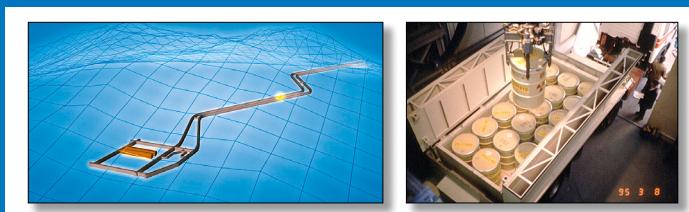


Metadata for Radioactive Waste Management



Metadata for Radioactive Waste Management

© OECD 2018
NEA No. 7378

NUCLEAR ENERGY AGENCY
ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT

The OECD is a unique forum where the governments of 36 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, 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.

This work is published on the responsibility of the Secretary-General of the OECD.

NUCLEAR ENERGY AGENCY

The OECD Nuclear Energy Agency (NEA) was established on 1 February 1958. Current NEA membership consists of 33 countries: Argentina, Australia, Austria, Belgium, 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/publishing/corrigenda.

© OECD 2018

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.

Cover photos: Diagram of the underground storage chambers (Bátaapáti NRWR, Hungary); Drums being introduced into the concrete container before being disposed into the vaults (El Cabri site, ENRESA, Spain).

Foreword

In the current global context of data-rich environments, where a greater variety of data are being captured and stored in rapidly increasing quantities, adequate data management is essential for any organisation to maximise and maintain the value of its data assets so as to support operations and business processes. The use of metadata – which provide additional description, context or supplementary information about existing data – is a key aspect of effective data management. The capture, use and maintenance of metadata can enhance the usability and long-term value of data and information in an organisation.

Radioactive Waste Management Organisations (RWMOs) have specific needs in terms of data and information that are required to manage radioactive waste repositories. The capacity must be made available within an RWMO to maintain the usability of relevant data and information over the long time frames corresponding to the lifetime of a waste facility, offering adequate support to meet possible regulatory requirements. Metadata can support these needs, as long as they are carefully planned and implemented. RWMOs can also benefit from international collaboration, to learn from the experience of others, potentially with more advanced programmes, and to assist with potential information and data exchange.

The Nuclear Energy Agency (NEA) Radioactive Waste Repository Metadata Management (RepMet) initiative was set up to examine and recommend approaches and techniques for using metadata in radioactive waste repository management. This initiative is complementary to the Preservation of Records, Knowledge and Memory across Generations (RK&M) initiative, which focuses on the period after repository closure.

Acknowledgements

The Nuclear Energy Agency 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 (RWM, United Kingdom – RepMet Chair)
- József I. Fekete (PURAM, Hungary)
- Massimo Ciambrella (NEA)
- Hervé Bienvenu (ANDRA, France)
- Kevin McMahon (Sandia National Laboratories, United States – RepMet Vice-Chair)

And

- Simon Lambert (United Kingdom – Consultant)

Table of contents

| | |
|--|-----------|
| List of abbreviations and acronyms | 7 |
| Executive summary | 9 |
| 1. Introduction | 13 |
| 1.1 Background and objectives of the RepMet initiative | 13 |
| 1.2 RepMet deliverables..... | 14 |
| 1.3 Audience of RepMet documents | 15 |
| 1.4 Structure of this report..... | 17 |
| 2. Metadata | 19 |
| 2.1 What are metadata? | 20 |
| 2.2 Why metadata? | 23 |
| 2.3 Metadata for RWMOs..... | 27 |
| 3. Implementation of metadata | 31 |
| 3.1 Metadata policies | 31 |
| 3.2 Implementation phases and issues arising..... | 33 |
| 3.3 Common problems with metadata..... | 39 |
| 4. Using RepMet for metadata implementation | 41 |
| 4.1 Using the RepMet deliverables | 41 |
| 4.2 Technical bases of the RepMet deliverables..... | 43 |
| 4.3 RepMet Libraries..... | 43 |
| 4.4 RepMet Tools and Guidelines | 46 |
| 5. Closing remarks and recommendations | 51 |
| 5.1 Summary of RepMet | 51 |
| 5.2 Recommendations | 52 |
| 5.3 Future work..... | 53 |
| 6. References | 55 |
| Annex A. The RepMet approach and methodology..... | 57 |
| Annex B. Designated communities..... | 61 |

List of abbreviations and acronyms

| | |
|-----------|---|
| Andra | Agence nationale pour la gestion des déchets radioactifs (France) |
| CFD | Computational fluid dynamics |
| Enresa | Empresa Nacional de Residuos Radioactivos SA (Spain) |
| FEP | Features, events and processes |
| GDPR | European General Data Protection Regulation (EU) 2016/679 |
| GeoSciML | Geoscience Markup Language (Data model and data transfer standard for geological data) |
| ID | Identifier |
| IFEP List | International FEP List |
| IGSC | Integration Group for the Safety Case (NEA) |
| INSPIRE | Infrastructure for spatial information in Europe (Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007) |
| IT | Information technology |
| JAEA | Japan Atomic Energy Agency (Japan) |
| Nagra | National Co-operative for the Disposal of Radioactive Waste (Switzerland) |
| NEA | Nuclear Energy Agency |
| NGO | Non-governmental organisation |

| | |
|--------------|--|
| NWMO | Nuclear Waste Management Organization (Canada) |
| OAIS | Open archival information system |
| OECD | Organisation for Economic Co-operation and Development |
| ONDRAF/NIRAS | National Agency for Radioactive Waste and Enriched Fissile Material (Belgium) |
| O&M | Observations and Measurements (ISO 19156) |
| PFEP | Project-specific FEP List |
| PURAM | Public Limited Company for Radioactive Waste Management (Hungary) |
| RDF | Resource description framework |
| RepMet | Radioactive Waste Repository Metadata Management (NEA) |
| RWM | Radioactive waste management |
| RWM/NDA | Radioactive Waste Management of the Nuclear Decommissioning Authority (United Kingdom) |
| RWMC | Radioactive Waste Management Committee (NEA) |
| RWMO | Radioactive waste management organisation |
| SKB | Swedish Nuclear Fuel and Waste Management Company |
| SKOS | Simple knowledge organisation system |
| SÚRAO | Radioactive Waste Repository Authority (Czech Republic) |
| URI | Uniform resource identifier |
| UML | Unified modelling language |
| W3C | World wide web consortium |

Executive summary

The Nuclear Energy Agency (NEA) initiated the Radioactive Waste Repository Metadata Management (RepMet) initiative in 2014, starting with the basic tenet of long-term data, information and records management, which is that data should be collected and managed for use by others. Key to the collection and management of data is the effective use of *metadata* in information systems, which has numerous benefits. For example, it allows for the effective management of information and data while helping radioactive waste management organisations (RWMOs) to meet statutory requirements and maintain the long-term quality and value of this information and data. Anyone in an organisation which produces and works with data, as well as managers and communication specialists, needs to be aware of the usefulness of metadata and play their own part to implement them within the organisation.

The goal of RepMet is to recommend sets of metadata that can be used by national radioactive waste programmes to manage radioactive waste repository data, information and records in a way that is both harmonised internationally and suitable for long-term management and use (e.g. for the development of safety cases). The initiative has produced a set of guiding principles and practical advice on capturing and generating metadata, including the use of relevant standards.

Because the preliminary work of RepMet has been carried out at the international level within the NEA framework, radioactive waste management organisations will benefit from an internationally harmonised approach that has been developed in association with similar organisations. RepMet based its work on the best practices currently in use nationally, the exchange of experiences among RepMet participants, and discussions with external experts and constituencies. The RepMet initiative fills a unique and important niche in the broader programmes on knowledge management that are being conducted nationally and internationally by operators, regulators and other relevant stakeholders.

This report provides an overview of RepMet goals, deliverables, various aspects of metadata implementation and associated issues for consideration, outlining the deliverables produced by RepMet since 2014, including:

- The present publication, *Metadata for Radioactive Waste Management*, which introduces the concept of metadata, explaining why it is valuable in data management, and in particular for radioactive waste management organisations (RWMOs), and providing advice on issues arising when implementing metadata as part of data management.

- Three libraries which set out reusable metadata models in the areas of waste packaging, engineering and geoscience. These three libraries can be used as a basis for establishing and maintaining reliable, high-quality, reusable data through the proper use of metadata.
- The report “RepMet Tools and Guidelines” (NEA, forthcoming) which presents tools, techniques and standards used by RepMet in its libraries, and will be valuable for all RWMOs in adopting and developing their own metadata-based systems.

RepMet recognises that each RWMO’s activities are unique in many respects, and each site has characteristics that are exclusive to its particular circumstances. The outputs of RepMet are nevertheless intended to be generic and adaptable to the needs of virtually every RWMO.

As part of the present report, the RepMet initiative has formulated the following ten recommendations on metadata for RWMOs:

- R1. Establish a comprehensive metadata policy that covers the aspects listed in this report and the individual organisation's statutory and other requirements.
- R2. Analyse the potential needs for, and benefits of, metadata in the organisation's processes, considering long-term access and use of information and the risks of loss of information over time. The benefits should be balanced against the costs of implementation to assess the overall value of implementation.
- R3. Define designated communities who will make use of the information, and analyse the consequent needs for metadata to support their use.
- R4. Dedicate sufficient resources to the formulation of requirements, to the planning of metadata implementation, and to ongoing and supporting activities such as metadata maintenance, testing and user training. A need for training may arise to understand the importance of metadata and the procedures for capturing them and verifying their quality.
- R5. Establish processes that will instil a culture of high-quality metadata creation and maintenance during the entire lifecycle of the relevant data. Accountability for metadata entry and quality should be clearly defined.
- R6. Ensure, to the extent possible, that metadata are captured at the earliest opportunity, rather than being added retrospectively and, if possible, generated automatically. If metadata are entered manually, there should be clear supporting guidelines and, where possible, data validation to ensure their quality.

- R7. Consider the opportunities and benefits of using international standards when planning for metadata implementation so as to avoid duplication of work and ensure coherence and harmonisation across RWMOs.
- R8. Make use of controlled dictionaries as the basis for metadata, both throughout the organisation and ideally across RWMOs so as to achieve the benefits of interoperability and international harmonisation. Controlled dictionaries should be cascaded to external organisations in the supply chain. Procedures should also be established to ensure controlled dictionaries are kept up-to-date, are made available to all users and have well-defined periods of validity.
- R9. Establish metadata implementation on the basis of RepMet's libraries, if possible.
- R10. Make use of "RepMet Tools and Guidelines" (NEA, forthcoming) when implementing metadata.

The adoption of RepMet results will contribute to ensuring that RWMO data and information management efforts are successful and cost efficient. RWMOs will thus see a return on their financial and time investment through the better management of data and information, while reducing the risks of information loss and the need to recreate data, as well as any consequent reputational and financial costs.

In future, the RepMet members wish to deliver additional libraries in key areas of radioactive waste management, including in the area of safety case/assessment, site operations, transportation, site pre-closure operations and other, specific areas of interest to RWMOs.

1. Introduction

1.1 Background and objectives of the RepMet initiative

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 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, the conclusion was reached that there is a considerable need and potential for metadata management and harmonisation.

Radioactive waste management organisations (RWMOs) are required to manage very large amounts of data, which they produce and receive, in order to support their operational, pre- or post-closure safety cases and other requirements. A special characteristic of radioactive waste repositories is the long time frames, typically in excess of one hundred years, between facility construction and closure. Systems handling data and relevant metadata may go through unforeseen technological and other changes, data carriers and the data itself may no longer be, and the computer programs handling such data may become obsolete. In addition, different generations of workers will perform tasks on the site during this time with a high probability that not all knowledge will be handed down to new employees. Therefore, data handled by RWMOs requires specific treatment in order to be considered reliable and therefore usable for such long time periods.

In such a challenging environment, the main goal of RepMet, as defined in its vision document, was to create sets of metadata that can be used by national programmes to manage their repository data, information and records in a way that is both harmonised internationally and suitable for long-term management and use in safety cases and elsewhere. RepMet also aimed to formulate a consistent set of guiding principles for capturing and generating metadata.¹

Several worldwide RWMOs and research laboratories from NEA countries were involved in the RepMet initiative: the National Agency for Radioactive Waste and Enriched Fissile Material (ONDRAF/NIRAS, Belgium); the Nuclear Waste Management Organization (NWMO, Canada); the Radioactive Waste Repository

¹ It should be specified that RepMet does not intend to promote any commercial products or services for managing metadata.

