Radioactive Waste Management and Decommissioning
NEA/RWM/R(2019)5
March 2022
www.oecd-nea.org

RepMet Tools and Guidelines

A Report of the Radioactive Waste Repository Metadata Management (RepMet) Initiative

OECD
BETTER POLICIES FOR BETTER LIVES

NEA
NUCLEAR ENERGY AGENCY

Nuclear Energy Agency

NEA/RWM/R(2019)5

Unclassified **English text only**

1 March 2022

NUCLEAR ENERGY AGENCY

Cancels & replaces the same document of 3 November 2021

Radioactive Waste Management Committee

RepMet Tools and Guidelines

A Report of the Radioactive Waste Repository Metadata Management (RepMet) Initiative

This document is available in PDF format only.

JT03490464

This document, as well as any data and map included herein, are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area.

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT

The OECD is a unique forum where the governments of 38 democracies work together to address the economic, social and environmental challenges of globalisation. The OECD is also at the forefront of efforts to understand and to help governments respond to new developments and concerns, such as corporate governance, the information economy and the challenges of an ageing population. The Organisation provides a setting where governments can compare policy experiences, seek answers to common problems, identify good practice and work to co-ordinate domestic and international policies.

The OECD member countries are: Australia, Austria, Belgium, Canada, Chile, Colombia, Costa Rica, the Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Israel, Italy, Japan, Korea, Latvia, Lithuania, Luxembourg, Mexico, the Netherlands, New Zealand, Norway, Poland, Portugal, the Slovak Republic, Slovenia, Spain, Sweden, Switzerland, Turkey, the United Kingdom and the United States. The European Commission takes part in the work of the OECD.

OECD Publishing disseminates widely the results of the Organisation's statistics gathering and research on economic, social and environmental issues, as well as the conventions, guidelines and standards agreed by its members.

NUCLEAR ENERGY AGENCY

The OECD Nuclear Energy Agency (NEA) was established on 1 February 1958. Current NEA membership consists of 34 countries: Argentina, Australia, Austria, Belgium, Bulgaria, Canada, the Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Japan, Korea, Luxembourg, Mexico, the Netherlands, Norway, Poland, Portugal, Romania, Russia, the Slovak Republic, Slovenia, Spain, Sweden, Switzerland, Turkey, the United Kingdom and the United States. The European Commission and the International Atomic Energy Agency also take part in the work of the Agency.

The mission of the NEA is:

– to assist its member countries in maintaining and further developing, through international co-operation, the scientific, technological and legal bases required for a safe, environmentally sound and economical use of nuclear energy for peaceful purposes;

– to provide authoritative assessments and to forge common understandings on key issues as input to government decisions on nuclear energy policy and to broader OECD analyses in areas such as energy and the sustainable development of low-carbon economies.

Specific areas of competence of the NEA include the safety and regulation of nuclear activities, radioactive waste management and decommissioning, radiological protection, nuclear science, economic and technical analyses of the nuclear fuel cycle, nuclear law and liability, and public information. The NEA Data Bank provides nuclear data and computer program services for participating countries.

This document, as well as any data and map included herein, are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area.

Corrigenda to OECD publications may be found online at: *www.oecd.org/about/publishing/corrigenda.htm*.

© OECD 2021

You can copy, download or print OECD content for your own use, and you can include excerpts from OECD publications, databases and multimedia products in your own documents, presentations, blogs, websites and teaching materials, provided that suitable acknowledgement of the OECD as source and copyright owner is given. All requests for public or commercial use and translation rights should be submitted to *neapub@oecd-nea.org*. Requests for permission to photocopy portions of this material for public or commercial use shall be addressed directly to the Copyright Clearance Center (CCC) at *info@copyright.com* or the Centre français d'exploitation du droit de copie (CFC) *contact@cfcopies.com*.

Acknowledgements

The Nuclear Energy Agency (NEA) wishes to express its gratitude for the preparation of the report to all the members of the Integration Group for the Safety Case (IGSC) Radioactive Waste Repository Metadata Management (RepMet) initiative. A special thanks to:

- Alexander Carter (NDA/RWM, United Kingdom – RepMet Chair)
- Kevin McMahon (Sandia National Laboratories, United States – RepMet Vice-Chair)
- Jozsef Fekete (PURAM, Hungary)
- Massimo Ciambrella (NEA)

And:

- Simon Lambert (United Kingdom – Consultant)
- Mark Thorley (United Kingdom – Consultant)
- László Sőrés (BTIX, Hungary – Consultant)

Table of contents

Tables

Figures

Boxes

List of abbreviations and acronyms

Andra	Agence nationale pour la gestion des déchets radioactifs (National Radioactive Waste Management Agency, France)
CDM	Conceptual data model
DBMS	Database Management System
EBS	Engineered Barrier System
Enresa	Empresa Nacional de Residuos Radioactivos SA (National Radioactive Waste Company, Spain)
ERD	Entity Relationship Diagram
EU	European Union
IGSC	Integration Group for the Safety Case (NEA)
INSPIRE	Infrastructure for Spatial Information in Europe
ISO	International Organization for Standardization
MRMS	Minnesota Recordkeeping Metadata Standard
JAEA	Japan Atomic Energy Agency
Nagra	National Cooperative for the Disposal of Radioactive Waste (Switzerland)
NDA	Nuclear Decommissioning Authority (United Kingdom)
NDA/RWM	Radioactive Waste Management Ltd. (United Kingdom)
NEA	Nuclear Energy Agency
NWMO	Nuclear Waste Management Organization (Canada)
OECD	Organisation for Economic Co-operation and Development
OMG	Object Management Group
ONDRAF/NIRAS	National Agency for Radioactive Waste and Enriched Fissile Material (Belgium)
OWL	Web Ontology Language
O&M	Observations and Measurements
Posiva	Expert organisation in nuclear waste management (Finland)
PURAM	Public Limited Company for Radioactive Waste Management (Hungary)
RDF	Resource Description Framework
RDFS	Resource Description Framework Schema
RepMet	Radioactive Waste Repository Metadata Management (NEA)
RWM	Radioactive Waste Management
RWMC	Radioactive Waste Management Committee (NEA)

RWMO	Radioactive Waste Management Organisation
SKB	Nuclear Fuel and Waste Management Company (Sweden)
SKOS	Simple Knowledge Organisation System
SÚRAO	Radioactive Waste Repository Authority (Czech Republic)
UML	Unified Modelling Language
URI	Universal Resource Identifier
URL	Universal Resource Locator
W3C	World Wide Web Consortium

1. Introduction

1.1 The aim of the RepMet initiative

In order to support their operational, pre- or post-closure safety cases and other requirements, Radioactive Waste Management Organisations (RWMOs) manage very large amounts of data that they both produce and receive. A special characteristic of radioactive waste repositories is the long time between construction and closure of the facility – typically periods in excess of one hundred years. This means that systems handling data and relevant supporting information (metadata) will, in all likelihood, go through technological and other changes; data media and the data themselves may become unreadable and programmes handling such data may become obsolete. In addition, successive generations of workers will perform tasks on the site during this period with a high probability that not all knowledge will be handed down through the generations. Therefore, the data handling operations of RWMOs must enable the long-term, intergenerational reliability and usability of data.

Given this challenge, the main aim of RepMet has been to formulate a consistent set of guiding principles for capturing and generating metadata, in order to enable national programmes to create sets of metadata that can be used to manage their repository data, information and records in a way that is both harmonised internationally and suitable for long-term management and utilisation in safety cases and elsewhere.

Box 1.1: What is RepMet?

The Radioactive Waste Repository Metadata Management (RepMet) initiative was launched in 2014 by the Integration Group for the Safety Case (IGSC) of the Radioactive Waste Management Committee (RWMC) at the OECD Nuclear Energy Agency (NEA). RepMet analysed and investigated the application of metadata, a fundamental tool of modern data and information management, within national programmes for radioactive waste repositories. Based on this analysis it was realised that there is a great need and potential for metadata management and harmonisation.

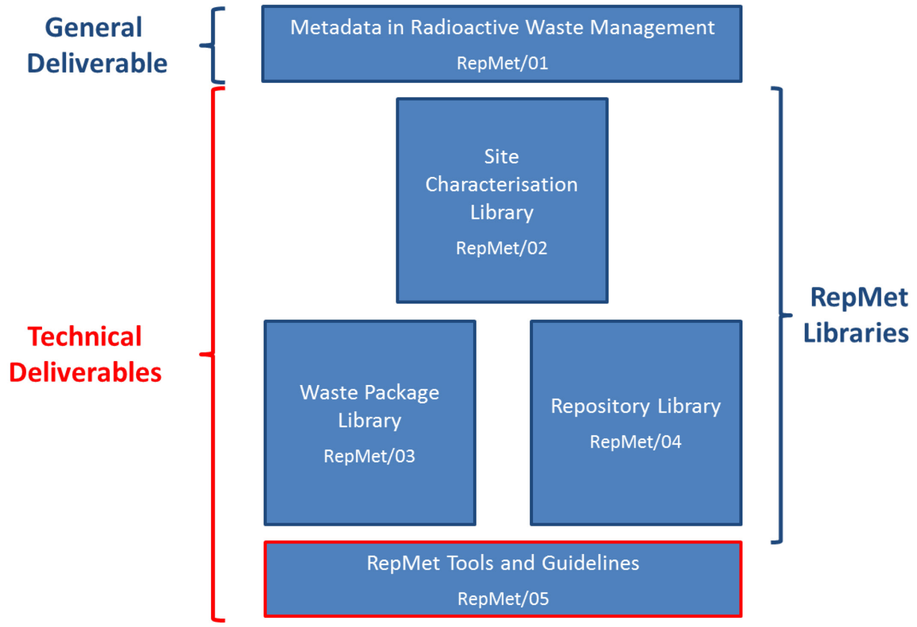
Several worldwide RWMOs and research laboratories from OECD NEA countries were involved in the RepMet initiative: Andra (France), Enresa (Spain), JAEA (Japan), Nagra (Switzerland), NDA/RWM (United Kingdom), NWMO (Canada), ONDRAF/NIRAS (Belgium), Posiva (Finland), PURAM (Hungary), Sandia National Laboratories (United States), SKB (Sweden) and SÚRAO (Czech Republic).

RepMet does not intend to promote any commercial products or services for managing metadata.

1.2 The products of the RepMet initiative and their intended audiences

RepMet has produced five key interrelated documents, summarised in Figure 1.1.

Figure 1.1: The RepMet Document Family



Source: NEA, 2019.

The information provided within these documents is primarily aimed at RWMOs that are considering developing information systems or establishing knowledge management practices related to geological disposal, or that are planning to renew or update their existing data management practices. This information is intended to be sufficiently generic to enable it to be adapted by almost any RWMO. The information may also be of use for other disciplines such as those related to developing inventory and decommissioning models.

The five documents[1] are as follows:

RepMet/01 – Metadata in Radioactive Waste Management (NEA, 2019) provides an overview of metadata and its application within RWMOs, discusses issues around the implementation of metadata, and outlines the outputs of RepMet and how they may be used. It also provides specific recommendations concerning metadata for RWMOs.

RepMet/02 – "Site Characterisation Library" (2021a) deals with data and related metadata that are considered during the characterisation of a site investigated and surveyed for suitability for radioactive waste disposal purposes, leading up to site selection.

RepMet/03 – "Waste Package Library" (2021b) deals with data and related metadata about packaged waste and spent nuclear fuel that, after proper treatment and conditioning processes, are ready for final disposal at the repository.

1. The documents are available in electronic form on the RepMet webpage of the NEA website. See www.oecd-nea.org/jcms/pl_61001.

RepMet/04 – “Repository Library” (2021c) deals with data and related metadata relating to the engineered structures and waste acceptance requirements of radioactive waste repositories.

RepMet/05 – “RepMet Tools and Guidelines” (this document) supports the libraries, providing a number of tools, methods, guidelines and approaches that were either used in developing the libraries or will be useful for RWMOs when adopting and implementing the libraries.

The documents are primarily designed for use by personnel in RWMOs with varying levels of expertise in information and data management. Please find more details on the intended audience in Table 1.1 below.

Table 1.1: Intended audiences for RepMet documents

Deliverable	Primary audience	Secondary audience
RepMet/01 – *Metadata for Radioactive Waste Management*	RWMO Managers and Decision Makers: • What metadata is and why it is valuable to their organisations; • Issues to consider in metadata implementation, and how RepMet proposals may be adopted; • High-level recommendations on metadata adoption and implementation at an organisational level. Information Systems Developers: • Awareness of benefits and risks in metadata implementation projects. • Identification of possible designated communities for metadata use.	Local and international regulators Other concerned authorities: • Awareness of role of metadata in ensuring audit trails and long-term reliability of data, information and records. Non-specialist audiences: • Understanding of best practices in information handling in RWM, and expectations on what information should be available over the long term.
RepMet/02 – Site Characterisation Library RepMet/03 – Waste Package Library RepMet/04 – Repository Library	Information Systems Developers: • Re-usable data models and controlled dictionaries developed and validated by RepMet. RWMO Engineers: • Awareness of attributes of interest to information systems for long-term access and use;	Academics: • Current best practice in metadata modelling for RWMOs, as basis for further development in future.

Table 1.1: Intended audiences for RepMet documents (Continued)

	• Agreed vocabulary for international harmonisation of terms.	
RepMet/05 – RepMet Tools and Guidelines	Information Systems Developers: • Tools and techniques for use during the implementation process; • Recommended existing standards and how they may be applied.	RWMO managers or decision makers interested in technical aspects (e.g. data modelling).

Source: NEA, 2019.

1.3 An introduction to RepMet/05 – RepMet Tools and Guidelines

The *RepMet/05* "RepMet Tools and Guidelines" aims to provide fundamental information on the metadata tools and techniques employed in the three RepMet Libraries in order to enable coherent understanding of the Libraries. However, the document should not be seen as a substitute for an appropriate professional technical grounding in information management techniques. A standard structure is used to describe each tool or technique:

- What the tool or technique is and why it is being used within RepMet;
- The fundamentals or key concepts of the tool or technique, supported by domain-specific examples;
- RepMet outputs that build on and use the tool or technique, and the use of these outputs.

Chapter 2 introduces data modelling techniques for representing the structure and the logical organisation of a database. It introduces the Entity Relationship Diagram (ERD) formalism used to design and visualise the data and metadata models in the RepMet Libraries, and explains the basic components of an ERD: entities, attributes, relationships and cardinalities. This chapter is essential preparation for Chapters 4 and 5, where the metadata standards selected for the library development are represented in data and metadata models according to the ERD formalism.

Chapter 3 explains the role, benefits and use of controlled dictionaries in information, data and knowledge management. The chapter focuses on the standards that the World Wide Web Consortium (W3C) recommends for use in the development of electronic controlled dictionaries. These have been used for the creation of original controlled dictionaries for the attributes of the RepMet Libraries' data models.

Chapter 4 illustrates the *ISO 19156 "Geographic information – Observations and Measurements"* (O&M) metadata standard. This standard is used to define a unique data model to present and encode data from any kind of observation and for features involved in sampling when making observations. The RepMet group developed an ERD formalisation of the O&M data model by defining entities, attributes, cardinalities and relationships based on the original standard. This chapter presents examples of how the

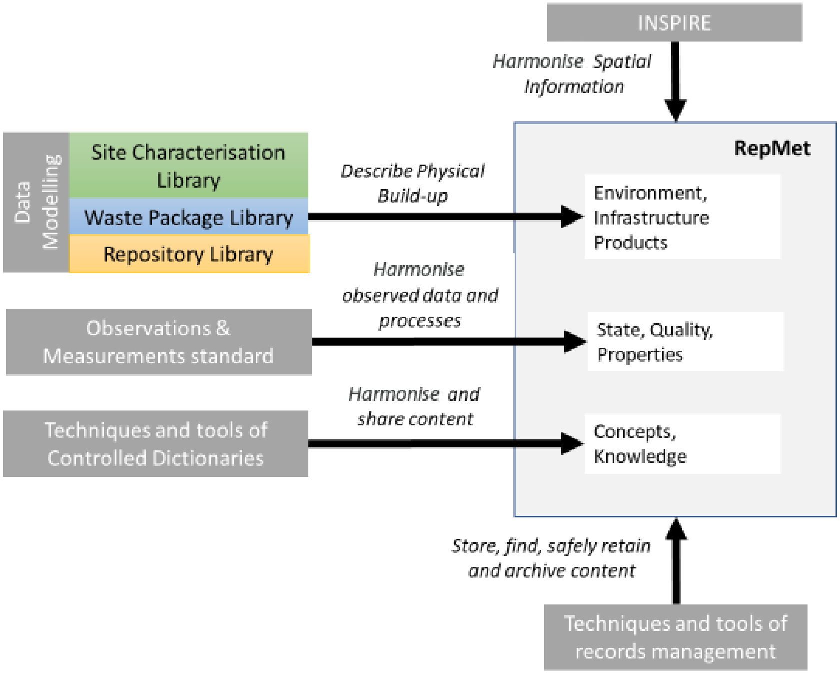
O&M data model can be applied to RWM observations by the use of a domain-specific controlled dictionary for the attributes used.

Chapter 5 deals with the long-term management of records. RepMet selected the Minnesota Recordkeeping Metadata Standard (MRMS) to show how metadata makes it easier to tackle this relevant aspect of information management. RepMet elaborated a data model based on the MRMS which is ready to be integrated into the O&M ERD data model. This chapter explains the MRMS ERD formalisation and provides domain-specific examples.

Chapter 6 provides a brief introduction of the basic elements of the EU INSPIRE Directive (Infrastructure for Spatial Information in Europe) that are considered relevant for RWMOs. It explains the benefit of applying INSPIRE in RWMO data management practices, offers concrete examples as to which INSPIRE components can be adopted by RWMOs and provides guidance on where RepMet based models can be positioned in the INSPIRE schema.

Figure 1.2 presents an overview of how the tools and guidelines presented in this report contribute to the areas of work of RepMet. Data modelling has been used to create the three libraries, which are themselves outputs of the RepMet initiative.

Figure 1.2: How the tools and guidelines contribute to RepMet



Source: NEA, 2019.

2. Data modelling

2.1. What is data modelling?

Data modelling is a method used in software engineering to define and analyse the data requirements needed to support business processes, and results in the production of a data model. Specific formal techniques and methodologies are used to produce a data model, with the data being modelled in a standard, consistent and predictable manner.

The data model is often used as an abstract representation of the structure and the logical organisation of a database. A database is an organised collection of data about a specific business areas of interest. Database Management Systems (DBMSs) are computer software systems for the interaction between the database and human or machine users. The DBMS allows users to access, modify, create and query the data content within a database. It is worth noting that there is a tendency to confuse databases and DBMSs, whereas they are two well-separated elements. If a data model is considered as the blueprint of a database, then data modelling can be considered as the database design process. Once the design is completed, it is possible to move on with the implementation of the database.

Data modelling is typically undertaken in three phases, each of which corresponds to a different level of detail contained in the model. In order, these phases are the conceptual, logical and physical data models. Each phase has a specific level of detail, expressive power and goals. Broadly speaking, the conceptual data model tackles the *what* about a database, whereas the logical and physical data models tackle its *how* at increasing levels of detail. Box 2.1 provides further details of these levels.

Box 2.1: The three levels of data models

A **conceptual data model** is a high-level model that is preliminary in structure, possibly abstract in content and sparse in attributes. It is intended to represent the semantics of an entire domain of interest, and is a collaborative product between domain experts and information system developers.

A **logical data model** is a data model that is independent of any specific DBMS platform, technology, data storage or organisational constraints. It typically describes data requirements from the point of view of information system users within the area of interest. Logical data models are created to be understood by the systems users and may be developed iteratively between the systems users and developers. Developers build the logical data model from input obtained in a collaborative approach with the systems users and concepts documented in the conceptual data model.

A **physical data model** is created from the logical data model in a process that is almost entirely within the realm of systems developers. The physical data model shows not only all of the data elements, relationships and properties of the logical model, but also the technical information to actually create the system on a particular computing platform within a particular software/networking environment. The physical data model illustrates how the information system is to be implemented on a real DBMS platform.

2.2. The need for data modelling in RepMet

The main aim of RepMet has been to formulate a consistent set of guiding principles for capturing and generating metadata. This is to enable national programmes to create sets of metadata that can be used to manage their repository data, information and records in a way that is both harmonised internationally and suitable for long-term management and utilisation in safety cases and elsewhere. Core to this aim is having well-defined data models, which will support a common understanding of the main concepts relating to radioactive waste management programmes.

RepMet created original data models and also adopted existing ones from consolidated standards such as the ISO19156 Geographic Information – Observations and Measurements (Chapter 4), or the Minnesota Recordkeeping Metadata Standard (Chapter 5).

The creation of new and specific data models aims to support the use of a consistent terminology and a common understanding of the main concepts related to radioactive waste management. This helps to address issues where waste management programmes vary from country-to-country, do not coincide exactly and may have different nuances of meaning. For example, the definition of the waste package or the engineered barrier system in a repository.

The adoption of existing and well-consolidated standards allowed RepMet to build on the extensive work of others, thus avoiding 'reinventing the wheel', and enabling easier use by RWMOs within their information systems, while also supporting data interoperability.

2.3. Using data modelling within RepMet

RepMet used data modelling in the development of the "RepMet Libraries". Each library includes both original and existing data models from selected metadata standards that are presented within this document. The use of concepts from the field of data modelling, such as Entity Relationship Diagrams, enabled RepMet to be consistent with existing methods and terminology that are well established within data and information management.

