

Reference model

Introduction defining context and scope

So, what is a Reference Model (RM)? A good place to start is with this [Wikipedia article on reference models](#). Its opening paragraph explains an RM as “an abstract framework consisting of an interlinked set of clearly defined concepts produced by an expert or body of experts in order to encourage clear communication. A reference model can represent the component parts of any consistent idea, from business functions to system components,”. It goes on to say that an RM can “... then be used to communicate ideas clearly among members of the same community”. This then, is the essence of an RM. It's a descriptive conceptual framework, establishing a common language of communication and understanding, about elements of a system and their significant relationships, within a community of interest. That's particularly important when, as in the environmental research infrastructures (RI) sector that community of interest brings together significant numbers of experts from vastly different scientific and technical backgrounds to talk about building distributed ICT infrastructures.

The present topic is concerned principally with the [ENVRI Reference Model](#) and is closely related to the topic of the [Linking model](#), which depends upon it. However, reference models are cutting across all aspects of infrastructure design and technology review. Thus, this topic relates to all the topics of the [technology review](#).

Change history and amendment procedure

Note: Do not record editorial / typographical changes. Only record significant changes of content.

Date	Name	Institution	Nature of the information added / changed
22 Dec 2015	Alex Hardisty	CU	Introduction, context, scope established. Sub-headings inserted
8 Feb 2016	Alex Hardisty	CU	More detail added; particularly on positioning the role of RMs in architectural design of research infrastructures
10 May 2016	Alex Hardisty	CU	Updated following completion of technology review, to match deliverable D5.1 A consistent characterisation of existing and planned RIs .

Sources of information used

Wikipedia provides general introductory level information on reference models and reference architectures. ISO/IEC publishes relevant international standards. Various Web resources have been used and are mentioned / linked in the text. Other sources are directly referenced from the text and listed in the bibliography.

Two-to-five year analysis

State of the art

The ENVRI Reference Model (ENVRI RM) ([envri.eu/rm](#)) is presently a work in progress. Version 1.1 has been published in summer 2013 as a deliverable of the ENVRI project. It is based on the requirements collected from 6 research infrastructures. In the ENVRIplus project there is a task 5.2 to review and improve the RM, based on new requirements analysis of 20 research infrastructures. At present the ENVRI RM is introduced through a sub-systems view of research infrastructure but this needs to shift to a data lifecycle oriented approach. The sub-systems perspective has to be more properly assigned only within the Engineering Viewpoint where it can support the complete lifecycle of research data (from design of experiments that produce new data through acquisition, curation and publishing of that data, to its use in processing and analysis to reach scientific conclusions) according to the specific scope and needs of individual RIs.

The ENVRI Reference Model (ENVRI RM) is presently work in progress. Based on RM-ODP [ISO/IEC 10746], version 1.1 has been published in summer 2013 as a deliverable of the ENVRI project. It is derived from commonalities of requirements collected from 6 research infrastructures. In the ENVRIplus project there is a task 5.2 to review and improve the RM, based on new requirements analysis of 20 research infrastructures. At present the ENVRI RM is introduced through a sub-systems view of research infrastructure but this needs to shift to a data lifecycle oriented approach. The sub-systems perspective has to be more properly assigned only within the Engineering Viewpoint where it can support the complete lifecycle of research data (from design of experiments that produce new data through acquisition, curation and publishing of that data, to its use in processing and analysis to reach scientific conclusions) according to specific scope and needs of individual RIs.

Moving forward with the RM in ENVRIplus

Use of reference models, and particularly viewpoint models such as ENVRI RM keeps the design discussion centred at the right level (see remarks on raising the level of discourse) while accommodating the perspectives of difference stakeholders. They allow moving from a high level description of RIs for researchers and sponsors that is founded on the science to be carried out, to a lower, more detailed design level for IT developers and technicians, concerning engineering and technology aspects. By using the ENVRI RM, RIs can create a set of models that separate concerns neatly but at the same time keep the consistency of the RI systems as a complete entity, as well as accommodating relevant policy constraints.

Validating the present ENVRI RM based on review of requirements from a wider set of RIs and completing and evolving the RM for easier use are main activities in the ENVRIplus project now. Another important activity is to explore ways in which RIs communities can be helped, and assisted to become self-sufficient. Working in conjunction with several use cases teams (see below) and producing specialised e-learning materials are two strands of planned activity. As well as delivering content specifically about the 'internals' of the ENVRI RM, training will also give guidance for different situations on how to use various parts of the RM. This will be very much driven by case examples and, over time we expect to see emergence of common re-usable patterns that can be applied elsewhere.

