This lesson presents the IV components that are collections of information objects and action types necessary to meet typical set of required functionalities.
- **Lesson 5:** Information Viewpoint – IV Object Lifecycle. This lesson will present the IV Object Lifecycle, which is a description of how information objects change as the RI operates on them, illustrated by showing the effects of the actions.
- **Lesson 6:** Information Viewpoint – IV Management Constraints. The IV of the ENVRI RM provides the means for specifying constraints – the rules governing data management. This lesson will present the Information Management Constraints which should be considered to ensure the integrity and preservation of information objects.

- **Lesson 7:** Computational Viewpoint – CV Objects. The Computational Viewpoint (CV) accounts for the computational objects in an environmental RI. These are aligned with a generic architecture of the RIs. Lesson 8 will introduce CV objects for each stage of the data lifecycle.
- **Lesson 8:** Computational Viewpoint – CV Objects and Subsystems. The SV roles include five subsystem roles which support each of the phases of the data lifecycle. In this lesson, the models of those subsystems are described using CV components.
- **Lesson 9:** Computational Viewpoint – CV Integration Points. This lesson will explain how Integration points support the movement of research data between phases.
- **Lesson 10:** DASSH use case. The DASSH Data Archive Centre[15], which is a national facility for the archival of marine species and habitat data, will be used as a source of specific examples throughout the lessons, though other examples will be drawn from other RIs. This lesson will draw these together into a walk-through of the experience when DASSH was modelled using the RM. The reader will then appreciate how the model concepts and the relationships between viewpoints were used and better understand the benefits from using the RM.

4.2 Specific skills to make use of the ENVRI Reference Model

An introductory level course for systems engineers. This course teaches the basics of the ENVRI RM (i.e., the core concepts), situated in a typical case study context. There are no pre-requisites for starting this course other than the student having a general understanding of the purpose and goals of research infrastructures, and a desire to learn about the ENVRI RM. In American parlance, this would be known as the "RM 101".

The criteria for designing this course were:

a) Select a typical scenario / use case from the ENVRI Community to act as a case study;
b) Present a simple method for applying the RM in the context of the case study;
c) Progress from a high-level description of the RI to a detailed description of the RI processes, and the systems supporting them; and,
d) Maintain the consistency of the presentation by following the modelling method at each step.

The context for the lessons has been provided by the DASSH use case (see Appendix 3 (page 30)). This use case has been selected both because it is a comprehensive example of the use of the RM in a real-world setting (thus presenting and building on experience) and because the scenario itself is not atypical of the kinds of scenario for which the RM is intended i.e., for addressing matters related to interfacing between RIs and / or organising matters internally within an RI.

The method for applying the RM in the context of the case study consists of five steps: identify, model, refine, review, and map. During the identify step, the user needs to look up existing objects in the RM and tries to map them to objects/entities in their domain. In the model step, the user builds an initial model using the identified objects. In the refine step, the model is enhanced by adding, removing or renaming RM objects to better fit the specific case. In the review step, the model is analysed against the use case to determine whether it addresses the needs of the use case. In the map step, a new set of objects or viewpoint is chosen to proceed with the modelling in greater detail.

From this general approach, a lesson syllabus has been designed to provide a structured introduction to the main concepts of the ENVRI Reference Model in its 3 logical viewpoints, Science Viewpoint, Information Viewpoint and Computational Viewpoint:

- **Lesson 1:** Introduces the ENVRI RM using the DASSH use case, and the research data lifecycle.

[15] DASSH the archive for marine species and habitats data: http://www.dassh.ac.uk/

- **Lessons 2 – 5:** Cover modelling of the DASSH use case from the perspective of the Science Viewpoint, introducing the concepts of research infrastructure communities, behaviours, roles and artefacts.
- **Lessons 6 – 7:** Continue the modelling of the DASSH use case from the perspective of the Information Viewpoint, introducing the concepts of information objects and information actions.
- **Lessons 8 – 9:** Continue the modelling of the DASSH use case from the perspective of the Computational Viewpoint, introducing the concepts of computational objects and object configurations.

The first version of the content for this syllabus, covered in detail in Appendix 3, aligns to version 2.2. of the RM, released end of October 2017 [ENVRI RM V2.2 2017].

Note that the course structure is open, allowing for additional lessons to be added at the end e.g., for the Engineering and Technology Viewpoints once these have been more fully developed and released as part of the RM. Lessons on other topics within any of the RM viewpoints, or with more detail on existing topics can be added in the middle of the course as well.

5 The training materials

5.1 Description of materials and access

Training materials for the first level of training (i.e., that described in 4.1 above) is presently being produced using the Xerte authoring tool, and is expected to be published by end 2017.

Training materials for the second level of training (i.e., that described 4.2 above) has been produced using the Xerte authoring tool and published to the ENVRI Community Training Platform.

Access to course materials can be gained:

- Via the knowledge transfer / training link on the left side of the ENVRI Community Platform website: http://envri.eu/ (which re-directs to the training platform, see next);
- Via the ENVRI Community training platform: https://training.envri.eu/my/; and,
- Directly to the website of the authors of the training materials.

Note 1: For access to the course materials via the ENVRI Community Platform, an ENVRI Community account is presently needed to be able to log in. This can be obtained by registering at http://envri.eu/participate/.

Note 2: When directly accessing the authors' website, the specific URL is needed. No login is needed but no record of training progress is retained when accessing via this method. This method is normally used for quick access by authors, reviewers, etc. Entry point URLs are presently available as follows:

  - ENVRI RM approach and the value of modelling (for RI professionals): https://xerte.cardiff.ac.uk/play_6726.
  - Specific skills to make use of the ENVRI Reference Model (for RI system engineers): https://xerte.cardiff.ac.uk/play_5230.

The training material is linked from the ENVRI RM website as well.

5.2 Testing and evaluation

The testing and evaluation of materials thus far has been focused on the content the specific skills training and illustration of it with the DASSH use case. For this, persons from the Marine Biological Association (MBA) responsible for DASSH have reviewed and provided feedback on how well the materials represent the processes, systems and data assets managed by DASSH. This process is ongoing and continuous, as each lesson modified is reviewed again. Currently (at time of writing the present report), DASSH is trying the material with two interns who started working at MBA, Autumn 2017. This will help provide a fresh view on the structure, content and effectivity of the learning materials.

Some specific evaluation activities have been planned to be carried out during the ENVRIweek meeting in Malaga, 6 – 10th November 2017.

5.3 Organised training sessions

The materials have been used at different stages with different partners in the ENVRIplus RI site visits, where some specific training events have been held. During these visits, the participants have been invited to try out some of the lessons and asked to provide feedback about them. Table 4 shows the visits that have taken place, and the activities performed to apply the designed training materials of the Practical introduction to the ENVRI RM in classroom settings.

TABLE 4 SUMMARY OF ENVRIPLUS STAFF EXCHANGE VISITS

Site Visit	Training/Demo	Feedback	Result
LTER/EUFAR 7 - 9 March 2017, Vienna	Provided first lesson for evaluation to participants from EUFAR and LTER	**IT staff:** Basic and repetitive, overly simplistic, not relevant. **Scientists:** High Level approach, need further examples, need more details on use.	Helped in review of initial material. Revision of content to minimise repetition
SeaDataNet /AtlantOS 5 - 8 April 2017, Brest	Demonstrated how to use the CV to model RIs	**IT staff:** Already provided by simpler models (Box models) **Scientists**: RM makes clear common design elements across RIs when modelled side by side	Helped in creating linking lessons on IV to CV
EPOS 13 - 16 September 2017, Rome	Demonstrated how to map EPOS architecture to CV Objects in the ENVRI RM	**IT staff**: Already provided by simpler models (Box models) **Scientists**: N/A	Tried CV material Helped in complementing CV lesson on configurations
AnaEE 11 - 13 October 2017, Paris	First demonstration of complete training course	**IT staff:** Need to harmonise design decisions made in previous phase and new challenges in implementation phase **Scientists:** It will help in further development during the implementation phase of AnaEE	Tried material on the learning platform Provided feedback about accessibility. Conversation about harmonising current design decisions and challenges posed by implementation phase requirements
ICOS Dates to be fixed.		Pending	Pending

5.4 Outcomes from validation and ideas for further awareness building

The results of these validation opportunities have been helpful to improve the content of the training course. However, they have also reinforced previous feedback and experience related to the perceived usefulness and applicability of the RM. We have begun to address that aspect by soliciting additional ideas for how to communicate, disseminate and train about the ENVRI RM.

Appendix 4 (page 34) contains a synopsis of ideas, gathered during the third ENVRIweek project meeting that took place in Prague, 14-18th November 2016. These ideas have been used to drive progress during 2017. This has included refining the key messages to disseminate to the wider ENVRI community.

One of the most frequent suggestions idea during our study sessions with RI colleagues was raising the motivation for using the model and awareness about it. Some preliminary work has been done toward that idea, presenting successful stories of using the model within a short professional movie. Concretely, a storyline has been conceived to serve as a basis for a movie aiming at presenting two use cases: DASSH and EUFAR [see Appendix 5].

6 Outlook and next steps

6.1.1 Organising on-line sessions for further dissemination of ENVRI RM

The learning materials developed so far will help in the further dissemination of the ENVRI RM. They enable the organisation of formal courses or seminars. These seminars would follow a format in which people subscribe to the course on the training platform, access the materials independently, and are able to receive feedback from a nominated course leader, if required. The seminar could run for two weeks, allowing students to access the material within certain deadlines and for completing the review. After, this they would complete an evaluation and receive a certificate of participation provided by ENVRI.

6.1.2 Anchoring learning outcomes with revision summaries

An important question is: "How well do you help students (RI trainees) anchor and structure the RM material in their minds?" In planning face-to-face teaching, investing time in revisiting and revising past material helps with this. This tends to emphasise the structure of the taught information by grouping and by choosing key points or key terms from which other concepts and methods hang. Someone who, for example has studied the Science Viewpoint, studied the Information Viewpoint and started to use it, and now starts the Computational Viewpoint finds a need to understand a CV-SV relationship, and needs a quick revision. Also, some people know they learn things by revisiting summary material and short exercises. So eventually, revision summaries with quick short tests may be needed. This is future work when new resources are available.