RepMet has mainly focused on the conceptual level of data modelling in order to produce data models that are independent from any database management system. RWMOs can then adapt and implement these conceptual models to fit within their specific information technology environments. Physical data modelling was considered to be outside of the scope of RepMet.

Conceptual data models are the backbones of the "RepMet Libraries", off which data and metadata are attached as explained in the following chapters. These data models are represented as Entity Relationship Diagrams using the notation detailed in Figure 2.5.

The exception to this approach was within the "Site Characterisation Library" (NEA, 2021a) where, because of the strong connection with the EU INSPIRE Directive, a logical data model was adopted. However, a conceptual level overview has been produced of the data models based on INSPIRE[2] to show how these could be utilised in the field of geological disposal.

2.4. Summary of technical basis of data modelling

A data model is defined in terms of entities, attributes and the logical relationships between the entities. A conceptual data model specifies the semantics of a business domain and the entities represent real world or abstract objects in this domain. The attributes illustrate the

2. RepMet produced data models as Entity-Relationship diagrams based on the Unified Modelling Language data models used in the INSPIRE directive.

main characteristics and properties of the entities, and the relationships assert a logical association between entities reflecting the nature of the objects within the domain.

A conceptual data model is the initial step in the database design process, where users, typically business domain experts, and the developers of the information systems collaborate to develop the definitions of the elements of interest in the specific domain[3].

A logical data model specifies the abstract data structures needed to reflect the logical schema supported by the type of DBMS being used. The most commonly used logical schemas are the hierarchical, network, relational and objected-oriented models, and further information about these can be found in specialist literature on data modelling and database design.

A physical data model illustrates how the data structures of the logical phase are physically implemented on a specific DBMS platform.

2.4.1. Entity Relationship Diagrams

The Entity Relationship Diagram (ERD) is a formal technique for visualising a data model using specific notations to depict data in terms of **entities**, the **attributes** of those entities, and the **relationships**[4] between entities. This technique is quite widespread for the production of conceptual data models.

In the RepMet framework, the entities can be both real world and abstract objects; they may also be composites of other entities. Attributes express the entity properties relevant for the database. Each entity may have one or more relationships with other entities determining some restrictions or constraints[5].

Entities

An entity is a class of persons, places, objects, events or concepts about which we need to capture and store data. Examples include:

- Persons: agency, contractor, customer, department, student.
- Places: sales region, building, room, branch office, campus.
- Objects: book, machine, product, raw material, software licence.
- Events: application, award, flight, order, requisition.
- Concepts: account, block of time, bond, course, fund.

Each entity should be well-defined so that users understand clearly what it refers to without ambiguity. For the simple example of a student, we would represent this entity on an ERD by the box shown in Figure 2.1.

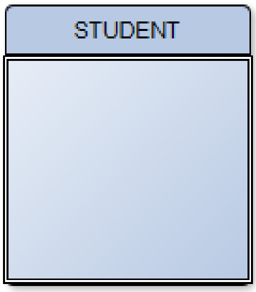
---

3. Conceptual data models are often used as an aid to communication between the users defining the requirements for an information system, and the systems engineers who develop an information system in response to those requirements.

4. This method of representing data models was developed by Peter Chen – see (Chen, 1976).

5. For example, in the Waste Package Library, *waste*, *wasteform*, *container* or *overpack* are real world entities; *waste package* is an example of composite entity. The number of overpacks that can be used with a given waste is a constraint for the relationship between these two entities. The *total activity* or the *gamma activity* is an example of attribute for the waste entity.

Figure 2.1: Example of an entity



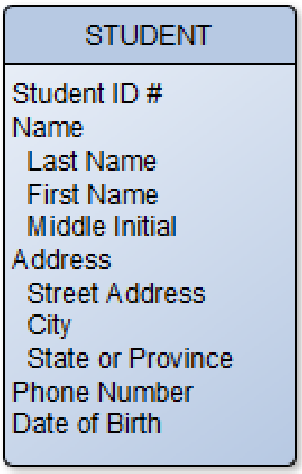
Source: NEA, 2019.

Attributes

An attribute[6] is a descriptive property or characteristic of an entity. Each attribute must be well-defined so that users understand clearly what it refers to without ambiguity. For the student example, attributes may be physical, such as hair colour, height, etc., or non-physical, such as first name, last name, address, etc. On the ERD, attributes can be represented as shown in Figure 2.2.

A special type of attribute is the compound attribute[7]. Compound attributes consist of two or more attributes, for example in Figure 2.2 the attribute entitled ‘Name’ is formed from a ‘First Name’, a ‘Middle Initial’ and a ‘Last Name’.

Figure 2.2: Example of the student entity with attributes



Source: NEA, 2019.

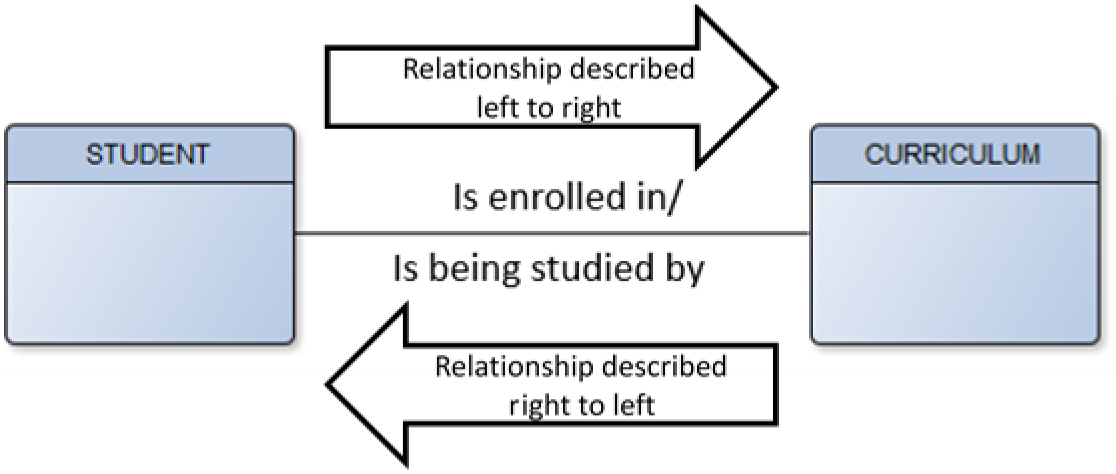
6. Synonyms include element, property and field.

7. In different data modeling languages a compound attribute can also be known (amongst other things) as a concatenated attribute, composite attribute or data structure.

Relationships

A relationship is a natural association that exists between one or more entities. A relationship may represent an event that links the entities or merely a logical affinity that exists between the entities. If the ERD containing the student entity also contained a curriculum entity, there would be a relationship between these two entities since each student is enrolled in a particular curriculum, and each curriculum is studied by students. On the ERD, relationships are represented by lines as shown in Figure 2.3. Where relationships are named, these names are often directional.

Figure 2.3: Relationship example between two entities

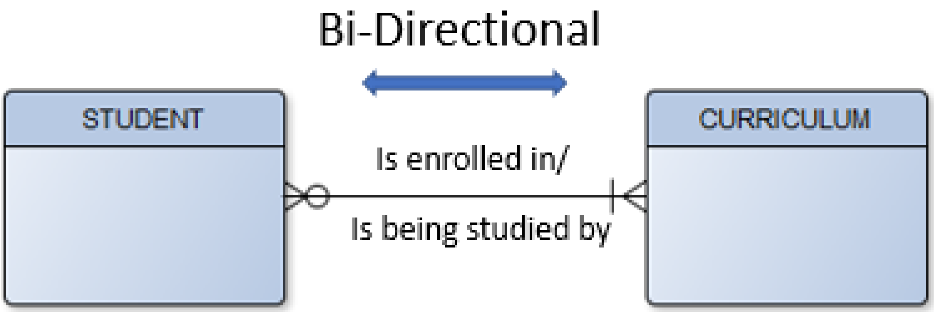


Source: NEA, 2019.

Cardinalities

Each relationship on an ERD has an associated cardinality. This describes the minimum and the maximum number of occurrences of one entity that may be related to a single occurrence of the other entity. Because all relationships are bidirectional, cardinality must be defined in both directions for every relationship. For the student-curriculum example, each student must be enrolled in one or more curriculum, but each curriculum may be studied by zero or more students (there may be a curriculum which no student has chosen to study this academic year). The cardinality is represented on the ERD through the use of a graphical marker on each end of the relationship as shown in Figure 2.4.

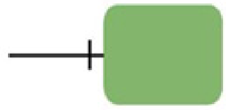
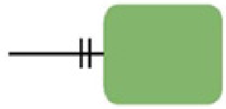
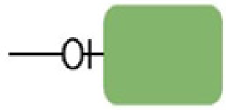
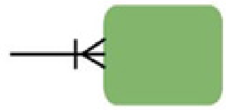
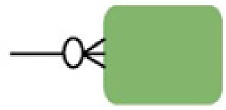
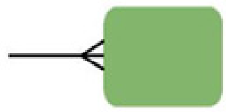
Figure 2.4: Cardinality example between two entities



Source: NEA, 2019.

Figure 2.5 shows all the possible cardinalities for a relationship between two entities in an ERD, together with this graphical marker.

Figure 2.5: Cardinalities for a relationship between two entities within RepMet

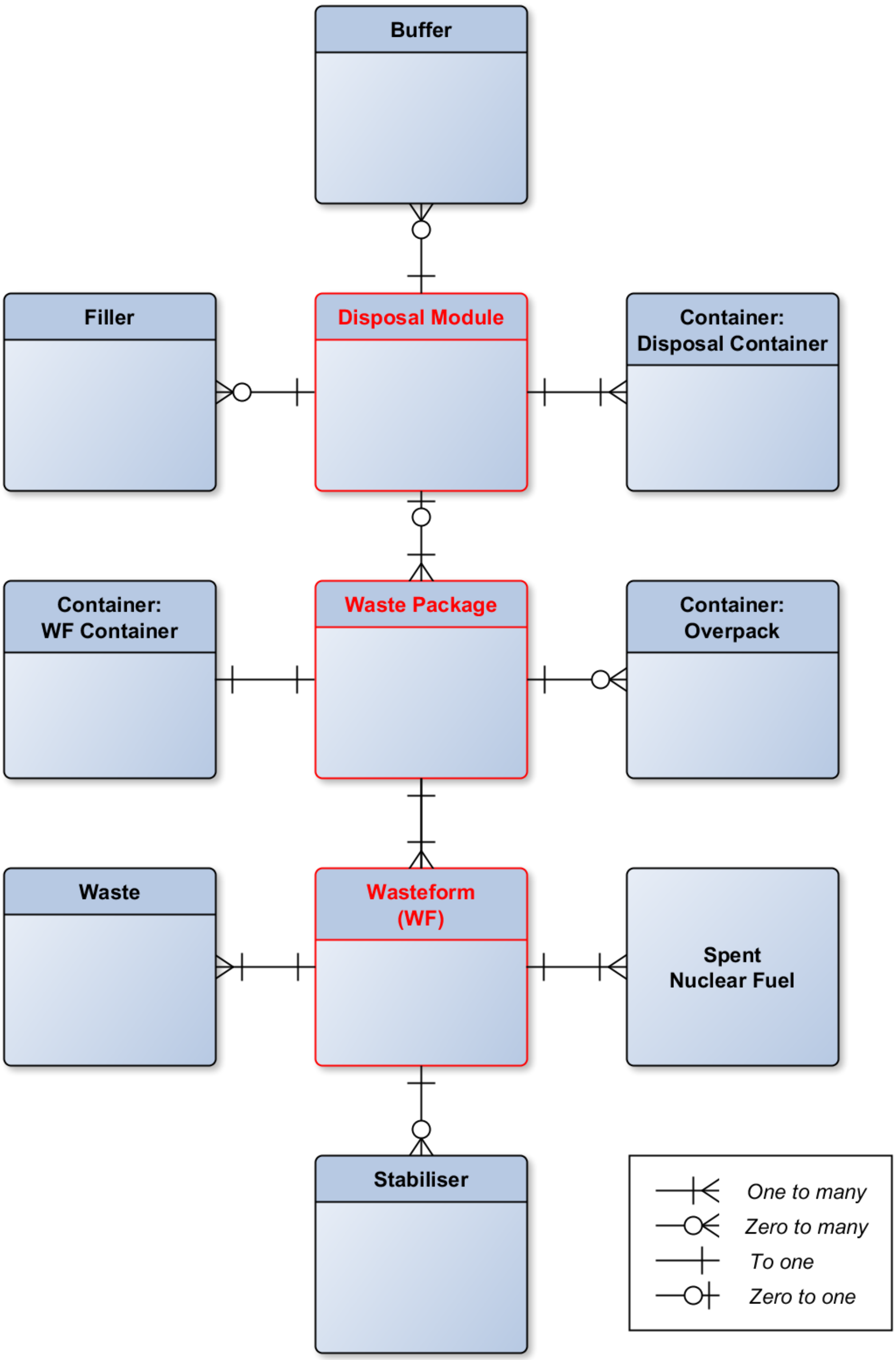
CARDINALITY INTERPRETATION	MINIMUM INSTANCES	MAXIMUM INSTANCES	GRAPHIC NOTATION
Exactly one (one and only one)	1	1	 – or – 
Zero or one	0	1	
One or more	1	many (>1)	
Zero, one, or more	0	many (>1)	
More than one	>1	>1	

Source: NEA, 2019.

2.4.2. Example of a RepMet Entity Relationship Diagram

The ERD for the Waste Package Library conceptual data model (Figure 2.6) provides a useful example of how the range of cardinalities are applied in a real world example.

Figure 2.6: ERD for the Waste Package Library conceptual data model



Source: NEA, 2019.

The entities in red (Wasteform, Waste Package and Disposal Module) are composite entities. Each of these composite entities is a logical unit, which includes everything physically contained within it. So for example, the Wasteform describes all the contents of a wasteform container (even if there are multiple wastes, spent nuclear fuel or stabilisers within). A Waste Package contains one or more Wasteforms, however, a Wasteform can only belong to one Waste Package. Similarly, a Disposal Module must contain one or more Waste Packages, however, a Waste Package does not necessarily need to be within a Disposal Module.

2.4.3. Unified Modelling Language vs Entity Relationship Diagrams

Unified Modelling Language (UML) provides the information systems community with a stable and common design language for developing applications. As well as data modelling, UML is used in many other aspects of system development.

The RepMet development team recognises that many information system specialists use UML in preference to Entity Relationship Diagrams to describe data models. However, as the RepMet documentation is primarily designed for use by personnel in RWMOs, regardless of whether they have a strong background or not in areas such as data modelling or database development, ERDs have been used in preference to UML as they are considered easier for non-technical specialists to understand.

For those who want to understand UML in more detail Booch et al (2005) provides a good introduction.

3. Controlled dictionaries

3.1. What are controlled dictionaries?

A controlled dictionary (also called a *controlled vocabulary*) is a collection of agreed terms that a community or an organisation uses, manages and maintains in a controlled way within a particular domain of interest. The terms will refer to entities within the domain and their attributes. All terms in a controlled dictionary have unambiguous and non-redundant identification, and may be connected to each other through clearly defined relationships declaring for example that one term is broader than another (so that every instance of the latter term will necessarily be an instance of the former). There may also be multilingual labels for terms, allowing consistent usage in different languages. Therefore the use of a controlled dictionary encourages clear terminology and communication within and across organisations.

Controlled dictionaries may present a logical structure of the terms. According to the logical relationships among the terms, it is possible to distinguish several kinds of controlled dictionaries such as *taxonomies, thesauri* and *ontologies*.

- A *taxonomy* is a controlled dictionary structured in a hierarchical way. The typical logical relationship is parent/child, broader/narrower, etc. (Example of relationship: *Football is a type of field sport*.)
- A *thesaurus* is a controlled dictionary where the logical relationships among the terms are not only hierarchical, but also associative or equivalent. (Example of relationship: *Soccer is a synonym of football*.)
- *Ontologies* are the most expressive controlled dictionaries. The logical relationships among the terms are very specific and customised according to the domain of use. Ontologies provide a generic way of data modelling with descriptions of objects and relations that are relevant in the specific domain. (Example of relationship: *A football team comprises 11 players*.)

A controlled dictionary might be represented as a simple glossary – a list of words with definitions – or for more elaborate cases there might be a visual representation of terms and the relationships between them. However, modern controlled dictionaries are often implemented using standards and technologies of the World Wide Web. This enables the use of modern data, information and knowledge management support in an effective way for collaborative long-term management of the information associated with the entities and attributes in the dictionaries.

Controlled dictionaries that are available on the Web may use Uniform Resource Identifiers (URIs) to specify their terms. The URI points to a resource on the World Wide Web with the description of the term: descriptions may be formatted in web pages readable by human users or in technical formats for machines. However, in both cases, they contain definitions, preferred, alternative labels in different languages, hierarchical relations to other dictionary terms, or any other types of feature.

A fundamental standard for controlled dictionaries on the World Wide Web is SKOS (Simple Knowledge Organisation System). This is a standard of the W3C to represent "knowledge organisation systems" (taxonomies, thesauri and other types of structured controlled dictionaries). SKOS is built on RDF (Resource Description Framework), another W3C standard for conceptual description or modelling of information about web resources – that is, anything that can be identified through a location on the Web. It is supported by numerous commercial and open source applications and widely used on the Web. It is a fundamental building block of the Semantic Web[8]. RDF is highly general and is not specifically intended for developing controlled dictionaries, but it is a foundational standard for others of more direct relevance such as SKOS. RDFS (RDF Schema) is a semantic extension of RDF that provides tools to create RDF vocabularies defining constraints and restrictions.

In general, the development of dictionaries requires a joint effort and needs to be controlled by some central body. There are hundreds of controlled dictionaries available on the World Wide Web developed by different communities. Scientific communities in biological, genetic research, environmental and earth sciences make extensive use of controlled dictionaries. Some important examples are listed here:

- NERC Vocabulary Server[9]
- UNESCO Thesaurus[10]
- EUROVOC Thesaurus[11]
- Geological Survey Austria[12] GBA-Thesaurus Projekt
- Statistical Linked Dataspaces[13]
- INSPIRE Codelist Register[14]

8. The Semantic Web is an extension of Web technologies to enable large-scale automated processing of resources located on the World-Wide Web, rather than only being aimed at human consumption.

9. National Oceanography Centre (2020) Resources: NERC Vocabulary Server, www.bodc.ac.uk/resources/products/web_services/vocab/ (accessed June 2019).

10. SKOS (n.b.) SKOS UNESCO Thesaurus, skos.um.es/unescothes/C00324/html (accessed June 2019).

11. European Union (2020) EuroVoc Thesaurus, https://op.europa.eu/en/web/eu-vocabularies/concept-scheme/-/resource?uri=http://eurovoc.europa.eu/100141 (accessed June 2019).

12. Geological Survey of Austria (2021) GBA Thesaurus, https://thesaurus.geolba.ac.at/ (accessed June 2019).

13. 270a.info (2014) 270a Linked Dataspaces, retrieved from https://270a.info/ (accessed June 2019).

14. EC (n.b.), INSPIRE code list register, European Commission, retrieved from https://inspire.ec.europa.eu/codelist (accessed June 2019).

3.2. The need for controlled dictionaries

The RepMet group has structured and organised the entities and attributes of the conceptual data models in the three RepMet Libraries using Web-based tools for creating and managing controlled dictionaries – most importantly the RDF/SKOS standard. These dictionaries play an important role in data harmonisation and application development in the domain of RWMOs. On the data provider side, controlled dictionaries help the development of uniform content, whereas, on the data user side, they support queries and understanding. There are several situations when using dictionaries is highly beneficial:

- When an application is used in a broad community, common terms must be harmonised between organisations or even countries.
- When an application requires the user to enter textual information, to avoid ambiguities a list of terms is provided for selection. These items are taken from a dictionary, which in this context may be called a *code list*.
- When using keywords to search a set of documents, if keywords are entered arbitrarily, the efficiency of search is greatly reduced. Using keywords from controlled dictionaries makes search more efficient.
- When textual elements on a web page are dictionary terms, a link may be provided for the user with multilingual definitions and descriptions to make the context more understandable.
- When an international service provides federated search against a distributed system, common names must be taken from common thesauri.

Furthermore, quite apart from the advantages in developing information systems, failure to adopt controlled dictionaries might lead to serious problems arising from incompatibility of terminology between organisations and consequent misunderstandings and errors. Controlled dictionaries do not need to be embedded in computerised systems to be of value; by simply reflecting agreed terminology, for example between RWMOs and their suppliers, problems of this kind can be avoided.

3.3. Using controlled dictionaries to build on the results of RepMet

3.3.1. Controlled dictionaries in the RepMet Libraries

The RepMet Libraries are sets of (meta)data models related to the main scientific and technical domains involved in the national programmes for radioactive waste management. The RWMOs can reuse the data models in the libraries to support the conceptual design of their own databases.

The backbone of each library is a conceptual data model expressed according to the techniques explained in Chapter 2. . The 11 fundamental entities relating to packaged waste and spent nuclear fuel, and their attributes, identify the main relevant properties of each object and they are structured in the form of controlled dictionaries.

The RepMet group developed its own controlled dictionaries for the entity attributes of the "Waste Package Library" and the "Repository Libraries" based on the members' experience and background, and it adopted and used the ones already available in the INSPIRE community (Chapter 6.) for the "Site Characterisation Library".

In the original RepMet controlled dictionaries, the attributes are structured in a hierarchical way. Then, for each attribute, the group identified some features such as the definition, the definition source, the reason for which a RWMO should collect data about such attribute (e.g. the attribute is a fundamental parameter for the development of a safety case) and a relevant comment.