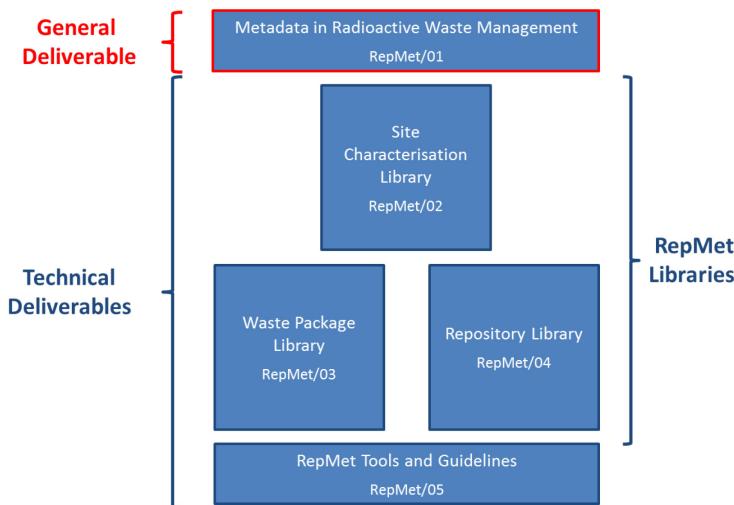
Authority (SÚRAO, Czech Republic); Posiva Oy (Finland); the National Radioactive Waste Management Authority (Andra, France); the Public Limited Company for Radioactive Waste Management (PURAM, Hungary); the Atomic Energy Agency (JAEA, Japan); Empresa Nacional de Residuos Radiactivos (Enresa, Spain); the Nuclear Fuel and Waste Management Company (SKB, Sweden); the National Co-operative for the Disposal of Radioactive Waste (Nagra, Switzerland); Radioactive Waste Management of the Nuclear Decommissioning Authority (RWM/NDA, United Kingdom) and Sandia National Laboratories (United States).

The RepMet group met twice yearly, with working groups composed of RepMet members and contractors furthering the initiative in the intervening periods.

1.2 RepMet deliverables

Figure 1.1 illustrates the relationships between the RepMet deliverables produced during the initiative from its beginning in 2014 through 2017.

Figure 1.1: RepMet deliverables



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

The three deliverables identified as “libraries” are technically more detailed. They outline the key aspects of data and related metadata for selected scientific and technical topics involved in the lifecycle 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 they are ready for use by RWMOs.

RepMet/02 – The “Site Characterisation Library” (NEA, forthcoming) 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.

RepMet/03 – The “Waste Package Library” (NEA, forthcoming) deals with data and related metadata related to packaged waste and spent nuclear fuel that, after proper treatment and conditioning processes, are ready for final disposal at the repository.

RepMet/04 – The “Repository Library” (NEA, forthcoming) deals with data and related metadata that are linked with the engineered structures and waste acceptance requirements of radioactive waste repositories.

The above libraries can be used independently of one another. However, using all of the libraries and the approach outlined in these documents provides the additional benefit of a uniform approach to metadata management.

RepMet/05 – “RepMet Tools and Guidelines” (NEA, forthcoming) supports the aforementioned libraries, providing a number of useful 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 deliverable documents are made available in electronic form on the RepMet webpage of the NEA website.²

1.3 Audience of RepMet documents

The information provided within the RepMet deliverables is primarily aimed at RWMOs that are considering developing information systems or establishing knowledge management practices related to radioactive waste management, or that are planning to renew or update their existing data management practices. The information provided is intended to be generic, meaning that it can be adapted by almost any RWMO. The information may also be of use for other disciplines such as developing inventory and decommissioning models.

The present report is primarily intended for managers and decision makers within RWMOs who wish to gain an understanding of the uses and benefits of metadata and of issues involved in their implementation. The other deliverables are aimed at technical specialists responsible for the implementation of metadata in their organisations; they present usable resources to assist with implementation.

2. www.oecd-nea.org/rwm/igsc/repmet

Table 1.1 outlines the audience groups and the intended use of RepMet deliverables.

Table 1.1: Audiences of RepMet deliverables

| Deliverable | Primary audiences | Secondary audiences |
|---|---|--|
| <i>Metadata for Radioactive Waste Management report</i> | <p>RWMO managers and decision makers:</p> <ul style="list-style-type: none"> • what metadata are and why they are valuable to organisations; • issues to consider in metadata implementation, and how RepMet results may be adopted; • high-level recommendations on metadata adoption and implementation at the organisational level. <p>Information system developers:</p> <ul style="list-style-type: none"> • awareness of benefits and risks in metadata; implementation projects • identification of possible designated communities for metadata use. | <p>Local and international regulators Other concerned authorities:</p> <ul style="list-style-type: none"> • awareness of the role of metadata in ensuring audit trails and long-term reliability of data, information and records. <p>Non-specialist audiences:</p> <ul style="list-style-type: none"> • understanding best practices in information handling in RMW, and expectations on what information should be available over the long term. |
| <i>RepMet Libraries</i> | <p>Information system developers:</p> <ul style="list-style-type: none"> • reusable data models and controlled dictionaries, developed and validated by RepMet. <p>RWMO engineers:</p> <ul style="list-style-type: none"> • awareness of the attributes of interest to information systems for long-term access and use; • agreed vocabulary for international harmonisation of terms. | <p>Academics:</p> <ul style="list-style-type: none"> • current best practice in metadata modelling for RWMOs, as a basis for further development in future. |
| <i>"RepMet Tools and Guidelines"</i> | <p>Information system developers:</p> <ul style="list-style-type: none"> • tools and techniques for use during the implementation process; • recommended existing standards and how they may be applied. | |

1.4 Structure of this report

This report is structured as follows:

- Chapter 2 introduces what metadata are, why they are important to radioactive waste management and the main benefits of their application.
- Chapter 3 provides guidelines for the implementation of metadata by RWMOs in their data management programmes.
- Chapter 4 describes the RepMet deliverables and how they may be used.
- Chapter 5 concludes the report with a short summary and makes ten specific recommendations concerning metadata for RWMOs.

2. Metadata

The simple definition of metadata is “data about data”. Although it may seem a superficial statement, it nonetheless captures an important point: that metadata provide some additional description, context or supplementary information about existing data. Existing data are often encapsulated in a dataset or record (for example a time series of measurements from an instrument, stored as a digital file), though it might be any digital object (such as an electronic document). In reality, the word “metadata” is often used more loosely – but without difficulty – to mean supplementary data about something that is not itself a digital object – an artefact in a museum, for example, or a radioactive waste package. Zeng (2016) provides an overview of metadata, although from the perspective of libraries and information scientists rather than those of business or technical fields.

The idea of metadata predates the digital world. A classic example is provided by the index cards in a library. Each of these cards relates to a book in the library collection and summarises its title, author, subject, year of publication, location on the shelves and other useful information. Herein is another important point: metadata are intended to be of use to people, in this case a library user searching for a book.

The purpose of metadata is to make the underlying data (or other object) easier to find, to understand and to use. Metadata may also help to give confidence in the data, by representing their provenance or chain of processing. The general principle is that data are being collected and managed for others to use them. In today’s digital world, with its deluge of data, metadata have a vital role to play. The trend towards open science (see for example European Commission [2017]) in particular requires increasing the accessibility and reuse of data on a large scale. In regulated environments, such as radioactive waste management, metadata can help to give assurance that correct procedures have been followed and that digital records are accessible.

A growing need in many fields is the long-term preservation of digital material. Many risks are involved in accessing and using material as time passes, and adequate metadata is one technique for countering such risks. The basic tenet of long-term data preservation is that “data are being collected and managed for others to use” – and metadata must support this tenet.

In order to be useful, metadata need managing. The amount and structure of the metadata must be defined. The processes of capturing and maintaining metadata must also be set up and implemented in ways that are appropriate for the organisation and its business. In some cases, simple unrestricted tagging is enough: tagging a photograph collection with the names of places or people is

adding a kind of metadata, and it is useful for retrieving images. In general, however, and especially in complex technical environments, much more is needed.

2.1 What are metadata?

It is difficult to formally define metadata, but relatively easy to understand in practice. Data are generally basic facts or figures in the form that they were collected or generated; metadata are then additional pieces of information which add something more to this data, either by giving a context to the data or by helping to organise it in some way. A simple example of a rain gauge in the field of hydrology illustrates the point: the data are the measured amount of rainfall on a series of days, while the metadata may include the units of measurement, the spatial co-ordinates of the gauge or the name of the field worker making the reading.

A formal distinction between data and metadata is not straightforward, primarily for two reasons. First, one person's data may be another's metadata. A set of bibliographic records are metadata for a library user looking for a book in the catalogue, but they are "data" for a librarian in charge of the entire collection, giving information, *inter alia*, on the coverage of subject areas and ages of books. Second, modern measuring equipment often contains processors and control systems and automatically produces output files that aggregate raw and calculated values with some associated metadata. The metadata do not have to be separate from the base data; they might be embedded within them.

Metadata may at any time be added or amended throughout the lifecycle of the data. For example metadata are added to a waste package record during its interim storage, and again at its final disposal.

For the purposes of RepMet, metadata are defined as *structured data providing or pointing to information about a digital or physical object*. In this report, the term *base data* or (more generally) *base object* is used to mean the digital or physical object to which the metadata relate.

The term "record" is also used when the emphasis is on information created or maintained by an organisation for legal obligations or as part of its business. The international standard ISO 15489-1:2016 is concerned with the concepts and principles of records management.

Metadata can be classified into various types according to their nature and purpose. A simple and clear classification identifies three types of metadata: *administrative metadata*, *structural metadata* and *descriptive metadata* (Riley 2017). It should be noted that this classification is intended to be a helpful way of thinking about the types of metadata, rather than a formal distinction, and there is some overlap between the types.

- *Administrative metadata* assist with managing a base object. For a digital object, they may include: technical information on file types and how to decode and render the data; electronic signatures; information on access rights; and copyright (if applicable). They also may include information that may be required by legislative prescriptions or business processes such as

the arrival date of a record, the date sent, registration number, delivery method, document subject, document administration deadline, registration numbers of preceding and subsequent documents, and retention period, among others.

- *Structural metadata* describe the relationships between base objects and parts of them, providing context for understanding. Examples are pages in a sequence of scanned documents, and links between different versions of a dataset that has been updated.
- *Descriptive metadata* give information about the base object that aids in finding or understanding it. An example is the library index card mentioned above. Descriptive metadata may be represented as keywords, tags or subject areas chosen from controlled dictionaries. Descriptive metadata may also put the base objects into a wider context, allowing them to be related and compared, or to understand their provenance. Examples include organisational details, or explanation of why a given record was produced. Descriptive metadata might also represent very specific technical attributes of an object, in the context of radioactive waste management; for example, information about the data or object such as type, weight, dimensions, source, origin of a waste package, or details as to how chemical and radionuclide inventory in the waste is derived and defined.

Box 2.1 illustrates how descriptive metadata may be used to give meaning to raw data, and, in a more elaborate example, to establish confidence in a dataset.

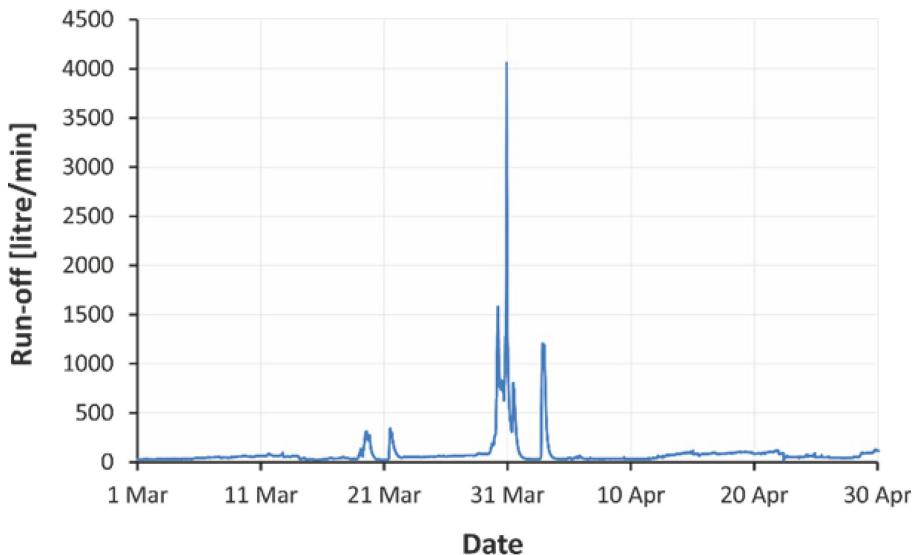
RepMet's libraries (discussed later) are primarily focused on descriptive metadata.