6.1.3 Further course development

Further course development is future work when funding becomes available.

In section 4 above we identified 2 further levels of more advanced training where it would be desirable to develop content. These are respectively, agile modelling with the RM for resource limited teams, and more in-depth training on specific aspects of the RM. To be effective at an advanced level, these levels must assume prior reasonable knowledge of the ENVRI RM.

Agile modelling for resource limited teams should focus on how to use specific parts of the RM in an agile way to suit an RI's own circumstances and to address specific problems. It's less about the RM details and much more about how to use it.

In-depth training on specific aspects of the RM should comprise a suite of individual, self-standing modules exploring specific aspects of the ENVRI RM in more depth. Topics should include, for example the incorporation of specific policies of the RI that must be reinforced through design, and cataloguing. In the latter case, the aim would be to teach how to use the relevant parts of the

RM to achieve catalogue integration across RIs and / or with an overarching canonical super-catalogue.

These advanced levels of training can address the needs of the personas we observe outside the RI context. This will become necessary as the remit and support for a reference model based on the ODP standard is broadened to deliver long-term support to the current beneficiaries, the ENVRI RIs and to extend the benefits to broader communities, such as to all users of the EOSC. A global open source development community will need such training to induct new members or to enable existing experts to address new remits and to collaborate on a global scale to meet multi-disciplinary challenges. The cost benefits and improved trust in continued availability from such global sharing, perhaps under the auspices of RDA, will warrant investment in developing the frameworks, tools and training for this to take place.

6.1.4 Preservation of the learning materials

The sustainability and availability of the learning materials is enabled by the possibility of exporting these materials as SCORM learning objects. Up to this point, the training materials developed as part of task 15.1 have been made available for the ENVRI community on the online learning platform. The advantage of using Xerte, is that the SCORM learning objects can be uploaded again into Xerte for editing and modification. This means that if there is a need for updating or further revision, there is no need to access the current source files hosted in the Cardiff Xerte installation.

Additionally, the SCORM packages containing the lessons are compressed archives of the html files that make each lesson. As such it is also possible to publish the learning objects as data files in online archives (e.g., Zenodo) with instructions for individual download for offline learning. The provision of these preservation and sustainability measures is in line with the ENVRIplus Data Management Plan.

7 References

[ADLI 2004] Advanced Distributed Learning Initiative. "Sharable Content Object Reference Model (SCORM™)." Advanced Distributed Learning, http://www. adlnet. org (2004).

[Atkinson 2009] Atkinson, M., De Roure, D., van Hemert, J., and Michaelides, D. (2010) Shaping Ramps for Data-Intensive Research At UK e-Science All Hands Meeting 2010. Retrieved 9th October 2017 from https://eprints.soton.ac.uk/271235/.

[Ball 2008] Ball, Simon, and Julian Tenney. "Xerte–a user-friendly tool for creating Accessible Learning Objects." International Conference on Computers for Handicapped Persons. Springer, Berlin, Heidelberg, 2008.

[Braun 2006] Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative research in psychology*, 3(2): 77-101.

[Cook 2015] Cook, Jo, "How to design a training session for the online classroom". Inside Learning Technologies and Skills, November 2015.

[Dix 2004] Dix, A., Finlay, J., Abowd, G., and Beale, R. Human-computer interaction. 3rd edition. 2004. England: Pearson Education Limited. ISBN 978-0-13-046109-4.

[ENVRI RM V2.2 2017] ENVRI Reference Model V2.2, 30th October 2017.Retrieved: 15th November 2016 from https://wiki.envri.eu/download/attachments/14454504/EC-151117-1525-2990.pdf?version=1&modificationDate=1510756051428&api=v2.

[Hardisty 2017] Hardisty, A., Nieva de la Hidalga, A., Lear, D., Magagna, B., Atkinson, M., Jeffery, K., Martin, P. and Zhao, Z. Reference-model guided RI design, Technical report D5.2, ENVRIplus project, 208 pages, January 2017. http://www.envriplus.eu/wp-content/uploads/2015/08/D5.2-A-definition-of-the-ENVRIplus-Reference-Model.pdf.

[Hetrick 2011] Hetrick, Tony. "ICT International. An Overview of Free Course Authoring Tools " TechTrends 55.6 (2011): 11-12.

[ISO/IEC 10746-1] ISO/IEC 10746-1:1998. Information technology -- Open Distributed Processing -- Reference model: Overview.

[ISO/IEC 10746-2] ISO/IEC 10746-2:2009 Information technology -- Open distributed processing -- Reference model: Foundations.

[ISO/IEC 10746-3] ISO/IEC 10746-3:2009 Information technology -- Open distributed processing -- Reference model: Architecture.

[Linington 2012] Linington, P. F., Milosevic, Z., Tanaka, A., & Vallecillo, A. (2012). Building enterprise systems with ODP: an introduction to open distributed processing. CRC Press.

[Marting 2017] Martin P, Chen Y, Hardisty A, Jeffery K, and Zhao Z. Computational Challenges in Global Environmental Research Infrastructures. In: Terrestrial Ecosystem Research Infrastructures: Challenges and Opportunities. Eds. Chabbi A, and Loescher HW. 2017 CRC Press ISBN 978-1-49-875131-5. doi: 10.1201/9781315368252-13.

[Moore 2016] Moore, Cathy, "Action mapping: A visual approach to training design" 2016, Available from: http://blog.cathy-moore.com/action-mapping-a-visual-approach-to-training-design/

[Singh 2003] Singh, Harvey. "Building effective blended learning programs." Educational Technology-Saddle Brook Then Englewood Cliffs NJ- 43.6 (2003): 51-54.

Appendix 1: Personas – Preliminary results and questionnaires analysis

This appendix provides information and analysis about personas following a series of interviews with RI representatives from LTER (Vienna, 7-9th March 2017) and the marine RIs (Brest, 6-7th April 2017)). All the interviews shared the same general aims: To gather information useful to create intellectual ramps [Atkinson 2009] customised to each persona's characteristics; to make it easy to understand the ENVRI RM; to quickly familiarise with the concepts; and use the model in any RI for various purposes.

The semi-structured interviews were focused on the participant's familiarity with the RM, the way they were introduced to it, and their goals and confidence when using the RM, as well as their perceptions about the benefits and challenges of using the RM. In addition, the participants were asked to provide suggestions for improving the RM, both at the conceptual and presentational levels. The data collected was first analysed using the top-down approach of thematic analysis [Braun 2006], which started with a set of pre-determined topics drawn from the questions used for the interviews. New themes emerging from the data were then added in line with the bottom-up/inductive approach [Braun 2006].

Familiarity with the RM varied broadly, from a few days to 7 years. Most of the people were introduced to the model via ENVRI/ENVRIplus projects, with one exception. One of the participants became interested in the RM during his PhD studies, when trying to find a model for sensor-data collection.

The interviewees' main goals for using the RM can be summarised as:

- Modelling the RI in a standard way, so as to understand it and to ease the communication within their community;
- Improving the interoperability between RIs;
- Helping establish requirements as input to decision making; and,
- Guiding the RI construction from scratch or the addition of new components or services.

People emphasised the benefits of the model, such as providing a common language, a common vision and standard solutions, enabling higher levels of interoperability, reuse and sharing of resources, as well as identifying bottlenecks and areas for improvement. The challenges perceived by the participants were mainly related to the communication about the model and motivating people to use it (e.g., difficulty in finding an entry point, too much content, difficulty to see the short-term benefits), but also to the conceptual aspects of the model (e.g., unclear concepts and definitions, difficulty understanding the diagrams).

Regarding suggestions made, most of them were focused on improving the communication about the RM (such as challenges related to the conceptual aspects of the model; clear and simple examples; simplify the material). Some suggestions referred to the improvement of the model (make concepts more generic, especially in the computational viewpoint, complete the model), whereas others referred to raising the motivation and awareness about the RM (e.g., by emphasising the benefits for each group of users and organizing special sessions or workshops to apply the model).

These interviews led to identifying three potential personas (categories of user), which were further refined by brainstorming and questionnaire analysis during an 'ENVRIweek' meeting (Grenoble, May 2017) and further during EPOS meeting (Rome, 14-15th September 2017).

During the ENVRIweek meeting in Grenoble, people who were familiar or not familiar with the RM to various degrees were invited to participate in a brainstorming interview and to answer a questionnaire related to the model. The aim of the questionnaire was to validate the 3 identified personas, as well as to obtain more ideas to refine them; whereas the brainstorming aim was to collect suggestions and ideas with respect to the improvement of communication about the RM. An additional aim for both the questionnaire and brainstorming was to find out possible questions, comments, and concerns related to the RM.

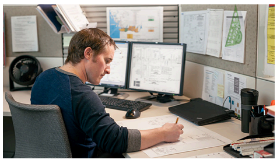
Data collected from questionnaires and stick-notes were analysed using thematic analysis in a way consistent with the semi-structured interview data. With only one exception, people identified themselves with one of the personas described based on the studies in Vienna and Brest. One person called himself a “software engineer”.

Several refinements were made to the persona descriptions. For example, some goals were added to each persona (e.g., “improving the interoperation with other RIs and providing services for other communities by making data available for RI Professional persona”; “identifying which services to outsource to e-infrastructures and what to develop/operate within the RI” for the strategist/manager persona). Also, some challenges were identified (e.g., “linking the RM with already existing RIs” for the CS specialist personas).

The suggestions for improving the RM were categorised into: tutorials, training, consulting, tools for design (e.g., templates instead of drawing tools), experiential reporting and case studies (e.g., clarification through success stories that can show clearly the value of the RM). They included improving the RM to make it more relevant (e.g., generalise meanings, uniform data and develop standards that communities can adopt without losing data accuracy)

From these results 3 personas can be identified with characteristics as illustrated in Figure 2 – Figure 4 below.

RI Professional
Philip Perrault
Age: 42
Location: Brest, France



Motto:
”Using a common language/ framework helps us finding our strong and weak points and improve our RI.”

Motivation

Philip is committed to making the aspects of his RI's for which he is responsible reliable and effective, with nothing going wrong in his part of the activities as he incorporates improvements whenever these are agreed by his community as justified by their research goals.

