IAEA Nuclear Energy Series

No. NG-T-6.15

Basic Principles

Objectives

Guides

Technical Reports

Exploring Semantic Technologies and Their Application to Nuclear Knowledge Management

IAEA
International Atomic Energy Agency

IAEA NUCLEAR ENERGY SERIES PUBLICATIONS

STRUCTURE OF THE IAEA NUCLEAR ENERGY SERIES

Under the terms of Articles III.A.3 and VIII.C of its Statute, the IAEA is authorized to “foster the exchange of scientific and technical information on the peaceful uses of atomic energy”. The publications in the **IAEA Nuclear Energy Series** present good practices and advances in technology, as well as practical examples and experience in the areas of nuclear reactors, the nuclear fuel cycle, radioactive waste management and decommissioning, and on general issues relevant to nuclear energy. The **IAEA Nuclear Energy Series** is structured into four levels:

(1) The **Nuclear Energy Basic Principles** publication describes the rationale and vision for the peaceful uses of nuclear energy.
(2) **Nuclear Energy Series Objectives** publications describe what needs to be considered and the specific goals to be achieved in the subject areas at different stages of implementation.
(3) **Nuclear Energy Series Guides and Methodologies** provide high level guidance or methods on how to achieve the objectives related to the various topics and areas involving the peaceful uses of nuclear energy.
(4) **Nuclear Energy Series Technical Reports** provide additional, more detailed information on activities relating to topics explored in the **IAEA Nuclear Energy Series**.

The IAEA Nuclear Energy Series publications are coded as follows: **NG** – nuclear energy general; **NR** – nuclear reactors (formerly **NP –** nuclear power); **NF** – nuclear fuel cycle; **NW** – radioactive waste management and decommissioning. In addition, the publications are available in English on the IAEA web site:

www.iaea.org/publications

For further information, please contact the IAEA at Vienna International Centre, PO Box 100, 1400 Vienna, Austria.

All users of the IAEA Nuclear Energy Series publications are invited to inform the IAEA of their experience for the purpose of ensuring that they continue to meet user needs. Information may be provided via the IAEA web site, by post, or by email to Official.Mail@iaea.org.

EXPLORING SEMANTIC TECHNOLOGIES AND THEIR APPLICATION TO NUCLEAR KNOWLEDGE MANAGEMENT

The following States are Members of the International Atomic Energy Agency:

AFGHANISTAN
ALBANIA
ALGERIA
ANGOLA
ANTIGUA AND BARBUDA
ARGENTINA
ARMENIA
AUSTRALIA
AUSTRIA
AZERBAIJAN
BAHAMAS
BAHRAIN
BANGLADESH
BARBADOS
BELARUS
BELGIUM
BELIZE
BENIN
BOLIVIA, PLURINATIONAL STATE OF
BOSNIA AND HERZEGOVINA
BOTSWANA
BRAZIL
BRUNEI DARUSSALAM
BULGARIA
BURKINA FASO
BURUNDI
CAMBODIA
CAMEROON
CANADA
CENTRAL AFRICAN REPUBLIC
CHAD
CHILE
CHINA
COLOMBIA
COMOROS
CONGO
COSTA RICA
CÔTE D'IVOIRE
CROATIA
CUBA
CYPRUS
CZECH REPUBLIC
DEMOCRATIC REPUBLIC OF THE CONGO
DENMARK
DJIBOUTI
DOMINICA
DOMINICAN REPUBLIC
ECUADOR
EGYPT
EL SALVADOR
ERITREA
ESTONIA
ESWATINI
ETHIOPIA
FIJI
FINLAND
FRANCE
GABON
GEORGIA
GERMANY
GHANA
GREECE
GRENADA
GUATEMALA
GUYANA
HAITI
HOLY SEE
HONDURAS
HUNGARY
ICELAND
INDIA
INDONESIA
IRAN, ISLAMIC REPUBLIC OF
IRAQ
IRELAND
ISRAEL
ITALY
JAMAICA
JAPAN
JORDAN
KAZAKHSTAN
KENYA
KOREA, REPUBLIC OF
KUWAIT
KYRGYZSTAN
LAO PEOPLE'S DEMOCRATIC REPUBLIC
LATVIA
LEBANON
LESOTHO
LIBERIA
LIBYA
LIECHTENSTEIN
LITHUANIA
LUXEMBOURG
MADAGASCAR
MALAWI
MALAYSIA
MALI
MALTA
MARSHALL ISLANDS
MAURITANIA
MAURITIUS
MEXICO
MONACO
MONGOLIA
MONTENEGRO
MOROCCO
MOZAMBIQUE
MYANMAR
NAMIBIA
NEPAL
NETHERLANDS
NEW ZEALAND
NICARAGUA
NIGER
NIGERIA
NORTH MACEDONIA
NORWAY
OMAN
PAKISTAN
PALAU
PANAMA
PAPUA NEW GUINEA
PARAGUAY
PERU
PHILIPPINES
POLAND
PORTUGAL
QATAR
REPUBLIC OF MOLDOVA
ROMANIA
RUSSIAN FEDERATION
RWANDA
SAINT LUCIA
SAINT VINCENT AND THE GRENADINES
SAMOA
SAN MARINO
SAUDI ARABIA
SENEGAL
SERBIA
SEYCHELLES
SIERRA LEONE
SINGAPORE
SLOVAKIA
SLOVENIA
SOUTH AFRICA
SPAIN
SRI LANKA
SUDAN
SWEDEN
SWITZERLAND
SYRIAN ARAB REPUBLIC
TAJIKISTAN
THAILAND
TOGO
TRINIDAD AND TOBAGO
TUNISIA
TURKEY
TURKMENISTAN
UGANDA
UKRAINE
UNITED ARAB EMIRATES
UNITED KINGDOM OF GREAT BRITAIN AND NORTHERN IRELAND
UNITED REPUBLIC OF TANZANIA
UNITED STATES OF AMERICA
URUGUAY
UZBEKISTAN
VANUATU
VENEZUELA, BOLIVARIAN REPUBLIC OF
VIET NAM
YEMEN
ZAMBIA
ZIMBABWE

The Agency's Statute was approved on 23 October 1956 by the Conference on the Statute of the IAEA held at United Nations Headquarters, New York; it entered into force on 29 July 1957. The Headquarters of the Agency are situated in Vienna. Its principal objective is "to accelerate and enlarge the contribution of atomic energy to peace, health and prosperity throughout the world".

IAEA NUCLEAR ENERGY SERIES No. NG-T-6.15

EXPLORING SEMANTIC TECHNOLOGIES AND THEIR APPLICATION TO NUCLEAR KNOWLEDGE MANAGEMENT

INTERNATIONAL ATOMIC ENERGY AGENCY
VIENNA, 2021

Marketing and Sales Unit, Publishing Section
International Atomic Energy Agency
Vienna International Centre
PO Box 100
1400 Vienna, Austria
fax: +43 1 26007 22529
tel.: +43 1 2600 22417
email: sales.publications@iaea.org
www.iaea.org/publications

Printed by the IAEA in Austria
May 2021
STI/PUB/1899

IAEA Library Cataloguing in Publication Data

Names: International Atomic Energy Agency.
Title: Exploring semantic technologies and their application to nuclear knowledge management / International Atomic Energy Agency.
Description: Vienna : International Atomic Energy Agency, 2021. | Series: IAEA nuclear energy series, ISSN 1995–7807 ; no. NG-T-6.15 | Includes bibliographical references.
Identifiers: IAEAL 21-01406 | ISBN 978-92-0-108719-5 (paperback : alk. paper) | ISBN 978-92-0-114520-8 (pdf) | ISBN 978-92-0-114620-5 (epub) | ISBN 978-92-0-114720-2 (mobipocket)
Subjects: LCSH: Semantic Web. | Semantic networks (Information theory). | Knowledge management. | Nuclear power plants — Management.
Classification: UDC 004.774.2:005.94 | STI/PUB/1899

FOREWORD

The IAEA's statutory role is to "seek to accelerate and enlarge the contribution of atomic energy to peace, health and prosperity throughout the world". Among other functions, the IAEA is authorized to "foster the exchange of scientific and technical information on peaceful uses of atomic energy". One way this is achieved is through a range of technical publications including the IAEA Nuclear Energy Series.