It would be interesting to find an early adopter RI prepared to invest in exploring the potential of the available tools (see above), casting a model in UML4ODP perhaps.

Problems to be overcome: Adoption

In the research infrastructures sector we have to move to an RM oriented approach for three reasons. Firstly, so that we can achieve interoperability within and between different infrastructures. Secondly, because there are multiple players and stakeholders in the sector that have to work together and talk to one another. And thirdly, so that the sector can achieve the economies of scale within and across infrastructures that we need for attracting the attention of industry. There is a role for bespoke design and development due to the unique attributes of individual infrastructures but wherever possible, off-the-shelf capabilities should be adopted first. We can do this more easily when we have a commonly accepted conceptual foundation upon which to base procurement. Achieving a shift in culture and mind set of the community is a significant issue to be overcome. It needs to balance the costs of replacing existing technology and the consequent impact on working practices with the long-term costs of support and maintenance – see Section 4.2.4

Problems to be overcome: Complexity

RMs are a systems modelling way of thinking that draws together all the conceptual elements and relationships in a large class of very complex distributed systems. Systems thinking gives us a means to cope with that complexity. It helps us to better deal with change in the (scientific) business, leading to more agile styles of thinking and response. Understanding relationships between the various parts of a research infrastructure helps us to understand the possible collective (emergent) behaviours of the infrastructure and to practically engineer and manage real systems. Thus (and according to APG) a reference model is really a framework from which a portfolio of services can be derived.

Complexity can be off-putting. [Hardisty 2015] has suggested ways to engage with RMs for the first time and how, particularly to get the best out of the ENVRI RM121. A Forbes article on Enterprise Architecture [Bloomberg 2014] also offers several suggestions that are transferable to the present context. You don't have to take reference models too literally. You don't have to "do" all of the RM to benefit from it. Just pick and choose what works for you. It's basically a toolkit. You can use it in several different ways - to baseline what you already have and to clean up; to target desired outcomes and plan out how to achieve them; or in combination to deal with a troublesome area (pain point) – first by baselining it, then by targeting it and then iterating until the pain has gone away.

Problems to be overcome: Tooling and skills development

Effective software systems engineering depends on having robust and capable Integrated Development Environment (IDE) within which all the processes of software design, implementation and test can take place. As noted above, industry-standard design tools are beginning to support the necessary concepts but their penetration and use in research infrastructures sector is still quite low. The level of architecting skills to be found among practitioners in research infrastructures is also quite low. This has to be addressed by targeted recruiting and specialised training.

Use of RMs in other sectors

RMs have been used widely in the telecoms, healthcare and defence sectors, as well as among architects of enterprise and public sector systems. All these sectors are characterised by their need for "infrastructure at scale". They involve multiple vendors who have to work, if not together then to a common framework of principles and concepts to bring about widespread interoperability. It's easy to make a phone call to more or less anywhere on the planet, or to receive streaming video there. That is the result of using reference models and standardising interfaces between sub-systems and components from different vendors.

One view of reference models, particularly expressed by practitioners at [Armstrong Process Group \(APG\)](#) is that they are a 'supporting capability' in the Enterprise Architecture value chain. Putting that into the ENVRI context is to say that RMs have relevance to and use for understanding and analysing the environmental science enterprise prior to and as part of planning and implementing (engineering) research infrastructures.

During 2013 the ESFRI cluster projects covering the biomedical sciences (BioMedBridges), physics (CRISP), social science and humanities (DASISH), and environmental sciences (ENVRI) came together to identify common challenges in data management, sharing and integration across scientific disciplines [Field 2013]. Reference models were identified as a common interest of all the clusters. Subsequently, RMs were ranked as one of the top three issues needing to be addressed jointly across all RIs at the European level.