A much more formal classification of metadata is given in (Jeffery, 1998), dividing them into *schema*, *navigational* and *associative* metadata.

Box 2.1: Some uses of descriptive metadata

Books in a library have associated titles, authors, years of publication, etc., and these data can be easily interpreted by a human with sufficient background knowledge. However, if the base data are entirely numerical, as in scientific and technical fields, some descriptive metadata are essential to interpret the data correctly, for example: what are the units? What are the co-ordinate systems or reference points?

As a more detailed example, in the field of hydrogeology, data are collected on water run-off using instruments at various locations. Looking at a data set from the beginning of March to the end of April, and displaying it in graph format (Figure 2.1), an expert may notice that Station B7 had an unexpected spike in run-off at the end of March, beginning of April. The expert may come to the conclusion that the data should be discarded because it seems likely that the measuring instrument was at fault.

Figure 2.1: Illustrative example of dataset without metadata**Station B7**

However, looking at the metadata in Table 2.1, the expert can observe that the instrument's calibration date is valid, the method applied for the measurements exists, the names of people verifying the measurement are correct, and confirmation that the range of measurements falls within the allowed limits (e.g. water levels can be negative numbers [below ground level], but water yield can only be positive). All of these factors strengthen confidence in the dataset and allow the presumption that it is valid and useful for further analysis in spite of the initial doubts – in other words the peak in yield was most likely caused by unusually high levels of precipitation.

Table 2.1: Illustrative example of descriptive metadata

| Measured data | | Calculated data | |
|-------------------------|-------------|----------------------|--|
| Date, time | Water level | Run-off (calculated) | |
| 01 March 2013, 00:00:00 | 0.04 | 25.76 | |
| 01 March 2013, 00:10:00 | 0.04 | 25.76 | |
| 01 March 2013, 00:20:00 | 0.04 | 27.40 | |
| 01 March 2013, 00:30:00 | 0.04 | 25.76 | |
| - | - | - | |
| - | - | - | |
| 30 April 2013, 23:00:00 | 0.08 | 128.61 | |
| 30 April 2013, 23:10:00 | 0.08 | 128.61 | |
| 30 April 2013, 23:20:00 | 0.08 | 128.61 | |
| 30 April 2013, 23:30:00 | 0.08 | 128.61 | |

Descriptive metadata

| Method name | Parameter name | Value | Unit | Med. gr. | Medium | Error type | Margin of error |
|---|------------------------|------------|-------|----------|--------|------------|-----------------|
| Water cruise gauging on Thomson segment | Water level | Measured | M | Liquid | Water | R | 0.01 |
| Water cruise gauging on Thomson segment | Water level corrected | Calculated | M | Liquid | Water | R | 0.01 |
| Water cruise gauging on Thomson segment | Run-off calculated | Calculated | l/min | Liquid | Water | R | 1 |
| Water cruise gauging on Thomson segment | Water level at templet | Measured | M | Liquid | Water | R | 0.01 |

| Instrument name | Serial No. | Instrument type | Manufacturer | Date of manufacture | Calibration date | Calibration validity |
|-------------------------------------|------------|-----------------|--------------|---------------------|------------------|----------------------|
| Automatic cruise gauging instrument | 201107020 | DA-S-LRB-122 | Dataqua Kft. | 2011 | 2011.07.22 | 2014.07.22 |

| Company | Name | Role |
|---------|---------------|------------------------------|
| PURAM | Technician1 | Data capture, data recording |
| PURAM | Technician2 | Data capture, data recording |
| PURAM | Data control1 | Control |

2.2 Why metadata?

Metadata should not be collected for their own sake; appropriate reasons should underpin the decision to capture and store metadata, and these should include financial considerations. Broadly speaking, the purpose of metadata is to make the base objects easier to find, understand or use, both now and possibly long into the future. The implication is that there is an agent doing the finding, understanding

and using; the agent might be a human being or an automated system through software. In the latter case “understand” and “use” have to be interpreted in terms of enabling interoperability and automated processing.

Examples of these three functions are as follows:

- **Find:** searching for datasets with a particular scope or records relating to a particular transaction; distinguishing similar resources; or carrying out a cross-system search.
- **Understand:** being able to read old data files; correctly interpreting data values; or assessing the context in which data were gathered.
- **Use:** having confidence in data; being aware of any restrictions on the data (such as intellectual property rights); or being able to combine data from different sources.

Thinking only of human users, the knowledge in the head of the creator of a dataset might not be shared by others, and the ability to reuse that dataset might critically depend on the capture of some of this knowledge in metadata. Even data that are restricted to a single user or organisation, however, may require metadata to maximise their value: consider a researcher who generates thousands of datasets and later needs to locate one particular instance; or how knowledge is lost over time within a large organisation as staff retire or move to other positions.

An important motivation for metadata is precisely this last point: ensuring the continued usability and value of datasets over time. This is the problem of *digital preservation* (see Box 2.2). The timescales do not need to be very long for this to become an important consideration, and of course for RWMOs with their long time horizons, digital preservation will certainly be a priority.

Box 2.2: Digital preservation and metadata

It has long been recognised that there are threats to the continued accessibility and usability of digital material, which must be countered if the value of the material is to be preserved. The basis of these threats is generally that things change over time such as hardware, software, the environment or tacit knowledge. Physical media, for example, might decay or become obsolete; file formats might change; software might become unusable; or knowledge about the interpretation of the data might be lost. All of these threats can be countered in some way, and indeed it is the job of a long-term digital repository to do just that.

Metadata have an important role to play in digital preservation by giving context to the base objects to assist with their continued usability: for example, information about file formats, the interpretation of numerical data, or information about calibration of instruments.

A key standard in digital preservation is OAIS (the Open Archival Information System), ISO 14721, developed for the space data community but of general applicability (CCSDS, 2012). The OAIS standard defines a reference model for an archive, setting out the concepts and frameworks for the responsibilities and functions relevant to carrying out long-term digital preservation. Of course RWMOs are not primarily archives in this sense, but some of the concepts are applicable and useful.

The OAIS standard introduces a number of important concepts, including:

- The *designated community*: an identified group of potential consumers of the material within the archive who should be able to understand a particular set of information. It is the archive's responsibility to ensure that they can indeed understand it.
- *Representation information*: the information that relates some base data to more meaningful concepts. Examples of representation information are given in Box 2.1, giving meaning to simple numbers to allow their correct interpretation.
- *Submission Information Package (SIP)*, *Archival Information Package (AIP)* and *Dissemination Information Package (DIP)*: these are, respectively, the material as it is submitted to the archive, the form in which it is stored and preserved (augmented with representation information) and the form in which it is distributed to users who request it.

The designated community is particularly relevant in the context of RepMet and is discussed further in Box 2.3.

An overview of many of these concepts is given by Giaretta (2011).

The functions of metadata lead to benefits that should be assessed in particular contexts and organisations. Some of them might be directly translatable to cost savings, while others might be harder to quantify but nonetheless have an impact in contributing to the success of the organisations' operations. The following list identifies some typical benefits to be expected from the capture and use of metadata in organisational processes:

- Reducing the likelihood that data will need to be captured multiple times as original datasets can be located and re-used with confidence. This reduces cost and unnecessary work.
- Facilitating the interchange of data between different systems automatically.

- Providing evidence during audits that effective management arrangements are in place and that legal obligations or requirements from regulators or stakeholders are being met.
- Enforcing confidentiality, access and other administrative requirements.
- Reducing the time employees (or other stakeholders) take to find data and information and therefore make the RWMO more efficient. According to various surveys employees spend up to one day per week searching for documents and information.
- Providing a way of handing down important knowledge over time, as staff change and newcomers arrive who do not possess the knowledge of those who have left. Metadata can be used to capture aspects of that tacit knowledge and enable its transmission to such staff.
- Contributing to a common understanding between organisations through standardisation, potentially at an international level.

Conversely, there are risks entailed in not making use of metadata, some of which are listed below. Since metadata play a key role in information preservation and contributing to establishing confidence in the existing information its loss may have a significant effect on an organisation – especially an organisation like a RWMO with a long time horizon.

- If data or records cannot be found, they can be of no value to support the organisation's future decisions. Likewise traceability of processes and decisions is important and may be at risk if adequate metadata do not exist to support provenance.
- Without adequate metadata it is possible that data may be misinterpreted or misused putting the organisation's reputation at risk.
- Serious financial costs can be incurred if data need to be recreated because existing records have been lost or cannot be adequately understood.

Of course there are costs associated with the effective use of metadata: one-off costs of designing or adopting a metadata model and integrating into organisational processes and systems, as well as ongoing costs of capturing the metadata and maintaining them over time. Capturing, storing and maintaining metadata may require more resources when compared with not capturing and storing metadata at all. The business case for implementing good metadata management generally lies in the savings that are produced as a result of the data being easier to locate and use, and to demonstrate that appropriate processes have been followed. In the worst case, poor metadata management can lead to far higher long-term reputational costs, for example if a safety case was challenged in court and evidence of appropriate processes having been followed could not be provided by the RWMO.

Chapter 3 provides guidance on what must be considered when implementing metadata in an organisational context.

2.3 Metadata for RWMOs

National programmes for radioactive waste repositories are collecting large amounts of data that have to be managed throughout the entire period of institutionalised oversight. This typically spans a considerable amount of time. The data and related records also increase in number, type, and quality as programmes proceed through the successive stages of repository development: pre-siting, siting, site characterisation, construction, operations, pre-closure and finally closure. Regulatory and societal approvals are included in this sequence. Current programmes are also documenting past repository programmes, so that current and future generations are able to understand actions carried out in the past, for example by retrospectively adding metadata to help organise and arrange programme records.

RWMOs, like any other organisation, can benefit from the use of metadata as outlined in Section 2.2. An ideal situation is where the available data, information and records are accessed and updated according to management systems, with users being able to locate what they require through searches of full text or the associated metadata. Metadata can also be a useful tool to help a RWMO to demonstrate that their programmes are appropriately driven. Such context setting information may include information on quality checking or approval; provenance; ownership; and similar.

When considering metadata for a particular domain (as for RepMet), metadata are typically associated with a set of predefined digital or physical objects, for example an ID for a waste package or signing date of a quality log. It is important to remember, however, that metadata also cover how these elements are to be constructed and related (for example how a waste package relates to a packaging campaign), the type and range of values they may take (for example, a package weight must be greater than 0 kg), etc.

Capturing all metadata relating to a particular object is highly impractical and also unnecessary. The idea of the designated community, introduced in the preceding section, is a very useful way of identifying the needs for metadata (see Box 2.3). Identifying the set of metadata that can be considered “sufficient metadata” is no easy task and trying to predict all possible future requirements is not an achievable or practical goal. The consideration of designated communities is a way of focusing on what is both necessary and practicable. RepMet has conducted a thorough analysis of the range of designated communities of interest to RWMOs, and the results of this analysis are provided in Annex B.

Box 2.3: Designated communities

In the OAIS standard, the designated community is defined as “an identified group of potential consumers who should be able to understand a particular set of information. The designated community may be composed of multiple user communities. A designated community is defined by the archive and this definition may change over time.”

A designated community is associated with a knowledge base, which allows members of the community to understand and use the base data. If there is a gap in the knowledge base, then it must be filled with representation information – more loosely, with metadata. The knowledge base of the community might change over time, and therefore the representation information (metadata) should be maintained and updated, and there should be processes in place for tracking the designated communities.

Without going so far as attempting a full implementation of a digital repository that conforms to the OAIS standard, the idea of the designated community and its associated knowledge base are very helpful in pinning down what metadata should be gathered to support the communities in their expected uses of the base objects.

Annex B lists a wide range of stakeholders identified by RepMet who could be designated communities for a RWMO. They include national and international organisations and authorities with an interest in waste disposal, the operators of nuclear facilities themselves and the general public with an interest in environmental safety in their local area.

It has already been noted that the acquisition and use of metadata is something that evolves over time. In particular, metadata must track the phases of the lifecycle of a radioactive waste repository. Data and metadata requirements may change over time, and newer data will be added. The significance and relevance of data can also change during the different phases of a waste or spent fuel management programme. Time phases are important from a metadata perspective as metadata can support choices at important decision points: for example, whether a waste package can begin to be accepted for storage in a given facility.