Goals

- Modelling the processes required or undertaken by his RI
- Finding a common language and understanding about how RI works to make the interaction easier within and outside the RI
- Identifying gaps in the organisation processes
- Identifying the strong and the weak points in their RI’s organisation
- Fixing the weak points
- Helping the RI to mature/develop
- Improving the interoperation with other RIs and providing services for other communities by making data available

Frustrations/Challenges related to the RM

- Lack of clear examples to guide him when applying the RM
- Difficulties with diagrams
- Difficult to see the benefits of the RM
- Difficult to start
- Difficult to allocate time while meeting his RI commitments

Background/IT skills

Education: Environmental Sciences

IT Skills: basic programming skills (e.g. writing scripts, using software for data analysis such as R, MATLAB, Python)

FIGURE 2: PERSONA RI PROFESSIONAL.

PHILIP IS COMMITTED TO MAKING THE ASPECTS OF HIS RI FOR WHICH HE IS RESPONSIBLE RELIABLE AND EFFECTIVE, WITH NOTHING GOING WRONG IN HIS PART OF THE ACTIVITIES AS HE INCORPORATES IMPROVEMENTS WHENEVER THESE ARE AGREED BY HIS COMMUNITY AS AS NEEDED THEIR RESEARCH.

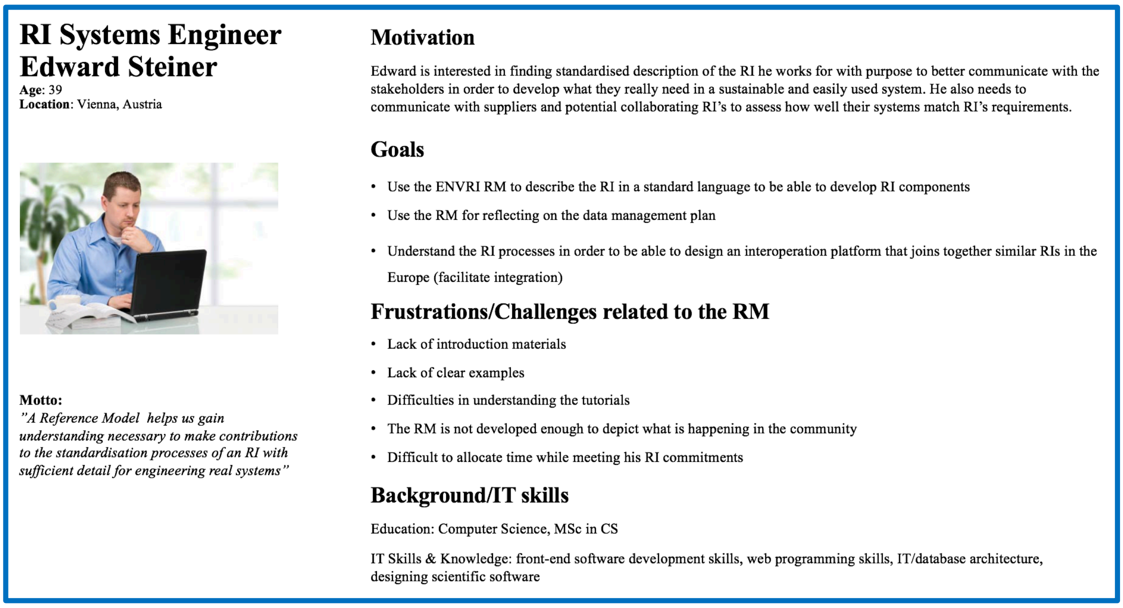


FIGURE 3: PERSONA RI SYSTEMS ENGINEER.

EDWARD IS INTERESTED IN FINDING STANDARDISED DESCRIPTION OF THE RI HE WORKS FOR WITH THE PURPOSE TO BETTER COMMUNICATE WITH THE STAKEHOLDERS TO DEVELOP WHAT THEY REALLY NEED IN A SUSTAINABLE AND EASILY USED SYSTEM.

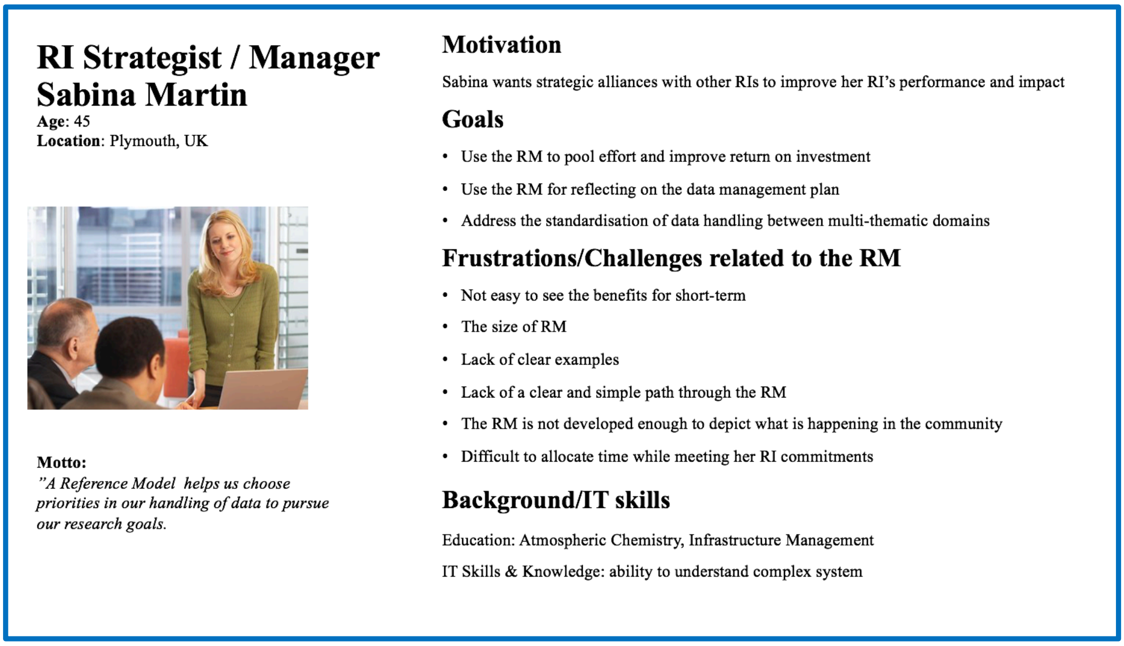


FIGURE 4: PERSONA RI STRATEGIST / MANAGER.

SABINA IS INTERESTED IN BUILDING AND IMPROVING HER RESEARCH INFRASTRUCTURE AND STEERING COLLABORATION WITH OTHER EU RESEARCH INFRASTRUCTURES TO HELP HER COMMUNITY GAIN EXTRA RESEARCH CAPABILITY THAT CN BE SUSTAINED FOR MANY YEARS.

A short ENVRI RM overview has been developed for the persona RI Professional based on the suggestions collected in LTER (Vienna, 7-9th March 2017), the marine RIs (Brest, 6-7th April 2017), and ENVRI week (Grenoble, May 2017) which was refined in a workshop with the participants from EPOS (Rome, 14-15th September 2017). This workshop aimed at: 1) understanding better whether the overview is appropriate for this persona; 2) getting feedback on the overview; 3) collecting suggestions to improve the overview and developing the tutorials for this persona.

The results were positive regarding the appropriateness of the overview for the RI professional persona. Only one participant considered it as being inappropriate. A list of suggestions was collected to help improve the overview (e.g., concrete benefits for the RI professional persona should be added, references to concrete tools, languages, etc. that make it possible to use this ENVRI RM for the tasks at hand). In addition, the discussion in Rome raised a new question: "Is there another persona(s) which describe computer scientists and systems providers who are not in an RI?". These questions led to other questions: "Is preparing RM training for them in the ENVRIplus remit?"; "And if it is should we allocate time and effort to developing RM training targeted for them?"[16].

[16] Serendipitously, they can already benefit from the examples and technical content of material prepared for RI persona, particularly that for the RI Systems engineers.

Appendix 2: Training website homepage

This appendix presents a mock-up of a possible design for organising access to collective training materials about the ENVRI Reference Model, to make them easier to find and relate to one another.

Some initial thought has been given to how to present the available training material for anyone wanting to learn about the ENVRI Reference Model (RM). Most learners will find the material they want at the time they want it, either by following a link they can easily find or have sent to them, or by conducting a search. We intend that they should land on a page that then helps them quickly select the material they need and head to it in one click.

Inspired by a well organised page for learning about data management developed by Edina, University of Edinburgh[17], one example of a possible design for an ENVRI RM training page is illustrated in Figure 5 below.

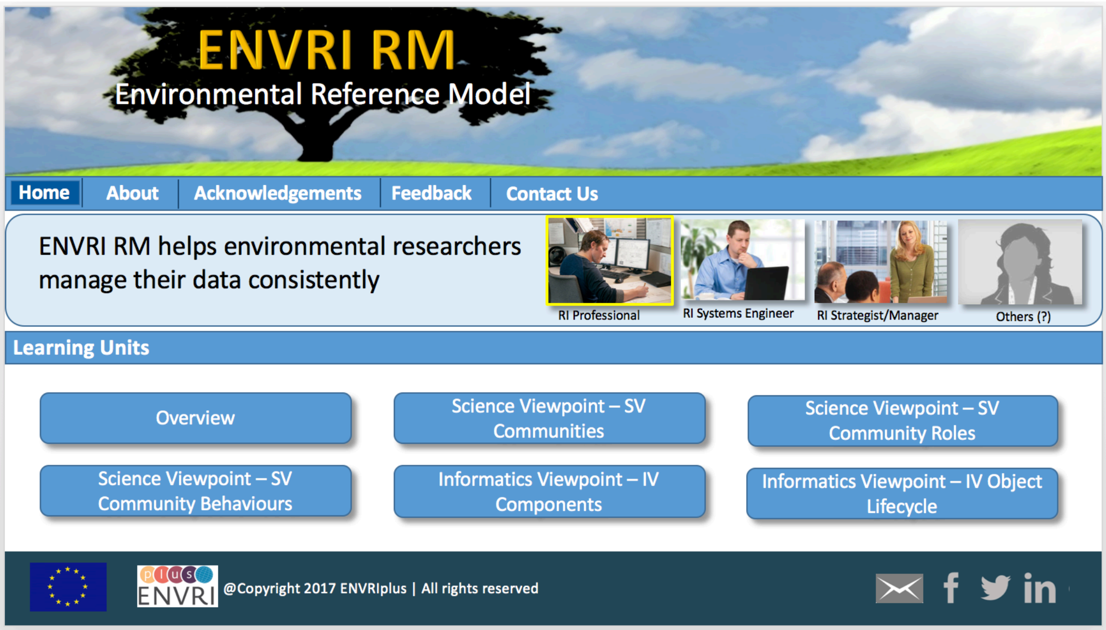


FIGURE 5: MOCK-UP OF A PAGE ORGANISING ACCESS TO ENVRI RM TRAINING MATERIALS

SELF-DIRECTED LEARNERS CAN USE AN APPROPRIATELY DESIGNED LANDING PAGE TO HELP THEM QUICKLY SELECT THE MATERIAL THEY NEED AND HEAD TO IT IN ONE CLICK.

