








The ENVRI Reference Model

Abraham Nieva de la Hidalga¹ , Alex Hardisty¹ , Paul Martin² ,
Barbara Magagna³ , and Zhiming Zhao² 

¹ Cardiff University, Cardiff, UK

{nievadelahidalgaa, hardistyar}@cardiff.ac.uk

² Multiscale Networked Systems, University of Amsterdam,
1098XH Amsterdam, The Netherlands

pwmartin.research@gmail.com, z.zhao@uva.nl

³ Environment Agency Austria, Vienna, Austria

barbara.magagna@umweltbundesamt.at

Abstract. Advances in automation, communication, sensing and computation enable experimental scientific processes to generate data at increasingly great speeds and volumes. Research infrastructures are devised to take advantage of these data, providing advanced capabilities for acquisition, sharing, processing, and analysis; enabling advanced research and playing an ever-increasing role in the environmental and Earth science research domain. The ENVRI community identified several recurring requirements in the development of environmental research infrastructures such as i) duplication of efforts to solve similar problems; ii) lack of standards to harmonise and accelerate development, and bring about interoperability; iii) a large number of data models and data information systems within the domain, and iv) a steep learning curve for integration complex research infrastructure systems. To address these challenges, the ENVRI community has developed and refined the Environmental Research Infrastructures Reference Model (ENVRI Reference Model or ENVRI RM), a modelling framework encoding this knowledge. The proposed modelling framework encompasses a language and a notation to describe the research domain, its systems and the requirements and challenges faced when implementing those systems. By adopting ENVRI RM as an integrative approach, the environmental research community can secure interoperability between infrastructures, enable reuse, share resources, experiences and common language, reduce unnecessary duplication of effort, and speed up the understanding of research infrastructure systems. This chapter provides a short introduction to the ENVRI RM.

Keywords: Reference Model · Research infrastructure · System modelling · Design framework

1 Motivation

The construction of a Research Infrastructure (RI) is often iterative, e.g. from simple functionality to more rich set of features, or from small scale to large scale. A large

© The Author(s) 2020

Z. Zhao and M. Hellström (Eds.): Towards Interoperable Research

Infrastructures for Environmental and Earth Sciences, LNCS 12003, pp. 61–81, 2020.

https://doi.org/10.1007/978-3-030-52829-4_4

RI is often an evolution of many iterations and can be typically characterised in terms of phases of concept development, design, preparation, implementation, operation and termination¹. The RIs in the ENVRIplus² project were in different phases when they joined the project. It is thus very challenging to develop those diverse RIs and make them interoperable.

During the past few years, interoperability between infrastructures has been extensively studied, e.g. between scientific models, workflow, metadata, semantics, middleware and infrastructure [1]. To enable interoperability among different systems, a common vocabulary for design descriptions is essential. The aim of the Environmental Research Infrastructures Reference Model (ENVRI RM) is to provide a framework for specifying and building the data management services required by environmental and Earth sciences research infrastructures.

The current version of the ENVRI RM³ was published in November 2017, following more than six years of work within the ENVRI [2] and ENVRIplus projects [3, 4]. These projects documented common practices and architectures supporting environmental research infrastructures, derived from the Reference Model for Open Distributed Processing (RM-ODP) [5–8].

The ENVRI RM provides the documentation of the basic concepts, the architectural model, and different examples of use with diagrams. The users of the ENVRI RM can be designers of RIs, but it is also intended to help people who build services to support RI activities, or who produce standards to capture best practice and reusable mechanisms. The ENVRI RM gives the designer a way of thinking about the system, and structuring its specification, but does not constrain the order in which the design steps should be carried out. The ENVRI RM can be used along with any type of design/development processes.

Since the design of an RI requires large collaborative efforts, it is likely that the actual process will be iterative, filling in detail in different parts of the specification as ideas evolve and requirements are better understood. The design of a new RI may follow a classical top-down, waterfall-style pattern, while the maintenance of an existing RI will start by capturing existing constraints. The development of services can follow an agile or rapid prototyping development model, stressing modularization and fine-grained iteration. The ideas for structuring specifications presented here can be applied within any of these methodologies. They remain valid if the design approach changes and provide a common framework and vocabulary for collaboration between designers using different processes.

Many competing architectural frameworks have recently been proposed; however, the ENVRI RM offers a set of distinguishing features that make it particularly relevant for the specification of an Environmental RI. First, it has the stability derived from continuous development during two successful European funded projects (ENVRI, and ENVRIplus) spanning more than six years (2011–2019) [1–4]; during this period the ENVRI RM has been reviewed and evaluated internally and externally by design experts

¹ <http://roadmap2018.esfri.eu/strategy-report/the-esfri-methodology/>.

² EU H2020 ENVRIplus project <http://www.envriplus.eu>.

³ The ENVRI RM and its associated derivatives are published and available online www.envri.eu/rm.

and by the research community. Second, it documents common requirements of environmental research infrastructures and best practices for fulfilling those requirements. Third, there has been an extended campaign of validation and refinement which used the ENVRI RM, analysing different infrastructures and services. And fourth, the discoveries have been formalised in the Open Information Linking for Environmental Research Infrastructures (OIL-E), an ontology framework designed to facilitate analysis, classification, and validation of RI designs; supporting the documentation of crosscutting requirements; and facilitating metadata exchange.

In this chapter, we will discuss the development of the reference model. The main aspects discussed include the context for the development of the ENVRI RM (Sect. 4.1), the main concepts supporting the modelling of environmental research infrastructure systems (Sect. 4.2), the modelling process (Sect. 4.3), and the outlook for the ENVRI RM and links to further chapters (Sect. 4.4).

2 Background of the ENVRI RM

Research Infrastructures are often complex distributed systems. Describing their structure and external properties is required to understand and manage these systems. When the system description concentrates on the distillation of general principles, it is called architecture. However, if the description is presented in a way that is useful for the derivation of a whole family of systems, it is called a framework. Hence, when describing a system supporting a broad range of applications, it is common to talk of an architectural framework. In this sense, the ENVRI RM is an architectural framework for the design of a distributed system for environmental research infrastructures.

The ENVRI RM was developed as a research infrastructure architecture framework based on the Reference Model for Open Distributed Processing (RM-ODP) [5–8]. The following sections describe the three concepts required for understanding the RM-ODP modelling paradigm: the object model, design viewpoints, and correspondences.

2.1 Object Model

RM-ODP system specifications are expressed in terms of objects. Objects are representations of the entities to be modelled. The specification and design of complex systems following the object paradigm makes use of two important object properties abstraction and encapsulation [9]. Abstraction allows highlighting aspects of the system relevant from a given perspective while hiding those of no relevance. Encapsulation is the property by which the information contained in an object is accessible only through interactions at the interfaces supported by the object [9]. In the ENVRI RM, objects are used to represent abstract entities (measurements, data sets, metadata, systems, services), physical entities (sensors, servers, networks) and social entities (institution, research group, researcher).