The time phases in the lifecycle of a radioactive waste repository discussed in RepMet are as follows:

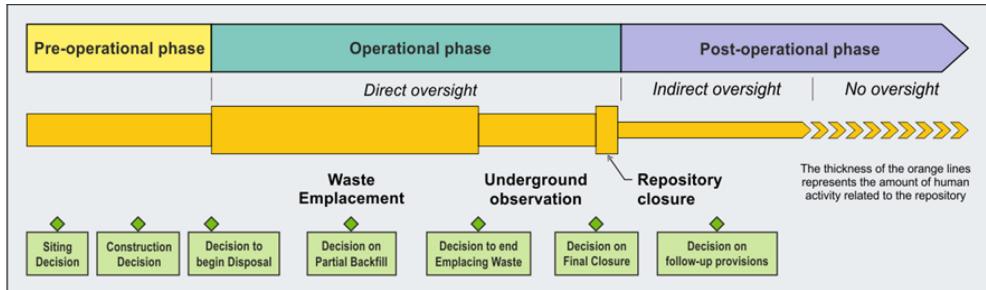
- **Pre-operational phase:** primarily the period of siting, obtaining required permits, and design and construction related activities.
- **Operational phase:** the period when radioactive waste is being disposed of in the repository, up to the time of final closure.
- **Post-operational phase:** the period after repository closure.

In each phase adequate metadata should be stored to support discovery and context setting in that phase and also in later phases. This includes sufficient information on: previous safety assessments; the ability to trace decisions of the implementation process; possible retrieving of material from the repository; and other activities.

Each phase also involves making decisions, for example on whether it is safe to close the facility, and these will generally be based on available data – which itself

is made findable and usable through the incorporation of high-quality metadata. The time phases are illustrated in Figure 2.2 with additional details.

Figure 2.2: Repository lifetime phases



3. Implementation of metadata

The preceding chapter outlined the nature of metadata and the many positive reasons for considering acquiring, maintaining and using them as an integral part of organisational processes. However, it is not a simple matter to implement metadata management. The scope of the metadata should be defined in relation to the goals and procedures of the radioactive waste management organisation (RWMO), and like any project, the implementation must be carefully managed. The objective of this chapter is to assist in approaching the task of implementing metadata management, which can seem complex. While there is no infallible recipe for success when implementing metadata (or any other system for that matter) within a RWMO, an understanding of the principles and practicalities outlined below should contribute to success in the implementation effort.

This chapter begins with a discussion of metadata policies, which are highly desirable at an organisational level for driving and constraining metadata implementation. This discussion is followed by a survey of the phases of metadata implementation, starting with planning and concluding with ongoing maintenance. Finally, some common problems with metadata are listed, with advice on how to avoid them. The intention is to provide helpful guidance on issues that arise in metadata implementation for RWMOs, based on the experience and informed discussions of the RepMet contributors. It is not intended to offer a comprehensive step-by-step procedure for implementation as there is too much diversity in the activities of RWMOs for this to be possible.

3.1 Metadata policies

RWMOs publish many policies and procedures in order to meet legislative, internal and other requirements. These policies cover numerous aspects of radioactive waste management, daily operations, safety, research, records management and other activities. The creation of quality metadata is an important activity that is vital to the production of a safe disposal facility, and metadata creation should be considered as one of the RWMOs core activities in the domain of records, knowledge and memory keeping. Adequate planning and resources must be devoted to this ongoing, mission-critical activity. For this reason a metadata policy should be put in place.

The metadata policy should cover the following areas.

- **Issuance and review dates** for the metadata policy.
- **Purpose of the metadata policy:** a declaration of the expected benefits of the use of metadata in the RWMOs operations and procedures. The motivating drivers include: enhanced retrieval of records, long-term access to, and the use of, operational and scientific data; compliance with statutory requirements; as well as giving assurance that the organisation takes seriously the issues of long-term access and use of its resources.

Simple example: The policy could explain that the value of metadata is recognised for supporting the present and future needs of users in finding, accessing and using the organisation's data, thereby helping to preserve the investment in data creation while assisting in complying with legislative and procedural requirements.

- **Scope of the metadata policy:** a statement of the high-level operations and procedures of the RWMO that are to be enhanced by use of metadata.

Simple example: The policy could define the scope of all processes concerned with the lifecycle of radioactive waste packages and state that the needs and opportunities for enhancement of the processes with metadata should be considered when designing or implementing the processes. This kind of definition makes it clear that other organisational processes, such as human resources, are not within the scope of the policy.

- **Requirements on metadata:** within the defined scope, a statement that metadata must be adequate to support the intended purposes. The base data/objects to be augmented with metadata should be identified. Metadata content and format will depend upon the type of data or information product (such as a publication or report) being described, but should be developed in a coherent way that is traceable back to the original purpose and scope.

Simple example: The policy could state that metadata must be designed to support data access and use across the lifecycle of radioactive waste packages, taking into account the needs of future users and the supplementary data or information to support those needs. In addition, appropriate international standards and best practices in other RWMOs should be considered.

- **Metadata creation and maintenance:** There must be procedures for capturing metadata that satisfy the above requirements. Metadata must be updated to reflect changes, which might arise from changes in the base data/object itself as it passes through its lifecycle, or from changes in the knowledge base of the designated community. There may also be a need for an approval procedure.

Simple example: It is generally accepted that metadata are best captured at the point at which they are generated, rather than being added retrospectively, and the policy could make this explicit. The policy could, for example, require automatic metadata capture from instruments (recording location with GPS-enabled equipment). It could state that quality checking must be done and who is qualified to carry it out, including both checking for compliance with metadata standards

(using a recommended metadata validation tool when available) and performing quality checks prior to use or permanent storage.

- **Responsibilities:** The policy should make clear that all RWMO employees, contractors engaged in data collection, operations, research and development activities, and in activities related to review, approval, and release of information products, are responsible for complying with the metadata policy. Lower level work instructions would then need to specify who was accountable for the population of metadata in each work area to comply with this policy. Responsibilities for metadata quality and for monitoring changing requirements should be defined.

The above themes should follow one from another, but need to be individually considered in terms of the RWMO's organisational and professional responsibilities.

The investment in time and effort to prepare and implement a policy regulating metadata management practices will give benefits in driving the organisation's implementation of metadata, and reduce the risk of incomplete implementations that might need to be revisited.

3.2 Implementation phases and issues arising

3.2.1 Planning

Implementation of metadata needs to start with careful planning and winning commitment for the project, both from management and the wider organisation. The starting point is of course the organisation's metadata policy: adoption of a metadata policy is a good way for the management of an organisation to demonstrate its commitment to metadata to all staff.

The approach to planning is dependent on the RWMO's internal culture and may be closely linked to overall strategy. In some organisations, strategic planning will be performed prior to preparing a metadata implementation plan. The strategic planning process may assess the suitability of existing systems, consider options to manage metadata centrally or individually in each system, and at the end of the process outline an agreed vision to accomplish more ambitious goals such as enterprise wide metadata management. Whether metadata implementation is specific to one information system or is enterprise wide, metadata should be designed at an enterprise level.

A plan to implement metadata within an RWMO should generally include:

- Identification of the users, specifically the designated communities (see Section 1.3) and their needs.
- Identification of the RWMO's objectives for metadata, including priorities for users and the statutory and organisational requirements for metadata (these two points can be seen as elaborations of the purpose and scope as defined in the metadata policy);

- A plan for implementation, specifying such elements as required internal and external resources, timelines, responsibilities and strategies for gaining commitment from stakeholders.

A high-level understanding of the importance of metadata and a commitment from management are essential for the successful implementation of a metadata strategy. A general understanding of principles by the managers and decision makers within an RWMO will make it easier to create adequate, consistent and appropriate metadata across all stakeholders (including internal users, the general public and expert researchers).

Once the plan receives the required commitment from management and the organisation, selection of metadata and systems development activities can begin.

3.2.2 Selection of metadata

Deciding which sets of metadata are important is a complex task. A basic rule of thumb is that metadata can be considered sufficient if they enable the finding and understanding of the base data (see Section 2 for definitions of these terms) and make the data available for use by designated communities.

The following questions provide guidance in selecting or developing appropriate sets of metadata:

- Who will be the users of the base data, and what are their present and possible future needs to access and use that base data?
- Which metadata are required to support the organisation in dealing with change over time, given that tacit knowledge might be lost?
- Which metadata are required to provide evidence of adequate quality control? Modern instruments can often record metadata along with the data, for example a data recorder used in the field can now automatically save data with GPS co-ordinates, a time stamp and the identity of the operator. Whenever possible, automatic processes should be introduced to validate data and metadata entry, these automatic processes can be supplemented by a manual approval at the end of the process.
- How will the stored information be accessed? Considering how metadata will be represented in the user interface of digital information systems, for example in a search screen, may help guide metadata requirements.

Once a set of metadata has been selected, resources required for metadata creation (both internal and external) should be assessed and allocated in order to create metadata that are accurate, consistent, sufficient and thus reliable.

3.2.3 System development

Whether the RWMO is planning to implement a new system or modify an existing one, system development activities will most likely be required. The steps of system development depend on the chosen development methodology. Traditional waterfall methods or more modern agile methods may be chosen, depending on

the practices of the organisation. It is not the purpose of RepMet to recommend a particular development method. However, whatever method is chosen, there will be some aspects of development activities that relate specifically to metadata and that should be considered in all cases. These aspects are listed below, with some key questions that should be raised during the development process.

- **Existing metadata:** Most RWMOs do not need to start their developments from scratch, as it is likely that there will be some existing metadata that are already captured and stored and can be reviewed and potentially adopted, for example existing data dictionaries. These should be assessed and may be incorporated into the metadata models.
- **Metadata modelling:** The selection of metadata has already been discussed, but this forms only part of the process of introducing metadata into system development. The metadata must form a coherent representation of the properties to be represented, and this is the purpose of a *metadata model*. The models will serve as the basis for population and querying of the metadata to meet the needs of the information system. The RepMet Libraries include such models at a high-level, tailored to the needs of RWMOs, and intended to be adaptable to many contexts and requirements. If modelling is to be done, it is necessary to agree a method to direct the modelling effort, choose tools to support the creation and documentation of models, and describe the data structures, and relationships between these. This should take into account business processes and seek to identify relevant standards to apply (see Box 3.1 below).
- **Legacy data:** Are there existing databases that must be imported into the system, possibly with data cleansing, and augmented retrospectively with metadata?
- **External dependencies:** Have possible third parties with whom metadata may need to be interoperable been identified and possible interoperability issues discussed?
- **User interfaces:** Which groups of users will require access to the system and for what purposes? For example, to carry out the entry of data or metadata, support validation and quality control, or search based on metadata.
- **Interfaces to other information systems:** Is there a need for interfaces to existing information systems in the organisation? For example, to automatically capture certain kinds of metadata or to feed metadata into records management systems?

Underlying many of these aspects of system development is the possible use of metadata standards. Metadata go hand-in-hand with standardisation. Although a simple tagging mechanism, without any constraints on the tags, can be considered a kind of metadata, in practice there is almost always a need for a more standardised approach to ensure data integrity.

Selecting metadata standards for application at an RWMO can be a confusing experience. Numerous international standards exist to help define metadata and

to operate with them; these provide guidance on what to store, define terminology (dictionaries, glossaries) and help in facilitating the exchange of information. Selecting an appropriate standard depends on the application field and may be guided or prescribed by legislation, directives or international practice. Selecting a standard that meets the requirements of an RWMO and its regulatory bodies is made even more difficult as standards are often inter-related, either linking to, or being based upon, other standards. Box 3.1 outlines RepMet's conclusions about metadata standards.

Box 3.1: Metadata standards in RepMet

RepMet has considered INSPIRE (INSPIRE, 2017), an EU Directive for European spatial information, that is based on a number of standards and prescribes metadata requirements for EU member states primarily relating to spatial information. However, INSPIRE can be extended and applied to the wider domain of radioactive waste management. Relevant standards are the Open Geospatial Consortium (OGC) and international standard (ISO) Observations and Measurements (O&M) (ISO, 2011) which defines a conceptual schema encoding for observations, and Minnesota Recordkeeping Metadata Standard (MRMS) (Minnesota State Archives, 2015); originally developed for records management in government. These have been adopted by RepMet due to their usability and positive feedback from RWMOs already applying them.

Recommended standards to be applied when capturing processes that may reach across multiple libraries are also discussed in the supporting documents. For more detailed information on standards see the "RepMet Tools and Guidelines" document.

The standards recommended by RepMet are a product of scientific debate and are based on professional considerations with a strong influence by INSPIRE. The needs of individual RWMO's may of course differ and these are not intended to be prescriptive.