As well as supporting instant self-selection by informative tokens and persona names, a landing page should also make learners subliminally aware of related material, so that they return later when they need it. This is also illustrated in Figure 5. When learners select a persona they identify with, the subset of material relevant to that persona and available for them pops up. Selecting a button takes them to it. Illustrated is the case where learners have identified themselves with the RI Professional persona.

[17] Mantra, http://mantra.edina.ac.uk

Appendix 3: Situating training in real-life case studies

This appendix explains the background to the case study example used in part of the training materials for the ENVRI RM.

The DASSH Data Archive

The DASSH Data Archive is hosted at the Marine Biological Association (MBA) and is part of the UK Marine Environmental Data and Information Network (MEDIN). Data are routinely supplied to the UK National Biodiversity Network, EMODNET Biology network and to the Global Biodiversity Information Facility (GBIF). MBA plans that DASSH e-Infrastructure will form part of the MBA's contribution to the European Marine Biological Resource Centre (EMBRC); a pan-European research infrastructure for fundamental and applied marine biology and ecology research.

In this context, MBA needed to review the current systems and working processes/practices of the DASSH Marine Data Archive[18] to be able to describe DASSH functions in a standard way, facilitating integration with other e-Infrastructures and assisting in identifying bottlenecks and areas for improvement. They recognised the possibilities offered by the ENVRI RM to help with this review.

However, given the previous lack of experience with Reference Models in any form, the initial challenge faced by MBA staff was to understand the overall concept of what a Reference Model represents, and the benefits of describing the DASSH infrastructure in this manner. Through participation as a partner in the ENVRIplus project, MBA could receive a comprehensive introduction to the RM, and how it can be applied. Later, they were supported in their analysis and modelling tasks by project experts.

The opportunity presented by this training and support scenario to both situate training content in a real-life practical example of how to make use of the ENVRI RM, and to use the scenario for testing and validation of the training approach was recognised early on in task T15.1.

Complementary feedback from survey and analysis work by task 5.2 also clearly delivered the message that more examples and use cases, illustrations, and tools linked to actual examples are important elements in ease of learning about the ENVRI RM.

Thus, an early decision was to situate the training material about the ENVRI RM in practical real-life examples, like the DASSH one.

Unpicking the DASSH scenario to design the training

DASSH infrastructure has grown organically since 2005 in response to technological changes, policy drivers and the requirements of funders. Staff involved in the DASSH Data Centre are typically marine biologists with some technical aptitude. They are not ICT specialists.

With basic understanding of the RM established by MBA staff, the first task was to pick a Viewpoint and begin to map existing data flow processes and systems to terms from the RM. The Science Viewpoint was selected as the most accessible point of entry, and the one that most closely resembled existing data flow and process diagrams developed by DASSH.

However, choosing the starting point of describing the infrastructure in the terms of the Science Viewpoint proved one of the hardest tasks. The available guidance does not help in the prioritisation or refinement of selecting a path through the viewpoints, and can leave the naïve user unsure or unclear of the optimal mechanism for transcribing an existing set of processes and behaviours into those terms available in the Reference Model. Whilst this lack of a prescribed methodology represents a strength and inherent flexibility for more confident users of the RM, it can leave new users with an intimidatingly "blank" canvas. Work on how best to render the DASSH infrastructure in terms of the RM was also needed. The lack of mature or easily accessible tools

[18] http://www.dassh.ac.uk/

necessitated an initial paper-based approach, with this subsequently being interpreted to an electronic form. It is unclear how to best represent the correspondences between viewpoints.

A further challenge, especially for a first-time user of the Reference Model is assessing the most appropriate level of detail at which to work. This should be informed by the challenge or tasks that the RM is being used to address. However, clearer guidance on both where to start and how much detail to include would be beneficial.

MBA opted to focus on the data publication elements of the DASSH data life cycle as this represents the interface with other e-Infrastructures. It is also the area where MBA have identified an existing bottle-neck in the data flow process. Once a starting point was selected, and with help from the RM team at Cardiff University, MBA split and transcribed the existing data flows into the corresponding RM viewpoints, moving from the Science Viewpoint to the Information and Computational Viewpoints.

Appendix 4: Practical Introduction to the ENVRI RM: Course details

The course "Practical Introduction to the ENVRI RM" is comprised of 9 lessons and is a structured introduction to the main concepts of the ENVRI Reference Model in its 3 logical viewpoints, Science Viewpoint, Information Viewpoint and Computational Viewpoint.

The course is built around a case study modelling a typical infrastructure activity; namely, the activity of making data / metadata from an archive available to users both directly and indirectly, via other partners' networks. The case study uses the example of a real archive – the Archive for Marine Species and Habitats Data (DASSH). On completion of the course, the student will have gained an understanding of the main concepts of the ENVRI Reference Model, and how the model can be used to carry out typical activities concerned with the design of infrastructure.

Course entry point: https://training.envri.eu/course/search.php?search=ENVRI+RM

The sections, titles, scopes and links to the 9 lessons are the following:

<u>Introduction</u>

One lesson introducing the ENVRI RM using the DASSH use case.

Lesson 1 – The ENVRI Reference Model and the research data lifecycle

Scope: Introduces the data lifecycle concept and its main phases, illustrating how this is used as the generic model that guides the use and understanding of the ENVRI Reference Model. Introduces the DASSH case study example, and a basic modelling process to enable the use of the models provided by the ENVRI Reference Model.

Link (ENVRI): https://training.envri.eu/mod/scorm/view.php?id=114
Link (direct): https://xerte.cardiff.ac.uk/play_5230

<u>Science Viewpoint</u>

The four lessons of this section continue the modelling of the DASSH use case from the perspective of the Science Viewpoint

Lesson 2 – Research Infrastructure Communities

Scope: Introduces the communities concept in the Science Viewpoint of the ENVRI RM and illustrates how the pre-defined communities in the RM can be selected and customised to the needs of specific infrastructures, including the definition of new communities.

Link (ENVRI): https://training.envri.eu/mod/scorm/view.php?id=115
Link (direct): https://xerte.cardiff.ac.uk/play_5714

Lesson 3 – Research Infrastructure Behaviours

Scope: Introduces the behaviours concept in the Science Viewpoint of the ENVRI RM and illustrates how pre-defined behaviours can be selected and customised to the needs of specific infrastructures, including the definition of new behaviours.

Link (ENVRI): https://training.envri.eu/mod/scorm/view.php?id=116
Link (direct): https://xerte.cardiff.ac.uk/play_6419

Lesson 4 – Research Infrastructure Roles

Scope: Introduces the roles concept in the Science Viewpoint of the ENVRI RM and illustrates how pre-defined roles can be selected and customised to the needs of specific infrastructures, including the definition of new behaviours.

Link (ENVRI): https://training.envri.eu/mod/scorm/view.php?id=117
Link (direct): https://xerte.cardiff.ac.uk/play_6421

Lesson 5 – Research Infrastructure Artefacts

Scope: Introduces the concept of artefacts in the Science Viewpoint of the ENVRI RM and illustrates how they can be defined for the needs of specific infrastructures.

Link (ENVRI): https://training.envri.eu/mod/scorm/view.php?id=118
Link (direct): https://xerte.cardiff.ac.uk/play_6427

Information Viewpoint

The two lessons of this section continue the modelling of the DASSH use case from the perspective of the Information Viewpoint.

Lesson 6 – Research Infrastructure Information Objects

Scope: Introduces the concept of information objects in the Information Viewpoint of the ENVRI RM and illustrates how they can be used to describe the information produced and consumed by processes in the DASSH case study example; also how they can be derived from the earlier Science Viewpoint models.

Link (ENVRI): https://training.envri.eu/mod/scorm/view.php?id=119
Link (direct): https://xerte.cardiff.ac.uk/play_6433

Lesson 7 – Research Infrastructure Information Actions

Scope: Introduces the concept of information actions in the Information Viewpoint of the ENVRI RM and illustrates how they can be used to describe the actions of the processes in the DASSH case study example; also how they can be derived from the earlier Science Viewpoint models.

Link (ENVRI): https://training.envri.eu/mod/scorm/view.php?id=120
Link (direct): https://xerte.cardiff.ac.uk/play_6437

Computational Viewpoint

The two lessons of this section continue the modelling of the DASSH use case from the perspective of the Computational Viewpoint.

Lesson 8 – Research Infrastructure Computational Objects

Scope: Introduces the concept of computational objects in the Computational Viewpoint of the ENVRI RM and illustrates how they can be used to model the processes in the DASSH case study example; also how they can be derived from the earlier Science Viewpoint models.

Link (ENVRI): https://training.envri.eu/mod/scorm/view.php?id=126
Link (direct): https://xerte.cardiff.ac.uk/play_6456

Lesson 9 – Research Infrastructure Object Configurations

Scope: Introduces the concept of object configurations in the Computational Viewpoint of the ENVRI RM and illustrates how they can be composed to model the sequences of behavours i.e., processes in the DASSH case study example; also how they can be derived from the earlier Science Viewpoint models.

Link (ENVRI): https://training.envri.eu/mod/scorm/view.php?id=127

Link (direct): https://xerte.cardiff.ac.uk/play_6572

Note that the course structure is open, allowing for additional lessons to be added at the end e.g., for the Engineering and Technology Viewpoints. Lessons on other topics of the RM, or with more detail can be added in the middle as well.