The IAEA Nuclear Energy Series comprises publications designed to further the use of nuclear technologies in support of sustainable development, to advance nuclear science and technology, catalyse innovation and build capacity to support the existing and expanded use of nuclear power and nuclear science applications. The publications include information covering all policy, technological and management aspects of the definition and implementation of activities involving the peaceful use of nuclear technology.

The IAEA safety standards establish fundamental principles, requirements and recommendations to ensure nuclear safety and serve as a global reference for protecting people and the environment from harmful effects of ionizing radiation.

When IAEA Nuclear Energy Series publications address safety, it is ensured that the IAEA safety standards are referred to as the current boundary conditions for the application of nuclear technology.

Over recent years semantic information technologies and Semantic Web standards have emerged as an efficient means to improve access to information and leverage information processing from the purely syntactic to the semantic level. Semantic technologies comprise a wide range of tools, methodologies and standards that allow for the processing of information according to its meaning in context. This is achieved by separating the data from metadata, thus representing the meaning of a certain object independently of the publication it is contained in.

Semantic Web standards play an important role in creating an infrastructure of interoperable data sources. These enable interoperability at the data and the system levels according to specific semantic integration design principles, which are also called 'linked data'. Linked data mark a transition from hierarchies to networks as an organizational principle for data and knowledge. Particularly in environments in which data are, or access to them is, distributed among different content providers or stakeholders, the standardization of data interchange is a prerequisite for managing distributed knowledge.

As part of its programme to assist Member States, the IAEA has issued this publication on applying semantic technologies to nuclear knowledge as an important contribution to knowledge management. This publication offers information for organizations dealing with nuclear knowledge and its management. It provides an introduction to semantic information technologies, the World Wide Web standards developed for interoperability, the construction of knowledge bases on the basis of distributed knowledge and the development of knowledge driven applications.

Appreciation is expressed to all the participants who contributed to the production of this publication. Particular thanks are due to T. Pellegrini (St. Pölten University of Applied Sciences, Austria) for the preparation of this report and the elaboration of the sections related to semantic technologies, and to D. Beraha (Spain) for contributing the section on specific nuclear related aspects and applications and revising the report.

The IAEA officer responsible for this report was M. Gladyshev of the Department of Nuclear Energy.

CONTENTS

1. INTRODUCTION

1.1. BACKGROUND

Within the past decade, the capability of machines to present, share and act upon knowledge has dramatically increased. This has become evident in the rapid growth of knowledge repositories such as Wikipedia and the omnipresence of search engines, which nowadays present a wealth of information related to the user's query, far exceeding their original goal of merely searching documents on the World Wide Web. The technologies that have enabled this development are customarily defined as 'semantic technologies'. From a knowledge management perspective, it is worthwhile to investigate these capabilities, which are most conspicuous in the public domain — within the more constrained scope of organizations — and to relate them to the specific tasks encountered in managing nuclear knowledge.

In particular, semantic technologies provide the ability to connect and relate content deriving from many different sources. Traditionally, most repositories of knowledge and information have been developed as stand-alone systems or systems with limited connectivity, often confined within the boundaries of a single organization. However, such information silos have prevented the gathering and relating of information needed for resolving the problem at hand from different, usually distributed and heterogeneous sources.

The ability to interlink data lies at the core of the enhanced capabilities of search engines not only to retrieve documents containing the requested keywords, but also to offer additional information relevant to the topic. This is achieved by querying an underlying knowledge base, which is constructed (off-line and/or on the fly) using knowledge organization systems from different provenances available on the World Wide Web, and linking them to form extensive knowledge structures, which can be further navigated by the user and used in the development of a variety of knowledge based services and applications.

Exploiting the capabilities of semantic technologies is of particular relevance in the nuclear field, which is fundamentally knowledge driven, and which strongly depends on sharing nuclear knowledge in and between all phases of designing, constructing, operating and decommissioning nuclear facilities. Although the introduction of new information technologies in nuclear organizations has to consider several issues, spelled out in more detail in Section 4, many nuclear organizations have realized the significant benefits that semantic technologies can provide, and accordingly have started programmes in various areas of interest.