When selecting metadata standards, a number of issues must be considered:

- The possible applicability of existing metadata standards.
- The completeness of the metadata standard: will it be sufficient for the intended purposes?
- The consistency of the metadata standard, at different levels, ranging from the consistent use of controlled dictionary terms to the logical integrity of the metadata model
- The relationship to business processes within the RWMO: is there a best point at which to capture a particular item of metadata?
- The standards identified as obligatory either by regulators or statutory requirements.
- The underlying perspective of the standard that is defined (is it a part of a wider family of standards or was the standard developed for one particular application?), and whether it meets the requirements of the RWMO
- Whether all parts of the standard are applicable to the RWMO or whether there are parts not relevant to the RWMO.

RepMet has provided comprehensive guidelines (see "RepMet Tools and Guidelines"). However, there is no "one-size-fits-all" metadata or data content standard suitable for all RWMOs. RepMet provides examples that are suitable for most RWMO's needs, however RWMOs must carefully develop a suite of metadata

models that are appropriate to their particular circumstances. These should utilise existing metadata standards, supported by appropriate localisation and internal guidelines.

3.2.4 Processes in the organisation

Considering the repository's long lifecycle, metadata creation should be viewed as an incremental process – though built on firm foundations in the policy and planning stages – and should be a shared responsibility across staff with responsibilities for the repository. No system can perform efficiently without proper organisational processes and procedures in place. Inefficiencies in metadata, poor-quality metadata and negative effects on metadata creation and workflow can be avoided by establishing and enforcing processes and procedures in all the participating organisational units throughout an RWMO, and by introducing accountability for metadata creation and quality.

Capturing metadata at the time of creation generally costs less than retrospectively adding metadata at a later time. Therefore, as a general principle, procedures should be put in place to ensure that metadata are captured alongside the corresponding data capture activities.

RWMOs should make an effort to automate or otherwise assist in metadata production and replace manual methods of metadata creation wherever possible and appropriate. Modern equipment can often be set to capture such metadata automatically to improve their reliability (for example by recording GPS coordinates as a disposal site is characterised and constructed). Where not automatable, RWMOs should give thought to how to make the process as reliable and robust as possible. The use of tools and systems which carry out automatic data validation, for example by making use of templates or controlled dictionaries in data entry forms, should be considered and implemented where appropriate.

RWMOs should make the creation of metadata a routine part of their workflows. Creation of consistent, standards-based, continuously refreshed and updated metadata enables RWMOs to handle information about their disposed radioactive waste and activities in a timely and efficient manner.

Collection of metadata during the period after the collection of data is sometimes necessary if an RWMO discovers that important metadata were not recorded at the time of data acquisition, but this should be avoided where possible.

3.2.5 Managing metadata quality

Data can become useless if it is associated with poor-quality metadata or, worse, no metadata at all. Processes and procedures should be employed to ensure that data and metadata are properly collected, handled, processed, used and maintained at all stages of the data lifecycle.

Manual entry by untrained personnel who are not familiar with the description of data by metadata or controlled dictionaries can contribute problems of quality associated with metadata. Metadata content terminology may be inconsistent, making it difficult to locate relevant information.

Metadata quality can briefly be described through the following main requirements:

- Completeness: all relevant metadata fields should be populated.
- Sufficiency and relevance: metadata fields should provide enough information to satisfy the needs of the designated communities.
- Consistency: there should be no contradictions, and the same information should not be captured in multiple places.
- Accuracy and timeliness: metadata should reflect reality and be up-to-date.
- Compliance with prescribed standards: mandatory elements prescribed by the metadata standard definitions must be present.

Metadata quality is not a once-and-for-all attribute. The accuracy of metadata may decline over time, as new knowledge becomes available, and furthermore new statutory or organisational requirements often arise that put new requirements on what metadata must be able to support.

3.2.6 User training and documentation

Staff competencies and their knowledge of the developed system's capabilities and requirements are essential for the successful implementation of a metadata plan. An adequate number of appropriately trained staff with a variety of expertise and skill sets is necessary for the successful implementation of a metadata plan across an RWMO. This includes: subject expertise; information management experience; technical knowledge; research skills; knowledge of intellectual rights issues; and data protection regulations.

Steps should be taken to retain this knowledge and keep it up to date with possible changes in staff and new requirements being introduced. A training plan should be in place for staff involved in creating and maintaining metadata.

The modified or newly developed system should be well documented, usually requiring not only the documentation for the system development, but also user manuals that provide knowledge at the user level, and operations manuals that enable the people in charge of IT operations to properly manage and maintain the system, including setting access rights, making backups and similar.

3.2.7 Testing

An activity often neglected during system development is testing. In the case of metadata implementation, testing is a twofold activity: the functionality of the system needs to be tested, and the metadata's appropriateness in meeting requirements also requires testing.

RWMO metadata and data records need to remain usable over long periods. Reviewing the associated metadata for completeness and accuracy is a good practice. Having a reviewer verify the metadata, ideally a person unfamiliar with the implementation project, contributes to the objective review of the metadata.

3.3 Common problems with metadata

There are many pitfalls in implementing metadata management, often resulting from a failure to recognise the changes that might happen over time or the need for completeness and precision in defining metadata for particular purposes.

The following is a list of common problems with metadata:

- Classification and naming systems change with time. For example, waste categories may change with legislation. In this situation, keeping references to older classifications is likely to prove to be useful.
- Geospatial co-ordinates without information about the reference coordinate system make the data unusable in future.
- References to measurement units may be missing or ambiguous. For example, a legacy database may describe the measurement simply as “Activity”. Without adequate metadata it cannot be known if the “Activity” is measured in Bq (Becquerel) or Ci (Curie).
- Metadata owners may not be data owners, leading to a divided responsibility and accountability. As collaborations grow, and as data are made more accessible, it is quite possible to have shared responsibility for metadata and the data with which they are associated, but careful management is needed within the organisation. The accountability should however always be clear.
- Metadata may not be sufficient to support the intended use. Sufficient metadata should support the needs of the designated communities. Deciding what level of metadata is sufficient will depend on the user; organisations are advised to consider which metadata elements will provide the most relevant information for the broadest user base.
- Metadata may not be complete. Completeness of metadata means that all relevant characteristics of the base object are captured (as far as practically and economically feasible and necessary for the application). Completeness can be enhanced through the use of automated systems, manual or automated data validation checks, processes and procedures; and most importantly that the organisation must provide staff with sufficient time to populate metadata while promoting an appropriate culture. Imposing extra effort on staff without providing time or rationale might lead to resistance and poor-quality metadata.
- Poor naming conventions and insufficient keywords or inadequately controlled dictionaries may hamper the use of metadata.

A further consideration, that might lead to serious problems if not anticipated, is data protection – that is, correctly handling data about identifiable persons. At the time of the writing of this report, there were a number of regulations that RWMOs needed to adhere to, primarily when dealing with personal data. Examples include the European General Data Protection Regulation, while the United States also has a number of federal and state level acts on the protection of personal data.

Other countries have their relevant national regulations as well. Such regulations often include a “right to know” clause, implying a need for metadata to enable the location of records that contain such data, and may include the “right to be forgotten” clause, implying a need to be able to completely erase personal information. There can be additional requirements preventing personal information from being retained for longer than the prescribed period without permission from the individual.

Attention should be placed on the ability of the recorded metadata to enable compliance with such regulations to ensure legal requirements for data protection are met. For example, an RWMO may record data related to doses received by personnel; these data are often regulated by data protection legislation as being health related with the need to be processed in a special manner.

Considerations related to the protection of personal data include:

- **Data categorisation:** what categories will aid processing (e.g. identifying personal data, data concerning health or similar)?
- **Confidentiality:** who is authorised to access or process the data?
- **Storage:** how long are the data required to be stored?
- **Disposal:** when are the data required to be deleted?
- **Right to rectification and erasure:** the right of the person (for example, a former employee) to demand that his or her records be rectified or deleted if the RWMO has no further statutory reasons to store the data. This requirement can be particularly difficult to implement in archiving and backup systems, if the design does not consider this requirement from the outset.

Data protection is a complex subject that is beyond the scope of RepMet, but each RWMO should ensure it complies with the relevant legislative requirements. This means that data protection should be considered when individual the RWMO's metadata requirements are analysed.

4. Using RepMet for metadata implementation

Chapter 3 outlined the main aspects of implementing metadata management within a radioactive waste management organisation (RWMO), with advice on the issues that arise. The RepMet initiative has produced several deliverables, including the present report, to provide assistance with aspects of implementation. These deliverables are intended to be useful resources for all RWMOs, without being prescriptive, since the various national RWMOs are governed by different legislation and have individual waste management programmes.

The RepMet deliverables were introduced in Section 1.2 and are summarised in Table 4.1.

Table 4.1: Summary of RepMet deliverables

| Deliverable | Key contents |
|---|---|
| Report: <i>Metadata for Radioactive Waste Management</i> (the present report) | Introduction to metadata and its benefits; advice on policies, planning and implementation within a RWMO; and recommendations for RWMOs to follow. |
| RepMet Libraries | Reusable libraries of data models and metadata models describing key entities in radioactive waste management and related domains. There are currently three libraries, each contained within its own report: Report: "Waste Package Library" Report: "Repository Library" Report: "Site Characterisation Library" The three reports are supplemented by RepMet web products, which provide engaging interactive web-based illustrations and descriptions of the libraries. |
| Report: "RepMet Tools and Guidelines" | Description of tools used to underpin and support the development of the RepMet Libraries, with guidelines on their use. |

4.1 Using the RepMet deliverables

The RepMet deliverables can be applied as part of a metadata implementation project and as part of ongoing activities. Table 4.2 maps the RepMet deliverables to the headings of Chapter 3, indicating specifically how they may be applied in those particular contexts.

Reference is also made in the table to the RepMet recommendations, which are presented in Chapter 5 and which, if adopted, will have a positive impact on the indicated areas.

Table 4.2: Mapping RepMet deliverables to implementation areas

| Area of metadata implementation | Metadata for Radioactive Waste Management report and its recommendations | | RepMet Libraries | "RepMet Tools and Guidelines" report |
|---|---|-----------------|---|---|
| <i>Policies</i> | Suggested coverage of metadata policies (3.1). | R1 | | |
| <i>Planning</i> | Identification of issues arising in all areas with advice (3.2). Proposed designated communities relevant to RWMOs (Annex B). | R2, R3, R4, R7 | Availability of libraries minimises the need for "starting from scratch" and will reduce resource requirements. | Availability of tools minimises the need for "starting from scratch" and will reduce resource requirements. |
| <i>Metadata selection</i> | | R3 | Libraries as starting points for the selection of metadata, and provide generic models intended to be adapted and extended as needed as part of the system development process. | Metadata standards validated in the RWMO context and ready for use. |
| <i>System development (including standards)</i> | | R6, R8, R9, R10 | | Data/metadata modelling techniques ready for use in adapting and implementing metadata based on libraries. Controlled dictionaries ready for adoption or possible modification. |
| <i>Organisational processes</i> | | R5, R6 | | |
| <i>Metadata quality</i> | | R6, R8 | | Controlled dictionaries available for user entry of metadata to encourage appropriate terms and quality of metadata. |
| <i>User training and documentation</i> | | R4 | | |
| <i>Testing</i> | | R3, R4 | | |

4.2 Technical bases of the RepMet deliverables

Before summarising the RepMet deliverables and how they may be used, some of the technical bases on which they are built should be described.

Data models are logically structured collections of relationships between types of data, metadata and the real-world objects that they represent. A data model defines a set of logical entities, their properties (alternatively called “attributes”) and the relationships between them, with a view to representing properties of interest (for example, about persons, organisations or waste packages) so that they can be captured and processed by computer-based systems. Data modelling is a standard technique in information systems development and is supported by many software tools.

In the remainder of this report, the term *data modelling* will be taken to include *metadata modelling*. As already noted, the data/metadata distinction is not clear-cut, and the modelling techniques are exactly the same.

Controlled dictionaries are collections of terms in a particular field that are defined, used, managed and maintained by a community or organisation in a controlled way. Often they are based on widely accepted international standards. Controlled dictionaries are useful for contributing to the development of uniform content in information systems; on the user side, they support queries and understanding. All terms in a controlled dictionary have unambiguous definitions. Within RepMet, such dictionaries are used to provide definitions for each entity on a data model and all the properties within those entities. Properties have also been organised into a hierarchy for each entity to provide further structure.

Controlled dictionaries allow communities and organisations to agree and use a common and well-defined terminology, without risk of ambiguity or misunderstanding. The entries in the dictionary typically include definitions and their sources, and may be multilingual, allowing for language translation. The entries might also be related to each other, as in RepMet, so that they start to become data models themselves, though with a specific focus on terminology.