2.2 Viewpoint Specification

The definition of objects is distributed in viewpoint specifications. The idea behind viewpoints is to break down a complex specification into a set of individual specifications

which consistently support and complement each other [10, 11]. The design of RM-ODP aimed at serving different stakeholders by introducing the idea of a set of linked viewpoints to maintain flexibility and avoid the difficulties associated with constructing and maintaining a single large system description. RM-ODP defines five viewpoints, as shown in Fig. 1, designed to appeal to different user groups [9].

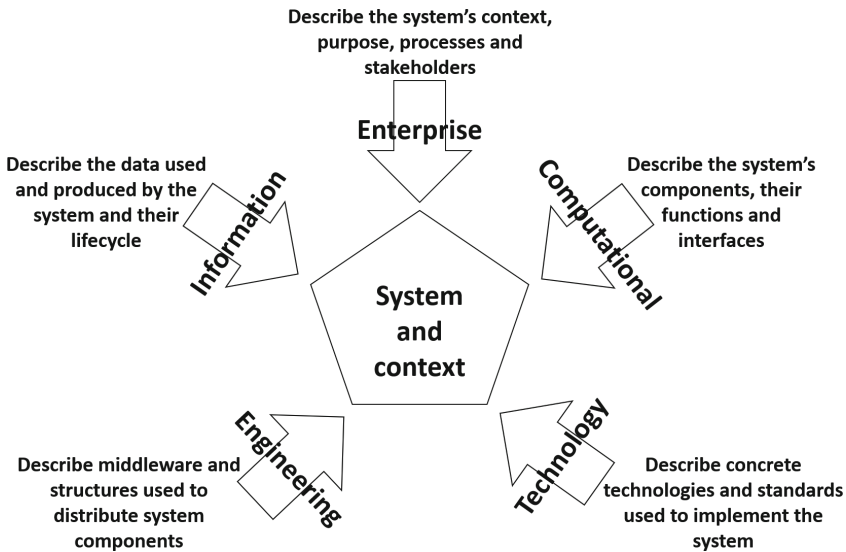


Fig. 1. The five viewpoints of RM ODP.

In the ENVRI RM, to better align the definition of viewpoints to the research domain, the *Enterprise Viewpoint* is renamed as the *Science Viewpoint*. The name change aims to acknowledge that the main type of systems modelled are intended for supporting scientific research. However, apart from this, the definition of the ENVRI RM Science Viewpoint respects the rationale, elements and structure of the RM-ODP Enterprise Viewpoint.

2.3 Correspondences

Dividing a system design in five viewpoint specifications facilitates the understanding of different groups of stakeholders. However, it is necessary to keep these specifications consistent with each other [9]. In RM-ODP, the consistency of the designs produced within each specification is maintained with the explicit mapping between elements defined in one viewpoint (e.g. objects, actions and constraints) to elements defined in other viewpoints. These mappings are formally defined as correspondence links between related elements. The correspondences can be one-to-one or one to many. A one-to-one correspondence allows mapping the representation of an element in one viewpoint to the representation of an element on another viewpoint. A one-to-many correspondence allows for an element representation in a viewpoint to be mapped to multiple elements in another viewpoint, providing a fine-grained description of that element (Fig. 2).

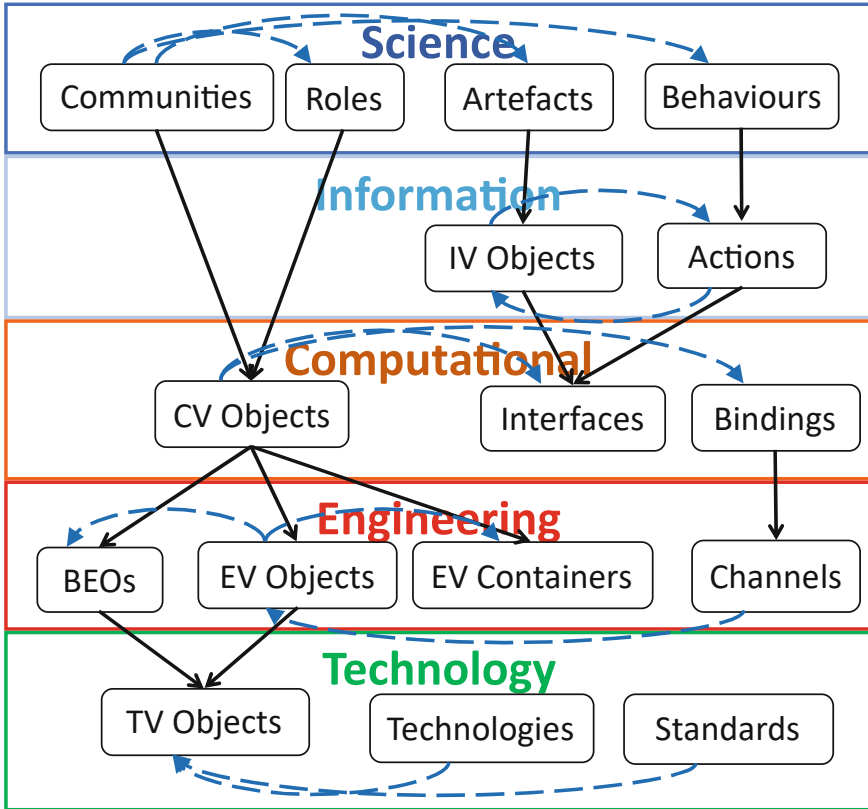


Fig. 2. Example of Viewpoint Correspondences in the ENVRI RM.

For the ENVRI RM, correspondences are formally defined in the Open Information Linking for Environmental Research Infrastructures (OIL-E) framework [12]. OIL-E is an ontology framework designed to facilitate analysis, classification, and validation of the design of a RI.

The three RM-ODP modelling mechanisms (objects, viewpoints, and correspondences) enable a complex system to be described as a set of interlinked viewpoint models. This set of models is equivalent to a single large and complex model with all viewpoints included; however, such a description is too complex to be useful. Instead, different groups of stakeholders will understand and use a subset of viewpoint specifications. A design team with members from all stakeholder groups is responsible for defining viewpoint correspondences when needed.

2.4 Domain Modelling Concepts

As stated previously, the environmental and Earth science research domain requires the development of complex systems to support data-intensive scientific research. Consequently, the systems and the data (namely research data) that they consume and produce are important modelling concepts. The explicit relationships among those concepts include the collection, curation, processing, publishing and use of research data, which is called the research data lifecycle. The following sections elaborate on these three concepts.

Research Infrastructure System. The main objective of the RI systems is the support of computational data analysis. These analyses are based on observation data collected, curated, stored and published by diverse research entities. For this reason, one of the main common characteristics of research infrastructures is that they all produce research data following a structured data lifecycle.

Research Data. Research data encompasses diverse data products derived from scientific research. The attributes which make research data stand out are that they are well-structured, carefully designed, goal-oriented, high value, and have a clearly defined lifecycle [13].