Appendix 5: Ideas to facilitate RIs engagement

This appendix contains a list of idea, generated during the third ENVRIweek project meeting that took place in Prague, 14-18th November 2016 for engaging the interest and commitment of the RIs technical experts to the ENVRI RM.

This appendix is organised into several categories of ideas, as follows:

- Primers, tutorials and practical examples
- Training, consulting, and helpdesk support
- Community building and self-help
- Tools for design
- Interaction more directly with the RIs
- KISS (Keep It Simple Stupid)
- Experiential reporting and case studies
- Focus
- Improve the RM to make it more relevant

Primers, tutorials and practical examples

Easy to understand 'primer' for the RM, to understand the basic concepts, with practical examples. = "How to start with it, simply". Easy to understand guides. Example based. Use of actual problems to illustrate.

More examples and use cases; illustrations, tools, linked to examples

Have a detailed list (glossary) with explanations of terms and acronyms.

Mapping the use cases to the concepts of RM.

Hold 30 minutes conversation with RM experts and see how RI business or community (n RI) fits in RM.

Designate some basic 'starting points' that allow users to see the correspondences arising between viewpoints for some core concepts without having to tackle the entire model at once.

Training, consulting, and helpdesk support

Consulting style training which shows targeted solutions.

Training workshops (physical or webinars) to introduce key concepts, terminology and barriers

Workshop for discussions.

Attend domain meeting to make sure developments are aligned.

Motivate the interest of the RI to the RM. If time is needed from them they must be convinced that they will benefit.

Training program, perhaps leading to certification

Supported from a helpdesk (ideally available 24/7).

Develop/deliver webinars that help people to understand patterns, tools, etc.

Community building and self-help

RM champions, within each RI one person with an understanding and experience.

Meetings (within RI staffs) to discuss how to make it easy to get started.

Make it easy to find and express the issues for broad range of users.

Online discussion forum. Google group to help build community and knowledge base / FAQ.

Centres of excellence e.g., this RI is good at acquisition and curation so it would be the poster boy for acquisition/curation integration.

Apply RM to an RI and gain immediate feedback from the group when problems pop up.

Tools for design

Use a tool with options to choose for each viewpoint. Visual tools.

Good example, visually enhanced on-line.

Share developed tools in github and try to converge to common code-base.

Share developed models in github and try to converge to common models.

In addition to UML, suggest to use less technical diagrams e.g., concept maps to avoid the message/impression that this is for IT people.

Provide ready-made templates that you think cover the kinds of topics they care about. Support these with tools.

Use linguistic notation as well as diagrams.

Good GUI plus case studies.

Develop a logic that is computable from the linguistic form of the RM.

Develop visualisation and consistency checking aids that depend on that logic, and cope with scale. Focus over the integrated viewpoints. People need to understand the integrated whole from the starting point they are at and to be able to drill down.

Create tools that allow RIs to build their descriptions incrementally, based on some underlying model.

Easy tool.

Interaction more directly with the RIs

Directly interact with the right people inside the different RIs.

Carry out a skills appraisal within RIs to identify where and whether system architects exist, thus to establish contacts and/or understand the gap.

Sit/lock the architects of different RIs in one room and collect their common ideas.

Speak to the domain specific experts and not only to the computer scientists.

Find out where RIs have problems with their current work and introduce tools, patterns and procedures that will be helpful.

KISS (Keep It Simple Stupid)

Keep it simple and abstract; details are for the implementation of RIs.

Focus on the simple but fundamental issues.

Set of basic ‘lego’ building blocks that connect with one ‘snap’.

Keep the documentation manageable to read and understand for groups from broad background.

Work towards a clean, simple structural RM with examples. It should be inviting.

Provide as many good examples as possible of applying the RM in different typical contexts; based on actual problems.

Describe processes / aspects that are critical/important to the people you are attempting to communicate with. In such descriptions, the terms link to concepts in the RM but you don’t need to say that. Then develop a dialogue refining what they do, and see how vocabulary and definitions develop.

Provide step-wise examples as technology will advance and examples will age/decay.

Template with different colours for each viewpoint. User must fill everything with help of online guidance. Compare with, for example DMPonline tool as an approach.

Experiential reporting and case studies

Ask non-RM 'insiders' to report on their experiences.

Cookbook of recipes for common scenarios.

How the RM can fit with already existing components of the RIs. The AtlantOS project/case is a very nice use case for this purpose. Similar to what DASSH use case is looking at.

AtlantOS use case is a typical data integration example/challenge that RIs are facing. In the near future, RI are asked to offer services on the aggregated data (beyond search and browse).

Explain different ways in which RM can be used e.g., to design new RI, when upgrading an RI, or other purposes. Practical guide to these possibilities is important.

Focus

Focus on the appropriate interfaces and stop considering that RIs have nothing in place.

Invite the RI to design the interface.

Tackle the system of RI services.

Push RIs data and metadata on EU cloud infrastructures and build inter-disciplinary operations.

Link RM with OIL-E with help of B2ANNO service. Annotate while reading RI data management documents.

Establish wiki platform of RM users.

Improve the RM to make it more relevant

Improve the RM from D5.4, D8.1, D8.3.

Consider that RI data system already exists/works and therefore focus on the publishing interfaces to facilitate interoperability with other RIs.

Make clear the functional and non-functional aspects of the interface, especially for interoperation.

Appendix 6: Storyline for promotional video

This appendix contains information relevant to developing a storyline as the basis for a promotional / dissemination video about the ENVRI Reference Model and its benefits.

The value of communicating by a short, carefully targeted video as an initial introduction, and as a means of advertising the value of the ENVRI work and the benefits from using the RM to organise that work, has been recognised. Ideally, a professional film maker / camera man should guide the filming, to conduct interviews and to edit the final product. To develop this potential the following story line has been prepared; with the key elements to be covered, the time and message to be allocated to each message and candidate leaders to communicate about that aspect of ENVRI's work. The storyline will be reviewed and refined when a means of acquiring the necessary professional help has been identified.

Topics / scenes to be covered:

1. Environmental sciences (20 s)

- style: narrator + movie/images
- present the complexity, with examples: atmosphere domain, solid earth, marine domain, biosphere
- highlight the need for multi-domain research

People: Ari Asmi (?)

2. Research Infrastructures (15s - 20s)

- Style: narrator + movie/images
- present the variety by animated representation (see ENVRIplus website for inspiration)

People: Ari Asmi (?)

3. **Highlight the need for RI collaboration (10 s)**
    - style: narrator + movie/images

4. **The metaphor (duration: 10 s)**
    - style: animation + narration over the image

Question: What do we want to emphasise?

a) The common language: use the “Babel Tower” – *this is the metaphor people preferred*

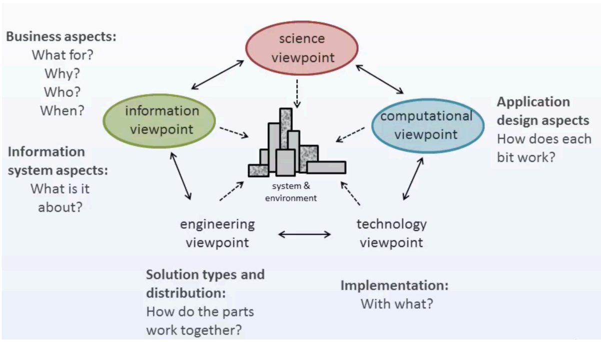
Suggestion: First present a word in several languages (e.g., air/sky, water/sea, life, earth) and then illustrate it with an image and translate into English

b) The enabler of interoperability aspect: the “Puzzle” metaphor



5. **Introduce ENVRI RM role/objectives (duration: 20 s – 30 s)**
    - style: narrator
    - Aims:
        a) providing a way of structuring thinking
        b) providing a common language
        c) discovering existing solutions for common problem
        d) identifying missing functionality

    People: Alex Hardisty



6. ENVRI RM successful stories (duration: 2 x 90s):

- DASSH story (People: Dan Lear)
- EUFAR story (People: Stefanie Holzwarth)
    - introduce the characters
    - state the primary ambition/aim(s)
    - name the conflict/problems

- explain how the conflict has been overcome
- what was the entry point (SV?)
- what tools have been used
- describe the path followed

***Note**: both stories can be included into the movie (see Biovel movie: https://www.youtube.com/channel/UCnnzon4RBMf4PoPgv8-lVkw)*

7. Conclude with lessons learned + Answers to key questions (duration: 35 s - 50 s)

(e.g., "How DASSH benefitted from the ENVRI RM?")

People: Sanna Sorvari

Appendix 7: Project Glossary

Project acronyms

AC: Active Collab (ENVRIplus Project Management System)

BEERi: Board of European Environmental Research Infrastructures - is an internal advisory board representing the needs of environmental Research Infrastructures

CA: Consortium Agreement - Legal contract between the ENVRIplus beneficiaries

DL: Deliverable / Deadline

DoW: Description of Work

DoA: Description of Action

GA:

1) Grant Agreement - Contract between Coordinator and Commission

2) General Assembly - GA is the ultimate decision-making body of the consortium

EB: Executive Board - supervisory body for the execution of the Project

EC: European Commission - is the executive body of the European Union responsible for proposing legislation, implementing decisions, upholding the EU treaties and managing the day-to-day business of the EU

ENV SWG ESFRI: the European Strategy Forum on Research Infrastructures - Strategic Working Group on Environment

ESFRI: the European Strategy Forum on Research Infrastructures

PM: Person Month

RI: Research Infrastructure

WP: Work Package

Organisational Acronyms

ACTRIS: Aerosols, Clouds, and Trace gases Research InfraStructure network

AQUACOSM : EU network of mesocosms facilities for research on marine and freshwater ecosystems open for

global collaboration

BEERI: Board of European Environmental Infrastructures

CEA: Commissariat a l Energie Atomique et aux Energies Alternatives

CINECA: Consorzio Interuniversitario CNR: Consiglio Nazionale Delle Richerche

CNRS: Centre National de la Recherche Scientifique

CODATA: Committee on data for Science and Technology

ConnectinGEO: Coordinating an Observation Network of Networks EnCompassing saTellite and IN-situ to fill the Gaps in European Observations

COOPEUS: Strengthening the cooperation between the US and the EU in the field of environmental research infrastructures