UML4ODP and tooling for software / systems engineering

Recently revised, UML4ODP [ISO/IEC 19793:2015] allows systems architects to express their systems architecture designs in a graphical and standard manner using UML notation. This is exciting because it means, for example that the ENVRI RM and all its concepts can be built into [software engineering IDEs](#) with all that implies for inheritance, compliance with agreements and standards, etc. This makes it possible for industry-standard model-based systems engineering tools, such as [Sparx Systems' Enterprise Architect](#), [IBM Rational Software Architect](#) or [MagicDraw](#) to deal with ODP based designs and thus to inherit concepts from an RM once that RM is encoded as a UML4ODP representation. (Note: The Information Viewpoint of the ENVRI RM has been cast in UML4ODP during its development in the ENVRI project. Work remains to be done to cast the other viewpoints in UML4ODP.) This has been explored, for example in the healthcare context by [Lopez 2009]. However, as far as we know there are no open-source IDE tools specifically supporting UML4ODP at this time. [Eclipse](#) has general support for UML but not specifically for UML4ODP.

On the other hand, the ODP and ENVRI reference models can also be represented as an ontology (see Section 3.9) expressed, for example in OWL and RDF, which means it can then be used in a knowledge base over which reasoning can take place. This has multiple applications.

Supporting the European Open Science Cloud (EOSC)

Early in April 2016 a High-Level Expert Group reported its strategic advice on the future [European Open-Science Cloud \(EOSC\)](#) to the European Commission. "By mapping the route to a European Open-Science Cloud", says expert group member Paul Ayris, "the group's ultimate goal is to create a trusted environment for hosting and processing research data to support world-leading EU science. Cloud computing can change the way that research in Europe is done. The creation of an open-science commons would allow European researchers to collaborate, share and innovate using shared infrastructures, tools and content."

Announced by the EC on 19th April 2016, EOSC is envisioned as a federated environment, made up of contributions from many stakeholders at both national and institutional levels. The desire for minimal international guidance and governance, combined with maximum freedom of implementation means that moving towards some kind of framework of reference as the basis of the **open science commons** is inevitable. Robust standards for exchanging information between different heterogeneous parts of the federated cloud environment will be paramount. Developing these in an open and transparent manner will be difficult and costlier without a framework of reference (such as the ENVRI Reference Model) within which to situate them.

The ENVRI RM can be used for describing the EOSC.

On one level, there is an implied assumption that cloud computing (as understood in common parlance) is the basis of the EOSC. This is a technology assumption (and therefore also partially an Engineering assumption). However, the true scope of EOSC has to be thought of in terms much wider than just technology and engineering; especially as the former is subject to rapid evolution. Consideration has to be given to the business of the EOSC, to the data and information it is expected to handle, and to the nature of the computation (in its widest sense) to be applied in order to create the *'trusted environment for hosting and processing'*.

EOSC implies more than is just meant by the term "cloud", as often used in common parlance to mean cloud computing. EOSC bundles: a) financial and business models, that are Science viewpoint; b) data and information to be handled, that are Information Viewpoint; c) shared provisioning, operations management, and systems support that is organisational, and involves multiple viewpoints; d) a hardware-level protection regime, involving Engineering and Technology viewpoints; e) a whole open-ended set of ways of building and deploying executable machine images; which has Computational, Engineering and Technology; f) a range of ways of allocating resources and scheduling work, again Computational, Engineering and Technology viewpoints; g) a variety of AAAI strategies; and h) a variety of collaboration and isolation regimes. EOSC will not be a single platform or a single technology but a heterogeneous collection of virtual and dynamic configurations responding to the circumstances of the moment. Initiatives such as **Kubernetes**, for example and our own ENVRI Linking Model (3.9) are exploring ways of developing smart mappings to cope with this.

Cloud is not easy, certainly if you're doing most of the things the RIs are expected to be doing. For them, using methods with the ENVRI RM to unpick the elements that make up cloud might be useful.

Alignment to Research Data Alliance (RDA)

By engaging the scientific communities to address the issues such as data identification and citation, discovery, access, sharing, etc., the **Research Data Alliance (RDA)** has a role to further promote the maturation and adoption of practices for open research data and open science.

One product of RDA thus far is the results from its Data Foundation and Terminology Working Group [DFT: Results RFC]. This is a set of core terms for classifying data objects and repositories, and a model of relationships between the terms. These DFT core terms correspond more or less with some main concepts in the Information Viewpoint of the ENVRI RM but the scope is limited to that.

Part of the envisaged evolution of the ENVRI RM during the ENVRIplus project will involve RDA alignment.