Environmental Science is observational, and currently most of the observations are made by sensors. This data is then translated into a digital representation creating research data. The increase in the number and diversity of sensor devices integrated with sensor networks has spurred an increase in the size and variety of data produced. Research data derived from these observations is a valuable asset which needs to be preserved and managed to derive the maximum value from it [13]. Although the size of the data sets produced is continuously growing, research data is different from what is known as big data. Big data usually includes data sets with sizes beyond the ability of commonly used software tools to capture, curate, manage, and process data within a tolerable elapsed time [14]. Big data encompasses unstructured, semi-structured and structured data, but the main focus is on unstructured data [15]. This difference comes from the processes that influence the creation of research data. In fact, research data are the product of carefully designed research projects. Moreover, taking advantage of big data requires the existence of well-structured datasets provided by research data (also called smart data) [13].

Research Data Lifecycle is the model of a process that covers the lifespan of research data products, from design to collection, curation, processing, publishing and reuse. Several data lifecycle models have been proposed in line with the importance assigned to research data products (for instance the data lifecycle models of the UK Data Service [16], Digital Curation Centre [17], and DataONE [18]). Inspired by these models and trying to find the most suitable for a wide range of cases presented by the institutions represented in the ENVRI consortium, the designers of the ENVRI RM looked at the commonalities of these models and produced a lightweight model of five stages (Fig. 3). The proposed lifecycle was designed to follow the main state changes to data (and metadata) as they are processed by RIs (acquired, curated, processed, published and used).

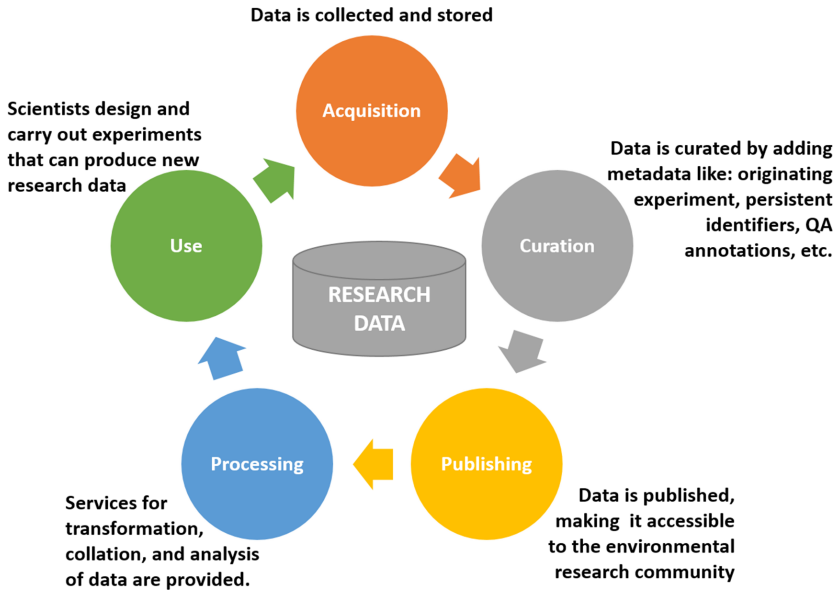


Fig. 3. The research data lifecycle model of the ENVRI RM.

The research data lifecycle model was refined with the analysis of the processes and practices for the management of research data of 26 research infrastructures (RIs) from four environmental areas (biosphere, lithosphere, atmosphere and hydrosphere) [1–4]. These analyses observed that the applications, services and software tools can be categorised following the five phases of the data lifecycle: acquiring data, storing and preserving data, making the data publicly available, providing services for further data processing, and using the data to derive other data products. The data lifecycle model was cross validated with an extended research campaign which visited seven research infrastructures. During these visits, it was observed that all the research infrastructures analysed exhibit behaviour that aligns with its phases. Furthermore, the campaign also served to validate structuring the ENVRI RM in line with the five phases of the data lifecycle.

3 The ENVRI Reference Model (ENVRI RM)

This section presents the ENVRI RM as the set of viewpoints, showing the main objects within each viewpoint, their structuring in line with the research data lifecycle and the correspondences to objects defined in other viewpoints. The ENVRI RM uses UML diagrams to produce the models of each viewpoint. UML4ODP [19] is the recommended notation for RM-ODP; however, this is not mandatory and different alternative notations can be used for each viewpoint as long as they can express equivalent concepts.

The viewpoint models proposed by the ENVRI RM aim to be as loosely coupled as possible, allowing parallel design and development among different teams. This approach allows some parts of the specification to reach a level of stability and maturity

before others. The idea of separating concerns by using a set of viewpoints can be applied to many design activities. However, components are more likely to be reused if the same set of viewpoints is accepted by many different teams. The largest possible degree of commonality is needed to support the creation of a useful architectural framework to cover a large and diverse domain, such as the development of systems for environmental research infrastructures. The ENVRI RM defines five viewpoints (Fig. 4), intended to appeal to five groups of stakeholders. The following subsections introduce the five viewpoints, describing on the objectives and areas of concern they cover.

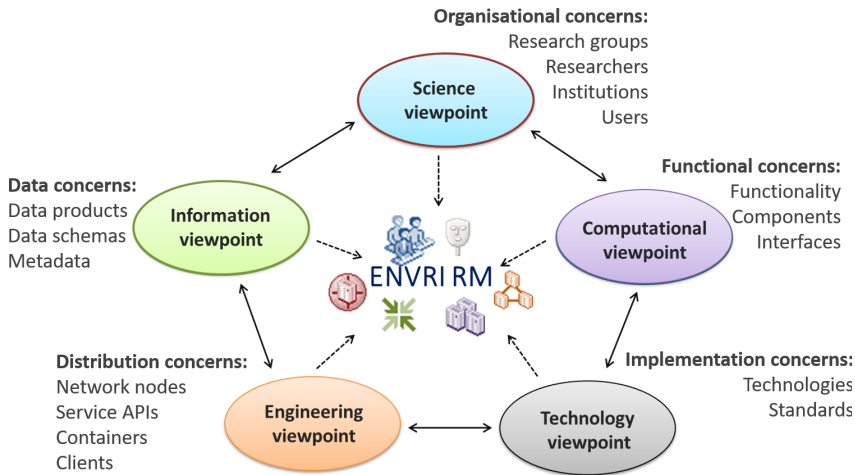


Fig. 4. The five viewpoints of the ENVRI RM

3.1 Science Viewpoint

The science viewpoint focuses on the institutional and social context of the domain in which the designed systems are intended to operate. This viewpoint concentrates on the objectives, processes, assets and policies that need to be supported by the system being modelled. The stakeholders to be satisfied are the research groups that promote the research processes, the managers making possible the operation of such processes, and the sponsors responsible for funding the research project. The emphasis is on the organisations, the research groups, their objectives, and on the environment within which the system operates.

The science viewpoint is intended to cover a wide range of operational setting; the target area can be whatever the designers are asked to describe. It can be a single experiment and its users, a research group, a larger institution, or a consortium with several partners.