3.3.2. The process to construct the controlled dictionaries

The RepMet team utilised a defined process to generate and make available online RDF/SKOS controlled dictionaries. The process comprises three main steps (see Figure 3.1):

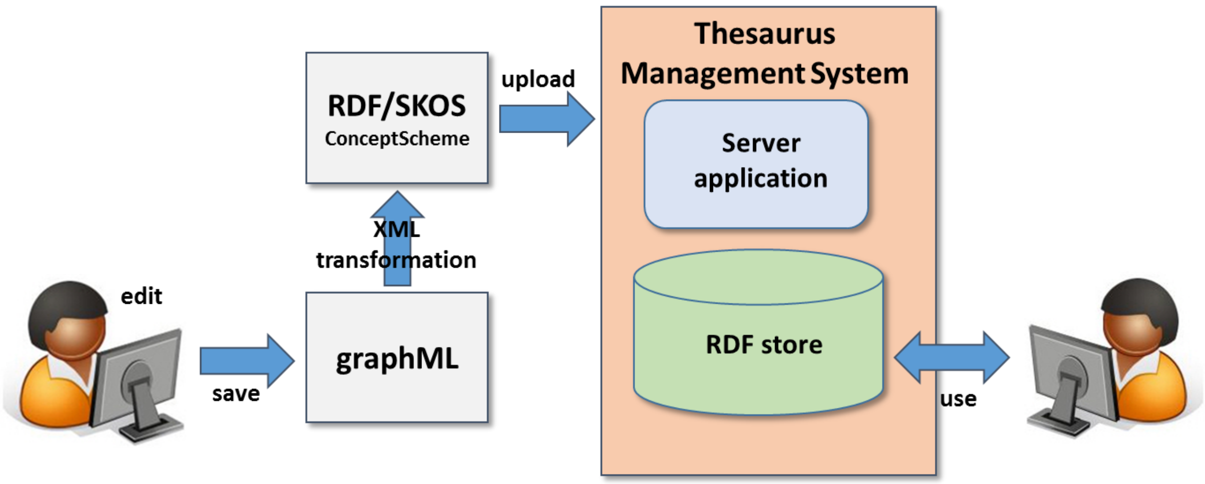
- edit and save;
- transform;
- upload for publishing.

The RepMet group used a general purpose graph editor[15] for structuring the attributes of the controlled dictionaries in mind-map format. Figure 3.3 shows a mind-map created in this way.

Once the mind-map was completed, it was saved in graphML format, an XML standard for graphics. Then, using the eXtensible Stylesheet Language (XSL) transformation, the graphML was itself converted into HTML and RDF/SKOS (XML serialisation) format.

Finally, the HTML and RDF/SKOS files were uploaded to an appropriate Thesaurus Management System. The RepMet group decided to use a static website with HTML links to the RDF resource descriptions. However, the management systems could be a dynamic service with RDF triple store, SPARQL endpoint and user front-end.

Figure 3.1: Process workflow for RDF/SKOS controlled dictionaries



Source: NEA, 2019.

15. The editor used was yEd – see www.yworks.com/products/yed. This is a free general-purpose graphical editor, that enables creation of mind-maps by adding attributes as elements to a chart and connecting them as needed. However, RepMet does endorse the use of any particular product.

3.3.3. Using the RepMet controlled dictionaries

It would be possible for information systems developers in RWMOs to follow the same steps as the RepMet team as outlined above to customise dictionaries for their own applications. However, it is obvious that editing the existing contents of these dictionaries is not desirable, as it would undermine the common understanding and definitions developed within RepMet, which are intended and expected to be of wide applicability. Nonetheless it is possible that some terms in the dictionary might be refined with the addition of sub-types relevant to particular RWMOs, or that completely new terms might be added as long as they are consistent with the existing dictionary.

A more likely use of the RepMet controlled dictionaries is to take them as they are and adopt them in development of organisational processes and information systems, fulfilling the functions listed in section 3.2 and others besides. A RWMO with well-developed and mature metadata models already in use might use the dictionaries for comparison with others and for export of its own data, while a RWMO still at an early stage might take them as they are as a ready-made basis for creation of processes and systems. A further application would be to provide multilingual translations with clear definitions.

The exact format to be used will depend on the application, as will the mechanisms for ensuring continued consistency if the RepMet dictionaries are later updated. The example given in section 3.4.4 illustrates different ways that a single term (In this case, “Cellulosic material”) may be represented, in either human readable or computer-processable forms.

3.4. Summary of technical bases of controlled dictionaries

3.4.1. Web-based standards for controlled dictionaries

In principle, the development and adoption of controlled dictionaries does not in itself require any special standards or technology. The essence is an agreement among a community on consistent vocabulary for the terms of interest to that community. A simple glossary to be printed off and displayed on an office wall might be sufficient in some situations. However, in practice much more than that will be needed. There will generally be a need to specify complex relationships such as equivalence of terms, “broader” and “narrower” relationships between terms, or translations of terms into other languages. Furthermore, there will need to be mechanisms for extending, updating and maintaining the controlled dictionary.

In today’s information technology environment, the obvious way to handle such intrinsically collaborative working is to use the technology of the World Wide Web. There are a number of standards that are especially relevant for controlled vocabularies, of which one, SKOS (Simple Knowledge Organisation System), has been adopted by RepMet for its own work. SKOS is specifically intended to support the use of knowledge organisation systems such as thesauri, classification schemes, subject heading lists and taxonomies within the framework of the Semantic Web. SKOS is built on the more basic standards Resource Description Framework (RDF) and RDF Schema, which are introduced below.

Annex A gives more technical detail about these standards and their use in RepMet, illustrating with examples.

3.4.2. Underlying standards: RDF and RDFS

Resource Description Framework (RDF) is a standard of the World Wide Web Consortium (W3C) for conceptual description or modelling of information about web resources.

Resources can be both physical things and abstract concept, as long as they can be denoted by a Uniform Resource Identifier (URI) (W3C, 2014).

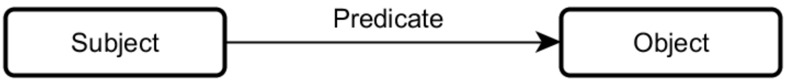
RDF permits statements about resources in the form of *subject-predicate-object*, known as a *triple*. The core idea of RDF is that any resource can be described by one or more triples, where for each triple:

- *Subject* is the resource to be described.
- *Predicate* is a resource expressing a property associated to the subject.
- *Object* is the value that the predicate assume: it may be another resource or a literal value such as a character string.

An alternative and similar way to understand a RDF triple is the following: it declares a relationship, which the predicate indicates, between two physical or abstract things that the subject and the object denote.

From a pictorial point of view, each triple can be associated to a RDF graph that can be visualised as a node and directed-arc diagram, in which each triple is represented as a node-arc-node link (Figure 3.2).

Figure 3.2: RDF graph



Source: NEA, 2019.

RDF can be considered as a technique of conceptual data modelling. If the subject-predicate-object triple is considered as a data model, then a resource is synonymous with entity. The RDF graph plays the role of the Entity Relationship Diagram.

The simplicity of RDF means it can be used to build complex systems of semantically related resources. In fact, large number of RDF statements can be assembled and queried to develop advanced information systems (see Box 3.1 for a summary of the query language). Structuring information in RDF allows it to be passed between computer applications in an interoperable way.

Box 3.1: RDF Query Language (SPARQL)

W3C developed the RDF Query Language known as (SPARQL) to extract and manipulate the information contained in databases of RDF triples. SPARQL is recognised as one of the key technologies of the Semantic Web. Using SPARQL it is possible to exploit information from the Web based on relations, conjunctions, disjunctions and logical patterns for queries. Websites may provide a "SPARQL endpoint" providing access to their datasets.

RDF provides a framework to make simple statements about resources. However, RDF does not make any assumptions or restrictions about the properties (e.g. what kind of property can be assigned to the resource? what value can they assume?) or the types of resources (e.g. what kind of resources can be used? is there any association between the allowed kinds of resources?). In other words, RDF lacks semantics. It misses a collection of terms (i.e. a vocabulary) with their own URIs and special meaning that can be used to formulate the RDF triples for a certain domain. In practice, RDF is typically used in combination with RDF vocabularies that include collections of URIs intended for use in the triples.

RDF Schema (RDFS) is a semantic extension of RDF. It provides tools to create RDF vocabularies defining constraints and restrictions about the properties and the type of resources to be used in triples. RDFS is a RDF vocabulary for its part composed of basic URI terms that can be used to articulate RDF vocabularies specific for the domain of interest.

RDFS works in a way similar to the object-oriented programming paradigm (i.e. it defines concepts or classes that then are instanced) and it is based on the fundamental notions of *class* and *propert*y:

- *Classes* are types (or categories) of resources sharing common properties.
- *Properties* are characteristics of the resources.

RDFS provides the semantic tools to define the classes and the properties of the RDF vocabularies. These tools are themselves a set of RDF resources that RDFS predefined as classes and properties.

Once the RDF vocabularies of interest are built, the user can create RDF triples where:

- The defined classes are typically used as subject or object; and
- The defined properties are typically used as predicates of the RDF triples.

Several RDF vocabularies have been created using RDFS. The first RDF vocabulary used worldwide was the Friend of a Friend (FOAF) for the description of resources about social networks. Other successful examples of RDF Vocabularies are the Simple Knowledge Organisation System, outlined below, and the Dublin Core Metadata Initiative (DCMI), which provides a RDF dictionary for creating RDF database about physical resources such as books, DVDs, but also about web resources such as pictures, web pages and videos.

3.4.3. Simple Knowledge Organisation System (SKOS)

Simple Knowledge Organisation System (SKOS) is a W3C standard to represent "knowledge organisation systems" (taxonomies, thesauri and other types of structured controlled dictionaries) using RDF. SKOS is a RDF vocabulary to create RDF databases about structured controlled dictionaries with their hierarchical and semantic relations (W3C, 2012).

SKOS is based on the notion of "concept", which is an abstract entity that does not depend on the term used to label it. For example, "dog" is an English term to label the underlying concept; "chien" is a French term to label the same concept. The aggregation of one or more SKOS concept is a SKOS "concept scheme", and this corresponds to a structured controlled dictionary.

Each SKOS concept is a RDF resource: it means that each SKOS concept is associated to a URI. SKOS defines several sets of properties that can be used to describe a concept: they form the so-called SKOS Core Vocabulary that is a kind of RDF vocabulary. Basic properties of the SKOS Core Vocabulary are "concepts", "lexical labels", "documentation properties" and "semantic relations". Following the RDF philosophy, using these properties, each SKOS concept can be described with RDF triples that express its different properties (e.g. hierarchical position, multilingual terms, etc.) in the controlled dictionary, or more appropriately in the concept scheme.

In common with RDF, SKOS concept schemes may be represented in different ways. A graphical representation like a mind-map would offer a visually appealing and easy to understand representation; the underlying RDF triples may be displayed directly; or the whole may be "serialised" in XML, that is, converted into a linear textual representation that is machine-readable.

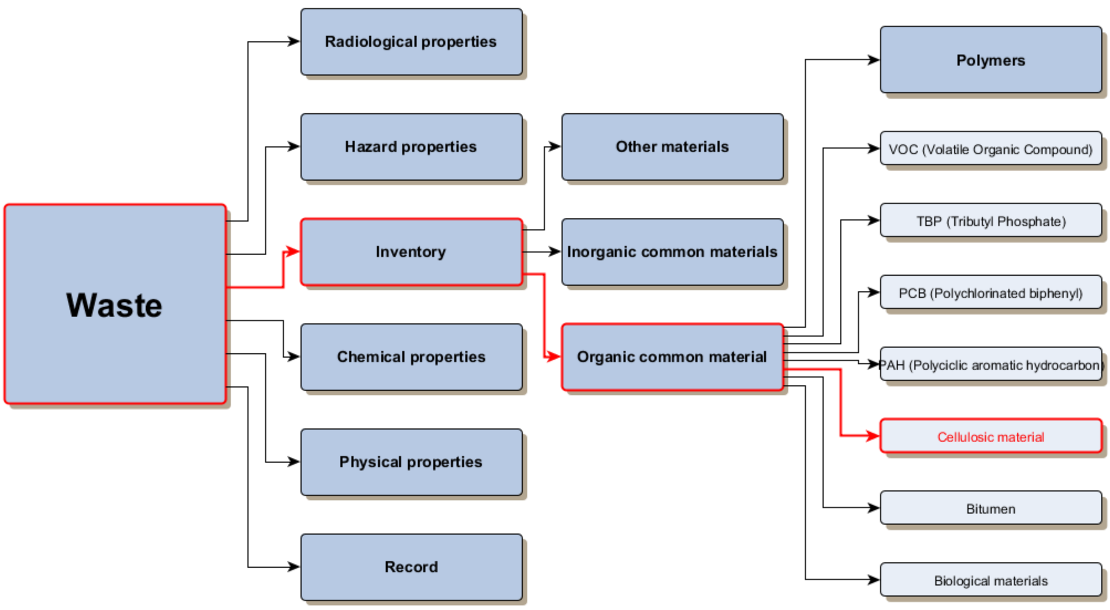
3.4.4. A specific example from RepMet

RepMet developed its controlled dictionaries following the RDF modelling and adopting the SKOS vocabulary in order to demonstrate how these tools can be applied in a real case and support the long-term management of the information stored in the electronic databases of the RWMOs.

This section provides an application example of the use of RDF predicates presented in Appendix A for the attribute "Cellulosic material" of the "Waste" entity in the Waste Package Library, illustrating how SKOS was used and the various ways in which the attribute may be represented for different purposes.

Figure 3.3 illustrates a limited portion of the controlled dictionary associated to the "Waste" entity in the Waste Package Library. The mind-map is the format that RepMet used to represent graphically the hierarchical organisation of the controlled dictionaries created for "Waste" and all the other entities.

Figure 3.3: Controlled dictionary for "Waste" (limited version) – Mind-map format



Source: NEA, 2019.

The level of detail is sufficient to note where the attribute "Cellulosic material" is located and what are the hierarchical connections with the other attributes within the "Waste" entity controlled dictionary. The arrows have different meanings, defined within the dictionary:

- *Cellulosic material* is a type of Organic common material.
- *Organic common material* is one kind of material in the Inventory.

- *Inventory* is one of the attributes of the general concept Waste, alongside others such as its radiological properties.

"Cellulosic material", as a RDF subject and SKOS concept, is a web-resource available online: the URL resource is "rpm:Waste/n53" [16]. Such URL points to an existing RepMet webpage that contains the description of the resource in terms of RDF triples using the predicates presented in Appendix A. Table 3.1 shows the application of such RDF triples for describing the RDF resource.

Table 3.1: "Cellulosic material" RDF triples[17]

***Cellulosic material* - RDF triples**		
Subject	**Predicate**	**Object**
https://www.oecd-nea.org/rwm/igsc/repmet/Waste/n53	skos:inScheme	https://www.oecd-nea.org/rwm/igsc/repmet/Waste/n0 (*Waste* resource)
	skos:broader	https://www.oecd-nea.org/rwm/igsc/repmet/Waste/n26 (*Organic common material* resource)
	skos:prefLabel	Cellulosic material @en
	skos:definition	Material containing cellule, an organic natural polymer with the formula $(C_6H_{10}O_5)n$.
	dc:source	-
	skos:scopeNote	Cellulosic material can degrade by microbial action and alkaline hydrolysis to form acidic species and to increase the pH environment.
	skos:comment	Examples of cellulosic material are: cotton, natural textile fibres, paper, wood.

Source: NEA, 2019.

Analysing Table 3.1 row by row, the RDF resource and SKOS concept "rpm:Waste/n53" has the following features:

- it belongs to the SKOS concept scheme identified as "rpm:Waste/n0" (i.e. the "Waste" entity);
- it has a broader SKOS concept available identified as "rpm:Waste/n26" (i.e. the "Organic common material");
- it has as preferred label in English "Cellulosic material";
- it is defined as "Material containing cellule, an organic natural polymer with the formula $(C_6H_{10}O_5)_n$, and this definition has not an external source;
- it has to be taken into account by the RWMOs for the waste chemical inventory since the "cellulosic material can degrade by microbial action and alkaline hydrolysis to form acidic species and to increase the pH environment";

16. "rpm:" is a namespace standing for "https://www.oecd-nea.org/rwm/igsc/repmet/"
17. The URLs provided in Table 3.1, Table 3.2 and Figure 3.4 are provided as examples only and do not currently link to any live resources.

- it comments that “examples of cellulosic material are cotton, natural textile fibres, paper, wood”.

The set of RDF triples describing the “rpm:Waste/n53” resource is available online under the NEA web domain in both human and machine-readable format:

Table 3.2 shows the HTML table illustrating in human readable format the features of the resource, while Figure 3.4 displays the XML serialisation to encode the resource features in a way that can be managed by a RDF Management System such as server application (e.g. SPARQL).

In conclusion of this example, the term *Cellulosic material* is clearly defined and positioned with respect to other terms associated with radioactive waste. The use of SKOS allows the generation of several different presentations of the term, some adapted for human readability while others (especially the XML serialisation) are suitable for computer processing. With this rigorous definition, the concept is now ready for use in applications developed by RWMOs, secure in the knowledge that the term is defined in a common and reusable way.

Table 3.2: “Cellulosic material” RDF description – Human readable format (HTML)

ID	http://www.oecd-nea.org/repmet/Waste/n53
RDF Type	http://www.w3.org/2004/02/skos/core#Concept
Broader term	Organic common material
Name	Cellulosic material
Definition	Material containing cellule, an organic natural polymer with the formula (C6H10O5)n.
Comment	Examples of cellulosic material are cotton, natural textile fibres, paper, wood.
Definition source	-
Purpose	Cellulosic material can degrade by microbial action and alkaline hydrolysis to form acidic species and to increase the pH environment.

Source: NEA, 2019.

Figure 3.4: “Cellulosic material” RDF description – Machine-readable format (XML serialisation)


```
<rdf:RDF
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:skos="http://www.w3.org/2004/02/skos/core#"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rpm="https://www.oecd-nea.org/rwm/igsc/repmet/">

  <rdf:Description rdf:about="https://www.oecd-nea.org/rwm/igsc/repmet/Waste/n53">
    <skos:inScheme rdf:resource="https://www.oecd-nea.org/rwm/igsc/repmet/Waste/n0"/>
    <skos:broader rdf:resource="https://www.oecd-nea.org/rwm/igsc/repmet/Waste/n26"/>
    <rdf:type rdf:resource="http://www.w3.org/2004/02/skos/core#Concept"/>
    <skos:prefLabel xml:lang="en">Cellulosic material</skos:prefLabel>
    <skos:definition xml:lang="en">
```

```
      Material containing cellule, an organic natural polymer with the formula (C6H10O5)n.
    </skos:definition>
    <dc:source>-</dc:source>
    <rdfs:comment xml:lang="en">
      Examples of cellulosic material are: cotton, natural textile fibres, paper, wood.
    </rdfs:comment>
    <skos:scopeNote>
      Cellulosic material can degrade by microbial action and alkaline hydrolysis to form acidic species and to decrease the pH environment.
    </skos:scopeNote>
  </rdf:Description>
</rdf:RDF>
```


Source: NEA, 2019.

4. Observations and Measurements standard

4.1. What is the Observations and Measurements standard?

Observation is a type of interaction between an observer and the external world that is used everywhere in science, technology and everyday life. In general, observation can be described as an action that follows a certain process and results in the acquisition of some data. In this sense, observations include measurements, calculations, numerical simulations and surveys.

Although the collection of information about the external world may be a highly complicated procedure, an observation itself can be modelled in a very simple and generic way using the ISO standard 19156 "Geographic Information – Observations and Measurements (O&M)" (Open Geospatial Consortium, 2013). The O&M standard was developed by the Open Geospatial Consortium[18], but its applicability is not limited to this field.

4.2. The need for the Observations and Measurements standard

The O&M standard defines a conceptual data model to represent and encode observations (and, as an extension, measurements based on sampling). In any technical or scientific area where description of observations is needed, it provides a simple and generic framework for structuring information, storing and sharing data. Instead of using hundreds of different data models for different kinds of observations, one single data model works for all. The standard establishes a basis for interoperability between separate information systems. The diversity of the real world is represented by controlled dictionaries used by the O&M data model. The addition of new fields or new kind of observations to the system is reduced to the act of improving or creating these controlled dictionaries.

The main reason to use the O&M model is to make observation data available in a well-organised and regular way. Use of the O&M standard helps support efficient search strategies, informs the user about essential metadata properties, and provides access routes to documentation and requested data.

Complex observations are often made by consecutive elementary observations: the result of one elementary observation becomes the input of another one. Such processing chains can be described by the same data model in the O&M standard.

The generic observation pattern of the O&M standard (see section 4.4.1) can also be applied to the calibration of instruments. Calibration is a special type of observation where the feature of interest is an instrument. During calibration instrument properties are measured

18. The Open Geological Consortium (OGC) is an international not-profit organisation working on the development of high-quality standards for the geospatial communities. The standards are the results of consensus process among the involved organisations and are freely available.

and changed. Measured qualifier values (e.g. offset, noise level) can be stored as results, with adjustments as process parameters.

4.3. Using the Observations and Measurements standard within RepMet

4.3.1. The RepMet Libraries and controlled dictionaries with O&M

The O&M standard provides a simple data model that is able to represent complex observation systems. It can be used for any type of observation since its structure is generic. This implies that complexity has to be transferred somewhere: it is defined within the data content, in other words in the values that these entity attributes can assume in the database.