4.3 RepMet Libraries

The RepMet Libraries are data models related to the main domains involved in the safety case for a radioactive waste repository, and their development was a core activity of the RepMet initiative. Each library is a directly reusable resource that may be adopted by RWMOS in the development of their own metadata-based information systems; in fact, the libraries can be seen as the core metadata models that were the original vision of RepMet (see Figure 4.1). The libraries describe real world or abstract entities of interest (such as “wasteform”, “spent nuclear form” or “disposal module”) with relevant attributes (such as “mass”, “content of anions” or “total alpha activity” of waste) in the form of controlled dictionaries.

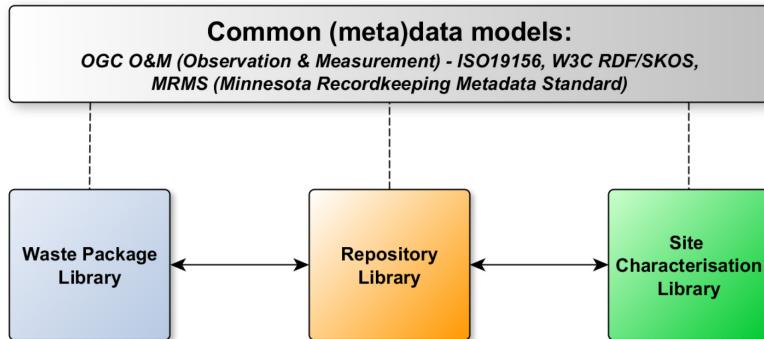
Figure 4.1: Overview of RepMet Libraries

Table 4.3 shows the topics selected for each library from the three domains.

Table 4.3: Topics of the RepMet Libraries

| Domain | RepMet Libraries | Topics |
|------------------------|-------------------------------|---|
| Waste packaging | Waste Package Library | Packaged waste and spent nuclear fuel ready for final disposal at the repository. |
| Engineering | Repository Library | Repository requirements and structure at closure. |
| Geoscience | Site Characterisation Library | Geological and geophysical characterisation of the repository site. |

Using the tools, techniques and resources described in the “RepMet Tools and Guidelines” report, the RepMet Libraries represent a common understanding between a range of organisations, giving assurance that the results are of wide applicability. These libraries are intended to be adopted and modified or extended as needed by RWMOs in the development of metadata-based information systems.

The “RepMet Web Products” are a web-based presentation and description of the RepMet Libraries, accessible from the RepMet initiative webpage³. They provide an attractive and immediate way to present the contents of the three RepMet Libraries, including conceptual data models and controlled dictionaries, together with introductory text. Users can navigate the web products to explore and understand the connections between different RepMet Libraries and their contents,

³ www.oecd-nea.org/rwm/igsc/repmet/

thereby providing a more integrated presentation. Furthermore, the user can more easily understand how the metadata models of the Open Geospatial Consortium (OGC) and international standard (ISO) Observation and Measurements (O&M)⁴, and the Minnesota Recordkeeping Metadata Standard (MRMS) standards can be applied to support each library. Controlled dictionaries are also provided in the Simple Knowledge Organisation System (SKOS) form via the “RepMet Web Products”.

4.3.1 Waste Package Library

The “Waste Package Library” presents a data model for the description of the packaged waste and spent nuclear fuel that are ready for disposal in a final repository. The library includes several examples of the application of the data model to real waste packages and disposal systems from the involved RWMOs.

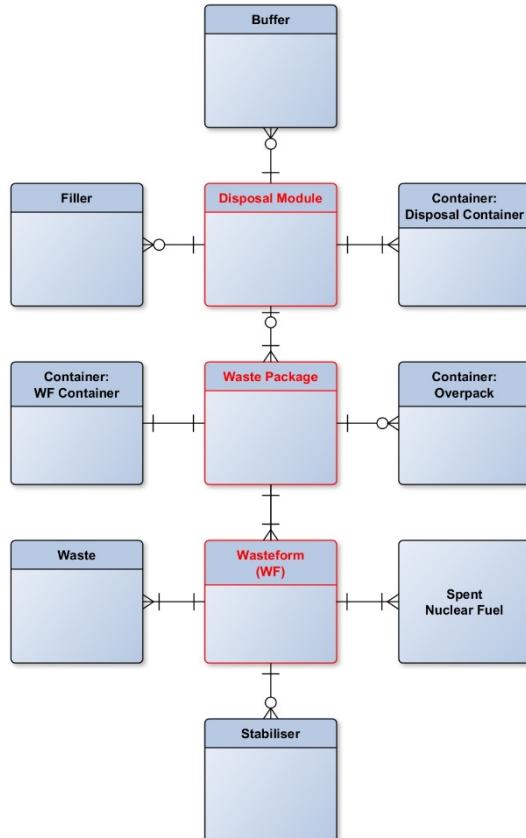
The top level of the “Waste Package Library” is shown in a simplified form in Figure 4.2. Any waste package can be described using combinations of the elements on the diagram and each of these elements can, in turn, be described using attributes defined for that element within the library (for the waste this would include, for example, activity and mass). All information may be formally encoded in a standardised form using the O&M standard. Justifications for needing to record each attribute have also been provided. In the figure the symbols on the lines joining elements indicate the possible relationships between these elements, for example, any given wasteform may comprise one or more (\leftarrow) types of waste or spent nuclear fuel, optionally (\bowtie) with stabilisers. Each element of the data model, together with all attributes associated with the element, are defined within the controlled dictionaries (see Section 4.4.2) produced by RepMet as part of the library. As this is a completely general model, it is applicable to the description of any waste package and may be adopted by any RWMO for their own use or to facilitate information exchange.

4.3.2 Site Characterisation Library

The “Site Characterisation Library” provides a data model that supports the description of geoscientific information for general radioactive waste disposal sites. It is based on open source international standards such as the Geoscience Markup Language (GeoSciML) and the OGC O&M that are fundamental building blocks of the European Spatial Infrastructure (INSPIRE).

As for the “Waste Package Library”, the top level of the model identifies key aspects to be represented, in this case geophysics, geology and environmental monitoring. The geophysical part provides entities for measurements and geophysical models. The geological part contains common features like geological units, tectonic elements, contacts that appear on geological maps or in 3D models. Boreholes and specimens are also represented in the model. The environmental monitoring part has elements to describe monitoring facilities and official monitoring programmes.

⁴ ISO 19156, Geographic Information – Observations and Measurements

Figure 4.2: Top-level data model of the Waste Package Library

4.3.3 Repository Library

The “Repository Library” presents a data model for radioactive waste repositories. Facilities for low, intermediate and high-level radioactive wastes, as well as commercial spent nuclear fuel, are all considered. This library is intended to describe the repository and its acceptance criteria, based on the practices currently in use internationally. As for the “Waste Package Library”, each data model entity is associated with a set of properties in the form of controlled dictionaries.

4.4 RepMet Tools and Guidelines

The team working on the RepMet initiative used a number of techniques and resources in their work, based on the experience of RepMet members and external advice. These techniques and resources provided the technical basis on which the RepMet Libraries described above were developed and expressed. Having been

successfully used for this purpose, they can now be recommended to all RWMOs as valuable support for their own work in adopting and implementing the results of RepMet in their own organisations.

The “RepMet Tools and Guidelines” report describes and explains the methods and techniques in detail, but a concise summary is given here of the three components:

- data modelling
- controlled dictionaries
- the use of existing standards.

The following sections provide a high-level summary of the contents of the “RepMet Tools and Guidelines” report, and highlights the possible areas of applicability during the metadata implementation process within a RWMO. The dedicated sections in the “RepMet Tools and Guidelines” report provide further details.

4.4.1 Data modelling

As noted in Section 4.2, data modelling is a standard technique that was adopted by RepMet to produce a coherent representation of the domains of interest, with the resulting models then serving as the basis for information systems. It should be noted that in the spirit of “one person’s data may be another’s metadata”, no distinction is made between *data modelling* and *metadata modelling* as the same techniques may be used to express either.

In RepMet data modelling was used for a number of reasons:

- To provide a framework for expressing a common shared understanding of key concepts in the field of radioactive waste management (RWM) by participating RWMOs and the relationships between these. In the development of the “Waste Package Library”, members from different countries understood and interpreted waste package terminology differently. Entities and relationships in a data model strictly depend on the formal definitions: this was the first step towards the development of a RepMet controlled vocabulary. Initially, RepMet based its definitions on the International Atomic Energy Agency (IAEA) Safety Glossary (IAEA, 2007), with individual definitions slightly modified in order to ensure self-consistency within the data model and to add flexibility for the diverse range of waste packaging solutions utilised by RWMOs.
- To allow the linkages between data and metadata to be clearly expressed, through the relationships between entities and properties. For example, every experimentally measured quantity may be associated with a set of metadata related to the observation and measurement process using a standardised approach.

RepMet used conceptual data modelling, and specifically Entity-Relationship Diagrams, to develop the scientific and technical content of the libraries, ensuring that they are well defined and can be easily customised and implemented by a

RWMO within its own information systems. The entities considered include “real-world objects” such as a “wasteform”, “container” or “barrier”, but also “abstract objects” such as “facility information” or “repository monitoring”. Furthermore, entities can be composites of other entities: for example, a waste package is a composite of a wasteform and one or more containers. The properties of entities are the pieces of information describing the entity (the “density” of a “wasteform”, or the “overall diameter” of a “container”) that have to be included in the data system for whatever purposes, for example safety case development. Furthermore, each entity may have one or more logical relationships with other entities; for example, there may be restrictions on which and how many “overpacks” can be used with a given waste package type.

The same modelling techniques and tools may be applied by RWMOs in developing their own data and information systems in a format which may be readily understood by IT professionals.

4.4.2 Controlled dictionaries

RepMet initially developed a simple glossary of terms, starting with definitions for the entities of each data model (for example, a waste encapsulant), but in time it became clear that a more structured approach was beneficial, leading to the adoption of formal controlled dictionaries, which were subsequently developed for all three RepMet libraries. Formal definitions such as these are essential for the development of an effective data management system and a robust safety case.

Controlled dictionaries can be reported in two different ways:

- a visual format using tabular formats with the support of mind-maps – hierarchical diagrams showing relationships between terms
- a technical format using the SKOS standard that the World Wide Web Consortium (W3C) recommends.

In the course of the RepMet initiative, the visual format was the first to be implemented and was found to be an effective format for visualising library properties, especially for non-IT staff.

As RepMet developed, it became clear that a technical format for expressing controlled dictionaries was also required. For this technical format, RepMet adhered to the W3C recommendation by adopting the Resource Description Framework (RDF) (W3C, 2014). This is a general method for conceptual description and modelling of information implemented in web resources. RDF is highly general, so RepMet adopted the SKOS, a specialisation of RDF, to develop its own controlled dictionaries. The mind-maps developed previously provided a pictorial basis for visualising controlled dictionaries expressed using SKOS.

Given the different audiences, RepMet has produced its controlled dictionaries using both the visual and technical formats described above.

RWMOs may use the controlled dictionaries developed in RepMet and relate them to their own terminology where it differs, thereby giving a common

vocabulary which allows for the benefits of harmonisation across organisations and guaranteed consistent usage of terms.

4.4.3 Standards

There are two established standards – not specific to the RWM field – that were examined and used by the RepMet initiative. They provide ready-made ways of representing important aspects of data and metadata relating to RWM.

4.4.3.1 Observations and Measurements (O&M)

O&M is an OGC and ISO standard (ISO 19156) which defines a conceptual schema for encoding observations, and for describing features involved when making observations. Although the O&M standard was developed in the context of geographic information systems, the model is intended to be generic and can easily be used for managing information from different fields, such as RWM.

The O&M standard provides a simple and generic framework for structuring information resulting from observations. The main benefit of its use is that instead of requiring multiple different models for different kinds of observations, one single model may be applied for all. It thereby establishes a wide ground for interoperability between different information systems, and makes database design much easier. The diversity of the “real world” is mapped through controlled dictionaries defined by the O&M data model so that adding new disciplines or types of observations to the system only involves updating these dictionaries.

The RepMet initiative has adopted the O&M standard and applied it to the RWM field. It does this by relating the standard’s five top-level concepts (“observer”, “feature of interest”, “observed property”, “process” and “result”) to RWM and providing examples of how the standard can be used when representing observations of interest (for example, physical, chemical and radiological properties of structural waste from spent nuclear fuel reprocessing). These are incorporated into the RepMet libraries and are available for reuse by RWMOs.

4.4.3.2 Describing records

RWMOs create large quantities of documentary material (records) throughout their lifetime including contractual agreements, reports and technical documentation. Publications are produced for different audiences and serve different purposes. Controlling and monitoring the life cycle of records is an important activity that is called recordkeeping. Recordkeeping at governmental or organisational level is a special issue having its own set of standards. One of these is the Minnesota Recordkeeping Metadata Standard (MRMS).