The main modelling concepts of the Science Viewpoint are communities, roles, actions and artefacts. The main modelling concepts of the Science Viewpoint are *communities, roles, actions* and *artefacts*:

- *Roles* are fulfilled by objects defined in a *community*, which represents the different system stakeholders, scientists, scientific institutions, evaluation and certification agencies, as well as the information systems that provide the supporting IT services.
- *Actions* describe how the roles interact.
- *Artefacts* represent the information exchanged among them.

The diagram in Fig. 5 is a UML activity diagram. This type of diagram represents the relationships of the objects as containment (communities contain roles, roles contain behaviour and artefacts), sequencing (‘take reading’ precedes ‘collect data’), and delegation (‘acquisition system’ performs ‘collect data’ producing a ‘[raw] data set’).

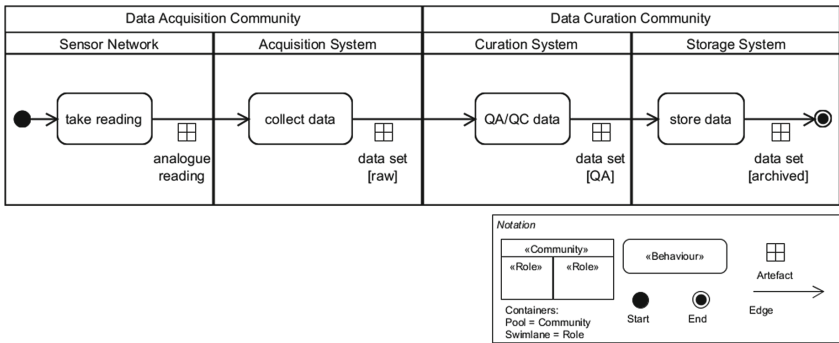


Fig. 5. The four main objects used to create science viewpoint specifications in the ENVRI RM: communities (outer container), roles (inner container), behaviours (rounded corner rectangles) and artefacts (small squares under the edge connectors (arrows)).

The science viewpoint specification enables the clear and concise representation of data processes at a high level. This specification is intended to be understood and shared by all the research infrastructure stakeholders.

3.2 Information Viewpoint

The information viewpoint specification enables the clear and concise representation of the data assets consumed and produced by the processes in the research infrastructure. The information viewpoint concentrates on modelling the data manipulated within the research infrastructure. Providing a common model that can be referenced from throughout a complete design specification assures that the same interpretation of information is applied at all levels. Common understanding about data and its interpretation minimises divergence and incomplete information among design, development and implementation teams.

The aim of the ENVRI RM information viewpoint is to achieve a shared model for the design activity, given that it can cover a federation of systems (possibly independently developed) and integration of legacy (preexisting) systems.

The main modelling concepts of the Information Viewpoint are *information objects* and *information activities*. The diagram in Fig. 6 is a UML activity diagram. This type

of diagram represents the relationships of the objects as sequencing of actions ('take reading' precedes 'collect data'), and information objects ('analogue reading' precedes '[raw] data set'). In the information viewpoint the emphasis is on the data, their evolution (change) and the activities which enable that evolution. In the information viewpoint, the artefacts specified at a high level in the science viewpoint are refined, providing a clear specification of the types, states and relationships between different data products. In addition to the activity diagrams, the specifications at this level also include class diagrams to specify the hierarchy of data assets (Fig. 7).

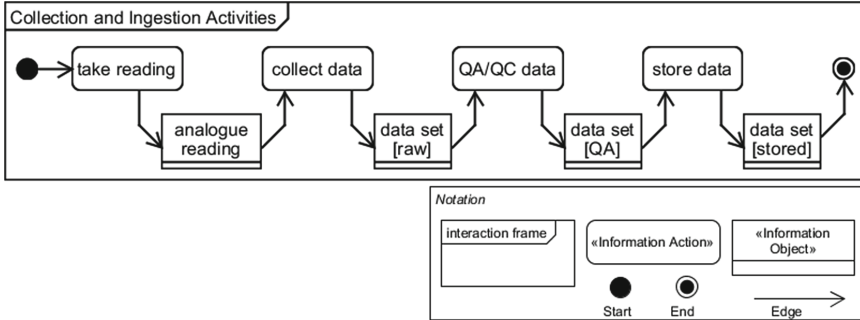


Fig. 6. The two main objects used to create information viewpoint specifications in the ENVRI RM: information objects (rectangles) and information actions (rectangles with rounded corners).

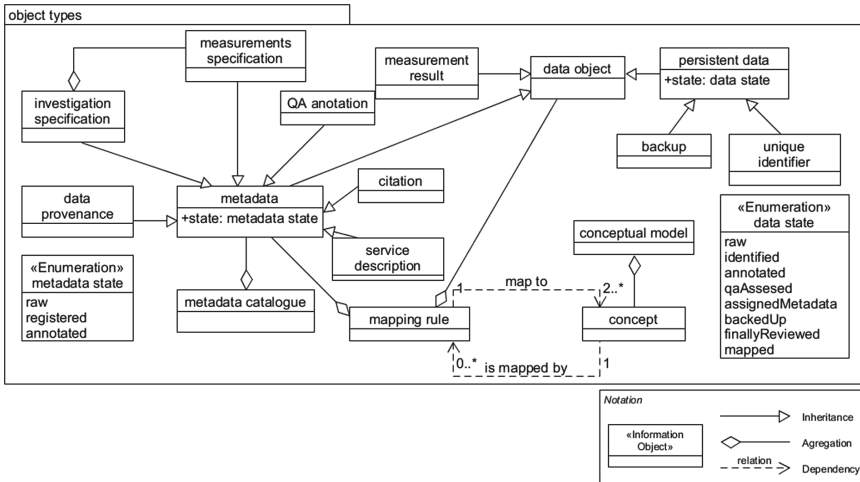


Fig. 7. A hierarchy of information objects. The class diagram emphasises the relationships of information objects such as composition, aggregation, generalisation, and multiplicity.

The correspondences between information viewpoint and science viewpoint objects can be seen directly by comparing this diagram with the one in Fig. 5. Artefacts in Fig. 5 correspond to information objects in Fig. 6 and behaviour in Fig. 5 can be mapped to information actions in Fig. 6.

3.3 Computational Viewpoint

The computational viewpoint specification models the units that provide different functionalities for processing data assets. The computational viewpoint is concerned with the development of the high-level design of the processes and applications supporting the RI research activities. This viewpoint expresses models in terms of objects with strong encapsulation boundaries, interacting at typed interfaces by performing a sequence of operations (or passing continuous streams of information). The computational viewpoint specification refers to the information viewpoint for the definitions of data objects and their behavioural constraints.

The main modelling concepts of the Computational Viewpoint are computing objects, their passive and active interfaces, and the relevant configurations in which objects are integrated to provide their services. The diagram in Fig. 8 is a UML component diagram. This type of diagram represents the relationships of the components as containment (nested subcomponents), and sequencing ('take reading' precedes 'collect data').