As discussed in Chapter 3. , controlled dictionaries are an ideal tool for this purpose. The complexity of all the possible processes for an observation such as the chemical characterisation of a radioactive waste or the definition of the geological features in a site can be represented in a specific controlled dictionary. Therefore, in the O&M standard the complexity of the real world observations is transferred into domain-specific controlled dictionaries. Moreover, the use of controlled dictionaries avoids ambiguity and ensures the interoperability between different data and information management system operating with the O&M data models.

The RepMet group developed controlled dictionaries for the attributes of the entities in the CDMs of the Waste Package and Repository Libraries, and based the Site Characterisation Library on INSPIRE. In the context of O&M, these RepMet products can be seen as controlled dictionaries for properties about features in their respective fields. These controlled dictionaries are available in human readable HTML and machine-readable RDF/SKOS (XML serialisation) formats.

From this perspective, the RepMet work represents a first attempt to map radioactive waste management to the metadata models of the O&M standard. The adoption of already existing controlled dictionaries, as in the Site Characterisation Library, and the creation of new ones, as in the Waste Package and Repository Libraries, went in the direction of the definition of a set of domain-specific dictionaries that the radioactive waste management communities could adopt and use in synergy with the O&M standard. The RepMet controlled dictionaries cover mainly the properties of specific features such as the waste package; in future, new dictionaries could be created to model and describe all the processes and parameters that are typical of the radioactive waste management.

The Waste Package Library includes the following application examples:

- Observation of waste – CSD-C Weighing, chemical tests, Calculation from alloys composition

The Site Characterisation Library contains the following application examples:

- Geologic Unit – Geologic Map
- Geologic Structure – Mapped Fault
- Geophysical Measurement – Borehole Logging Measurement
- Solid Model – Seismic Volume
- Monitoring Facility – Air Pollution Monitoring Station

4.3.2. High-level data flows

Another application of the O&M standard of relevance to RepMet is in representing high-level data flows, going beyond simple observations to model chains of observations and results.

Figure 4.1 shows two datasets from a process chain as related Sampling Features (see section 4.4.8). The blue box represents a composite log dataset from borehole geophysics, the red box the result of the geophysical inversion. Two observations are involved: field data acquisition, and global optimisation by simulated annealing. The results are geophysical logs of three different properties, and a layer model, the final result showing lithology and depth. The two sampling features form a Sampling Feature Complex connected as source and target by the data flow. Borehole logging is used as input by the inversion. A high-level flow chart can easily be drawn by analysing the relations between the Sampling Features. A complete description of the data flow can be given in the layer model result using SensorML procedural language. It allows the exact definition of input and output files, processing software and parameter settings.

Figure 4.1: Geophysical log and layer model as related Sampling Features

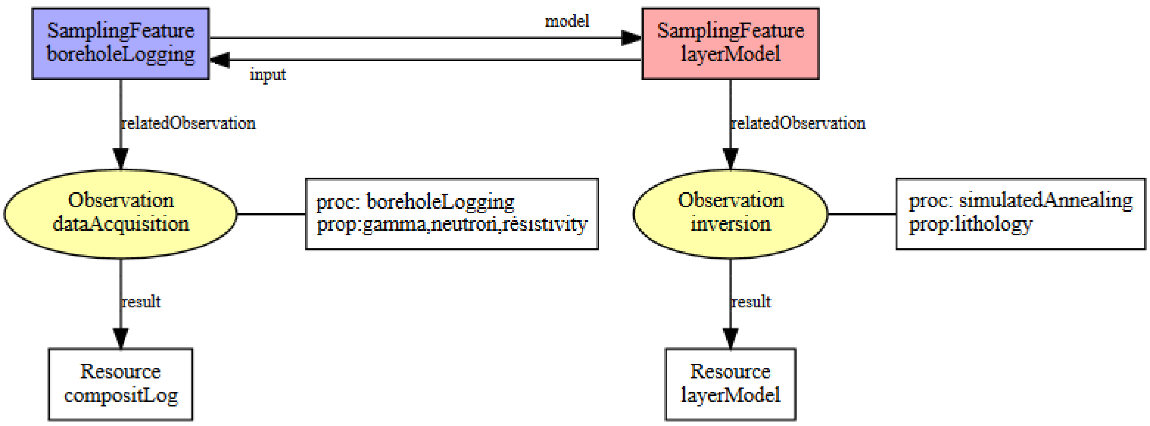
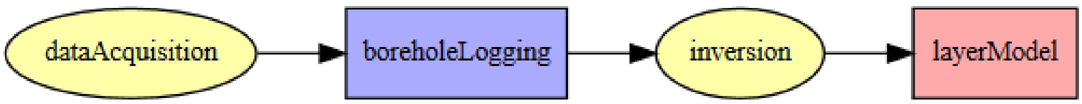


Figure 4.2: Data flow



Source: NEA, 2019.

The output of the data acquisition is a borehole logging dataset (composite log) that is used as input for the inversion. The result of inversion is a layer model.

4.4. Summary of technical basis of the Observations and Measurements standard

4.4.1. Fundamental characteristics of observations

The O&M standard identifies several fundamental characteristics of any observation, shown in Table 4.1, while Figure 4.3 illustrates the conceptual data model as an Entity Relationship Diagram. Whether the observation is simple like discerning the colour of an apple, or complex like collecting satellite images, these characteristics are sufficient to document the observation. In the first case the responsible party is a human, the feature of interest is an apple, the observed property is colour, the process is discerning colour, and the result is red.

In the second case, the responsible party may be an institution such as the *European Space Agency*, the feature of interest a *rectangular area scanned by sensors*, the observed property the *reflectance at a given wavelength*, the process the *hyper-spectral imaging* and the result the following *hyper-spectral cube*. Of course, it is an intricate procedure carried out by the computer systems on board the satellite, but to document the measurements it is enough to name the observed properties, uniquely identify the process (providing links to the proper specification) and reference the results (link to the repository where the image data are available).

Table 4.1: Observations and Measurements CDM – Definitions

Entity	Definition
Observation	Act that results in the estimation of the value of a feature property, and involves the application of a specified procedure, such as a sensor, instrument and algorithm or process chain.
Process	Procedure that is followed during the observation process to generate an observation result[19].
Result	The estimated value of a feature property generated by the procedure that was used for the observation.
Property	Any particular attribute of the feature that is estimated during the observation process[20].
Feature	Any entity that has the property for which a value is provided through the observation process.
Parameter	Variable of the process that characterise the observation[21].
Responsible Party	Person or organisation in charge of the observation[22].

Source: NEA, 2019.

Process and Result are often interfaces to other observations, allowing complex representations to be constructed.

19. Instances of this entity may be a measurement instrument or a sensor, but also a calculator (e.g. physics simulation), or an algorithm using as inputs more primitive results coming from previous observations.

20. In the Observations and Measurements standard the term 'Property' has a specific usage. It refers to an observable (chemical, physical, etc.) property. An Object Property called 'observedProperty' points to a Class called 'Property', as shown in Figure 4.3.

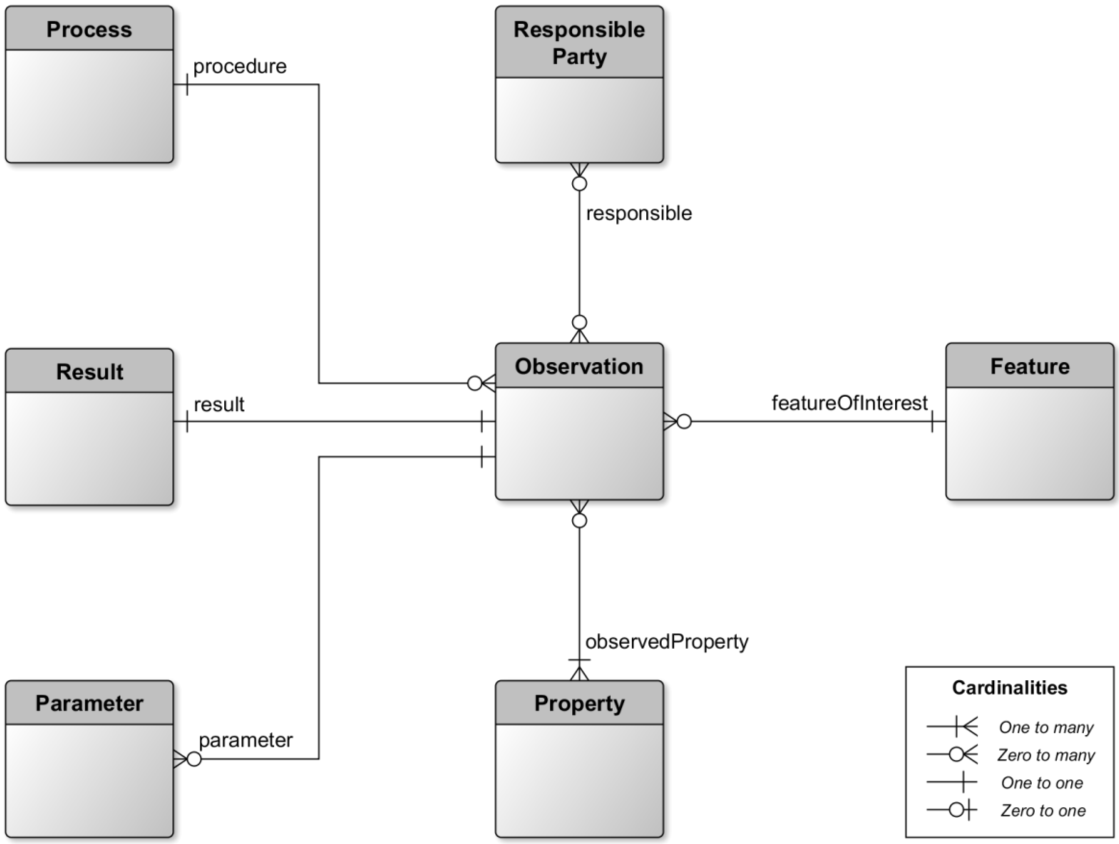
21. This entity is a simple data structure to store parameters in the name-value pair form (the corresponding data type is indicated with "Namedvalue"). The names are usually constrained by controlled dictionaries and the corresponding values are any suitable allowed for the name in the dictionary.

22. Instances of this entity may be operators, processors, reviewers, etc.

When represented as a conceptual data model, the entities have a set of attributes that are extracted from the original O&M standard: they are listed in Tables 4.2 to 4.7. For each entity attribute, these tables provide the type of associated data ("Data Type" column) and a short description ("Description" column). The same attribute may be associated with an entity more than one time. This cardinality is indicated in UML style in the "Cardinality" column: [0..1] means "zero or one", [0..N] means "zero or many", and [1..N] is "one or many" and [1..1] means "exactly one". Moreover, an entity may also include and refer to the attributes of a different entity: the table shows the name of the relationship between the two entities (in this case, the "Attribute column" report the name of the relationship between two.

The "Feature" is a special case since it may be replaced by any kind of entity that is being observed. "Feature" may be the entities of the RepMet Libraries CDMs such as, for example, the "Waste", the "Wasteform" or the "Waste Package" of the Waste Package Library. For this reason the following sections do not report the attributes for the "Feature" entity.

Figure 4.3: Observations and Measurements CDM – Entity Relationship Diagram



Source: NEA, 2019.

4.4.2. The "Observation" entity

The "Observation" entity serves as container for basic metadata, and as bridge to the other entities (Table 4.2).

Table 4.2: Attributes of the "Observation" entity

Entity	Attribute	Cardinality	Data Type	Description
Observation	**identifier**	[1..1]	*Identifier*	Unique identifier for the observation.
	name	[1..1]	*String*	Human readable name of the observation.
	responsible	[0..N]	*"Responsible Party"*	Reference to the attributes of the "Responsible Party" entity *(See Table 4.7) that* performed the observation.
	observedProperty	[1..N]	*"Property"*	Reference to the attributes of the "Property" entity *(See Table 4.6)* that is evaluated in the observation.
	phenomenonTime	[1..1]	*dateTime*	Time instant that the result applies to the property of the feature of interest.
	resultTime	[1..1]	*dateTime*	Time instant when the result becomes available.
	parameter	[0..N]	*"Parameter"*	Reference to the attributes of the "Parameter" entity *(see Table 4.5)* representing an arbitrary event-specific process parameter[23].
	procedure	[1..1]	*"Process"*	Reference to the attributes of the "Process" entity *(see Table 4.3)* used to generate the result.
	result	[1..1]	*"Result"*	Reference to the attributes of the "Result" entity *(see Table 4.4)* generated by the observation procedure.

Source: NEA, 2019.

Several attributes are connected to the other entities: for example, "processParameters" of the "process" entity are name-value pairs giving an indication about how the observation process was carried out.

4.4.3. The "Process" entity

The "Process" entity (Table 4.3) describes the procedure for observations of the same type. Process name identifies the procedure with a controlled dictionary term, allowing the usage of domain-specific terms. Examples of process names, in the case of the Site Characterisation Library, may be "2D Seismic Data Acquisition" or "Borehole Data Acquisition", whereas, in the Waste Package Library, they may be "gamma spectroscopy" or "chemical test" or "passive neutronic counting".

The "type" attribute is used for further specifying the procedure. The "documentation" attribute makes available a citation of a full procedure description: it may be human

23. This might be an environmental parameter, an instrument setting or input, or an event-specific sampling parameter that is not tightly bound either to the feature of interest or to the observation procedure. The sampling interval (e.g. 0.1m) in a borehole may be a parameter in the observation procedure.

readable text or a machine-readable SensorML document that contains a detailed description of the procedure and the related data flow. The "processParameter" attribute is a GenericName defining the name of a Parameter that is going to be used in Observations described by the Process.

Table 4.3: Model of the "Process" entity

Entity	Attribute	Cardinality	Data Type	Description
Process	**identifier**	[1..1]	*Identifier*	Unique identifier for the process.
	name	[1..1]	*GenericName*	Reference to a controlled dictionary for the possible process name.
	type	[1..1]	*GenericName*	Reference to a controlled dictionary for the possible process type.
	documentation	[1..1]	*Citation*	Citation of process documentation, (title, date, URL) It may refer to human readable text or a procedural language description, e.g. SensorML.
	processParameter	[0..N]	*GenericName*	Reference to a controlled dictionary for the possible process parameters.
	responsible	[0..N]	*"Responsible Party"*	Reference to the attributes of the "Responsible Party" entity is related to the procedure.

Source: NEA, 2019.

There is a difference between "processParameter" and the "parameter" attribute of an Observation. The latter also has a value. The set of "processParameters" is mainly used as a checklist. As an example: a "Borehole Data Acquisition" process has "samplingInterval", "depthMin" and "depthMax". An observation has samplingInterval=0.1m, depthMin=1.0m, depthMax=150m. Dictionaries of "processParameters" are expected to contain names and also descriptions. The "responsibleParty" attribute reports the list of people who are related to the procedure (such as custodian, or author of the documentation).

4.4.4. The "Result" entity

The "Result" entity is used to model and make available in the same structure different types of results from observations (Table 4.4). This model allows representation of results that are spatially and temporally both dependent and invariant.

Table 4.4: Model of the "Result" entity

Entity	Attribute	Cardinality	Data Type	Description
Result	**value**	[0..1]	*Any*	Suitable and recommended data types are provided by the SWE Common Data Model (Open Geospatial Consortium Inc., 2011).
	coverage	[0..1]	*Coverage*	Reference to GML coverage[24]
	resource	[0..1]	*CI_OnlineResource*	Reference to any external source for the result data.

Source: NEA, 2019.

The "value" attribute points to simple values encoded according to the "Sensor Web Enablement (SWE) Common Data Model" (Open Geospatial Consortium Inc., 2011) in the case the result does not depend on space and time. The "coverage" attribute encodes the result in the case it varies spatially and temporally. The "resource" attribute is a reference to some external data source such as a file, a web service or even a hardcopy document that contains the result in some format relevant to domain experts.

4.4.5. The "Parameter" entity

The "Parameter" entity is used to model the variables that characterise the observation. The two attributes that Table 4.5 shows are in form of name-value pair. The "name" attribute is constrained by a specific controlled dictionary, where the "value" attribute assumes any data type that the controlled dictionary or other constraints allow.

Table 4.5: Model of the "Parameter" entity

Entity	Attribute	Cardinality	Data Type	Description
Parameter	**name**	[1..1]	*GenericName*	Name of parameter referring to a specific controlled dictionary.
	value		*Any*	The value data type depends on the type of the parameter.

Source: NEA, 2019.

4.4.6. The "Property" entity

The "Property" entity is used to model the traits and the qualities of the feature of interest that is estimated through the observation. Properties may be quite complex in some applications, but they are generally expressed through a single attribute with the "GenericName" data type as illustrated in Table 4.6. It is essential to develop controlled dictionaries of property names specific to the scientific and technical discipline involved in the observation in order to improve the interoperability and to make observation metadata comparable and queryable.

The RepMet group developed some new controlled dictionaries (see Chapter 3.) according to the RDF/SKOS format for the properties about the features of interest of the Waste Package and Repository Library topics, that are, respectively, the "packaged waste and

24. The INSPIRE Directive, in the case of geographic data, recommends adopting the GML data model for coverage results.

spent nuclear fuel ready for final disposal at the repository" and "repository requirements and structure at closure".

Table 4.6: Model of "Property" entity

Entity	Attribute	Cardinality	Data Type	Description
Property	**name**	[1..1]	GenericName	Name of observed property referring to a specific controlled dictionary.

Source: NEA, 2019.

It is worth noting the difference between "name" attributes of "process", "parameter" and "property" entities. These types of terms are often mixed in practice, but in the framework of the O&M data model a rigorous distinction is necessary.

4.4.7. The "Responsible Party" entity

The "Responsible Party" entity is used to model people and organisations carrying out the observations. Table 4.7 reports the attribute associated to the "responsible party" entity. The "identifier" attribute identifies the responsible party. The "type" attribute specifies its nature (e.g. "radiation protection expert", "seismographic technician) with reference to a specific controlled dictionary. The "name" attribute is the formal identifying name of the responsible party.

Table 4.7: Model of "Responsible Party" entity

Entity	Attribute	Cardinality	Data Type	Description
Responsible Party	**identifier**	[1..1]	*Identifier*	Unique identifier for the "responsible party".
	type	[1..1]	*GenericName*	Type of responsible party. A specific controlled dictionary constrains its value.
	name	[1..1]	*GenericName*	Name of responsible party.

Source: NEA, 2019.

4.4.8. The "Sampling Feature" extension

The data model presented in the previous section can be used to model any kind of *direct* observations. In order to use the O&M standard to model *indirect* observations, it is necessary to adopt its extension: the "Sampling Feature" (Open Geospatial Consortium, 2007).

Indirect observations include for example observations involving sampling techniques. For such kind of observations, between the observer and the observation target there is one more feature: the "proximate feature of interest".

The example of satellite observation proposed in the previous section can be considered. The radiation intensity detected by the sensors is not a property of the Earth, but rather characteristic of the observation setup. In this case, "the observation procedure obtains values for properties that are not characteristic of the type of the ultimate feature" (Open Geospatial Consortium, 2007) (i.e. the Earth), but characteristics of an intermediate or "proximate" feature (i.e. the rectangular area scanned by sensors). However, the information obtained in such indirect way can be used to estimate property values that belong to the ultimate feature of interest and that cannot be observed directly. In this case, the ultimate feature of interest, the Earth, may be considered as a sampled feature, whereas the proximate feature, the rectangular area scanned by the satellite sensors, is a sampling feature.

This kind of observation mechanism often generates a chain of proximate features for the estimation of the value of the ultimate feature property. The same observation may be modelled in different ways, depending on the requirements. In fact, the direct observation example can be interpreted in a more complex way: from the point of view of medical science, the perception of the apple colour may be a very complex observation involving optical sensing, neuron activity and pattern recognition processes in the brain.

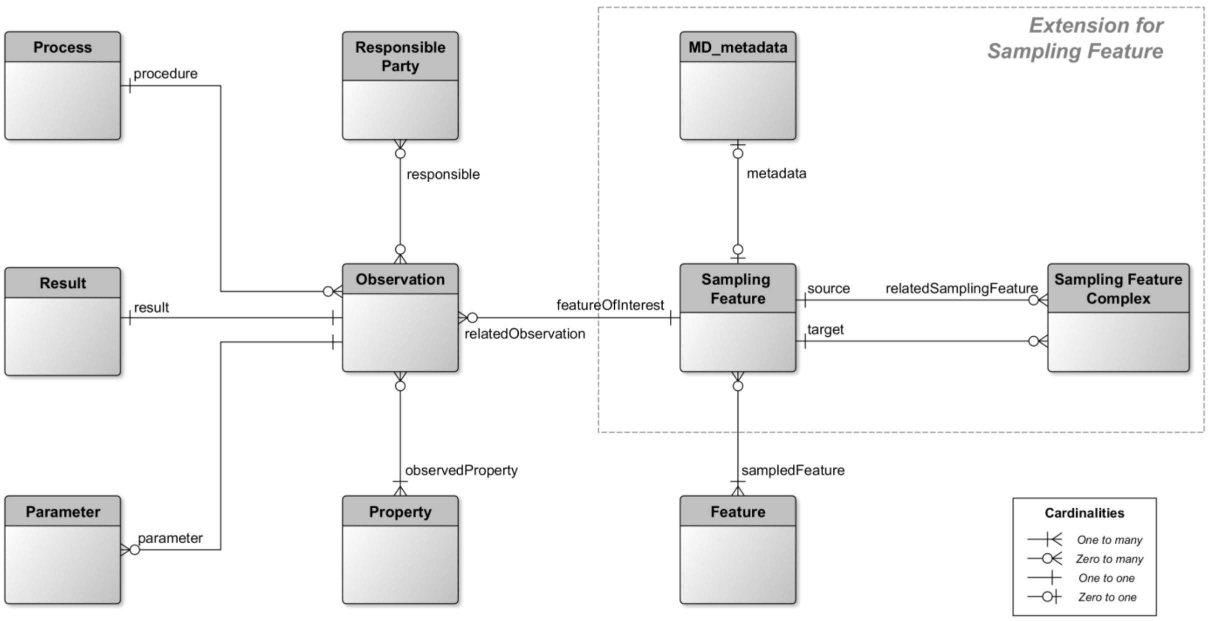
Table 4.8 shows the entities that have to be integrated in the O&M data model to handle indirect observations, while 4 illustrates the conceptual data model.