The MRMS was examined, adopted and modified by the RepMet initiative to respond to the need for a recordkeeping approach applicable to the volume and diversity of records from waste generation to disposal. A slight modification of MRMS is recommended in the “RepMet Tools and Guidelines” document. The elements of the original standard have been separated into two main parts, record and resource:

- Record has all the attributes that are required for official recordkeeping (for example, agent, right management, date, type and management history).
- Resource contains the attributes required for access (identifier, format, location).

With the above model, the linking of MRMS and O&M is possible, so information from two different domains can be handled in one unified data model. In other words, the same resource can be used as the object of recordkeeping, and at the same time be regarded as the result of an observation.

RWMOs may adopt the modified model for their own purposes, thereby avoiding the need for duplicating effort in this important area.

5. Closing remarks and recommendations

5.1 Summary of RepMet

The benefits of radioactive waste management organisations (RWMOs) using metadata are numerous and make a compelling case for the introduction or extension of metadata management both within and across organisations. Metadata enable RWMOs to manage their data and information in a well-organised manner, meeting statutory requirements and ensuring that data quality is in line with requirements. Metadata are a basis for maintaining the long-term value of an organisation's digital assets, helping it to remain effective and support future decisions and operation efforts, and to meet the requirements of designated communities now and in the future. Metadata have a role to play at all stages of the lifecycle of a radioactive waste repository.

RWMOs participating in the Nuclear Energy Agency (NEA) RepMet initiative have worked together to provide advice and establish a set of common libraries for RWMOs, which can help to map their own systems for these deliverables.

RepMet has formulated a consistent set of guiding principles for metadata management relating to radioactive waste repositories. The results will be useful to operators, regulators and other stakeholders. The RepMet Libraries and associated documents have been prepared with the intent to provide generic models, processes and descriptions that can be tailored to the needs of virtually any RWMO in order to develop metadata. The "RepMet Tools and Guidelines" can be adopted by RWMOs in their implementation of metadata and systems using the libraries.

RepMet recognises that each RWMO's activities are unique in many respects, and that each radioactive waste site has characteristics that are exclusive to that particular site. However, there are also many common and general aspects of radioactive waste management and final disposal. Each RWMO would need to adapt the provided models, processes and descriptions to ensure that the constituents meet the requirements of local regulations, the RWMO's individual needs, and the individual characteristics and applied technology of the individual radioactive waste management operations.

The joint development effort of RepMet is an attempt to ease the burden on each RWMO and a move towards interoperability and international harmonisation. A shared set of principles, controlled dictionaries and data model libraries facilitates data exchange with common stakeholders such as international peer review groups, NGOs, collaborating regulators, local communities and others. This approach will allow less mature programmes to benefit from the advances made by other organisations and for all organisations to learn from the experience of others.

Adoption of RepMet's results will contribute to making an RWMO's data and information management efforts more successful and cost efficient. The RWMO will see a return of its financial and time investment through the better management of data and information, while reducing the risks of information loss and the need to recreate data, as well as the consequent reputational and financial costs.

5.2 Recommendations

The following ten recommendations are being proposed by RepMet for the use of metadata by any RWMO. The recommendations follow from foregoing discussions, and distil the most important aspects of discussions into a concise form, presented here with references to the related sections of this report.

- R1. Establish a comprehensive metadata policy that covers the aspects listed in this report and the individual organisation's statutory and other requirements (Section 3.1).
- R2. Analyse the potential needs for, and benefits of, metadata in the organisation's processes, considering long-term access and use of information and the risks of loss of information over time. The benefits should be balanced against the costs of implementation to assess the overall value of implementation (Sections 2.2, 2.3, 3.2.1).
- R3. Define designated communities who will make use of the information, and analyse the consequent needs for metadata to support their use (Sections 2.3, 3.2.1).
- R4. Dedicate sufficient resources to the formulation of requirements, to the planning of metadata implementation, and to ongoing and supporting activities such as metadata maintenance, testing and user training. A need for training may arise to understand the importance of metadata and the procedures for capturing them and verifying their quality (Chapter 3).
- R5. Establish processes that will instil a culture of high-quality metadata creation and maintenance during the entire lifecycle of the relevant data. Accountability for metadata entry and quality should be clearly defined (Section 3.2.4).
- R6. Ensure, to the extent possible, that metadata are captured at the earliest opportunity, rather than being added retrospectively and, if possible, generated automatically. If metadata are entered manually, there should be clear supporting guidelines and, where possible, data validation to ensure their quality (Sections 3.2.4, 3.2.5).
- R7. Consider the opportunities and benefits of using international standards when planning for metadata implementation so as to avoid duplication of work and ensure coherence and harmonisation across RWMOs (Section 3.2.3).

- R8. Make use of controlled dictionaries as the basis for metadata, both throughout the organisation and ideally across RWMOs so as to achieve the benefits of interoperability and international harmonisation (Section 3.2.3). Controlled dictionaries should be cascaded to external organisations in the supply chain. Procedures should also be established to ensure controlled dictionaries are kept up-to-date, are made available to all users and have well-defined periods of validity.
- R9. Establish metadata implementation on the basis of RepMet's libraries, if possible (Chapter 4).
- R10. Make use of "RepMet Tools and Guidelines" (NEA, forthcoming) when implementing metadata (Chapter 5).

5.3 Future work

During the course of the RepMet initiative, work was identified that would be completed in the period beyond the initial four years of the initiative:

- Development of a data model to describe the process of safety assessment, including linkages between the safety case and underlying data and modelling activities, and the use of Features, Events and Processes (FEP) lists, as well as safety functions within the safety case.
- Creation of new RepMet Libraries for additional topics during the operational period, for example to describe plant operations or waste treatment and conditioning processes.
- Testing and further development of the current RepMet Libraries.
- Better integration of RepMet controlled dictionaries with the NEA FEP Database⁵ that illustrates the NEA International FEP List and some project-specific FEP (PFEP) Lists (i.e. the FEP list that the RWMOs involved in the NEA Integration Group for the Safety Case [IGSC] have developed for their own safety assessment studies).
- Production of metadata principles to allow improved interoperability of information systems within an RWMO (for example to allow a single query to carry out a search across multiple bespoke or commercial systems).
- Formulation of data dictionaries to support numerical modelling within RWMOs.

⁵ www.oecd-nea.org/feadb

6. References

- CCSDS (2012), Reference Model for an Open Archival Information System (OAIS), Consultative Committee for Space Data Systems, <https://public.ccsds.org/pubs/650x0m2.pdf>.

European Commission (2017), EOSC Declaration, European Commission, https://ec.europa.eu/research/openscience/pdf/eosc_declaration.pdf.

Giaretta, D. (2011), Advanced Digital Preservation, Springer, Berlin.

IAEA (2007), IAEA Safety Glossary, International Atomic Agency, www-ns.iaea.org/standards/safety-glossary.asp.

INSPIRE (2017), INSPIRE website, <https://inspire.ec.europa.eu/>.

ISO (2011), ISO 19156:2011 Geographic information – Observations and measurements, www.iso.org/standard/32574.html.

Jeffery, K. (1998), “Metadata: an overview and some issues”, 11th ERCIM Database Research Group Workshop on Metadata for Web Databases, ERCIM, www.ercim.eu/publication/ws-proceedings/11th-EDRG/jefferey.pdf.

Minnesota State Archives (2015), Minnesota Recordkeeping Metadata Standard, www.mnhs.org/preserve/records/metadatastandard.php.

NEA (forthcoming), “Site Characterisation Library”, NEA/RWM/R(2018)3, OECD Publishing, Paris.

NEA (forthcoming), “Waste Package Library”, NEA/RWM/R(2018)4, OECD Publishing, Paris.

NEA (forthcoming), “Repository Library”, NEA/RWM/R(2018)5, OECD Publishing, Paris.

NEA (forthcoming), “RepMet Tools and Guidelines”, NEA/RWM/R(2018)6, OECD Publishing, Paris.

Riley, J. (2017), Understanding Metadata, NISO, Baltimore, http://groups.niso.org/apps/group_public/download.php/17446/Understanding%20Met%E2%80%A6.

W3C (2014), Resource Description Framework (RDF), www.w3.org/RDF/

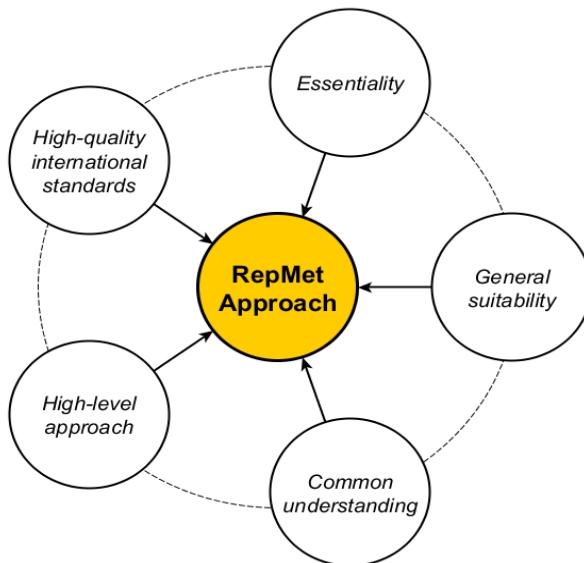
Zeng, M.L., (2016), Metadata, Facet Publishing, London.

Annex A. The RepMet approach and methodology

The main goal of RepMet was the identification of metadata libraries that Radioactive Waste Management Organisations (RWMOs) can adapt and use for managing their repository data, information and records in a way that is both suitable for long-term management and use, and harmonised internationally. This annex illustrates the approach and the methodology that RepMet adopted for the development of such libraries.

Figure A.1 shows the principles that the RepMet initiative team followed for the development of the libraries and other deliverables.

- **High-level approach:** RepMet did not consider the choice of the particular IT systems for data management and the related technology.
- **Essentiality:** The content of the libraries (for example, the information about the final waste package for the safety case) is not exhaustive or universal. The group defined minimal and essential information about the single, selected topics;
- **General suitability:** All the RWMOs, independently of the maturity of their waste management programmes, should be able to apply the results of RepMet's work in a manner appropriate to their specific circumstances and needs. National organisations just beginning their own national programmes might look at the current content as the core information that it is recommended to collect, at least in the first stages. More experienced RWMOs can extend the information according to their requirements and specifications;
- **Common understanding:** Controlled dictionaries for library contents such as entities, data, metadata and properties are at the basis of the initiative, to promote and facilitate the dialogue among the different organisations;
- Use of **international**, well-consolidated and high-quality **standards**.

Figure A.1: Principles adopted by RepMet in library production

Approach to metadata standards

The last of the above listed principles of the RepMet approach for the library development requires additional explanation. There are a large number of international and local standards relating to metadata management, and different organisations also favour, or are prescribed to use, different standards.

As a principle, RepMet has attempted to select existing standards, rather than define its own to avoid escalating this number further. Selecting the best standard from the numerous standard candidates that often overlap or are subsets of each other is complex. Different standards may focus on different themes.

The principles followed by RepMet when selecting relevant standards were the following:

- Recommend the essential minimum and do not overwhelm.
- Recommend only standards that would work for all RWMOs.
- Give preference to less complex standards if suitable.
- Favour standards proven to work and already widely used by the profession.

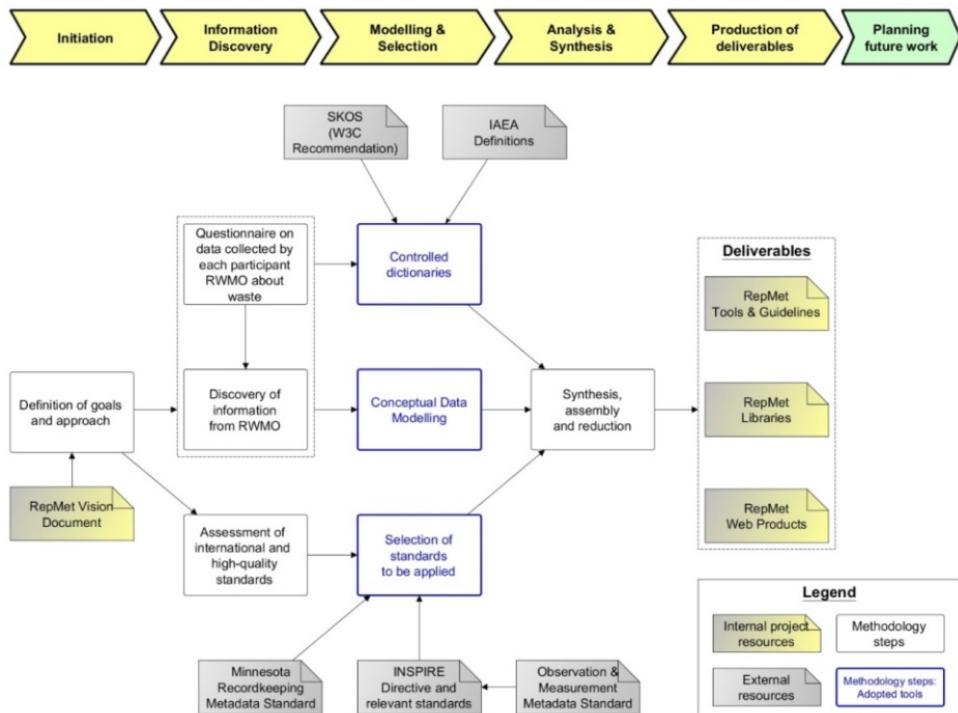
For a discussion on standards, see Section 4.4.3.