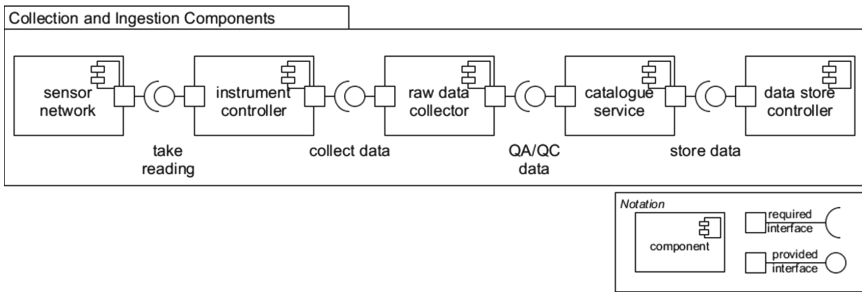


Fig. 8. Component objects and their interfaces. Component diagrams like this are used to create computational viewpoint specifications in the ENVRI RM.

3.4 Engineering Viewpoint

The main goal of the engineering viewpoint is to represent the distribution of components among different hardware and software systems. For instance, containers representing subsystem can be nested inside containers representing hardware platforms (servers and/or networks). The engineering viewpoint tackles the problem of diversity in infrastructure provision, and it gives the prescriptions for supporting the necessary abstract computational interactions in a range of different situations. It thereby offers a way to avoid lock-in to specific platforms or infrastructure mechanisms. An interaction may involve communication between subsystems, or between objects hosted in various servers, and accordingly different engineering solutions will be used.

The engineering viewpoint is also concerned with providing a set of guarantees (called transparency) to the designer. Providing a transparency involves taking responsibility for a distribution problem, so that the computational design does not need to worry

about it. The transparency mechanisms needed are provided in the form of standard middleware or web services components, simplifying the engineering specification, since it can reference the existing solutions and merely state how they are combined to meet the infrastructure needs of the system.

The main modelling concepts of the Engineering Viewpoint are engineering objects, containers and channels. In Fig. 9, the diagram represents two subsystems (acquisition and curation) which in turn contain (host) different basic engineering objects. The objects in one subsystem can communicate with other objects using standard interfaces (e.g. APIs).

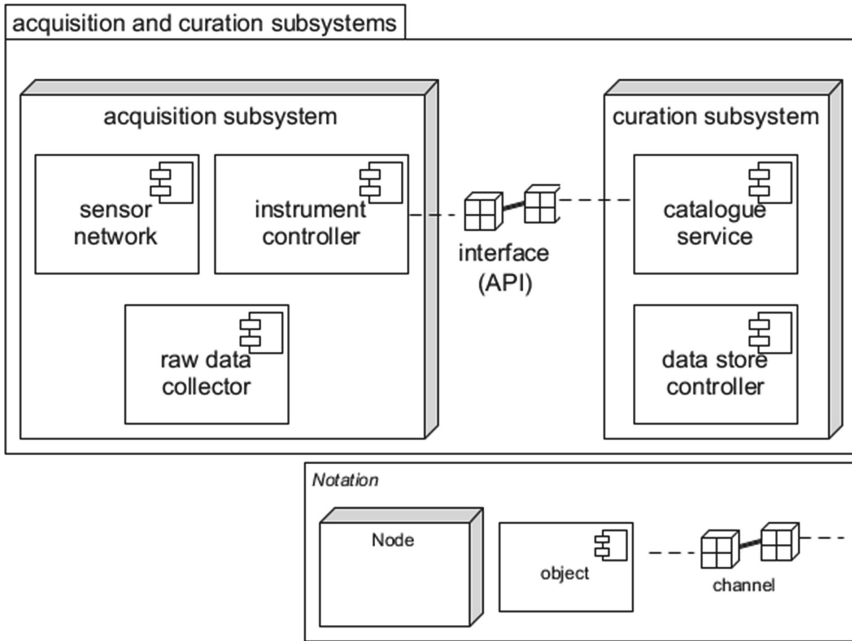


Fig. 9. Deployment diagrams are used to create engineering viewpoint specifications in the ENVRI RM. This type of diagram represents the relationships of the engineering objects as containment (nested node containers), and interfaces (communication channels).

3.5 Technology Viewpoint

Technology Viewpoint specifications are intended to represent the concrete dependencies between design and implementation. The technology viewpoint is concerned with managing real-world constraints, such as restrictions on the hardware available to implement the system within budget, or the existing application platforms on which the applications must run. The designer never really has the luxury of starting with a green-field, and this viewpoint brings together information about the existing environment, current procurement policies and configuration issues. It is concerned with selection of ubiquitous

standards to be used in the system, and the allocation and configuration of real resources. It represents the hardware and software components of the implemented system, and the communication technology that provides links between these components. Bringing all these factors together, it expresses how the specifications for an ODP system are to be implemented.

This viewpoint also has an important role in the management of testing conformance to the overall specification because it specifies the information required from implementers to support this testing. The main modelling concepts of the Technology Viewpoint are *conformance points* and *standards*. In Fig. 10, the diagram represents a system component (catalogue service) and the technology constraints which condition its operation. The diagram shows three conformance points each paired with a corresponding standard or implementation constraint. For instance, the catalogue service API is a conformance point to be provided as part of the service, and its corresponding constraint indicates that the corresponding API definition should use a standard such as Open API.

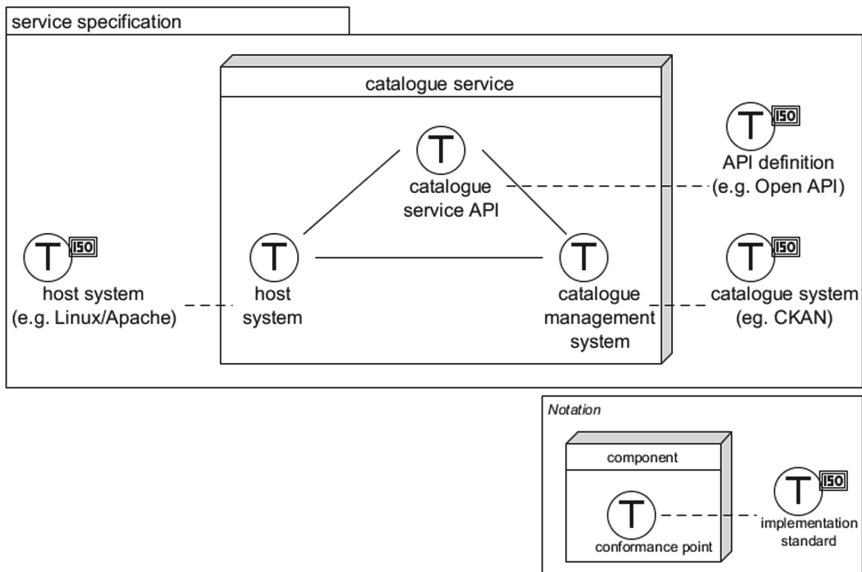


Fig. 10. Deployment diagrams are used for technology viewpoint specifications in ENVRI RM. This type of diagram represents the relationships of the objects and their implementation constraints as relationships to requirements, system configurations and services.