Table 4.8: Entities of the Sampling Feature model

Entity	Definition
Sampling Feature	Artefacts of an observational strategy, with no significant function outside of their role in the observation process. Sampling Features provide links between features of scientific or technical domains and observation metadata.
Sampling Feature Complex	Sampling features are frequently related to each other, as parts of associated sets (known as *complexes*), through sub-sampling and in other ways.
MD_metadata	Set of most common metadata attributes that are often used for spatial features.

Source: NEA, 2019.

Figure 4.4: CDM for O&M including Sampling Feature extension – Entity Relationship Diagram



Source: NEA, 2019.

The "Sampling Feature" entity

The "Sampling Feature" entity (Table 4.9) serves to connect the domain feature and the observation metadata records to provide standard access routes for searching and well-defined connection points for encoding. The list of attributes for this entity includes an identifier, an optional definition of its shape and the connection with at least one observation. The "sampledFeature" attribute contains a link pointing to the ultimate feature of interest of the observation.

Table 4.9: Model of "Sampling Feature" entity

Entity	Attribute	Cardinality	Data Type	Description
Sampling Feature	**identifier**	[1..1]	*Identifier*	Unique identifier.
	shape	[0..1]	*Geometry*	Shape of Sampling Feature.
	sampledFeature	[1..N]	*"Feature"*	Reference to sampled "Feature" entity attributes.
	relatedObservation	[0..N]	*"Observation"*	Reference to "Observation" entity attributes *(see Table 4.2).*
	relatedSamplingFeature	[0..N]	*"Sampling Feature Complex"*	Reference to a "Sampling Feature Complex" entity attributes *(see Table 4.10).*
	metadata	[0..1]	*"Metadata_MD"*	Reference to "MD_Metadata" entity attributes *(see Table 4.11).*

Source: NEA, 2019.

The "Sampling Feature Complex" entity

The "Sampling Feature Complex" (Table 4.10) models the associations between Sampling Features that, as explained, can be related to each other. This entity has two references to the associated Sampling Features: source and target. The type of relation between them is specified by the role attribute. For example: A process creates some measurement result that is used as input in a modelling process. In this case there are two Sampling Features involved: measurement and model. Measurement (source) has a relatedSamplingFeature that is the model (target) and the association role is "model". In the opposite direction model (source) points to the measurement (target). The association role is "input". Terms for role should be controlled dictionary items. Sampling Feature Complexes are also useful in describing process chains.

Table 4.10: Model of "Sampling Feature Complex" entity

Entity	Attribute	Cardinality	Data Type	Description
Sampling Feature Complex	**role**	[1..1]	*GenericName*	Dictionary reference to role names.
	source	[1..1]	*"Sampling Feature"*	Reference to the source "Sampling Feature" entity *(see Table 4.9).*
	target	[1..1]	*"Sampling Feature"*	Reference to the target "Sampling Feature" entity *(see Table 4.9).*

Source: NEA, 2019.

The "MD_metadata" entity

The "MD_metadata" entity (Table 4.11) is used to associate feature level metadata with Sampling Features. It provides useful attributes such as "title", "description", "responsible parties", "security classifications". The entity contains metadata elements that are found in other metadata standards such as ISO 19115 and the Dublin Core Metadata Initiative (DCMI).

Table 4.11: Model of "MD_metadata" entity

Entity	Attribute	Cardinality	Data Type	Description
MD_Metadata	**title**	[1..1]	*MlString*	Title describing the referencing entity.
	description	[0..1]	*MlString*	Short description of the referencing entity.
	otherIdentifier	[0..N]	*Identifier*	Identifiers used in different data systems or registries.
	timeRange	[0..1]	*TimeRange*	Time range characteristic of the referencing entity.
	keyword	[0..N]	*GenericName*	Keyword from controlled dictionary.
	contact	[0..N]	*"ResponsibleParty"*	Contact person with responsibility. See attributes of Table 4.7.
	classification	[0..1]	*GenericName*	Classification code for security constraints.

Source: NEA, 2019.

5. Records and record management

5.1. What are records and record management?

Records have a central role in data and information management. The exact definition of a record can vary depending on the application area, however, a common concept behind the different uses of the term *record* is that a record involves pieces of information bound and handled together. For example, a set of records may represent the employees of an organisation, with each record recording the same information (e.g. name, age and salary) for each individual.

A useful definition is provided by the framework of the NEA *Preservation of Records, Knowledge and Memory (RK&M) across Generations* initiative (NEA, 2014), where a record is: *a usually unique and original object or a selected piece of data/information that has been committed to a medium and that is kept, together with the appropriate context and structure, for later use*.

In data management, a *record* can be a row in a table or a particular occurrence of a database entity. Within the RepMet a *record* may, for example, be a collection of attributes describing a waste package, a monitoring unit or one of its' observations.

Records management are the activities during which records are being created, used, maintained and disposed of[25]. Within governments or large organisations, especially those working in safety critical domains, records management is a significant issue having its own standards. One such is the *Minnesota Recordkeeping Metadata Standard* (MRMS) (RMDC, 2015).

The MRMS specification contains two definitions for a record: a long and very specific definition for strict administrative use, and another definition that is more useful for broader purposes. According to the latter, a record is*: information that is inscribed on a tangible medium or that is stored in an electronic or other medium and is retrievable in [a useable] form.*

The RK&M and MRMS definitions are well aligned, and by these definitions, any piece of information can be called a record, regardless of its content and physical form. What makes a record special in MRMS is that it is subject to *record-keeping*, where record-keeping means *the act or process of creating, maintaining and disposing of records*. The aim is to keep track of all information that is important in the life cycle of a record from its creation to its disposal.

25. Record disposal involves removing a record either physically and/or logically from the system in which it is held.

5.2. The need for Records and Record Management in RepMet

Large amounts of data and information are created during the course of radioactive waste management activities. This ranges from the results of observation and monitoring activities through to documents such as contracts, reports, technical documentation and publications produced for a range of audiences or purposes. Programmes generate large amounts of data across multiple disciplines (e.g. geoscience, waste management or engineering) throughout the cradle-to-grave life cycle of radioactive waste from generation to disposal. This occurs within a variety of activities including site selection and characterisation, numerical modelling, repository design and construction, repository operation, repository licensing, waste packaging, safety case production and environmental impact assessment.

A special characteristic of radioactive waste repositories is the long time between construction and closure of the facility – typically periods in excess of one hundred years. This means that systems handling data and relevant supporting information (metadata) will, in all likelihood, go through technological and other changes; data media and the data themselves may become unreadable if not actively managed; and programmes handling such data may become obsolete. In addition, successive generations of workers will perform tasks on the site during this period with a high probability that not all knowledge will be handed down through the generations.

Therefore, the data handling operations of RWMOs must enable the long-term, intergenerational reliability and usability of data and information. Records management techniques provide a framework for establishing common practices for the maintenance and administration of this material.

5.3. Using Records and Record Management within RepMet

5.3.1. The Minnesota Recordkeeping Metadata Standard (MRMS)

To facilitate record management at the governmental level the *Recordkeeping Metadata Development Committee* of the US State of Minnesota developed the Minnesota Recordkeeping Metadata Standard (MRMS), releasing version 1.3 of MRMS in 2015. It shares many of its elements with other metadata standards, such as the Dublin Core[26] and ISO 19115[27]. Apart from information on format, location and access, MRMS provides elements to describe responsible parties, management, preservation history, and all administrative details that are relevant for the life cycle of material in hardcopy, analogue or digital form. See reference (RMDC, 2015) for more details.

RepMet considered that the use of MRMS for record-keeping at the government level provides a good basis for recordkeeping within RWMOs. It has also already been tested by PURAM (Hungary). RepMet therefore adopted and adapted the MRMS to provide the framework for record-keeping integrated into its metadata initiative.

26. Dublin Core: The Dublin Core Metadata Initiative provides a simple model for general-purpose metadata. There is significant overlap with ISO19115. See (DCMI Usage Board, http://dublincore.org).

27. ISO 19115 (Geographic information – Metadata) is a generic spatial-metadata standard (Open Geological Consortium [n.d.], retrieved from www.iso.org/obp/ui/#iso:std:iso:19115:ed-1:v1:en).

5.3.2. Use of the MRMS in RepMet

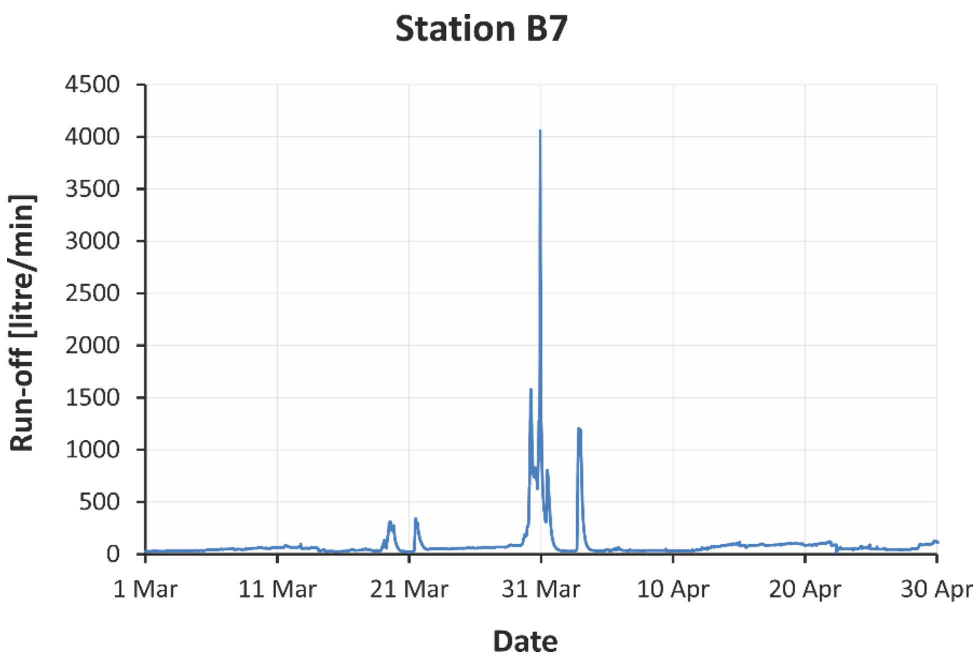
While a records manager (or data provider) is likely to be interested in all of the administrative details required to ensure correct record-keeping, a data user is generally more interested in the results arising from an observation (see Chapter 4). Information relating to the maintenance and administration of records may be of little value in interpreting the scientific content of a record.

For this reason, RepMet has found it practical to categorise metadata in two types: metadata required by a *records manager / data provider* and metadata required by the *data user*.

An overview of metadata and its application within RWMOs is provided by the RepMet report *Metadata in Radioactive Waste Management.* Using a content-based approach, it identifies three metadata classification types - *administrative, structural* and *descriptive*. Within this scheme, the main focus of MRMS is *administrative* metadata. *Descriptive* metadata is also included in MRMS, but is represented at a very high level, using titles, keywords and a short description. Detailed descriptive metadata is typically provided by dedicated data models similar to those found in the RepMet Libraries and Observations and Measurements (O&M) standard.

Metadata in Radioactive Waste Management illustrates the importance of metadata in validating and interpreting a data set using an example from hydrogeology[28]. Water run-off data are presented in a graph which would be stored in a digital document which is itself described by an MRMS record (see Figure 5.1).

Figure 5.1: Illustrative example of a dataset without metadata



Source: PURAM, 2019.

The MRMS record holds the administrative information (e.g. where the document is located, the history of the document and who is responsible for it), though, this

28. See Figure 2.1 and Table 2.1 from Metadata in Radioactive Waste Management.

administrative information provides little help in interpreting the scientific content. Fortunately, scientific interpretation is possible within RepMet through the use of the O&M standard, which supports detailed descriptive metadata including expressing the time series, together with suitable context (for example unit definitions and observed property names).

This example shows how effective management and understanding of information can be supported by using a range of descriptive metadata elements, based on a range of standards designed to support future use and management.

In the remainder of this chapter a model will be developed which combines the MRMS standard with Observations and Measurements, thereby allowing scientific data to be encoded while satisfying the need to manage the record appropriately. To facilitate a more general use of the MRMS data model, which allows integration with the O&M standard, RepMet has grouped the MRMS metadata elements into two complementary sets: *Resource* and *Record*.

Resource is the core part of the MRMS record, composed of the metadata elements to support data usage, distribution and access. While *Record* contains the remainder of the MRMS metadata elements to support record-keeping activities. These are formally defined in Table 5.1.

Table 5.1: Entities of the MRMS CDM

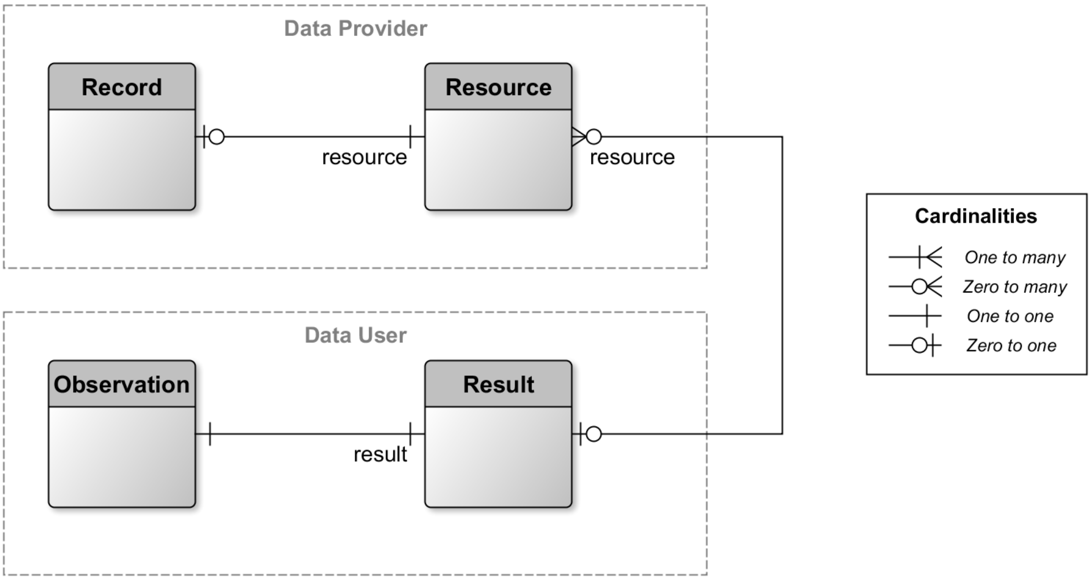
Entity	Definition
Resource	The set of metadata elements needed to find and access information stored on a tangible, electronic or other medium that is retrievable in a useable form.
Record	The set of administrative metadata elements and the reference to the Resource that is subject of record-keeping.

Source: NEA, 2019.

RepMet has integrated the Resource and Record MRMS groupings with the O&M standard via the O&M *Result* entity - see Figure 5.2.

From Figure 5.2 it can be seen that an *Observation* has a *Result*, and this result may be stored in and available from a number of physical sources described in the *Resource* entity. Each physical source has an associated *Resource* entity, each of which is associated with a *Record* entity.

Figure 5.2: Connection between Records and Observations



Source: NEA, 2019.

In the O&M standard, the *Result* of an *Observation* is often provided as a data *Resource* (a digital file or a hard copy document). This means that the concept of *Resource* within the O&M standard is the same as the concept of *Resource* within the MRMS.

If the MRMS *Resource* is subject to organisational record-keeping requirements, then it must have a supporting *Record*. However, if there are no formal record-keeping requirements in place, then this is optional and can be decided by the organisation holding the data.

Similarly, if the *Result* of an *Observation* is not subject to organisational record-keeping requirements, the *Record* can be ignored, and only the *Resource*, with details of how to access to the *Result* is required. The *Resource* can then be directly referenced from an *Observation*.

Figure 5.3 illustrates a global conceptual data model that RepMet has developed to merge the O&M including the Sampling Feature extension and the MRMS.

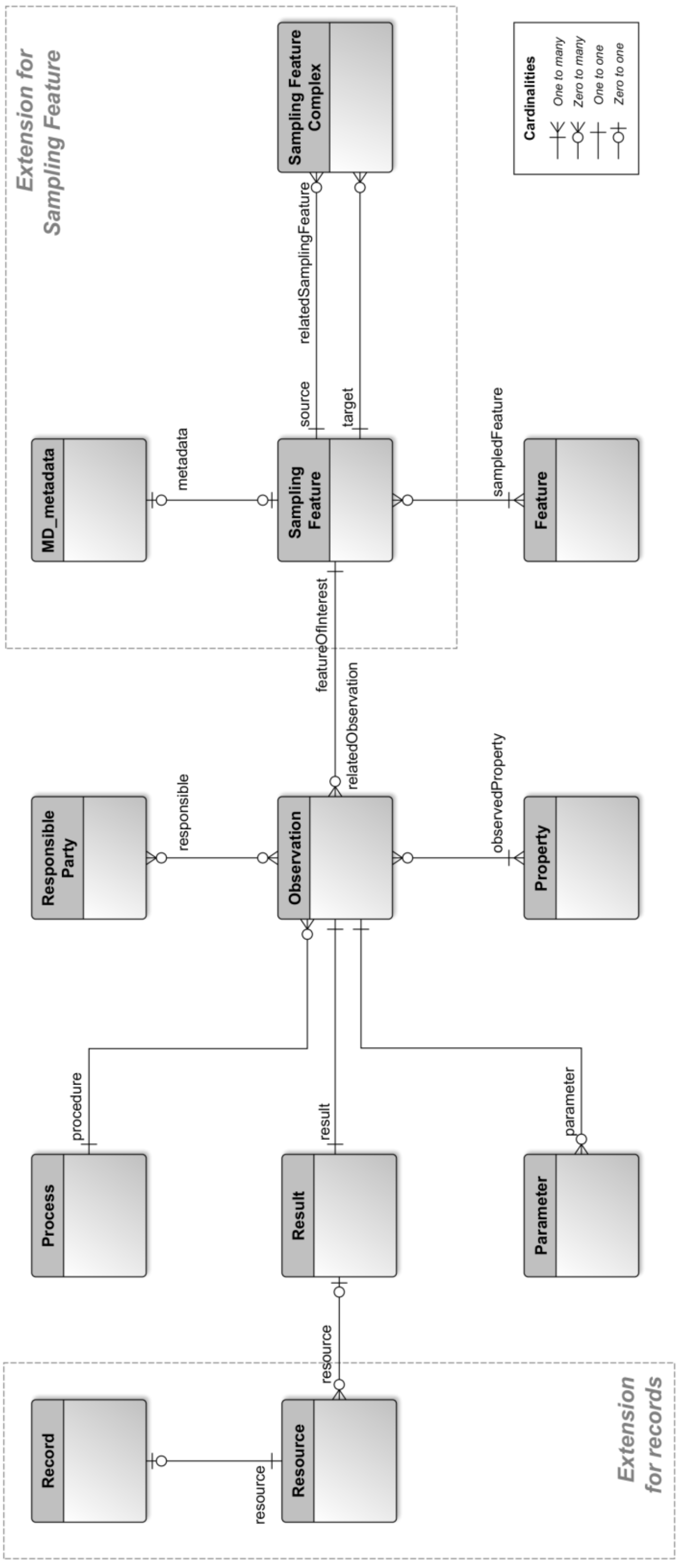
5.3.3. Cross-compliance between MRMS and ISO 19115

Most MRMS metadata elements are also found in the ISO 19115 metadata standard. This means that when using ISO elements it is relatively straight forward to set up a metadata profile for Resource and Record entities based on the ISO schema. However, as record-keeping is an organisational issue, and on the whole, details of how a resource are managed, preserved and used are not usually shared with the wider community, RepMet decided there was currently no requirement to develop an ISO compliant profile for the *Resource* and *Record* entities.

Given the efficiency and simplicity of using the MRMS concept, RepMet recommends that it should be used by RWMOs in their local implementations; although RWMOs may also need to take national standards and regulations into consideration.

Details of the cross-compliance are given in the following sections that consider the *Resource* and *Record* entities in more detail.

Figure 5.3 CDM for O&M (including Sampling Feature extension) connected with the MRMS – Entity Relationship Diagram



Source: NEA, 2019.

The lists of metadata elements for the Resource and Record entities do not correspond exactly with version 1.3 of MRMS. The aim is to provide an example of the essential attributes for each of the entities based on the MRMS, as summarised in Figure 5.2. This has been done to illustrate how the MRMS can be integrated with the O&M standard. However, it would be possible to develop a more detailed implementation for RepMet if required.

5.3.4. The Resource entity

A RepMet *Resource* entity provides a gateway to the data and it uses three of the MRMS metadata elements, two of which are mandatory. The MRMS also defines metadata sub-elements, which equate to sub-attributes within the RepMet *Resource* entity. These are described in Table 5.2, and the cross-compliance between MRMS and ISO 19115 for the *Resource* entity is described in Table 5.3.