Within this general approach, Figure A.2 illustrates the methodology that RepMet followed.

The methodology started with an analysis of the information that RepMet members reported collecting to support their safety case or other reporting requirements. A questionnaire about data and information related to waste packages ready for the final disposal was also produced and issued by RepMet.

In parallel, RepMet investigated several current, high-quality, international metadata standards. As stated before, RepMet did not create a new metadata standard but selected and recommends a number of existing ones that suit RWMO's requirements and help meet the initiatives objectives.

Figure A.2: RepMet methodology



After the “Information Discovery” phase, RepMet considered different methods and techniques from within the field of data, information and knowledge management, adopting those with the potential to assist RWMOs in the long-term management of information for radioactive waste repository programmes. These include the use of tools such as data modelling and controlled dictionaries together with the selected metadata standards. These set of tools were used for the development of the libraries in the initiative.

Annex B. Designated communities

RepMet has identified stakeholders who are anticipated to be interested in the data produced as part of a radioactive waste management programme and would therefore benefit from appropriate discovery or context metadata. These stakeholders, together with the information they are likely to require and the active time period, are given below:

| Designated communities | What will they want to do? Which questions will they want answered? | Which time phases are they active? | How can RepMet help them? | Which RepMet deliverables may help them? |
|--|--|--|---|---|
| National institutions responsible for radioactive waste management – RepMet members and other Radioactive Waste Management Organisations (RWMOs) | Metadata are very important to ensure the usability and traceability of all data and information including knowledge transfer between the different generations of staff members | Pre-operational | Conception/development of data management | <ul style="list-style-type: none"> • <i>Metadata for Radioactive Waste Management</i> • Site Characterisation Library • Waste Package Library • Repository Library • RepMet Tools and Guidelines |
| | | Operational phase and Post-operational phase | Improve management of data using common definition (metadata) | <ul style="list-style-type: none"> • <i>Metadata for Radioactive Waste Management</i> • Site Characterisation Library • Waste Package Library • Repository Library • RepMet Tools and Guidelines |

| Designated communities | What will they want to do? Which questions will they want answered? | Which time phases are they active? | How can RepMet help them? | Which RepMet deliverables may help them? |
|---|---|--|--|---|
| Nuclear facilities' operators – national institutions responsible for radioactive waste management – RepMet members and other RWMOs | The operators have to deliver data and information on the existing and future waste to the RWMOs until the power plants has been dismantled | Pre-operational and operational phase (up to the waste emplacement) | Help understanding (and use) of data, for national/international assessment (peer review) | <ul style="list-style-type: none"> • <i>Metadata for Radioactive Waste Management</i> • Site Characterisation Library • Waste Package Library • Repository Library • RepMet Tools and Guidelines |
| Nuclear facilities' operators. Stakeholders interested in projects' development from a nuclear programme/RWM perspective | Require a clear timeline of repository's loading programme, to be continuously updated and communicated | Pre-operational and operational phase (as long as waste is disposed) | Help understanding (and use) of data, for national/international assessment (peer review) | <ul style="list-style-type: none"> • <i>Metadata for Radioactive Waste Management</i> • Site Characterisation Library • Waste Package Library • Repository Library • RepMet Tools and Guidelines |
| Policy makers – Policy making and legislative actors, for example local and national government or its departments | Information requirements include elements for assessing strategic decisions about disposal, common ground (safety) for sustainable decision making, and for official information and statements | All periods – from pre-operational to post-operational phase | Help understanding (and use) of data, but only for available data. Guarantee of "good management" of data | <i>Metadata for Radioactive Waste Management</i> |

| Designated communities | What will they want to do? Which questions will they want answered? | Which time phases are they active? | How can RepMet help them? | Which RepMet deliverables may help them? |
|--|--|--|---|---|
| Local communities – communities living or working in surrounding areas near radioactive waste repositories | The perceived primary concern for these groups would be environmental safety and will require continuous and reliable information concerning the safety of their living- or workspace. These groups are likely made up of non-experts but with more information than the general public, and will require key information or a set of selected information that is easily comprehensible with a focus on monitoring data | All periods, from pre-operations to post-operational phase | Help understanding (and use) of data, but only for public data (open data) | <ul style="list-style-type: none"> • Site Characterisation Library • Waste Package Library • RepMet Tools and Guidelines |
| General public – all the individuals and groups who are concerned | This group will likely require similar information to that of the local communities, probably with less detail | All periods, from pre-operations into post-operational phase | Will benefit indirectly by having access to improved radioactive waste repository information | Will benefit indirectly by having access to improved radioactive waste repository information |
| Future generations are “passive stakeholders” in a sense that they are not able to influence the current decision-making processes | Information required by future generations are likely to be a complete description of the decision-making process, and a clear demonstration that decisions were made to avoid undue burden upon them | Post-operational phase | Help understanding (and use) of data, but only for available/archived data (retrievable data) | <ul style="list-style-type: none"> • <i>Metadata for Radioactive Waste Management</i> • Site Characterisation Library • Waste Package Library • Repository Library • RepMet Tools and Guidelines |

| Designated communities | What will they want to do? Which questions will they want answered? | Which time phases are they active? | How can RepMet help them? | Which RepMet deliverables may help them? |
|---|---|--|---|---|
| Local institutions – elected officials and local sections of national institutions/organisations | Similar to local communities, providing a clear framework of how dealing with roles and responsibilities in monitoring operations and post-closure safety | All periods, from pre-operations to post-operational phase | Help understanding (and use) of data, but only for public data (open data) | <ul style="list-style-type: none"> • Site Characterisation Library • RepMet Tools and Guidelines |
| International bodies (e.g. OECD/NEA, IAEA) – international agencies that regulate nuclear sector on behalf of member states | Interested in data to be used to verify compliance with international guidelines, data to be shared in peer-review inquiries and documents | All periods, from pre-operations to post-operational phase | Help understanding (and use) of data for comparison or for national/international assessment (peer review) | <ul style="list-style-type: none"> • <i>Metadata for Radioactive Waste Management</i> • Site Characterisation Library • Waste Package Library • Repository Library • RepMet Tools and Guidelines |
| NGOs – national or international associations, such as environmental sector and policy think tanks | Requirements for information are similar to general public above, adding robustness in order to prevent and/or to match a negative attitude | All periods, from pre-operations to post-operational phase | Will benefit indirectly by having access to improved radioactive waste repository information | Will benefit indirectly by having access to improved radioactive waste repository information |
| Advisory and consultative bodies – Organisations/groups nominated by public authorities to advise in specific topics | Requirements similar to policy makers, regulators and technical support organisations except probably prefer a less technical language/style | Pre-operational phase | <ul style="list-style-type: none"> Help understanding (and use) of data, for national/international assessment (peer review) But only for available (public) data | <ul style="list-style-type: none"> • RepMet Tools and Guidelines |

| Designated communities | What will they want to do? Which questions will they want answered? | Which time phases are they active? | How can RepMet help them? | Which RepMet deliverables may help them? |
|---|--|---------------------------------------|---|---|
| Scientific and technical societies (learnt societies) – organisations providing technical support in specific issues related to RWM | Requirements similar to policy makers, regulators and technical support organisations, using a less formal language/style | Pre-operational and operational phase | Help understanding (and use) of data, for national /international assessment (peer review) But only for available (public) data | <ul style="list-style-type: none"> • <i>Metadata for Radioactive Waste Management</i> • Site Characterisation Library • Waste Package Library • Repository Library • RepMet Tools and Guidelines |
| RWMOs – organisations responsible for radioactive waste management | During the pre-operational phase (including construction) these organisations can draw on the libraries and supporting documents for information and what data are required, how they can be organised and how they can be complemented with metadata. During the operational and closure phases RWMOs may perform benchmarking to improve their data management, and assistance in preparing their data for closure | Pre-operational and operational phase | Improve management of data using common definition (metadata) Help understanding (and use) of data, for national/ international assessment (peer review) | <ul style="list-style-type: none"> • <i>Metadata for Radioactive Waste Management</i> • Site Characterisation Library • Waste Package Library • Repository Library • RepMet Tools and Guidelines |

| Designated communities | | What will they want to do? Which questions will they want answered? | Which time phases are they active? | How can RepMet help them? | Which RepMet deliverables may help them? |
|--|--|---|--|---|---|
| Organisations providing independent oversight – regulators and technical support organisations | | Concerned with data to be used to verify compliance with international guidelines and national rules/prescriptions | Pre-operational and operational phase | Improve management of data using common definition (metadata) Help understanding (and use) of data, for national/ international assessment (peer review) | <ul style="list-style-type: none"> • <i>Metadata for Radioactive Waste Management</i> • Site Characterisation Library • Waste Package Library • Repository Library • RepMet Tools and Guidelines |
| Developing RWMOs – agencies and organisations implementing similar projects in their own countries | | Data and experiences may be shared in order to foster mutual learning and to peer review inquiries and documents | All periods, from pre-operations to post-operational phase | Improve management of data using common definition (metadata) Help for conception / development of data management | <ul style="list-style-type: none"> • <i>Metadata for Radioactive Waste Management</i> • Site Characterisation Library • Waste Package Library • Repository Library • RepMet Tools and Guidelines |
| Member of the public | Present-day local communities, communities living or working in the areas surrounding a waste repository | The perceived primary concern for these groups is likely to be environmental safety and the impact of the facility on their community (for example operational noise, numbers of road shipments or the provision of new jobs) | Pre-operational and operational phase | Will benefit indirectly by having access to improved radioactive waste repository information | Will benefit indirectly by having access to improved radioactive waste repository information |

| Designated communities | What will they want to do? Which questions will they want answered? | Which time phases are they active? | How can RepMet help them? | Which RepMet deliverables may help them? |
|--|--|--|---|---|
| Future generations – “passive stakeholders” in the sense that they are not able to influence current decision-making processes | Information required by future generations are likely to include a complete description of the decision-making process, and a clear demonstration that the as-built facility is safe | Operational and post-operational phase | Will benefit indirectly by having access to improved radioactive waste repository information | Will benefit indirectly by having access to improved radioactive waste repository information |

NEA PUBLICATIONS AND INFORMATION

The **full catalogue of publications** is available online at www.oecd-nea.org/pub.

In addition to basic information on the Agency and its work programme, the NEA website offers free downloads of hundreds of technical and policy-oriented reports. The professional journal of the Agency, **NEA News** – featuring articles on the latest nuclear energy issues – is available online at www.oecd-nea.org/nea-news.

An **NEA monthly electronic bulletin** is also distributed free of charge to subscribers, providing updates of new results, events and publications. Sign up at www.oecd-nea.org/bulletin.

Visit us on **Facebook** at www.facebook.com/OECDNuclearEnergyAgency or follow us on **Twitter** @OECD_NEA.



Metadata for Radioactive Waste Management

National programmes for radioactive waste management require very large amounts of data and information across multiple and disparate disciplines. These programmes tend to run over a period of many decades resulting in a serious risk of data and information loss, which in turn can threaten the production and maintenance of robust safety cases.

Metadata and associated tools and techniques play a crucial role in modern data and information management. The Radioactive Waste Repository Metadata Management (RepMet) initiative has prepared the first international study on the application of metadata to the field of radioactive waste management. This report introduces the concept of metadata, explains how metadata can help to facilitate data management, and gives advice on the issues arising when developing metadata within radioactive waste management programmes. It is aimed at readers looking to obtain a high-level overview of metadata, and associated tools and techniques, and the strategic importance they can play in Radioactive Waste Management Organisations (RWMOs).