4 The Modelling Process

Diagrams can help understand part of the operation of a RI. However, a single diagram without context can invite many interpretations and needs to be complemented with further information when presented to different stakeholders. Different stakeholder

groups can be interested in issues such as standards, data and metadata formats, chains of responsibility, communication protocols, software and hardware dependencies and many other issues which are hard to convey on a single representation. Moreover, it is expected to find multiple sources describing how many of those concerns are addressed.

During the period from April 2017 to January 2018, the ENVRI Reference Model development team, consulted with nine environmental research infrastructures from different domains about their status and development plans⁴. The interactions during those consultations served to define a structured modelling method [20].

The proposed modelling method is recursive and consists of five steps: identification, modelling, refinement, review-revision, and mapping (Fig. 11). In this method, the designer is free to select a starting viewpoint, model the characteristics of interest within that viewpoint and then model additional details by mapping the specification to other viewpoints. The advantage of modelling using the ENVRI RM in this way is that the designer can add detail to the models while keeping consistency at different levels of abstraction. The following sections will elaborate on each of the modelling steps illustrating them with an example.

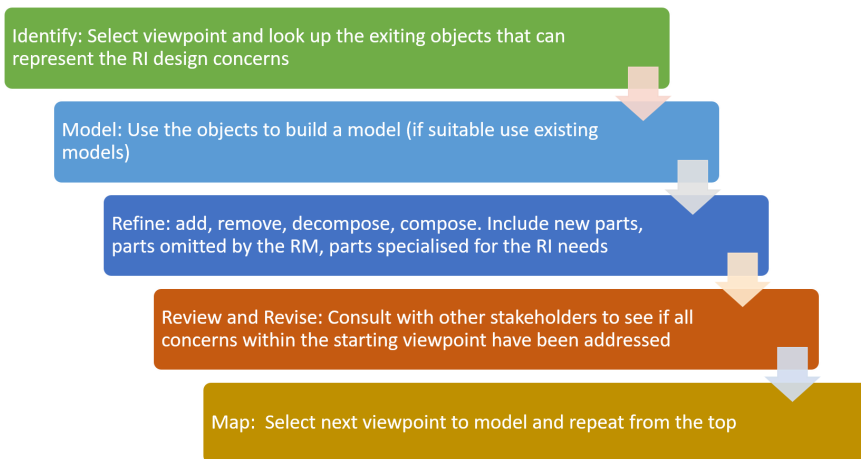


Fig. 11. The ENVRI RM modelling method.

4.1 Identify

The identification step requires gathering existing RI documentations and use it to determine the viewpoint from which to start modelling. The main representation of a system coincides with the main interest of the system designers. For instance, if the system must provide data with well-established formats, the information viewpoint might be the best described specification of the system. Similarly, if the main challenge is the integration

⁴ As part of the ENVRI visits [19].

of processing components, then a computational specification that describes the operations to be supported might contain the most complete description of the system. In this scenario, the recommendation is to identify the most complete specification of the system and start by mapping it to one of the existing viewpoints. This will help in further understanding the systems and discovering which attributes of the system are common (shared with other RIs, domain independent) and which are special (unique, domain dependent).

In the case of EPOS, the main model describes the architecture of the RI systems using a block diagram (Fig. 12). This description is complemented with the definition of the functions of each of the components [21]. The description of components, their functionalities, and integration matches the concepts described by the computational viewpoint of the ENVRI RM, which is designated as the starting viewpoint to model. After deciding to start with the computational viewpoint, the viewpoint objects are revised to select the ones that can be used to represent the concepts of the initial model. Figure 13 shows how computational viewpoint components can be used to build a model equivalent to the EPOS architecture. The mapping is not one to one, there are components which cannot be mapped to existing computational viewpoint components, such as the Thematic Core Services and Workspace Connector, these are addressed by creating custom models, as explained in the next section.

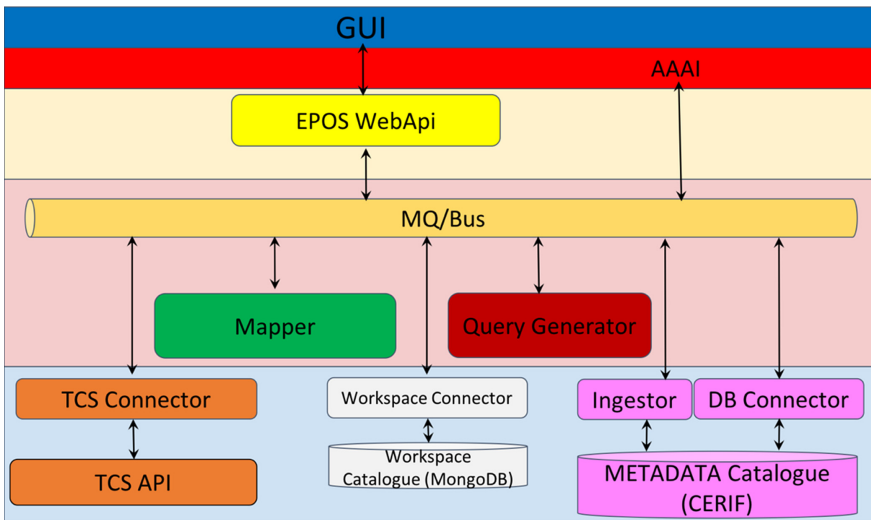


Fig. 12. EPOS Integrated Core Services Layered Architecture [21].

4.2 Model

The ENVRI RM is not expected to cover all possible cases, consequently some of the entities described in the infrastructure design will not have equivalent viewpoint object representations. In these cases, new objects can be defined and modelled to implement

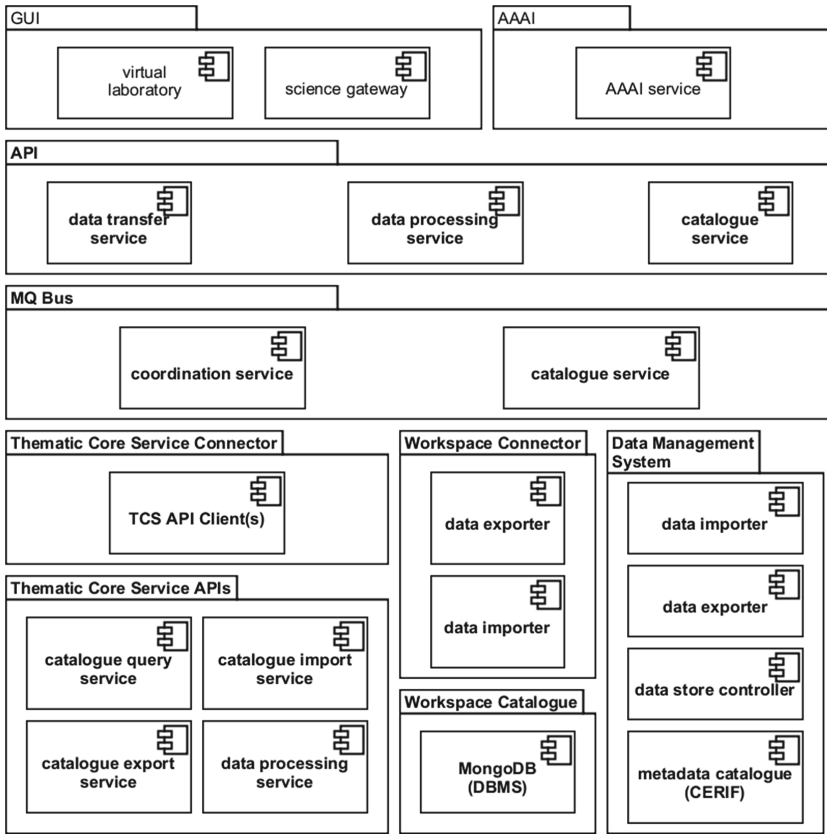


Fig. 13. Initial mapping of Integrated Core Services Layered Architecture using the ENVRI RM.