Note: When describing both the *Resource* and *Record* entities, the numbers of occurrences of an attribute or sub-attribute is indicated in square brackets as follows: [1..1] is a single, mandatory occurrence, [1..N] means one or more mandatory occurrences, [0..1] means a single, optional occurrence and [0..N] is zero or more (optional) occurrences.

Table 5.2: Attributes of the Resource entity

Entity	**RESOURCE**		
Attribute / Sub-attribute		**Cardinality**	**Description**
Identifier		[1..1]	A unique identifier for a Resource record.
Title		[1..1]	The representative title of the record.
	Official title	[1..1]	
	Alternative title	[0..1]	
Format		[0..1]	-
	Content medium	[1..1]	The logical form of the resource – the form that the resource is presented in and the format of the data.
	Data format	[1..1]	
	Storage medium	[1..1]	The physical form of the resource – the media type and the size.
	Extent	[0..1]	
Location		[1..1]	The exact location of the resource (hardcopies, analogue recordings, digital copies).
	Home location details	[1..1]	Organisation name, address of home location
	Home storage details	[1..N]	Detailed description to find the resource, such as filenames or URL to download resource or access authorised services.
	Record-keeping system	[0..1]	The identifier used by the record-keeping system.
	Current location	[1..1]	Organisation name, address of current location
Resource Type		[0..1]	Classification of the type of a resource, using a controlled dictionary.

Source: NEA, 2019.

Table 5.3: Cross-compliance between MRMS and ISO 19115 for the *Resource* entity

Attribute in MRMS Resource entity	Element in ISO 19115	Note
Title	CI_Citation/title	-
Format	MD_Distribution_Type	ISO 19115 provides the following useful code lists as part of the standard: CI_PresentationFormCode, MD_MediumNameCode, MD_MediumFormatCode
Location	MD_Distributor, MD_Distribution, CI_OnlineResource, MD_DigitalTransferOptions/offLine	-

Source: NEA, 2019.

5.3.5. The Record entity

There are 17 MRMS elements that relate to the Record entity, of which 8 are mandatory in the MRMS. These are described in Table 5.4 and the cross-compliance between MRMS and ISO 19115 for the Record entity is described in Table 5.5.

Table 5.4: Attributes of the Record entity.

Entity	RECORD		
Attribute / Sub-attribute		**Cardinality**	**Description**
Agent		[1..N]	An attribute documenting the role and identity of organisations and individuals that perform some action on a record or use a record in some way.
	Name	[1..1]	
	Address	[1..1]	
	Email	[0..1]	
	Etc…	-	*Not all potential sub-attributes are listed here.*
Rights management		[1..N]	Legislation, policies, and caveats governing or restricting access to or use of records.
	Access conditions	[1..1]	
	Use conditions	[1..1]	
	Encryption details	[0..1]	
	Etc…	-	*Not all potential sub-attributes are listed here.*
Subject		[1..N]	The subject matter or topic of a record.
Description		[0..N]	An account, in free text, of the content and/or purpose of the record.

Language		[0..N]	The language of the content of the record, according to the ISO 639-2 standard[29].
Relation		[0..N]	Attribute for documenting associations between individual Records or other information sources.
	Identifier	[1..1]	Of the referenced record.
	Nature	[1..1]	Of the relationship.
	Description	[0..1]	Of the relationship.
Coverage		[0..1]	Spatial and temporal extent or jurisdictional scope of the record.
	Coverage type	[1..1]	-
	Coverage name	[0..1]	-
Function		[0..N]	The general or agency-specific business function(s) and activities documented in the record. The use of a controlled vocabulary is recommended.
Date		[1..1]	The dates and times at which fundamental record-keeping actions occur.
	Creation Date	[1..1]	The date and time at which a record or a record series is created.
	Other date & Description	[0..N]	The date and time and description of other significant record-keeping events.
Record Type		[0..1]	The recognised form or genre a record takes, which governs its internal structure. The MRMS provides a specific dictionary for this attribute. Record/Record Type may have the same value, or more specific about the content using narrower terms of the same concept. (e.g. Report – processingReport).
Aggregation level		[1..1]	The level at which the record(s) is/are being described and controlled. The value can be 'Record' or 'Record Series', in the latter case this describes a collection of records.
Record identifier		[1..N]	A unique identifier code for the specific instance of the record.
Management history		[1..N]	The dates and descriptions of all records management events from the registration of a record into a record-keeping system until its disposal.
	Event date	[1..1]	Date and time of the event.
	Event type	[1..1]	MRMS provides a code list with elements including *access reviewed*, *checked out* or *unauthorised access attempt*.

29. ISO 639 is the International Standard for language codes. Its purpose is to establish codes to represent languages or language families. ISO 639-2 provides a code list based on 3 letter characters, for example, eng – English, fre/fra – French.

	Event description	[1..1]	A description of the event that is being recorded.
Use history		[0..1]	The dates and descriptions of both legal and illegal attempts to access and use a record, from the time of its registration into a record-keeping system until its disposal.
	Use date	[1..1]	Data and time of the access/use event.
	Use type	[1..1]	MRMS provides a code list with the following elements: *Accessed*, *Checked out*, *Unauthorised access attempted*.
	Description	[0..1]	-
Preservation history		[0..N]	The dates and descriptions of all actions performed on a record after its registration into a record-keeping system which ensure that the record remains readable (renderable) and accessible for as long as it has value to the organisation and to the community at large.
	Action date	[1..1]	Data and time of the action performed on a record.
	Action type	[1..1]	MRMS provides a code list with elements including: *Backed up*, *Condition checked*, *Compressed*, *Converted*.
	Description	[1..1]	A description of the action that is being recorded.
	Next action	[0..1]	-
	Next action due date	[0..1]	-
Disposal		[1..1]	Information about policies and conditions which pertain to or control the authorised disposal of records. Information about the current retention schedule and disposal actions to which the record is subject.
	Retention schedule	[1..1]	-
	Retention period	[1..1]	-
	Disposal actions	[1..1]	MRMS provides a code list to define what to do after the retention period has expired. This code list includes: *Permanent*, *Destroy*, *Transfer to archives*, *To be determined*.
	Disposal due date	[1..1]	-
Mandate		[0..N]	The source of record-keeping requirements. For example, a piece of legislation, formal directive, policy, standard, guideline, set of procedures, or community expectation which (explicitly or implicitly) imposes a requirement to create, keep, dispose of, or control access to and use of a record.
	Mandate type	[1..1]	The MRMS code list for the *Mandate type* includes *Legislation*, *Regulation* and *Formal Directive*.
	Refers to	[1..1]	Values include: *Creation*, *Retention*, *Access* and *Usage.*
	Mandate name	[1..1]	-
	Mandate reference	[0..1]	-

	Requirement	[0..1]	-
Resource		[1..1]	A unique identifier code referencing the specific instance of the *Resource* entity.

Source: NEA, 2019.

Table 5.5: Cross-compliance between MRMS and ISO 19115 for the Record entity

Attribute in MRMS Record entity	Element in ISO 19115	Note
Agent	CI_ResponsibleParty (overlapping)	Sub element 1.1 Agent type is coincident with the codelist CI_Role Code from the ISO standard. Dictionary harmonisation is required.
Rights management	MD_Constraints (overlapping)	Coincident codelists: MGDPA classification and MD_ClassificationCode
Subject	MD_Keywords	-
Description	MD_DataIdentification/abstract	-
Language	MD_DataIdentification/language	-
Relation	MD_AggregateInformation	-
Coverage	EX_SpatialTemporalExtent (applicable)	Jurisdiction is not covered by ISO
Function	MD_DataIdentification/purpose (similar)	Free text element
Date	CI_Date	-
Record Type	-	-
Aggregation level	MD_Metadata/hierarchyLevel	Available codelist is MD_ScopeCode, provided as part of the ISO standard.
Record identifier	RS_Identifier	Codespace and code together for globally unique identification
Management history	LI_Lineage/processStep	It could be used for the same purpose with date, rationale, description. Rationale is free text attribute.
Use history		-
Preservation history	LI_Lineage/processStep	It could be used for the same purpose with date, rationale, description. Rationale is free text attribute.
Disposal	-	-
Mandate	-	-
Resource	CI_OnlineResource (parts of it)	-

Source: NEA, 2019.

5.4. Example application of the MRMS framework

The following examples demonstrate the use of the MRMS framework for the present "RepMet Tools and Guidelines" report, where Table 5.6 is the MRMS Record entity and Table 5.7 the MRMS Resource entity. These examples are illustrative.

Table 5.6: MRMS *Record* entity applied to the "RepMet Tools and Guidelines" report

Entity	Attribute	Sub-attribute	Value
Record	**Agent**	Name	OECD Nuclear Energy Agency
		Address	46 Quai Alphonse le Gallo, 92100 Boulogne-Billancourt, France
	Rights Management	Access conditions	Public
		Use conditions	NA
	Subject	First subject term	Radioactive waste disposal
		Enhanced subject term	Metadata
	Creation date	Date created	2017.06.31
	Record Type	-	Report
	Aggregation level	-	Record
	Record identifier	-	REP_001
	Management history	Event Date	2017.12.15
		Event type	Publication
		Event description	Published on OECD NEA RepMet website
	Disposal	Retention schedule	Not Scheduled
		Retention period	Permanent
		Disposal action	To Be Determined
	Resource	Identifier	RTG-001

Source: NEA, 2019.

Table 5.7: MRMS *Resource* entity applied to the “RepMet Tools and Guidelines” report

Entity	Attribute	Sub-attribute	Value
Resource	**Identifier**	-	RTG-001
	Title	Official title	RepMet Tools and Guidelines
	Format	Content medium	Compound
		Data format	Office Open XML document (docx)
		Storage medium	Hard disk
	Location	Current location	HOME
		Home location details	OECD NEA, Paris
		Home storage details	host=10.01.01.123
		Home storage details	filename=//repmet/2017/reports/RTG-001.docx
		Home storage details	url=http://oecd.org/nea/repmet/RTG-001.html

Source: NEA, 2019.

The attributes of Table 5.2 and Table 5.4 are the same as those found in Table 5.6 and Table 5.7, however, the two examples only include the mandatory MRMS elements. These mandatory elements still provide basic information about the “RepMet Tools and Guidelines” report such as author, classification, creation and publication dates, title, format and access.

6. INSPIRE

6.1. What is INSPIRE?

The Infrastructure for Spatial Information in Europe (INSPIRE) is a European Union (EU) Directive (formally, Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community) with the aim to create a European spatial data infrastructure for the purposes of policies and activities which may have an impact on the environment. This European spatial data infrastructure will enable the sharing of environmental spatial information among public sector organisations, facilitate public access to spatial information across Europe and assist in policymaking across boundaries (About INSPIRE, 2018).

It is important to recognise that INSPIRE is not a *standard* or a set of standards, but a set of principles and common implementing rules that have been translated into legislation in the member states of the EU. However, INSPIRE has dedicated data models to help data providers in sharing the information under their custody in a uniform way.

RepMet's adoption of a number of INSPIRE principles, standards and models has a twofold purpose: the first being the logical step of not *reinventing the wheel* when a suitable existing approach is available; and the second is its obvious usefulness to RWMOs of being able to exchange their data. These advantages make the adoption of certain elements of INSPIRE desirable, even by RWMOs which are not required to comply with EU Directives.

It is beyond the scope of RepMet and this section to give a thorough introduction to INSPIRE, but some of its most relevant models will be briefly introduced below.

6.2. Why is INSPIRE useful to RepMet?

Data providers in the environmental domain in the EU are obliged to provide information to the public in accordance with the INSPIRE Directive with the goal of sharing content in a harmonised structure, making spatial information more transparent and interoperable. The consolidated data model framework provided by INSPIRE contains a large number of well-designed elements that can be adopted and reused.

Datasets within the scope of INSPIRE are classified into 34 spatial data themes[30] as set out in the INSPIRE Directive. Entities defined in the RepMet Libraries overlap with INSPIRE entities. It was found that sharing the common attributes with similar fundamental INSPIRE entities and adding missing ones that are required for radioactive waste management not only saves effort but also contributes to a harmonised structure. In fact, the content of RepMet Libraries can be published, shared and converted into INSPIRE

30. Examples include Geology, Elevation, Buildings, Habitats and biotopes, Energy resources, Land use, Environmental monitoring facilities and similar.

conformant data. This not only brings the benefit of conformity with the INSPIRE Directive but permits standardised data sharing.

The Site Characterisation Library has been designed taking account of the models of INSPIRE's Geology (GE) and Environmental monitoring facilities (EF) themes. For future data modelling activities, it would be desirable to continue to take INSPIRE into consideration in the creation of other libraries as well.

Analysing the existing RepMet Libraries, the following three INSPIRE themes (of the 34) were recognised as being potentially useful for RepMet activities[31]:

- Production and industrial facilities (PF)
- Utility and governmental services (US)
- Environmental monitoring facilities (EF)

Relationships between the above themes and RepMet are briefly explained below.

6.3. INSPIRE themes relevant to Radioactive Waste Management

INSPIRE provides a dedicated theme for *Production and Industrial Facilities* with a relatively complex data model. Waste repositories can be regarded as an example of a special industrial facility that can be described in detail by using the production and industrial facilities generic model. The Annex III *Production and Industrial Facilities* theme contains the following entities:

- **Production Site.** *Definition:* All land at a distinct geographic location where the facility was, is, or is intended to be located. This includes all infrastructure, equipment and materials.
- The entity attributes include "identifiers", "geometry (2D)", "name", "site plan", "description" and "status".
- **Production Facility**. *Definition:* One or more installations on the same site operated by the same natural or legal person, designed, built or installed to serve specific production or industrial purposes, comprehending all infrastructure, equipment and materials.

  The entity attributes include "identifiers", "geometry (2D)", "status", "activity", "input", "output", "description". To define the nature of "activity", "input" and "output" several INSPIRE code lists are available:

  - EU Waste Statistics Economic Activity Classification
  - EU Waste Recovery Disposal Classification
  - EU Waste Classification

At the time of publication of this report these dedicated code lists are empty, and reference web pages of the Official Journal of the European Union. Values to be used are usually defined in the Annexes of the legal texts.

Production Facilities may contain "production plots", "production buildings" and "production installations", each of which has its own definition and attributes.

31. Official INSPIRE theme abbreviations are shown in parentheses.

The INSPIRE theme for Utility and governmental services contains the entity:

- **Environmental Management Facility.** *Definition:* A physical structure designed, built or installed to serve specific functions in relation to environmental material flows, such as waste or waste water flows, or a delimited area of land or water used to serve such functions.

  As for the case of "Production Facility" entity, the attributes include "identifiers", "geometry", "status", "activity", "input", "output", "description". In addition, the "Environmental Management Facility" entity has the following attributes: "type (site or installation)", "service hours", "facility description", "physical capacity", "licences".

The theme for Environmental monitoring facilities contains the entity:

- **Environmental Monitoring Facility.** *Definition:* A georeferenced object directly collecting or processing data about objects whose properties (e.g. physical, chemical, biological or other aspects of environmental conditions) are repeatedly observed or measured. An environmental monitoring facility can also host other environmental monitoring facilities.

  This entity is explained in more detail in the Site Characterisation Library.

6.4. Positioning RepMet in INSPIRE

Several entities in the RepMet Repository Library having strong connections with existing INSPIRE features were identified: disposal and emplacement elements, boreholes, engineered barrier, surface and underground facilities. These entities can be categorised into three groups corresponding to three INSPIRE entities: "Production Installation", "Production Facility", and "Environmental Monitoring Facility". Table 6.1 shows the relationships between the entities that could be used in the geological disposal and the recommended INSPIRE spatial data types.

Some elements may belong to more than one INSPIRE category. e.g. Disposal Drift and Disposal Room fit both into Building and Production Building. An Emplacement Borehole is a Production Installation, but can also be described as Borehole in the INSPIRE geology theme.

The INSPIRE entities have generic attributes to characterise facilities and installations, buildings and so forth of any kind. Entities in the RepMet Libraries have specific attributes for the waste disposal case. It would be desirable to have both sets of attributes available in a unified structure; in data modelling subclassing could be used for that purpose.

Using the proposed model elements, a waste disposal site can be fully described utilising INSPIRE methods, allowing data providers to make important public information available through and for the INSPIRE community, and also provide specific RepMet attributes for the radioactive waste domain.

Conversely, RepMet can also provide useful information to the INSPIRE community by filling presently empty code lists in INSPIRE with relevant concepts. One example is the list of repository monitoring systems that would complement the Specialised Environmental Monitoring Facility type.

INSPIRE's Environmental Management Facility is the entity corresponding to waste repositories within the "INSPIRE Annex III Utility and Governmental Services" theme. In the future, further extending the RepMet facility entity with the attributes of Environmental Management Facility would be desirable.

In conclusion, the extension of the existing INSPIRE controlled dictionaries, following the generic INSPIRE guidelines and the above recommendations, is anticipated to be sufficient for establishing INSPIRE services for spatial information related to radioactive waste repositories.

In RepMet it was not possible to fully develop the synergies between RepMet and INSPIRE. In the future, it is recommended that the existing RepMet Libraries be extended taking these recommendations into consideration.

Table 6.1: Examples of entities for geological disposal and comparable INSPIRE Spatial Data Types

Geological disposal	**INSPIRE**	
Emplacement	Theme	Spatial Data Type
DisposalDrift	AnnexIII/PF	ProductionBuilding
DisposalRoom	AnnexIII/PF	ProductionBuilding
DiposalDrift	AnnexIII/BU	Building
DisposalRoom	AnnexIII/BU	Building
EmplacementBorehole	AnnexIII/PF	ProductionInstallation
EmplacementBorehole	AnnexII/GE	Borehole
EmplacementVault	AnnexIII/PF	ProductionInstallation
EngineeredBarrier	AnnexIII/PF	ProductionInstallation
BarrierInstallation	AnnexIII/PF	ProductionInstallation
BarrierInstallation	AnnexIII/PF	ProductionInstallationPart
Facility	Theme	Spatial Data Type
Site	AnnexIII/PF	ProductionSite
NearSurfaceFacility	AnnexIII/PF	ProductionFacility
DeepRepositoryFacility	AnnexIII/PF	ProductionFacility
Monitoring	Theme	Spatial Data Type
RepositoryMonitoringSystem	AnnexIII/EF	EnvironmentalMonitoringFacility

Source: NEA, 2019.

7. Concluding remarks

The “RepMet Tools and Guidelines” report is a technical product of the Integration Group for the Safety Case (IGSC) Radioactive Waste Repository Metadata Management (RepMet) initiative. The purpose of this report is to provide the minimal technical background to enable the use and the understanding of the three “RepMet Libraries”. It aims to provide, within a single document, an adequate and sufficient amount of information for each technique and metadata-based tool adopted for the RepMet Libraries, allowing them to be consistent and self-explanatory.

The three deliverables identified as *“libraries”* are more technically detailed. They discuss the key aspects of data and related metadata for selected scientific and technical topics involved in the life cycle of a radioactive waste repository. The libraries include high-level conceptual data models, descriptions of data entities, attributes, associated metadata and other relevant information, and are ready to support the activities of RWMOs. The libraries can be used independently of each other; however, utilising all of the libraries and the approach outlined in these documents helps provide the additional benefit of a uniform approach to metadata management.

The documents are primarily designed for use by personnel in RWMOs, regardless of whether they have a strong background or not in such areas as database management, database development, data modelling or any other area of information and/or computing systems. The documents provide high-level overviews and summaries suitable for RWMO Managers and Decision Makers, and include more detailed, implementation specific information targeted at Information Systems Developers working within a RWMO environment.

This report aims to complement the existing, more specialised literature by presenting the information management techniques and tools within the context of radioactive waste disposal.

Although the RepMet initiative has now finished, there is further work that can be done. This includes the improvement of the controlled dictionaries included in the Site Characterisation Library. The controlled dictionaries could then become an international resource curated by the NEA. Other activities include:

- Further development of the scientific and technical content of the controlled dictionaries (e.g. more details for “definition” and “purpose” features for each attribute).
- Definition of a strong connection between the attributes of the controlled dictionaries and the NEA International Features, Events and Processes (IFEP) List included in the NEA FEP Database. This is because each item of the NEA IFEP List reports and explains their eventual relevance for safety assessment.
- Elaboration of controlled dictionaries for attributes of entities in the O&M and MRMS standards.

References

270a.info (2014), 270a Linked Dataspaces, retrieved from https://270a.info/ (accessed June 2019).

EC (n.b.), INSPIRE code list register, European Commission, retrieved from https://inspire.ec.europa.eu/codelist (accessed June 2019).

EC (2018), "About INSPIRE", European Commission, retrieved from https://inspire.ec.europa.eu/about-inspire/563 (accessed June 2019).

Booch, G., J. Rumbaugh and I. Jacobson (2005), The Unified Modelling Language User Guide (Second Edition), Addison-Wesley.

Chen, P. P.-S. (1976), "The entity-relationship model: toward a unified view of data," ACM Transactions on Database Systems, vol. 1, no. 1, pp. 9-39.

DCMI Usage Board, "Dublin Core Metadata Initiative", retrieved from http://dublincore.org (accessed June 2019).

Geological Survey of Austria (2021) GBA Thesaurus, https://thesaurus.geolba.ac.at/ (accessed June 2019).

Open Geospatial Consortium Inc. (n.d.), "Geographic information – Metadata", retrieved from www.iso.org/obp/ui/#iso:std:iso:19115:-1:ed-1:v1:en (accessed June 2019).