COPERNICUS: previously known as GMES (Global Monitoring for Environment and Security), is the European Programme for the establishment of a European capacity for Earth Observation

CSC: CSC - IT Center for Science

CU: Cardiff University

D4Science: is an organisation offering a Hybrid Data Infrastructure service and a number of Virtual Research Environments

DANUBIUS: The international center for Adavanced studies on river-sea systems

DASSH: Data archive for seabed species (a UK marine biology resource centre)

DIRAC : Distributed Infrastructure with Remote Agent Control

DiSSCo: Distributed Systems of Scientific Collections

DKRZ: Deutsches Klimarechenzentrum GmBH

EAA : Umweltbundesamt GmbH - Environment Agency Austria

EduGAIN: is an international interfederation service interconnecting research and education identity federations

EEA: European Environment Agency

EGI : European Grid Infrastructure

EGLEU:

EINFRA-1-2014:H2020 Call for e-infrastructures (Managing, preserving and computing with big research data)

EISCAT: EISCAT Scientific Association

EMBL: European Molecular Biology Laboratory

EMBRC: European Marine Biological Resource Centre a consortium of research organisations interested in marine biology

EMODNET: The European Marine Observation and Data Network

EMRP: European Metrology Research Programme

EMSC: Euro-Mediterranean Seismological Centre

EMSO: European Multidisciplinary Seafloor and Water Column Observatory

ENVRI : FP7 project on Implementation of common solutions for a cluster of ESFRI infrastructures in the field of environmental Sciences

EPOS: The European Plate Observing System

EUDAT : H2020 project on Research Data Services, Expertise & Technology Solutions (previously funded by FP7)

EUFAR : European Facility for Airborne Research

EUROCHAMP2020 : European atmospheric simulation chambers

EURO-ARGO: European ARGO programme (ARGO are a type of marine survey device)

EUROFLEETS:New operational steps towards an alliance of European research fleets

EUROGOOS: European Global Ocean Survey System

EuroSITES: European Ocean Observatory Network

ERIS: Environmental Research Infrastructure Strategy 2030

ESONET Vi: European Seafloor Observatory NETwork

ETHZ: Eidgenoessische Technische Hochschule Zurich

ESFRI: European Strategy Forum on Research Infrastructures

FIM4R: Federated Identity Management for Research collaborations

FMI: Ilmatieteen Laitos (Finnish Meteorological Institute)

FZJ: Forschungszentrum Juelich GmbH

FixO3: Fix point open ocean observatories (survey programme)

GBIF: Global Biodiversity Information Facility

gCube: is an open-source software toolkit used for building and operating Hybrid Data Infrastructures enabling the dynamic deployment of Virtual Research Environments by favouring the realisation of reuse oriented policies

GEO : The Group on Earth Observations coordinates international efforts to build a Global Earth Observation System of Systems (GEOSS)

GEOMAR: Helmholtz Zentrum Für Ozeanforschung Kiel

GEOSS : Global Earth Observation System of Systems coordinated by GEO (The Group on Earth Observations)

GROOM: Gliders for research ocean observation and management

H2020: Horizon 2020, European level research funding scheme

HELIX Nebula: partnership between big science and big business in Europe that is charting the course towards the sustainable provision of cloud computing - the Science Cloud

IAGOS - In-service Aircraft for a Global Observing System

ICOS : Integrated Carbon Observation System

ICSU: The International Council for Science

INFREMER : Institute Francais de Recherche Pour l’Exploitation de la Mer

INGV: Instituto Nazionale di Geofisica e Vulcanologia

INSPIRE : Integrated Sustainable Pan-European Infrastructure for Researchers in Europe

INRA: Institut National de la Recherche Agronomique

IS-ENES: RI for the European Network for Earth System Modelling

INTERACT: International Network for Terrestrial Research and Monitoring in the Arctic

IPBES: Intergovernmental Platform on Biodiversity & Ecosystem Services

I3: Integrated Infrastructures Initiative (I3) combines several activities essential to reinforce research infrastructures and to provide an integrated service at the European level

JERICO: Towards a joint European research infrastructure network for coastal observatories

LifeWatch:European e-Science infrastructure for biodiversity and ecosystem research

LU: Lund University

LTER: The Long-term Ecological Research Network

LTER-EUROPE : European Long-term Ecosystem Research network of 21 national LTER networks

MBA: Marine Biological Association of the United Kingdom

NERC: Natural Environment Research Council

NILU: Norsk Institutt for Luftforskning (Norwegian Institute of Air Research)

NMI: National Metrological Institutes

PANGAEA: Data Publisher for Earth & Environmental Science (Open Access library aimed at archiving, publishing and distributing georeferenced data from earth system research)

PLOCAN : Consorcio Para el Diseno, Construccion, Equipamiento y Explotacion de la Plataforma Oceanica de Canarias

RCN: Norges Forskningsrad (Research Council of Norway)

RDA: Research Data Alliance

RI: Research Infrastructures – facilities, resources and related services used by the scientific community to conduct top-level research in their respective fields, ranging from social sciences to astronomy, genomics to nanotechnologies.

SCAPE: SCAlable Preservation Environments (FP7 project)

SeaDataNet: Pan-European infrastructure for ocean & marine data management

SIOS: Svalbard Integrated Arctic Earth Observing System

SME: small and medium-sized enterprises

UCPH: Kobenhavns Universitet (Copenhagen University)

UEDIN: University of Edinburgh

UGOT: Goeteborgs Universitet (University of Gothenburg)

UHEL: Helsingin Yliopisto (University of Helsinki)

UiT: Universitetet i Tromsoe (University of Tromso)

UniHB: Universitaet Bremen (University of Bremen)

UNILE: Universita del Salento (University of Salento)

UNITUS: Universita Degli Studi della Tuscia

USTAN : The University Court of the University of St. Andrews (University of St Andrews)

UvA : Universiteit van Amsterdam (University of Amsterdam)

Important Technical Terms/Acronyms

API: Application Program Interface, is a set of routines, protocols, and tools for building software applications

Biodiversity: is the variety of different types of life found on earth

Biodiversity metrics: measurements of the number of species and how they are distributed

CERIF: Common European Research Information Format

CIARD RING: A global directory of information services and datasets in agriculture

Data stream: is a sequence of digitally encoded coherent signals used to transmit or receive information that is in the process of being transmitted

Data pipeline: In computing, a pipeline is a set of data processing elements connected in series, where the output of one element is the input of the next one.

DCAT: is a resource description format vocabulary designed to facilitate interoperability between data catalogues

DOI: Digital Object Identifier

E-infrastructure: can be defined as networked tools, data and resources that support a community of researchers, broadly including all those who participate in and benefit from research

HPC: High Performance Computing

HTC: High Throughput Computing

IoT: The Internet of Things - is a scenario in which objects, animals or people are provided with unique identifiers and the ability to transfer data over a network without requiring human-to-human or human-to-computer interaction.

ICT: Information and Communications technology

INFRADEV-4: Subcall of H2020 INFRADEV call for Implementation and operation of cross-cutting services and solutions for clusters of ESFRI and other relevant research infrastructure initiatives

IPR: Intellectual Property Rights

KOS: Knowledge Organization Systems - is a generic term used in Knowledge organization about authority lists, classification systems, thesauri, topic maps, ontologies etc.

LOD: Linked open data is linked data that is open content

LOV: Linked Open Vocabularies

Metadata : is data that describes other data. Metadata summarizes basic information about data, which can make finding and working with particular instances of data easier

NGI: National Grid Initiative

NREN: National Research and Education Network

NRT: Near Real Time - refers to the time delay introduced, by automated data processing or network transmission, between the occurrence of an event and the use of the processed data (For example, a near-real-time display depicts an event or situation as it existed at the current time minus the processing time, as nearly the time of the live event)

OASIS: Advancing Open Standards for the Information Society (non-profit consortium)

ODP: Online Data Processing

OIL-E: The Open Information Linking model for Environmental science - is a semantic linking framework

Ontology: (In computer science and information science) an ontology is a formal naming and definition of the types, properties, and interrelationships of the entities that really or fundamentally exist for a particular domain of discourse

QoE: Quality of user experience

over dispersion : a statistical characteristic of data such that the data have more clusters than compared to what might be expected if the data were distributed randomly in proportion to the time/space available.

NetCDF: a file format.

OceanSITES: s a worldwide system of long-term, open-ocean reference stations measuring dozens of variables and monitoring the full depth of the ocean from air-sea interactions down to the seafloor

OOI: Ocean Observatories Initiative

RDA: Resource Description and Access, a standard for descriptive cataloguing

RM: Reference Model - is an abstract framework or domain-specific ontology consisting of an interlinked set of clearly defined concepts produced by an expert or body of experts in order to encourage clear communication

SensorML - The primary focus of the Sensor Model Language is to provide a robust and semantically-tied means of defining processes and processing components associated with the measurement and post-measurement transformation of observations

Semantics : is the study of meaning

Syntax: In computer science, the syntax of a computer language is the set of rules that defines the combinations of symbols that are considered to be a correctly structured document or fragment in that language

SLA: Service Level Agreement

UV: unmanned vehicles

VCP: (ENVRI) Virtual Community Platform

VL: Virtual Laboratory

VLDATA: this was the name of the failed project proposal so I think it can be deleted

VRE: Virtual Research Environments, web based package tailored to a specific community

Definitions

Intradisciplinary: working within a single discipline.

Crossdisciplinary: viewing one discipline from the perspective of another.

Multidisciplinary: people from different disciplines working together, each drawing on their disciplinary knowledge.

Interdisciplinary: integrating knowledge and methods from different disciplines, using a real synthesis of approaches.

Transdisciplinary: creating a unity of intellectual frameworks beyond the disciplinary perspectives.