1.2. OBJECTIVE

The rapid development of knowledge bases and knowledge based applications is changing the way that knowledge is managed profoundly. As machines increasingly 'understand' the meaning of the concepts involved by evaluating the attributes characterizing them and the relationships between them, developing knowledge bases by deploying semantic technologies becomes an important issue in managing knowledge. This publication provides insight into the fundamentals of semantic technologies and the tools and methods used to develop knowledge bases.

Within the nuclear field, the vast body of knowledge, involving many scientific, technical and managerial fields, is distributed among many organizations of different types. Managing and provisioning distributed knowledge is therefore becoming one of the major challenges in federated organizational environments. This is certainly a relevant issue for the IAEA, whose mostly knowledge based activities have to be coordinated among 172 Member States (as of September 2020), more than 2500 employees and several special interest groups and stakeholders who produce and consume millions of database records, web sites and other sources of information.[1]

In addition to providing insight into the development of distributed knowledge bases, this publication provides examples of applications of semantic technologies specifically in the nuclear field.

1.3. SCOPE

This publication gives an insight into semantic technologies, their use in building knowledge bases and the development of knowledge driven services and applications. As knowledge in the nuclear field is hosted in many different organizations around the world, a dominant concern is the management of distributed knowledge.

This publication is applicable to all organizations concerned with managing knowledge for knowledge sharing, preservation and use, thereby covering the whole life cycle of knowledge. As knowledge bases are increasingly used as foundations for applications and services, the development of such applications is described, and examples are given that assist knowledge management and informatics units in the design of such systems.

1.4. STRUCTURE

This publication is structured as follows:

Section 2 introduces the impact of semantic technologies on knowledge management and states the problem of managing knowledge under the conditions of distributed knowledge, pointing out the challenges encountered in an organization, and introducing semantic technologies and the standardization efforts by the World Wide Web Consortium (W3C).

Section 3 gives a general overview of knowledge organization. It discusses current trends in knowledge organization, provides a classification scheme that enables the scale and scope of various knowledge organization systems to be defined in relation to each other, and discusses the topic of interoperability as an important design principle in the organization of distributed knowledge. It further discusses Semantic Web technologies, a set of standards defined by the W3C to improve data processing and to allow sharing of rich metadata sources for various purposes. Additionally, the technological and organizational benefits of designing systems according to Semantic Web guidelines are discussed. In this context the concept of linked data is introduced, which is commonly referred to as ‘Semantic Web done right’.

The section also gives a hands-on introduction to simple knowledge organization system (SKOS), a Semantic Web standard for the creation and sharing of controlled vocabularies. SKOS is commonly used to support knowledge bases at the terminological level but it also grounds highly expressive and customizable knowledge organization systems that go beyond the expressivity of a thesaurus.

Section 4 provides an overview and the current status of nuclear knowledge and information management, and further describes the use of semantic technology specifically in the nuclear domain. Several fields are highlighted for which applications based on knowledge models and semantic technologies are being developed.

Section 5 provides a brief conclusion.

The Appendix contains a list of publicly available thesauri, which may be reused and linked to any knowledge organization system.

The Annex depicts two use cases from the nuclear sector. The first describes a system that has been implemented at the Indira Gandhi Centre for Atomic Research (IGCAR), providing enhanced search

[1] For IAEA by numbers see https://www.iaea.org/about/by-the-numbers

capabilities within a knowledge portal. The second addresses the work in progress to construct a database for learning objects, which utilizes semantics for structuring and tagging the content.

2. OVERVIEW OF SEMANTIC TECHNOLOGIES IN KNOWLEDGE MANAGEMENT

2.1. THE IMPACT OF SEMANTIC TECHNOLOGIES ON KNOWLEDGE MANAGEMENT

Semantic technologies provide methods for uncovering the meaning and context within data and text by utilizing different approaches such as formal machine reading, text mining, natural language processing and advanced artificial intelligence. The application of semantic technology benefits a large number of industries and academic disciplines in managing and developing their knowledge. The Semantic Web is a focal point in standardizing this new-generation technology so that data, information and knowledge in different forms and within heterogeneous systems can more easily be discovered, processed and shared.