Open Geospatial Consortium (2007), "Observations and Measurements – Part 2 - Sampling Features", retrieved from www.opengeospatial.org/standards/om#downloads (accessed June 2019).

Open Geospatial Consortium Inc. (2011), "OGC SWE Common Data Model Encoding Standard", retrieved from www.opengeospatial.org/standards/swecommon (accessed June 2019).

Open Geospatial Consortium (2013), "Geographic Information: Observations and Measurements, OGC Abstract Specification Topic 20, OGC 10-004r3 and ISO 19156", www.opengeospatial.org/standards/om#downloads (accessed June 2019).

National Oceanography Centre (2020) Resources: NERC Vocabulary Server, www.bodc.ac.uk/resources/products/web_services/vocab/ (accessed June 2019).

NEA (2021a), "Site Characterisation Library", OECD Publishing, Paris.

NEA (2021b), "Waste Package Library", OECD Publishing, Paris.

NEA (2021c), "Repository Library", OECD Publishing, Paris.

NEA (2019), *Metadata for Radioactive Waste Management*, OECD Publishing, Paris.

NEA (2014), "Glossary of Terms. Preservation of Records, Knowledge and Memory (RK&M) across Generations," Paris (France).

European Union (2020) EuroVoc Thesaurus, Publications Office of the European Union https://op.europa.eu/en/web/eu-vocabularies/concept-scheme/-/resource?uri=http://eurovoc.europa.eu/100141 (accessed June2019).

RMDC (2015), “Minnesota Recordkeeping Metadata Standard”, retrieved from www.mnhs.org/preserve/records/metamrms.php (accessed June 2019).

SKOS (n.b.) SKOS UNESCO Thesaurus, skos.um.es/unescothes/C00324/html (accessed June 2019).

W3C (World Wide Web Consortium) (2014), “RDF 1.1 Concepts and Abstract Syntax” retrieved from www.w3.org/TR/rdf11-concepts/ (accessed June 2019).

W3C (World-Wide Web Consortium) (2012), “Introduction to SKOS (Simple Knowledge Organisation System)”, retrieved from www.w3.org/2004/02/skos/intro (accessed June 2019).

W3C (World Wide Web Consortium) (2009a), “SKOS (Simple Knowledge Organisation System) Reference”, retrieved from www.w3.org/TR/skos-reference/ (accessed June 2019).

W3C (Wold Wide Web Consortium) (2009b), “SKOS (Simple Knowledge Organization System) Primer”, retrieved from www.w3.org/TR/skos-primer (accessed June 2019).

Annex A. Further details and examples of controlled dictionaries

A1 Example of use of RDFS

In order to provide an application example of RDFS, we assume that a bookshop has planned to create a RDF database about the books available in the store. For each book, the database manager in the bookshop is interested in:

- Title and author; and
- Name and family name of the author.

The first step is the creation of the RDF vocabulary for the information about the book-resource that the bookshop wants to capture in the RDF triples. Table A.1 illustrates how RDFS classes and properties can be used to create a RDF vocabulary customised for the information that the bookshop is interested in storing in the RDF triples.

Table A.1: Example of RDF vocabulary through RDFS[32,33]

Subject	Predicate	Object	Comment
bs:Book	rdf:type	rdfs:Class	It declares bs:Book as a class.
bs:Person	rdf:type	rdfs:Class	It declares bs:Person as a class.
bs:familyName	rdf:type	rdf:Property	It declares bs:familyName as a property.
bs:name	rdf:type	rdf:Property	It declares bs:name as a property.
bs:familyName	rdf:domain	rdf:Person	It declares bs:familyName as a property of the class bs:Person.
bs:name	rdf:domain	rdf:Person	It declares bs:name as a property of the class bs:Person.
bs:title	rdf:type	rdf:Property	It declares bs:title as a property.
bs:title	rdf:domain	bs:Book	It declares bs:title as a property of the class bs:Book.
bs:hasAuthor	rdf:type	rdf:Property	It declares bs:hasAuthor as a property.
bs:hasAuthor	rdf:domain	bs:Book	It declares bs:hasAuthor as a property of the class bs:Book.
bs:hasAuthor	rdf:range	bs:Person	It declares that bs:hasAuthor points to an instance of the class bs:Person (including the properties of its domain).

Source: NEA, 2019.

By using the RDF vocabulary in Table A.1, it is possible to create RDF triples about particular books containing the information that the database managers of the bookshop are interested in collecting. Table A.2 shows an application example for a particular book. This book is considered as a RDF resource identified with bs:Thermodynamics.

32. We assumed “bs:” as the namespace of the bookshop.

33. The rdfs:domain and rdfs:range assign the quality of property to the RDF subjects and the quality of class to the RDF objects. Thus, it would be possible to avoid the declarations of the RDF subjects and objects as, respectively, classes and properties, but we decided to keep for illustrative purpose.

Table A.2: Example for the RDF database built with the RDF vocabulary in Table A.1

Subject	Predicate	Object	Comment
bs:Thermodynamics	rdf:type	bs:Book	The resource bs:Thermodynamics is declared as an instance of the class bs:Book. Thus, it must have the property bs:title and bs:hasAuthor.
bs:Thermodynamics	bs:title	Thermodynamics	The bs:title property for bs:Thermodynamics has value equal to "Thermodyanmics".
bs:EnricoFermi[34]	rdf:type	bs:Person	The resource bs:EnricoFermi is declared as an instance of the class bs:Person. Thus, it must have the property bs:familyName and bs:name.
bs:EnricoFermi	bs:familyName	Fermi	The bs:familyname for bs:EnricoFermi has value equal to "Fermi".
bs:EnricoFermi	bs:name	Enrico	The bs:familyname for bs:EnricoFermi has value equal to "Enrico".
bs:Thermodynamics	bs:hasAuthor	bs:EnricoFermi	The bs:has Author property for bs:Thermodynamics is equal to bs:EnricoFermi that is an istance of the class bs:Person.

Source: NEA, 2019.

A2 Principles of SKOS

It is out of the scope of this document to provide an exhaustive explanation of the SKOS standard. More details are available on the SKOS Reference (W3C, 2009a) and SKOS Primer (W3C, 2009b) webpages. The following paragraphs provide a brief explanation of the main SKOS elements that the RepMet group adopted for its work.

The W3C has created permanent webpages containing the description of each property. They are available at the namespace "skos:"[35]. For example, the "prefLabel" property is available online at "skos:prefLabel"[36]. In this sense, the SKOS properties are RDF resources for predicates in RDF triples.

Table A.3: SKOS Core Vocabulary

Concepts	Concept, ConceptScheme, inScheme, hasTopConcept, topConceptOf
Lexical Labels	prefLabel, altLabel, hiddenLabel
Documentation properties	note, changeNote, definition, editorialNote, example, historyNote, scopeNote
Semantic relations	semanticRelation, broader, narrower, related, broaderTransitive, narrowerTransitive

Source: NEA, 2019.

34. Databases working according to the same principles are interoperable. In the provided example, we assumed that bs:EnricoFermi is a RDF resource available on the bookshop website. However, it could be possible to use RDF resource available in other databases such as DBpedia. In this case, the resource db:EnricoFermi contain much information about Enrico Fermi and it is the result of a work that someone has already done. This way of thinking avoids to reinvent the wheel each time and to duplicate the efforts in vain.

35. www.w3.org/2004/02/skos/core#

36. www.w3.org/2004/02/skos/core#prefLabel

Table A.4 illustrates the items of the SKOS Core Vocabulary to describe at high level the SKOS concepts and the SKOS concept schemes.

Table A.4: Items for concepts is SKOS Core Vocabulary

Lexical labels	URI	Definition
Concept	skos:Concept	Class of the SKOS concepts.
ConceptScheme	skos:ConceptScheme	Class of the SKOS concept schemes.
inScheme	skos:inScheme	Property for a concept to state that it belongs to a certain concept scheme.
hasTopConcept	skos:hasTopConcept	Property identifying a top concept of a concept scheme
topConceptOf	skos:topConceptOf	Property identifying the parent concept scheme of a top concept

Source: NEA, 2019.

Table A.5 illustrates the items of the SKOS Core Vocabulary to tag the SKOS concepts: the values of these kind of properties are strings of UNICODE characters in a given natural language.

Table A.5: Item for lexical labels in SKOS Core Vocabulary

Lexical labels	URI	Definition
prefLabel	skos:prefLabel	Property for a concept declaring its preferred lexical label in a given language.
altLabel	skos:altLabel	Property for a concept declaring an alternative lexical label in a given language.
hiddenLabel	skos:hiddenLabel	Property for a concept declaring a lexical label that should be hidden when generating visual displays of the concept, but should still be accessible to free text search operations.

Source: NEA, 2019.

Table A.6 illustrates the items of the SKOS Core Vocabulary to provide the information related to the SKOS concepts. There is no restriction on the nature of this information. It could be plain text, hypertext or an image. Moreover, it could be a definition, information about the scope of a concept, editorial information or any other type of information.

Table A.6: Definitions for SKOS documentation properties

Documentation properties	URI	Definitions
note	skos:note	Property for a concept defining a general note, for any purpose.
changeNote	skos:changeNote	Property for a concept defining a note about a modification to a concept.
definition	skos:definition	Property for a concept stating a formal explanation of its meaning.
editorialNote	skos:editorialNote	Property for a concept defining a note for an editor, translator or maintainer of the controlled dictionary.
example	skos:example	Property for a concept providing an example of its use.
historyNote	skos:historyNote	Property for a concept defining a note about its past state/use/meaning.
scopeNote	skos:scopeNote	Property for a concept defining a note that helps to clarify its meaning.

Source: NEA, 2019.

Table A.7 illustrates the items of the SKOS Core Vocabulary to create semantic relations (i.e. links) between SKOS concepts, where the link is inherent in the meaning of the linked concepts.

Table A.7: Definitions for SKOS semantic relations

Semantic relations	URI	Definitions
semanticRelation	skos:semanticRelation	Property of a concept related to another concept by meaning.
broader	skos:broader	Property of a concept that is more general in meaning[37] than another concept.
narrower	skos:narrower	Property of a concept that is more specific in meaning[38] than another concept.
related	skos:related	Property of a concept that has an associative semantic relationship with another concept.
broaderTransitive	skos:boraderTransitive	As broader with transitive value.
narrowerTransitive	skos:narrowerTransitive	As narrower with transitive value.

Source: NEA, 2019.

A3 Application example of RDF/SKOS Dictionary

This section provides an example how to build a controlled dictionary with rich semantic contest, complex hierarchies and multilingual features in the format of a SKOS Concept Scheme. The domain selected for the controlled dictionary example is nuclear plants.

Figure A.8.1 illustrates the items that are parts of the controlled dictionary including how they are hierarchically structured and what are their multilingual features. The limited level of details of the provided example is due to its illustrative purposes. The legend on the right side of Figure A.8.1 suggests how the SKOS standard is applied to the proposed controlled dictionary.

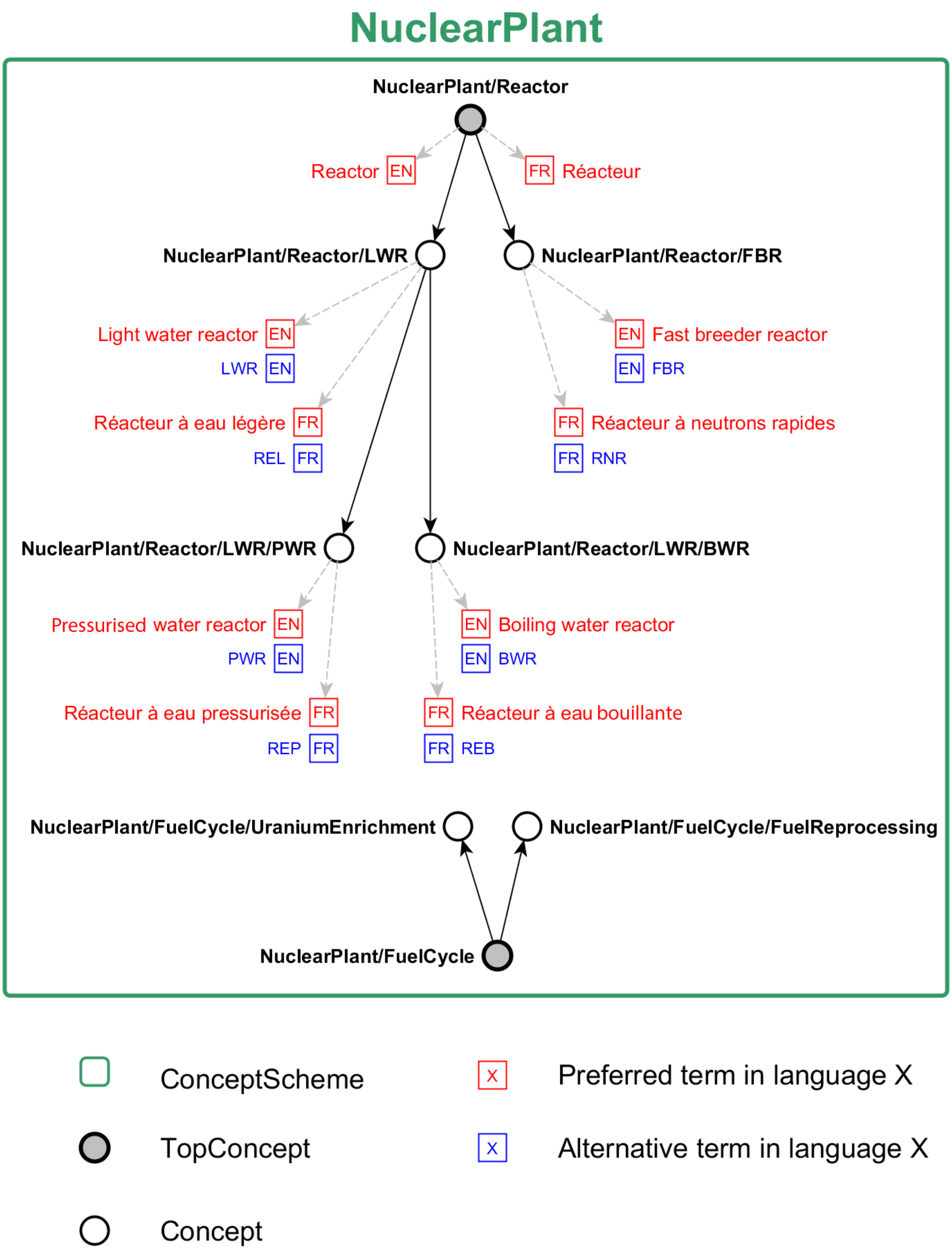
37. Broader concepts are typically rendered as parents in a concept hierarchy (tree).

38. Narrower concepts are typically rendered as children in a concept hierarchy (tree).

The controlled dictionary (i.e. the SKOS concept scheme according to a more appropriate use of the SKOS terminology and methodology) in Figure A.8.1 is “NuclearPlant”. It has two top concepts: “NuclearPlant/Reactor” and “NuclearPlant/FuelCycle”.

- The first top concept, “NuclearPlant/Reactor” has preferred terms in English (i.e. “Reactor”) and French (i.e. “Réacteur). Moreover, it has two narrower concepts: “NuclearPlant/Reactor/LWR” and “NuclearPlant/Reactor/FBR”.
    - The concept “NuclearPlant/Reactor/LWR” has preferred terms in English (i.e. “Light water reactor”) and in French (i.e. “Réacteur à eau légère”); it has also an alternative term for both languages (i.e. “LWR” for English and “REL” for French). Furthermore, it has two narrower concepts: “NuclearPlant/Reactor/LWR/PWR” and “NuclearPlant/Reactor/LWR/BWR”.
        - The concept “NuclearPlant/Reactor/LWR/PWR” has preferred terms in English (i.e. “Pressurised water reactor”) and in French (i.e. “Réacteur à eau pressurisée); it has also an alternative term for both languages (i.e. “PWR” for English and “REL” for French).
        - The concept “NuclearPlant/Reactor/LWR/BWR” has preferred terms in English (i.e. “Boiling water reactor”) and in French (i.e. “Réacteur à eau bouillante”); it has also an alternative term for both languages (i.e. “BWR” for English and “REB” for French).
    - The concept “NuclearPlant/Reactor/FBR” have preferred terms in English (i.e. “Fast Breeder Reactor”) and in French (i.e. “Réacteur à neutrons rapides”); it has also an alternative term for both languages (i.e. “FBR” for English and “RNR” for French).
- The second top concept, “NuclearPlant/FuelCycle” has two narrower concepts: “NuclearPlant/FuelCycle/UraniumEnrichment” and “NuclearPlant/FuelCycle/FuelReprocessing”.

Figure A.0.1: Structure controlled dictionary for "NuclearPlant" according to SKOS standard

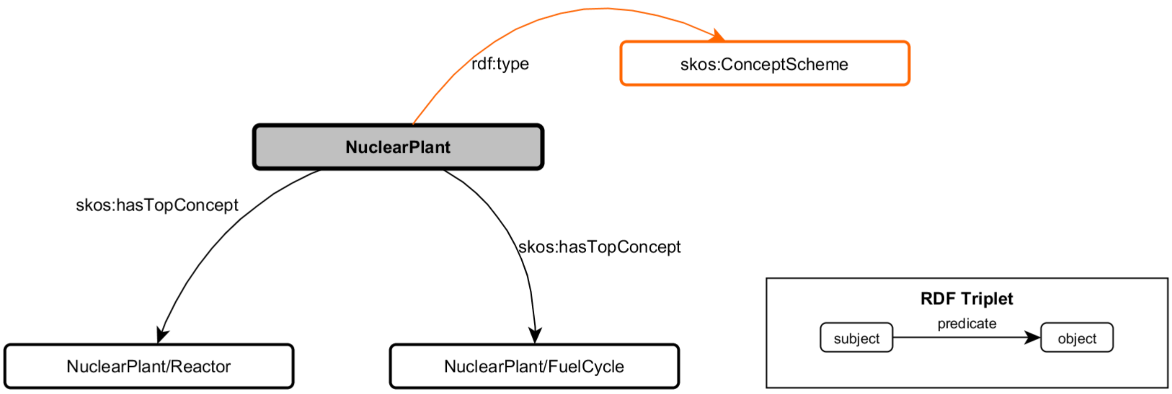


Source: NEA, 2019.

The following figures and tables help in understanding how the structured controlled dictionary in Figure A.8.1 can be translated into RDF triples according to the SKOS Core Vocabulary (i.e. the RDF dictionary that the SKOS standard provides). For the concept scheme and each concept in Figure A.8.1, it is possible to define a set of RDF triples using the SKOS classes and properties (that are reported from Table A.4 to Table A.7) in order to represent the semantic, hierarchical and multilingual features of the controlled dictionary.

Figure A.8.2 shows the RDF graph for the controlled dictionary item "NuclearPlant": this item is the subject of three different RDF triples shown in Table A.8.

Figure A.0.2: RDF graph for "NuclearPlant" concept



Source: NEA, 2019.

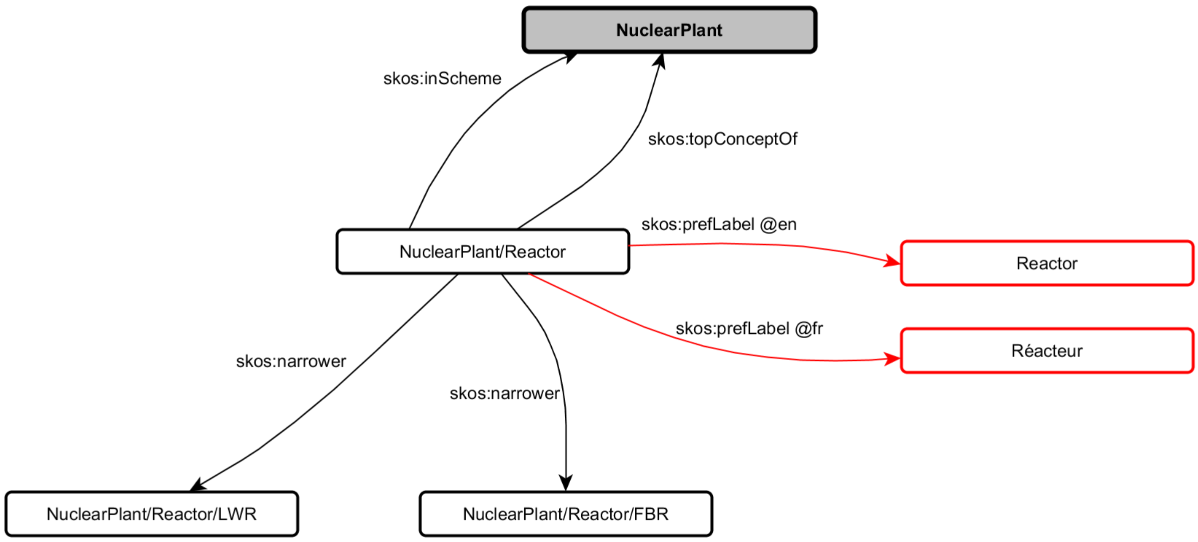
Table A.8: RDF triples for "NuclearPlant"

RDF Triples			RDF/SKOS Explanations
Subject	**Predicate**	**Object**	
NuclearPlant	rdf:type	skos:ConceptScheme	The RDF resource "NuclearPlant" is declared as an instance of the SKOS Concept Scheme class.
NuclearPlant	skos:hasTopConcept	NuclearPlant/Reactor	"NuclearPlant" has two top concepts: "NuclearPlant/Reactor" and "NuclearPlant/FuelCycle".
NuclearPlant	skos:hasTopConcept	NuclearPlant/FuelCyclePlant	

Source: NEA, 2019.