ENVRI Reference Model Glossary

CCSDS	Consultative Committee for Space Data Systems
CMIS	Content Management Interoperability Services
CERIF	Common European Research Information Format

DDS	Data Distribution Service for Real-Time Systems
ENVRI	Environmental Research Infrastructure
ENVRI_RM	ENVRI Reference Model
ESFRI	European Strategy Forum on Research Infrastructures
ESFRI-ENV RI	ESFRI Environmental Research Infrastructure
GIS	Geographic Information System
IEC	International Electrotechnical Commission
ISO	International Organisation for Standardization
OAIS	Open Archival Information System
OASIS	Organization for the Advancement of Structured Information Standards
ODP	Open Distributed Processing
OGC	Open Geospatial Consortium
OMG	Object Management Group
ORCHESTRA	Open Architecture and Spatial Data Infrastructure for Risk Management
ORM	OGC Reference Model
OSI	Open Systems Interconnection
OWL	Web Ontology language
SOA	Service Oriented Architecture
SOA-RM	Reference Model for Service Oriented Architecture
RDF	Resource Description Framework
RM-OA	Reference Model for the ORCHESTRA Architecture
RM-ODP	Reference Model of Open Distributed Processing
UML	Unified Modelling Language
W3C	World Wide Web Consortium
UML4ODP	Unified Modelling Language For Open Distributed Processing

- Terminology

Access Control: A functionality that approves or disapproves of access requests based on specified access policies.

Acquisition Service: Oversight service for integrated data acquisition.

Active role: A active role is typically associated with a human actor.

Add Metadata: Add additional information according to a predefined schema (metadata schema). This partially overlaps with data annotations.

Annotate Data: Annotate data with meaning (concepts of predefined local or global conceptual models).

Annotate Metadata: Link metadata with meaning (concepts of predefined local or global conceptual models). This can be done by adding a pointer to concepts within a conceptual model to the data. If e.g. concepts are terms in and SKOS/RDF thesaurus, published as linked data then this would mean entering the URL of the term describing the meaning of the data.

Annotation: (verb) The action of annotating or making notes. (noun) A note added to anything written, by way of explanation or comment.

Annotation Service: Oversight service for adding and updating records attached to curated datasets.

Assign Unique Identifier: Obtain a unique identifier and associate it to the data.

Authentication: A functionality that verifies a credential of a user.

Authentication Service: Security service responsible for the authentication of external agents making requests of infrastructure services.

Authorisation: A functionality that specifies access rights to resources.

Authorisation Service: Security service responsible for the authorisation of all requests made of infrastructure services by external agents.

Backup: A copy of (persistent) data so it may be used to restore the original after a data loss event.

Behaviour : A behaviour of a community is a composition of actions performed by roles normally addressing separate business requirements.

Build Conceptual Models: Establish a local or global model of interrelated concepts.

Capacity Manager: An active role, which is a person who manage and ensure that the IT capacity meets current and future business requirements in a cost-effective manner.

Carry out Backup: Replicate data to an additional data storage so it may be used to restore the original after a data loss event. A special type of backup is a long term preservation.

Catalogue service: Oversight service for cataloguing curated datasets.

Check Quality: Actions to verify the quality of data.

Citation: from the ENVRI RM perspective, citation is defined as a pointer from a publication to:

- data source(s)
- and/or the owner(s) of the data source(s)
- a description of the evaluation process, if available
- a timestamp marking the access time to the data sources, thus reflecting a certain version

Citizen (synonyms: General Public, Media): An active role, a person, who is interested in understanding the knowledge delivered by an environmental science research infrastructure, or discovering and exploring the Knowledge_Base_Glossary enabled by the research infrastructure.

Citizen Scientist: An active role, member of the general public who engages in scientific work, often in collaboration with or under the direction of professional scientists and scientific institutions (also known as amateur scientist).

Community: A collaboration which consists of a set of roles agreeing their objective to achieve a stated business purpose.

Concept: Name and definition of the meaning of a thing (abstract or real thing). Human readable definition by sentences, machine readable definition by relations to other concepts (machine readable sentences). It can also be meant for the smallest entity of a conceptual model. It can be part of a flat list of concepts, a hierarchical list of concepts, a hierarchical thesaurus or an ontology.

Conceptual Model: A collection of concepts, their attributes and their relations. It can be unstructured or structured (e.g. glossary, thesaurus, ontology). Usually the description of a concept and/or a relation defines the concept in a human readable form. Concepts within ontologies and their relations can be seen as machine readable sentences. Those sentences can be used to establish a self-description. It is, however, practice today, to have both, the human readable description and the machine readable description. In this sense a conceptual model can also be seen as a collection of human and machine readable sentences. Conceptual models can reside within the persistence layer of a data provider or a community or outside. Conceptual models can be fused with the data (e.g. within a network of triple stores) or kept separately.

Coordination Service: An oversight service for data processing tasks deployed on infrastructure execution resources.

Data Acquisition Community. A community, which collects raw data and bring (streams of) measures into a system.

Data Acquisition Subsystem: A subsystem that collects raw data and brings the measures or data streams into a computational system.

Data Analysis: A functionality that inspects, cleans, transforms data, and provides data models with the goal of highlighting useful information, suggesting conclusions, and supporting decision making.

Data Assimilation: A functionality that combines observational data with output from a numerical model to produce an optimal estimate of the evolving state of the system.

Data Broker: Broker for facilitating data access/upload requests**.**

Data Cataloguing: A functionality that associates a data object with one or more metadata objects which contain data descriptions.

Data Citation: A functionality that assigns an accurate, consistent and standardised reference to a data object, which can be cited in scientific publications.

Data Collection: A behaviour performed by a *Data Collector* that control and monitor the collection of the digital values from a *sensor* instrument or a human sensor, such as a *Measurer* or a *Observer*, associating consistent time-stamps and necessary metadata.

Data Collector: Active or passive role, adopted by a person or an instrument collecting data.

Data Consumer: Either an active or passive role, which is an entity who receives and use the data.

Data Curation Community: A community, which curates the scientific data, maintains and archives them, and produces various data products with metadata.

Data Curation Subsystem: A subsystem that facilitates quality control and preservation of scientific data.

Data Curator: An active role, which is a person who verifies the quality of the data, preserve and maintain the data as a resource, and prepares various required data products.

Data Discovery & Access: A functionality that retrieves requested data from a data resource by using suitable search technology.

Data Exporter: Binding object for exporting curated datasets.

Data Extraction: A functionality that retrieves data out of (unstructured) data sources, including web pages, emails, documents, PDFs, scanned text, mainframe reports, and spool files.

Data Identification: A functionality that assigns (global) unique identifiers to data contents.

Data Importer: An Oversight service for the import of new data into the data curation subsystem.

Data infrastructure: a collection of data assets, organisations that operate and maintain them and guides describing how to use and manage the data. A data infrastructure is sustainably funded and has oversight that provides direction to maximise data use and value by meeting the needs of society. Data infrastructure includes technology, processes and organisation.

Data management: a process development and execution of architectures, policies, practices and procedures in order to manage the data lifecycle needs of a specific research community.

Data management plan (DMP): a formal document that outlines how data are to be handled both during a research project and after the project is completed.

Data Mining: A functionality that supports the discovery of patterns in large data sets.

Data Originator: Either an active or a passive role, which provide the digital material to be made available for public access.

Data Processing Control: A functionality that initiates the calculation and manages the outputs to be returned to the client.

Data Processing Subsystem: A subsystem that aggregates the data from various resources and provides computational capabilities and capacities for conducting data analysis and scientific experiments.

Data Product Generation: A functionality that processes data against requirement specifications and standardised formats and descriptions.

Data Provenance: Information that traces the origins of data and records all state changes of data during their lifecycle and their movements between storages.

Data Provider: Either an active or a passive role, which is an entity providing the data to be used.

Data Publication: A functionality that provides clean, well-annotated, anonymity-preserving datasets in a suitable format, and by following specified data-publication and sharing policies to make the datasets publically accessible or to those who agree to certain conditions of use, and to individuals who meet certain professional criteria.

Data Publication Community: A community that assists the data publication, discovery and access.

(Data Publication) Repository: A passive role, which is a facility for the deposition of published data.

Data Publishing Subsystem: A subsystem that enables discovery and retrieval of data housed in data resources.

Data Quality Checking: A functionality that detects and corrects (or removes) corrupt, inconsistent or inaccurate records from data sets.

Data Service Provision Community: A community that provides various services, applications and software/tools to link, and recombine data and information in order to derive knowledge.

Data State: Term used as defined in ISO/IEC 10746-2. At a given instant in time, data state is the condition of an object that determines the set of all sequences of actions (or traces) in which the object can participate.

Data Storage & Preservation: A functionality that deposits (over long-term) the data and metadata or other supplementary data and methods according to specified policies, and makes them accessible on request.

Data Store Controller: A data store within the data curation subsystem.

Data Transfer Service: Oversight service for the transfer of data into and out of the data curation subsystem.

Data Transmission: A functionality that transfers data over communication channel using specified network protocols.

Data Transporter: Generic binding object for data transfer interactions.

Data Use Community: A community who makes use of the data and service products, and transfers the knowledge into understanding.

Data Use Subsystem: A subsystem that provides functionalities to manage, control, and track users' activities and supports users to conduct their roles in the community.

Describe Service: Describe the accessibility of a service or process, which is available for reuse, the interfaces, the description of behaviour and/or implemented algorithms.

Design of Measurement Model: A behaviour that designs the measurement or monitoring model based on scientific requirements.

Do Data Mining: Execute a sequence of metadata / data request --> interpret result --> do a new request

e-Infrastructure: a combination and interworking of digitally-based technology (hardware and software), resources (data, services, digital libraries), communications (protocols, access rights and networks), and the people and organisational structures needed to support modern, internationally leading collaborative research be it in the arts and humanities or the sciences.

Educator (synonym: Trainer): An active role, which is a person who makes use of the data and application services for education and training purposes.

Engineer (synonym: Technologist): An active role, which is a person who develops and maintains the research infrastructure.

Environmental Scientist: An active role, which is a person who conduct research or perform investigation for the purpose of identifying, abating, or eliminating sources of pollutants or hazards that affect either the environment or the health of the population. Using knowledge of various scientific disciplines, may collect, synthesize, study, report, and recommend action based on data derived from measurements or observations of air, food, soil, water, and other sources.

ENVRI Reference Model: A common ontological framework and standards for the description and characterisation of computational and storage systems of ESFRI environmental research infrastructures.

Experiment Laboratory: Community proxy for conducting experiments within a research infrastructure.

Field Laboratory: Community proxy for interacting with data acquisition instruments.

Final review: Review the data to be published, which will not likely be changed again.