In general terms, the "digital transformation agenda" (encompassing cloud infrastructure, continuous delivery of IT services, DevOps, agile software development, etc.) acts as a significant driver. Bots, services, APIs and apps - this is a catch-all for the general trend in consumer computing towards a world of smart applications, interacting with services (both bot and human) via a range of APIs. Knowing all the APIs, where they are and how they relate to one another in terms of compatibility and composition potential will be a crucial development to watch as it spills over from mainstream consumer computing into enterprise and academic/research sectors. To what extent do current RMs overtly accommodate this trend? To what extent do RIs realise the impact it will have for them? One possible argument is that it's just engineering and that all the logical stuff is already provided for.

Wider uptake and dependence on RMs for design, planning and change management becomes apparent. Design patterns, based on a widely accepted conceptual understanding of the archetypical architecture(s) of research infrastructures become more prominent.

Architectures become agile and dynamic, requiring continuous re-appraisal and evolution of RMs to suit new circumstances.

Relationships with requirements and use cases

TC_16 Description of a National Marine Biodiversity Data Archive Centre seeks to integrate the **DASSH Data Archive Centre** (a UK national facility for archival of marine species and habitat data) with other European marine biological data (e.g., data curated by EMSO, SeadataNet, JERICO and EMBRC) as a joint contribution to EMODNET Biology, the COPERNICUS provider. This is a typical test case for the ENVRI Reference Model.

Using the ENVRI Reference Model (RM), **IC_12 Implementation of ENVRI(plus) RM for EUFAR and LTER** seeks to describe two RIs with (in part) very different framework requirements. EUFAR (European Facility for Airborne Research) is an emerging RI to coordinate the operation of instrumented aircraft and remote sensing instruments for airborne research in environmental and geo- sciences. LTER (Long-Term Ecosystem Research) is a global effort aiming at providing information on ecosystem functioning and processes as well as related drivers and pressures on ecosystem scale (e.g. a watershed).

A number of other use cases (for example: **SC_3**, **TC_2**, **TC_4**, **IC_3**) would probably also benefit from applying RM thinking and concepts in their analysis and design. Each of these use cases contains one or more detailed scenario descriptions and explanations that could benefit from being thought about from the different viewpoints of science ("the business"), information and computation. Ultimately, engineering and technology aspects also become important.

Issues and implications

Reference Models (RM) and the ENVRI RM in particular have a significant role to play in fostering the use of common language and understanding in the architectural design of environmental research infrastructures. Adoption and use contributes significantly towards the goal of interoperability among research infrastructures. However, there are social barriers to be overcome. These have to be addressed by marketing, education and training.

Lack of training is a key issue, and with it the lack of skilled architects.

RMs have been ranked by the first round of ESFRI research infrastructure cluster projects as one of the top three issues needing to be addressed jointly across all RIs at the European level.

Further discussion of the reference model technologies can be found in Section 4.2.13. This takes a longer term perspective and considers relations with strategic issues and other technology topics

Reference Models (RM) and the ENVRI RM in particular have a significant role to play in fostering the use of common language and understanding in the architectural design of environmental research infrastructures. Adoption and use contributes significantl

Bibliography and references to sources

Bloomberg, J. (2014) *Enterprise Architecture: Don't be a fool with a tool*. Forbes 7th August 2014. .

DFT WG – RDA (2015). RDA Data Foundation and Terminology - DFT: Results RFC. Eds. Gary Berg -Cross, Raphael Ritz, Peter Wittenburg. Date: 29/06/2015. Consulted on: 04/03/2016. Available at: <https://rd-alliance.org/system/files/DFT%20Core%20Terms-and%20model-v1-6.pdf>

ENVRI Reference Model <http://envri.eu/rm>

Field, Suhr et al. (2013) *Realising the full potential of research data: common challenges in data management, sharing and integration across scientific disciplines*. doi: [10.5281/zenodo.7636](https://doi.org/10.5281/zenodo.7636)

Hardisty, A., (2015) *Reference models: What are they and why do we need them?* Accessed 18 April 2016.

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

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

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

ISO/IEC 10746-4:1998 Information technology -- Open Distributed Processing -- Reference Model: Architectural semantics.

ISO/IEC 19793:2015 Information technology -- Open Distributed Processing -- Use of UML for ODP system specifications.

Lopez, D. M., and Blobel, B. 2009. *A Development Framework for Semantically Interoperable Health Information Systems*. International Journal of Medical Informatics, Volume 78, Issue 2, Pages 83-103, February 2009. doi: [10.1016/j.ijmedinf.2008.05.009](https://doi.org/10.1016/j.ijmedinf.2008.05.009).