Knowledge management deals with the different forms of knowledge embedded in an organization's products, processes and people. It ensures that the organization's knowledge related assets are continuously improved and effectively utilized. The processes of knowledge management involve all aspects of knowledge acquisition, capture, creation, refinement, storage, transfer, sharing and utilization. The knowledge management framework in the organization operates on these processes, providing the methodologies to define and sustain them. Knowledge management systems effectively apply the organization's information and communication technology to support these processes. Conventional knowledge management systems are typically less automated and require substantial manual support and involvement. By applying semantic technologies, various tasks associated with knowledge management processes can be automated or semi-automated to significantly reduce human effort and time.

Semantic technologies have implications for a wide range of knowledge management applications and services such as context sensitive search engines, professional metadata management, indexing and tagging. They provide additional value to the existing IT infrastructure that enables the meaningful interconnection of data, content and processes. Furthermore, semantic technologies and tools add a new level of depth to provide intelligent, more relevant and interactive interfaces during all stages of managing knowledge in the process life cycle.

The knowledge acquisition activities of searching, recognizing and assimilating potentially valuable knowledge from internal and external sources can be efficiently supported by applying semantic technologies. These can effectively be utilized for the creation of new explicit knowledge by merging, categorizing and synthesizing existing knowledge sources. As a part of the refinement and storage of knowledge, semantic technologies play a major role in improving the process of explicating, codifying, extracting, indexing, organizing, integrating and selecting the contents for inclusion in a repository. Semantic technologies enhance the means of transferring or sharing knowledge across individuals, groups and organizations through standardization, interoperability and context aware query retrieval systems.

In summary, the impact of semantic technologies will deeply affect the way knowledge is managed in organizations, enhancing the efficacy, legitimacy, validity and relevance of the contributed knowledge.

2.2. KNOWLEDGE MANAGEMENT UNDER CONDITIONS OF DISTRIBUTED KNOWLEDGE

Organizations, especially those operating in the nuclear sector, are often characterized by a high degree of complexity resulting from procedural, structural and functional interdependencies between workers, institutional units and infrastructures. Information technology in general and knowledge management systems in particular can be applied to decrease certain types of complexity associated with the management and logistics of information under highly networked circumstances. But information access does not provide a value *eo ipso*, if information cannot be found and/or retrieved according to the needs and interests of its users. Hence modern information systems need to support "a form of organised complexity, providing meaning in context, and promoting understanding" [1] among the actors involved. This is achieved by providing an infrastructure that allows for integrating, feeding and curating federated information sources, while promoting principles of decentralization, compliance standards and shared principles of knowledge organization [2]. Under such circumstances interoperable metadata and controlled vocabularies have become key to leveraging information as a valuable resource in knowledge intensive organizations [3].

This publication defines knowledge as: "A mix of experiences, values, contextual information and expert insight for acquiring, understanding and interpreting information. Together with attitudes and skills, it forms a capacity for effective actions".

According to Ref. [4], the technological design of systems for the organization of distributed knowledge has to take account of the following challenges:

Agility and dynamics of an organization and its environment: Institutions in a networked society are embedded in various subsystems that frequently change their structural behaviour and are continually being further developed. From a data model perspective, more flexible means than relational databases are needed to represent these structures and dynamics and their inherent ideas of putting schemas in place to leave them untouched over decades.

Heterogeneity of sources and complexity of information management: The level of distribution and ever increasing degree of decentralization of organizational structures and their corresponding subsystems to gather, link, distribute and interpret information has led to a paradoxical situation: The task of managing information in organizations is strongly decentralized, in many cases even as a 'personal responsibility', but existing technologies do not provide the means to share and reuse this information. Under the current situation an increasing number of data silos are produced that lack the necessary interoperability to integrate various data formats and systems necessary to provide single and uniform access to federated data sources (i.e. as 'knowledge as a service').