Free text annotation: to add a short explanation or opinion to a text or drawing (equivalent to the dictionary definition of annotation).

Instrument Controller: An integrated raw data source.

Knowledge Base: (1) A store of information or data that is available to draw on. (2) The underlying set of facts, assumptions, and rules which a computer system has available to solve a problem.

Knowledge infrastructure: robust networks of people, artifacts, and institutions that generate, share, and maintain specific knowledge about the human and natural worlds.

Mapping Rule: Configuration directives used for model-to-model transformation.

(Measurement Model) Designer: An active role, which is a person who design the measurements and monitoring models based on the requirements of environmental scientists.

Measurement result: Quantitative determinations of magnitude, dimension and uncertainty to the outputs of observation instruments, sensors (including human observers) and sensor networks.

Measurer: An active role, which is a person who determines the ratio of a physical quantity, such as a length, time, temperature etc., to a unit of measurement, such as the meter, second or degree Celsius.

Metadata: Data about data, in scientific applications is used to describe, explain, locate, or make it easier to retrieve, use, or manage an information resource.

Metadata Catalogue: A collection of metadata, usually established to make the metadata available to a community. A metadata catalogue has an access service.

Metadata Harvesting (Publishing Community Role): A behaviour performed by a metadata harvester to gather metadata from data objects in order to construct catalogues of the available information. A functionality that (regularly) collects metadata (in agreed formats) from different sources.

Metadata Harvester (Publishing Community Role): A passive role performed by a system or service collecting metadata to support the construction/selection of a global conceptual model and the production of mapping rules.

Metadata State:

- raw: are established metadata, which are not yet registered. In general, they are not shareable in this status
- registered: are metadata which are inserted into a metadata catalogue.
- published: are metadata made available to the public, the outside world. Within some metadata catalogues registered.

Passive Role: A passive role is typically associated with a non-human actor.

Perform Mapping: Execute transformation rules for values (mapping from one unit to another unit) or translation rules for concepts (translating the meaning from one conceptual model to another conceptual model, e.g. translating code lists).

Persistent Data: Term (data) used as defined in ISO/IEC 10746-2. Data is the representations of information dealt by information systems and users thereof. Data which are persistent (stored).

Perform Measurement or Observation: Measure parameter(s) or observe an event. The performance of a measurement or observation produces measurement results.

PID Generator: A passive role, a system which assigns persist global unique identifiers to a (set of) digital object.

PID Registry: A passive role, which is an information system for registering PIDs.

PID Service: External service for persistent identifier assignment and resolution.

Policy Maker (synonym: Decision Maker): An active role, a person, who makes decisions based on the data evidences.

Process Control: A functionality that receives input status, applies a set of logic statements or control algorithms, and generates a set of analogue / digital outputs to change the logic states of devices.

Process Controller: Part of the execution platform provided by the data processing subsystem.

Process Data: Process data for the purposes of:

- converting and generating data products
- calculations: e.g., statistical processes, simulation models
- visualisation: e.g., alpha-numerically, graphically, geographically

Data processes should be recorded as provenance.

Provenance: The pathway of data generation from raw data to the actual state of data.

Publish Data: Make data public accessible.

Publish Metadata: Make the registered metadata available to the public.

QA Notation: Notation of the result of a Quality Assessment. This notation can be a nominal value out of a classification system up to a comprehensive (machine readable) description of the whole QA process.

Quality Assessment (QA): Assessment of details of the data generation, including the check of the plausibility of the data. Usually the quality assessment is done by predefined checks on data and their generation process.

Query Data: Send a request to a data store to retrieve required data.

Query Metadata: Send a request to metadata resources to retrieve metadata of interests.

Observer: An active role, which is a person who receives knowledge of the outside world through the senses, or records data using scientific instruments.

Raw Data Collector: Binding object for raw data collection.

Reference Mode: A reference mode is an abstract framework for understanding significant relationships among the entities of some environment.

Register Metadata: Enter the metadata into a metadata catalogue.

Research Infrastructure: means facilities, resources and related services that are used by the scientific community to conduct top-level research in their respective fields and covers major scientific equipment or sets of instruments; knowledge-based resources such as collections, archives or structures for scientific information; enabling Information and Communications Technology-based infrastructures such as Grid, computing, software and communication, or any other entity of a unique nature essential to achieve excellence in research. Such infrastructures may be "single-sited" or "distributed" (an organised network of resources) [**41**].

Resource Registration: A functionality that creates an entry in a resource registry and inserts resource object or a reference to a resource object in specified representations and semantics.

Role : A role in a community is a prescribing behaviour that can be performed any number of times concurrently or successively.

Science Gateway: Community portal for interacting with an infrastructure.

Scientific Modelling and Simulation: A functionality that supports the generation of abstract, conceptual, graphical or mathematical models, and to run an instance of the model.

Scientist (synonym: Researcher): An active role, which is a person who makes use of the data and application services to conduct scientific research.

(Scientific) Workflow Enactment: A specialisation of Workflow Enactment, which support of composition and execution a series of computational or data manipulation steps, or a workflow, in a scientific application. Important processes should be recorded for provenance purposes.

Security Service: Oversight service for authentication and authorisation of user requests to the infrastructure.

Semantic Annotation: link from an information object (single datum, data set, data container) to a concept within a conceptual model, enabling the discovery of the meaning of the information object by human and machines.

Semantic Broker: Broker for establishing semantic links between concepts and bridging queries between semantic domains.

SV Community Behaviour: A behaviour enabled by a *Semantic Mediator* that unifies similar data (knowledge) models based on the consensus of collaborative domain experts to achieve better data (knowledge) reuse and semantic interoperability.

Semantic Laboratory: Community proxy for interacting with semantic models.

Semantic Mediator: A passive role, which is a system or middleware facilitating semantic mapping discovery and integration of heterogeneous data.

Sensor: A passive role, which is a converter that measures a physical quantity and converts it into a signal which can be read by an observer or by an (electronic) instrument.

Sensor Network: A passive role, which is a network consists of distributed autonomous sensors to monitor physical or environmental conditions.

Service: Service or process, available for reuse.

Service Consumer: Either an active or a passive role, which is an entity using the services provided.

Service Description: Services and processes, which are available for reuse, be it within an enterprise architecture, within a research infrastructure or within an open network like the Internet, shall be described to help avoid wrong usage. Usually such descriptions include the accessibility of the service, the description of the interfaces, the description of behavior and/or implemented algorithms. Such descriptions are usually done along service description standards (e.g. WSDL, web service description language). Within some service description languages, semantic descriptions of the services and/or interfaces are possible (e.g. SAWSDL, Semantic Annotations for WSDL)

Service Provider: Either an active or a passive role, which is an entity providing the services to be used.

Service Registry: A passive role, which is an information system for registering services.

Setup Mapping Rules: Specify the mapping rules of data and/or concepts.

Specification of Investigation Design: This is the background information needed to understand the overall goal of the measurement or observation. It could be the sampling design of observation stations, the network design, the description of the setup parameters (interval of measurements) and so on... It usually contains important information for the allowed evaluations of data. (E.g. the question whether a sampling design was done randomly or by strategy determines which statistical methods that can be applied or not).

Specification of Measurements or Observations: The description of the scientific measurement model which specifies:

- what is measured;
- how it is measured;
- by whom it is measured; and
- what the temporal design is (single /multiple measurements / interval of measurement etc.)

Specify Investigation Design: specify design of investigation, including sampling design:

- geographical position of measurement or observation (site) -- the selections of observations and measurement sites, e.g., can be statistical or stratified by domain knowledge;
- characteristics of site;
- · preconditions of measurements.

Specify Measurement or Observation: Specify the details of the method of observations/measurements.

Stakeholder (synonyms: Private Investor, Private Consultant): An active role, a person, who makes use of the data and application service for predicting market so as to make business decision on producing related commercial products.

Storage: A passive role, which is memory, components, devices and media that retain digital computer data used for computing for some interval of time.

Storage Administrator: An active role, which is a person who has the responsibilities to the design of data storage, tune queries, perform backup and recovery operations, raid mirrored arrays, making sure drive space is available for the network.

Store Data: Archive or preserve data in persistent manner to ensure continuing accessible and usable.

Subsystem: a set of capabilities that collectively are defined by a set of interfaces with corresponding operations that can be invoked by other subsystems. Subsystems can be executed independently, and developed and managed incrementally.

Technician: An active role, which is a person who develop and deploy the sensor instruments, establishing and testing the sensor network, operating, maintaining, monitoring and repairing the observatory hardware.

Track Provenance: Add information about the actions and the data state changes as data provenances.

Unique Identifier (UID): With reference to a given (possibly implicit) set of objects, a unique identifier (UID) is any identifier which is guaranteed to be unique among all identifiers used for those objects and for a specific purpose.

User Behaviour Tracking: A behaviour enabled by a Data Use Subsystem that to track the Users. User Behaviour Tracking is the analysis of visitor behaviour on a website. The analysis of an individual visitor's behaviour may be used to provide options or content that relates to their implied preferences; either during a visit or in the future visits. Additionally, it can be user to track content use and performance.

User Group Work Supporting: A behaviour enabled by a Data Use Subsystem that to support controlled sharing, collaborative work and publication of results, with persistent and externally citable PIDs.

User Profile Management: A behaviour enabled by a Data Use Subsystem that to support persistent and mobile profiles, where profiles will include preferred interaction settings, preferred computational resource settings, and so on.

User Working Space Management: A behaviour enabled by a Data Use Subsystem that to support work spaces that allow data, document and code continuity between connection sessions and accessible from multiple sites or mobile smart devices.

User Working Relationships Management: A behaviour enabled by a Data Use Subsystem that to support a record of working relationships, (virtual) group memberships and friends.

Virtual Laboratory: Community proxy for interacting with infrastructure subsystems.

Virtual Research Environment (VRE, sysnomyms:Science Gateway, Collaboratory, Digital Library, Inhabited Information Space, e-Infrastructure, Cyberinfrastructure): a web-based working environment tailored to serve the needs of a research community. A VRE is expected to provide an array of commodities needed to accomplish the research community’s goal(s); it is open and flexible with respect to the overall service offering and lifetime; and it promotes fine-grained controlled sharing of both intermediate and final research results by guaranteeing ownership, provenance and attribution.

END.