the required functionalities. Continuing with the EPOS example, Thematic Core Services and Workspace Connector are two cases in which components described in the architecture do not map one-to-one to existing reference model objects. For instance, the diagram in Fig. 14 shows the components required to provide the functionality of the thematic core services components.

4.3 Refine

The refinement of the models requires integrating the components in different configurations to provide additional functionalities. Continuing with the example, ENVRI RM components can be composed as shown in Fig. 15. The diagrams show the composition of the catalogue export service. Notice that the model is built using existing ENVRI RM components.

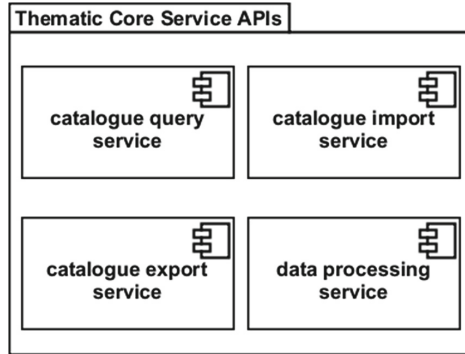


Fig. 14. ENVRI RM model components selected providing the functionality of Thematic Core Services (TCS). TCS require components for cataloguing and data processing (four services).

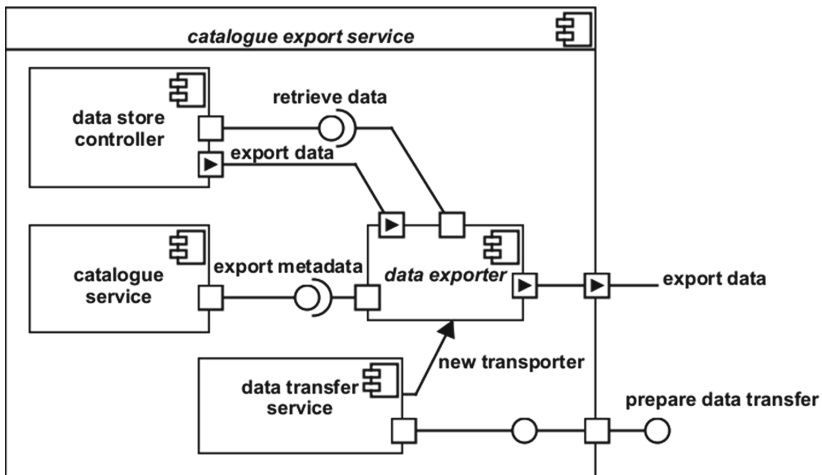


Fig. 15. Model of the Catalogue Export Service component, required for implementing the export data functionalities required by the Thematic Cores Services of the EPOS Architecture. The model is a refinement of the component specified in Fig. 14.

4.4 Review

In the review step, the models and compositions are discussed with the relevant stakeholders to determine if the models are complete and represent the entities considered in the original RI representation. To facilitate the discussion, further configuration diagrams can be produced, to show how the components are supposed to interact. For example, Fig. 16 shows a configuration describing how the components can be integrated to support importing data from different thematic core services for the EPOS case example.

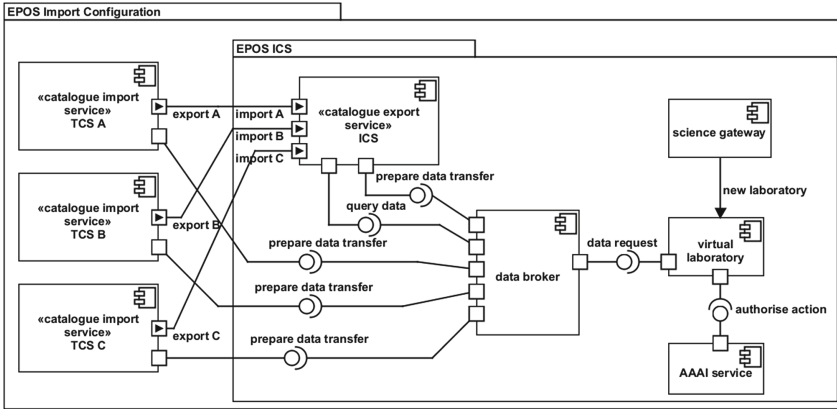


Fig. 16. Model of the configuration of components to support importing data from different thematic core services the configuration uses both the custom components designed to provide the functionality required by EPOS (Catalogue Import and Export Services) and with standard ENVRI RM components (data broker, virtual laboratory, AAAI service, and science gateway).

4.5 Map

The next stage requires determining the next viewpoint to model and using the correspondences to produce the initial models for that viewpoint. If the system stakeholders require to a concrete definition of the data assets consumed and produced by the computational components, the ideal next viewpoint would be the information viewpoint. Alternatively, if the stakeholders need to visualise the way in which components are distributed across the resources i.e. servers, databases, and sites (existing or to be sourced). For instance, the diagrams in Fig. 16 show the catalogue query service and its corresponding mapping to an engineering viewpoint model.

4.6 Complete Modelling

The basic modelling process (identify, model, refine, review, map) can be repeated several times to obtain models covering complementary design concerns. The point at which the process should stop varies according to the intended use of the models (documentation, reporting, validation, etc.). The modellers should evaluate the benefits of creating models for each viewpoint with the rest of the stakeholders and stop the modelling process once a sufficiently fit for the purpose set of models has been obtained (Fig. 17).