Figure A.8.3 shows the RDF graph for the controlled dictionary item "NuclearPlant/Reactor": this item is the subject of six different RDF triples shown in Table A.9.

Figure A.0.3: RDF graph for "NuclearPlant/Reactor" concept



Source: NEA, 2019.

Table A.9: RDF triples for "NuclearPlant/Reactor" concept

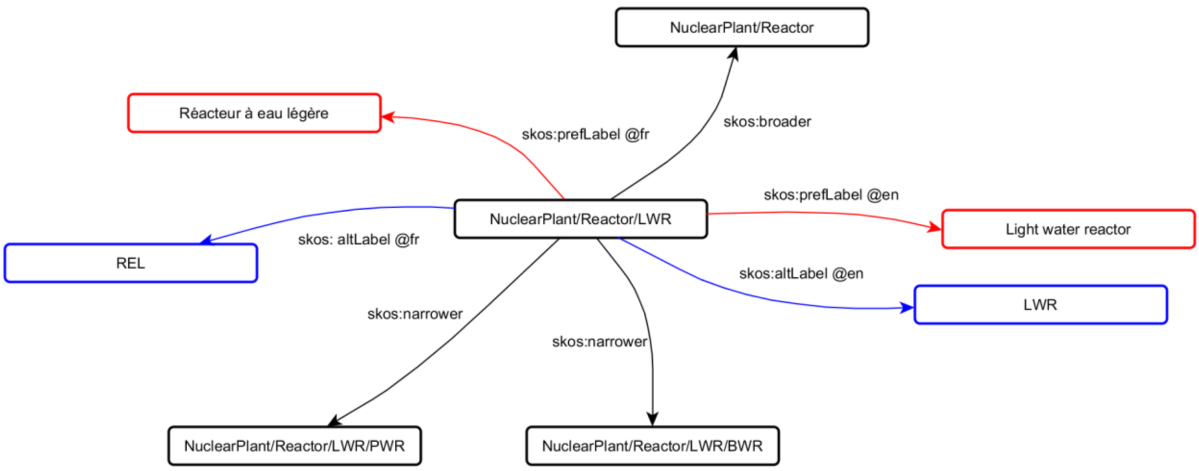
RDF Triples			RDF/SKOS Explanations
Subject	**Predicate**	**Object**	
NuclearPlant/Reactor	skos:inScheme	NuclearPlant	"NuclearPlant/Reactor" is a concept of the "NuclearPlant" concept scheme.[39]
NuclearPlant/Reactor	skos:topConceptOf	NuclearPlant	"NuclearPlant/Reactor" is a top concept of the "NuclearPlant" concept scheme.
NuclearPlant/Reactor	skos:prefLabel	"Reactor"@en	"NuclearPlant/Reactor" has preferred labels in English ("Reactor") and French ("Réacteur").
NuclearPlant/Reactor	skos:prefLabel	"Réacteur"@fr	
NuclearPlant/Reactor	skos:narrower	NuclearPlant/Reactor/LWR	"NuclearPlant/Reactor" has two narrower concept: "NuclearPlant/Reactor/LWR" and "NuclearPlant/Reactor/FBR"
NuclearPlant/Reactor	skos:narrower	NuclearPlant/Reactor/FBR	

Source: NEA, 2019.

39. In other words, "NuclearPlant/Reactor" is a term of the "NuclearPlant" dictionary.

Figure A.8.4 shows the RDF graph for the controlled dictionary item "NuclearPlant/Reactor/LWR": this item is the subject of eight different RDF triples shown in Table A.10.

Figure A.0.4: RDF graph for "NuclearPlant/Reactor/LWR" concept



Source: NEA, 2019.

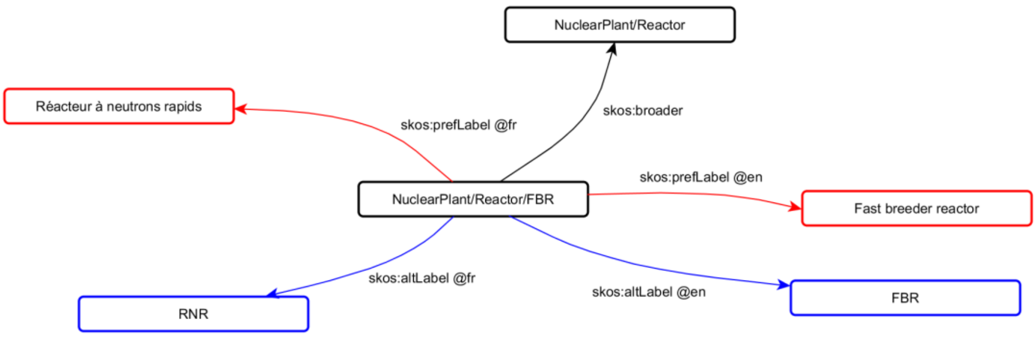
Table A.10: RDF triples for "NuclearPlant/Reactor/LWR" concept

RDF Triples			RDF/SKOS Explanations
Subject	Predicate	Object	
NuclearPlant/Reactor/LWR	skos:inScheme	NuclearPlant	"NuclearPlant/Reactor/LWR" is a concept of the "NuclearPlant" concept scheme.
NuclearPlant/Reactor/LWR	skos:prefLabel	"Light water reactor"@en	"NuclearPlantReactor/LWR" has two preferred labels: "Light water reactor" in English and "Réacteur à eau légère" in French.
NuclearPlant/Reactor/LWR	skos:prefLabel	"Réacteur à eau légère"@fr	
NuclearPlant/Reactor/LWR	skos:altLabel	"LWR" @en	"NuclearPlantReactor/LWR" has two alternative labels: "LWR" in English and "REL" in French.
NuclearPlant/Reactor/LWR	skos:altLabel	"REL" @fr	
NuclearPlant/Reactor/LWR	skos:broader	NuclearPlant/Reactor	The broader term of "NuclearPlant/Reactor/LWR" is "NuclearPlant/Reactor".
NuclearPlant/Reactor/LWR	skos:narrower	NuclearPlant/Reactor/LWR/PWR	"NuclearPlant/Reactor/LWR" has two narrower terms: "NuclearPlant/Reactor/LWR/PWR" and "NuclearPlant/Reactor/LWR/BWR"
NuclearPlant/Reactor/LWR	skos:narrower	NuclearPlant/Reactor/LWR/BWR	

Source: NEA, 2019.

Figure A.8.5 shows the RDF graph for the controlled dictionary item “NuclearPlant/Reactor/LWR”: this item is the subject of six different RDF triples shown in Table A.11.

Figure A.0.5: RDF graph for “NuclearPlant/Reactor/FBR” concept



Source: NEA, 2019.

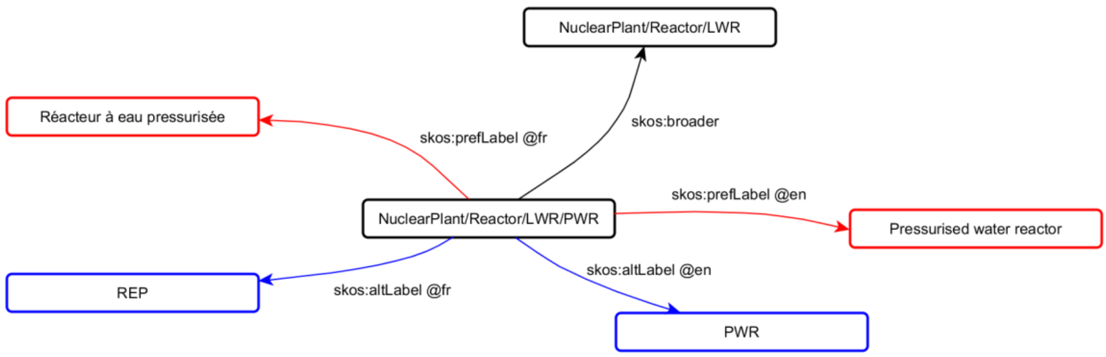
Table A.11: RDF triples for “NuclearPlant/Reactor/FBR” concept

RDF Triples			RDF/SKOS Explanation
Subject	Predicate	Object	
NuclearPlant/Reactor/FBR	skos:inScheme	NuclearPlant	“NuclearPlant/Reactor/FWR” is a concept of the “NuclearPlant” concept scheme.
NuclearPlant/Reactor/FBR	skos:prefLabel	”Fast breeder reactor”@en	“NuclearPlantReactor/FBR” has two preferred labels: “Fast breeder reactor” in English and “Réacteur à neutron rapids” in French.
NuclearPlant/Reactor/FBR	skos:prefLabel	”Réacteur à neutrons rapids”@fr	
NuclearPlant/Reactor/FBR	skos:altLabel	“FBR” @en	“NuclearPlantReactor/FBR” has two alternative labels: “FBR” in English and “RNR” in French.
NuclearPlant/Reactor/FBR	skos:altLabel	“RER” @fr	
NuclearPlant/Reactor/FBR	skos:broader	NuclearPlant/Reactor	The broader term of “NuclearPlant/Reactor/FBR” is “NuclearPlant/Reactor”.

Source: NEA, 2019.

Figure A.8.6 shows the RDF graph for the controlled dictionary item "NuclearPlant/Reactor/LWR": this item is the subject of six different RDF triples shown in Table A.12.

Figure A.0.6: RDF graph for "NuclearPlant/Reactor/LWR/PWR" concept



Source: NEA, 2019.

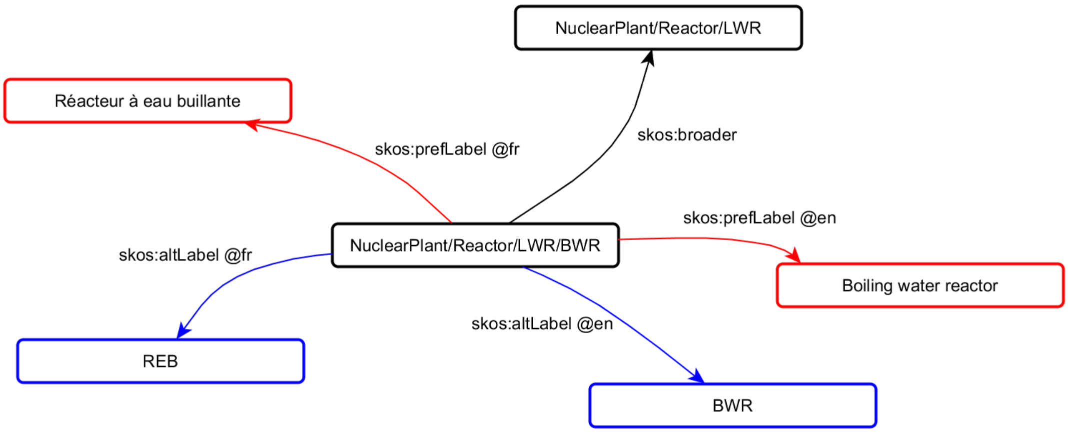
Table A.12: RDF triples for "NuclearPlant/Reactor/LWR/PWR" concept

RDF Triples			RDF/SKOS Explanation
Subject	Predicate	Object	
NuclearPlant/Reactor/LWR/PWR	skos:inScheme	NuclearPlant	"NuclearPlant/Reactor/LWR/PWR" is a concept of the "NuclearPlant" concept scheme.
NuclearPlant/Reactor/LWR/PWR	skos:prefLabel	"Pressurised water reactor"@en	"NuclearPlantReactor/LWR/PWR" has two preferred labels: "Pressurised water reactor" in English and "Réacteur à eau pressurisée" in French.
NuclearPlant/Reactor/LWR/PWR	skos:prefLabel	"Réacteur à eau pressurisée"@fr	
NuclearPlant/Reactor/LWR/PWR	skos:altLabel	"PWR" @en	"NuclearPlantReactor/LWR/PWR" has two alternative labels: "PWR" in English and "REP" in French.
NuclearPlant/Reactor/LWR/PWR	skos:altLabel	"REP" @fr	
NuclearPlant/Reactor/LWR/PWR	skos:broader	NuclearPlant/Reactor/LWR	The broader term of "NuclearPlant/Reactor/LWR/PWR" is "NuclearPlant/Reactor/LWR".

Source: NEA, 2019.

Figure A.8.7. shows the RDF graph for the controlled dictionary item "NuclearPlant/Reactor/LWR": this item is the subject of six different RDF triples shown in Table A.13.

Figure A.0.7: RDF graph for "NuclearPlant/Reactor/LWR/BWR" concept



Source: NEA, 2019.

Table A.13: RDF triples for "NuclearPlant/Reactor/LWR/BWR" concept

RDF Triples			RDF/SKOS Explanation
Subject	Predicate	Object	
NuclearPlant/Reactor/LWR/BWR	skos:inScheme	NuclearPlant	"NuclearPlant/Reactor/LWR/BWR" is a concept of the "NuclearPlant" concept scheme.
NuclearPlant/Reactor/LWR/BWR	skos:prefLabel	"Boiling water reactor"@en	"NuclearPlantReactor/LWR/BWR" has two preferred labels: "Boiling water reactor" in English and "Réacteur à eau buillante" in French.
NuclearPlant/Reactor/LWR/BWR	skos:prefLabel	"Réacteur à eau buillante"@fr	
NuclearPlant/Reactor/LWR/BWR	skos:altLabel	"BWR" @en	"NuclearPlantReactor/LWR/BWR" has two alternative labels: "BWR" in English and "REB" in French.
NuclearPlant/Reactor/LWR/BWR	skos:altLabel	"REB" @fr	
NuclearPlant/Reactor/LWR/BWR	skos:broader	NuclearPlant/Reactor/LWR	The broader term of "NuclearPlant/Reactor/LWR/BWR" is "NuclearPlant/Reactor/LWR".

Source: NEA, 2019.

The RDF triples can be serialised[40] in several different formats such as Turtle, JSON-LD, RDF/XML. RepMet decided to follow the RDF/XML serialisation for the development of its original RDF/SKOS controlled dictionaries included in the libraries.

40. Serialisation is a process to convert an object with its own data structure in a digital format that can be easily stored or transmitted.

Box A.8.1 illustrates the RDF/XML serialisation of the RDF/SKOS triples within the "NuclearPlant" concept scheme that the figures and the tables above report.

Box A.0.1: XML serialisation of the NuclearPlant SKOS concept scheme


```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF
<!— DECLARATION OF THE RDF SYNTAX NAMESPACES -->
xmlns:dc=http://purl.org/dc/elements/1.1/
xmlns:fo=http://www.w3.org/1999/XSL/Format
xmlns:gr="http://graphml.graphdrawing.org/xmlns"
xmlns:nea="http://oecd.org/nea/"
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
xmlns:skos="http://www.w3.org/2004/02/skos/core#"
xmlns:ss="urn:schemas-microsoft-com:office:spreadsheet"
xmlns:y="http://www.yworks.com/xml/graphml">

<!— DEFINITION OF THE SKOS CONCEPT SCHEME AND ITS PROPERTIES -->
        <skos:ConceptScheme rdf:about="http://oecd.org/nea//NuclearPlant">
            <dc:creator>OECD/NEA RepMet</dc:creator>
            <skos:prefLabel xml:lang="en">Nuclear Plants</skos:prefLabel>
            <skos:hasTopConcept rdf:resource="http://oecd.org/nea//NuclearPlant/Reactor"/>
            <skos:hasTopConcept rdf:resource="http://oecd.org/nea//NuclearPlant/FuelCyclePlant"/>
         </skos:ConceptScheme>

<!— DEFINITION OF THE SKOS CONCEPTS AND THEIR PROPERTIES -->
        <rdf:Description rdf:about="http://oecd.org/nea//NuclearPlant/Reactor">
            <rdf:type rdf:resource="http://www.w3.org/2004/02/skos/core#Concept"/>
            <skos:inScheme rdf:resource="http://oecd.org/nea//NuclearPlant"/>
            <skos:topConceptOf rdf:resource="http://oecd.org/nea//NuclearPlant"/>
            <skos:narrower rdf:resource="http://oecd.org/nea//NuclearPlant/Reactor/LWR"/>
            <skos:narrower rdf:resource="http://oecd.org/nea//NuclearPlant/Reactor/FBR"/>
            <skos:prefLabel xml:lang="en">Reactor”</skos:prefLabel>
            <skos:prefLabel xml:lang="fr">Réacteur”</skos:prefLabel>
         </rdf:Description>
         <rdf:Description rdf:about="http://oecd.org/nea//NuclearPlant/Reactor/LWR">
            <rdf:type rdf:resource="http://www.w3.org/2004/02/skos/core#Concept"/>
            <skos:inScheme rdf:resource="http://oecd.org/nea//NuclearPlant"/>
            <skos:broader rdf:resource="http://oecd.org/nea//NuclearPlant/Reactor"/>
            <skos:narrower rdf:resource="http://oecd.org/nea//NuclearPlant/Reactor/LWR/PWR"/>
            <skos:narrower rdf:resource="http://oecd.org/nea//NuclearPlant/Reactor/LWR/BWR"/>
            <skos:prefLabel xml:lang="en">Light water reactor”</skos:prefLabel>
            <skos:prefLabel xml:lang="fr">Réacteur à eau légère”</skos:prefLabel>
         </rdf:Description>
         <rdf:Description rdf:about="http://oecd.org/nea//NuclearPlant/Reactor/FBR">
            <rdf:type rdf:resource="http://www.w3.org/2004/02/skos/core#Concept"/>
            <skos:inScheme rdf:resource="http://oecd.org/nea//NuclearPlant"/>
            <skos:broader rdf:resource="http://oecd.org/nea//NuclearPlant/Reactor"/>
            <skos:prefLabel xml:lang="en">Fast breeder reactor”</skos:prefLabel>
            <skos:prefLabel xml:lang="fr">Réacteur à neutrons rapids”</skos:prefLabel>
```

```
        </rdf:Description>
        <rdf:Description rdf:about="http://oecd.org/nea//NuclearPlant/Reactor/LWR/PWR">
            <rdf:type rdf:resource="http://www.w3.org/2004/02/skos/core#Concept"/>
            <skos:inScheme rdf:resource="http://oecd.org/nea//NuclearPlant"/>
            <skos:broader rdf:resource="http://oecd.org/nea//NuclearPlant/Reactor/LWR"/>
            <skos:prefLabel xml:lang="en">Pressurised water reactor”</skos:prefLabel>
            <skos:prefLabel xml:lang="fr">Réacteur à eau pressurisée”</skos:prefLabel>
        </rdf:Description>
        <rdf:Description rdf:about="http://oecd.org/nea//NuclearPlant/Reactor/LWR/BWR">
            <rdf:type rdf:resource="http://www.w3.org/2004/02/skos/core#Concept"/>
            <skos:inScheme rdf:resource="http://oecd.org/nea//NuclearPlant"/>
            <skos:broader rdf:resource="http://oecd.org/nea//NuclearPlant/Reactor/LWR"/>
            <skos:prefLabel xml:lang="en">Boiling  water reactor”</skos:prefLabel>
            <skos:prefLabel xml:lang="fr">Réacteur à eau buillante”</skos:prefLabel>
        </rdf:Description>
</rdf:RDF>
```

A4 Use of RDF predicates in RepMet

Each attribute of the RepMet controlled dictionaries can be considered as a SKOS concept and all the attributes of an entity as a SKOS concept scheme[41]. In fact, the SKOS classes and the properties presented in the previous sections of this chapter are ideal for representing the hierachical structure of the dictionary and the features that the group identified for each attribute in a way that is both harmonised internationally, since it is a W3C standard, and suitable for long-term management and use.

Table A.14 illustrates the RDF predicates of the SKOS Vocabulary used to represent and encode the features that RepMet selected for each attribute.

Table A.14: RDF predicates for attribute feature

Attribute features in the RepMet controlled dictionary	Corresponding RDF predicates
Name	skos:prefLabel @en
Definition	skos:definition @en
Definition source	dc:source
Purpose	skos:scopeNote @en
Comment	skos:comment @en

Source: NEA, 2019.

Table A.15 illustrates the other RDF predicates of the SKOS Vocabulary used to encode the hierarchical structure of the concept scheme associated to the entity to which the attribute belongs.

Table A.15: RDF predicates for describing the attribute dictionary structure

Attribute features in the RepMet controlled dictionary	Corresponding RDF predicates
Entity association	skos:inScheme
Hierarchical position with respect to other attributes	skos:broader
	skos:narrower

Source: NEA, 2019.

Most of the RDF predicates used are from the SKOS Vocabulary developed by RepMet. However, the RepMet group decided to use RDF properties coming from other RDF vocabularies such as the DCMI Vocabulary (such as the "definition source" property). This is a classic example of how different RDF vocabularies can be integrated and the work of a user community can be reused. Indeed, a different option for the RepMet group would

41. At the end of the RepMet activities, the elaborated controlled dictionaries, or SKOS concept schemes, have inhomogeneous set of attributes. It means that they have different semantic roles. There are typical dictionary terms that are pure skos:Concept (e.g. "explosive agent", "exotic agent", "corrosive agent") and other that are more owl:ObjectProperties (e.g. "waste owner", "international classification", "origin of the waste") that represent more property with range of some other classes. For the future activities, one of the possible tasks is to improve the RepMet dictionaries by separating better the ontology and the dictionary items.

have been to create a specific RDF vocabulary, but the use of consolidated RDF vocabularies such as the SKOS and DCMI vocabularies avoided the duplication of efforts in “reinventing the wheel”. That approach could be easily implemented by the RWMOs.