5 Outlook

The ENVRI RM was designed and developed to support understanding emerging and established research infrastructures, and their operation environments (processes, systems and assets). The main goals of this research effort were to (1) discover common operations, (2) describe the systems and services which they provide and depend-on, and

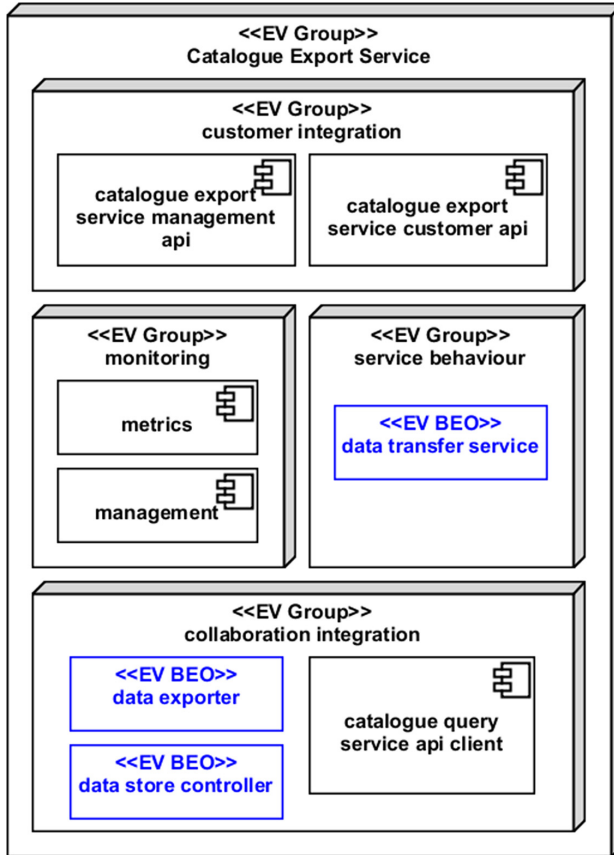


Fig. 17. Engineering Viewpoint Model of the Catalogue Export Service. This model includes the three components used in the corresponding computational model shown in Fig. 13.

(3) identify the requirements and challenges of integrating (required services, standards, and coordination).

The recommendation for the engineering viewpoint follows a microservice architecture model which allows the definition API interfaces that support flexible integration of services and systems. The recommendation for the Technology Viewpoint allows the use of templates for defining conformance points to verify the suitability of technologies and standards.

The ENVRI RM serves as a reference architecture for the evolution of the services offered and consumed by different research infrastructures into a coherent software product line. During the past years, ENVRI RM has not only been used by the RIs within ENVRIplus projects, but also application outside, e.g. for a Chinese agricultural data management infrastructure [22]. This software product line can facilitate:

- Creating client libraries for commonly used services Identifier services are a good use case, they are likely to connect to existing third-party Services (ORCID, DOI and ePIC.);
- Creating service Templates for commonly implemented services. Cross-cutting services such as cataloguing, provenance, processing, and AAI services are candidates for service templates;
- Creating engineering tools supporting the selection and use of services;

Facilitating the profiling of exiting complex solutions which may be considered for adoption, for instance, VRE implementations.

Acknowledgements. This work was supported by the European Union’s Horizon 2020 research and innovation programme via the ENVRIplus project under grant agreement No 654182.

References

1. Zhao, Z., Grosso, P., de Laat, C.: OEIRM: an open distributed processing based interoperability reference model for e-science. In: Park, J.J., Zomaya, A., Yeo, S.-S., Sahni, S. (eds.) NPC 2012. LNCS, vol. 7513, pp. 437–444. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-35606-3_52
2. ENVRI Project: Common Operations of Environmental Research Infrastructures. Grant agreement ID: 283465. FP7 programme. Start: 01 November 2011 End: 31 October 2014. <https://cordis.europa.eu/project/rcn/101244/factsheet/en>
3. ENVRIplus Project: Environmental Research Infrastructures Providing Shared Solutions for Science and Society. Grant agreement ID: 654182. H2020 programme. Start: 01 May 2015 End: 31 July 2019. <https://cordis.europa.eu/project/rcn/194947/factsheet/en>
4. Zhao, Z., et al.: Reference model guided system design and implementation for interoperable environmental research infrastructures. In: 2015 IEEE 11th International Conference on e-Science, Munich, Germany, pp. 551–556. IEEE (2015). <https://doi.org/10.1109/eScience.2015.41>
5. ISO/IEC 10746-1:1998 Information technology – Open Distributed Processing – Reference model: Overview
6. ISO/IEC 10746-2:2009 Information technology – Open Distributed Processing – Reference model: Foundations
7. ISO/IEC 10746-3:2009 Information technology – Open Distributed Processing – Reference model: Architecture
8. ISO/IEC 10746-4:1998 Information technology – Open Distributed Processing – Reference model: Architecture Semantics
9. Linington, P.F., Milosevic, Z., Tanaka, A., Vallecillo, A.: Building Enterprise Systems with ODP. CRC Press, Boca Raton (2012)
10. IEEE, 1471-2000, IEEE Recommended Practice for Architectural Description of Software Intensive Systems (2000)
11. ISO/IEC/IEEE Standard 42010:2011 Systems and Software Engineering – Architecture description
12. Martin, P., et al.: Open information linking for environmental research infrastructures. In: 2015 IEEE 11th International Conference on e-Science, Munich, Germany, pp. 513–520. IEEE (2015). <https://doi.org/10.1109/eScience.2015.66>

13. Guo, Y.K.: Assimilated learning-bridging the gap between big data and smart data. In IoTBDS, p. 7 (2017)
14. Snijders, C., Matzat, U., Reips, U.D.: Big data: big gaps of knowledge in the field of internet science. *Int. J. Internet Sci.* 7(1), 1–5 (2012)
15. Dedić, N., Stanier, C.: Towards differentiating business intelligence, big data, data analytics and knowledge discovery. In: Piazzolo, F., Geist, V., Brehm, L., Schmidt, R. (eds.) *ERP Future 2016*. LNBIP, vol. 285, pp. 114–122. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-58801-8_10
16. UK Data Service: Research Data Lifecycle. <https://www.ukdataservice.ac.uk/manage-data/lifecycle.aspx>. Accessed 28 Apr 2019
17. Digital Curation Centre: Curation Lifecycle Model. <http://www.dcc.ac.uk/resources/curation-lifecycle-model>. Accessed 28 Apr 2019
18. Data One: Data Life Cycle. <https://www.dataone.org/data-life-cycle>. Accessed 28 Apr 2019
19. Rec. ITU-T X.906 (10/2014) | ISO/IEC 19793 ITU-T: 2015 Information technology – Open Distributed Processing – Use of UML for ODP system specifications
20. Nieva de la Hidalgo, A., Hardisty, A., Magagna, B., Martin, P., Zhao, Z.: Use of the ENVRI Reference Model to Support the Design of Environmental Research Infrastructures. *Geophysical Research Abstracts*, vol. 20, EGU2018-18552, 2018 EGU General Assembly (2018)
21. EOS: Integrated Core Services Architecture. <https://www.epos-ip.org/data-services/ict-architecture/ics-architecture>. Accessed 28 Apr 2019
22. Zhao, Z., Liao, X., Wang, X., Ruan, C., Zhu, Y., Feng, D.: A Reference Model approach for developing agriculture big data infrastructures. *J. East China Normal Univ. (Nat. Sci.)* (2) 77–96 (2019). <http://xbk.ecnu.edu.cn/EN/10.3969/j.issn.1000-5641.2019.02.009>

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