Diversity of stakeholders and perspectives: Information systems serve multiple stakeholders with differing interests, needs and levels of expertise. Each stakeholder takes a unique perspective on the information of interest and has differing skills in retrieving it. Thus, the pragmatic values of an (business) object changes with its context in use as defined by the stakeholder. Existing information systems hardly support multidimensional and cross-divisional access to information owing to legacies deriving from the information architecture. Hence, knowledge organization systems are needed that provide an integrated view of objects by presenting them in context.

In recent years semantic technologies have emerged as a means to improve access to information and leverage information processing from the purely syntactic to the semantic level. Semantic technologies comprise a wide range of tools, methodologies and standards that allow for processing information according to its meaning in context. This is achieved by separating the data from the metadata, thus representing the meaning of a certain object independently of the document it is contained in.

An important development in the area of semantic technologies is the Semantic Web initiative[2] governed by the W3C. The Semantic Web initiative aims to provide the standards for representing and sharing machine readable information. It emerged from the research traditions of knowledge

[2] See http://www.w3.org/standards/semanticweb/

representation, artificial intelligence and information retrieval and has spread into areas such as knowledge management, library and documentation science, enterprise information management, and more recently into e-commerce and e-learning. According to Ref. [5], Semantic Web technologies can considerably improve the information sharing process by overcoming the limitations of current semantic technologies in terms of interoperability and portability of data.

The Semantic Web marks a transition from hierarchies to networks as an organizational principle for knowledge [6]. This opens up opportunities for new modes of value creation but also challenges established organizational philosophies and practices to capture its full potential (i.e. from the large amount of information produced every day just about 5% is 'structured' [7]). But 92% of all analytical activities are exercised on top of structured data [8]. The remaining data are currently hardly utilized or discarded as a whole. Hence new approaches are being developed to improve the machine processability of available data ideally not just by creating more structured data but also by applying structural principles that support interoperability at the structural, syntactic and semantic level alike.

The increasing availability of Semantic Web enabled data on the Web deriving from open data strategies by governments, companies or crowdsourcing initiatives drives the question of how knowledge workers can benefit from this newly emerging ecosystem. To do so it is necessary to apply Semantic Web standards to one's own sources, provide semantically enriched resources to one's stakeholders of interest and — very appropriately — reuse third party resources to enrich one's knowledge bases with additional information. Anecdotal evidence[3] from sectors such as automotive, media and publishing or the life sciences supports the hypothesis that Semantic Web technologies ease access to information, improve workflow efficiency and trigger diversification at the organizational and business level [14].

3. MANAGING DISTRIBUTED KNOWLEDGE

3.1. ORGANIZING DISTRIBUTED KNOWLEDGE[4]

3.1.1. Trends in knowledge organization

The accelerating pace of change in the economic, legal and social environment combined with tendencies towards increased decentralization of organizational structures have had a profound impact on the way we organize and utilize knowledge [15]. Organizational and legal transformations coincide with the emergence of integrated communication systems as universal platforms for the creation and distribution of information [16]. The Internet as we know it today and especially the World Wide Web as the multimodal interface for the presentation and consumption of multimedia information are the most prominent examples of these developments. To illustrate the impact of new communication technologies on information practices, Saumure and Shiri [17] conducted a survey on knowledge organization trends in the library and information sciences before and after the emergence of the World Wide Web. Table 1 shows their results.

---

[3] See Ref. [9] for a BBC use case or Ref. [10] for a Guardian case. Additional use cases can be found in Refs [11], [12] and [13]. See also the collection of use cases at http://www.w3.org/2001/sw/sweo/public/UseCases/

[4] The material in Sections 3.1.1–3.1.3 was originally provided by T. Pellegrini for this publication. It has since been published online at https://semantic-web.com/2015/05/05/thoughts-on-kos-part-3-trends-in-knowledge-organization/; https://semantic-web.com/2015/04/21/classifying-knowledge-organisation-systems/; and https://semantic-web.com/2015/04/10/ getting-to-grips-with-semantic-interoperability/

TABLE 1. RESEARCH TRENDS IN LIBRARY AND INFORMATION SCIENCE [17]

Research area	Pre-Web	Post-Web
Metadata applications/uses	–	16%
Cataloguing/classification	14%	15%
Classifying Web information	–	14%
Interoperability	–	13%
Machine assisted knowledge organization	14%	12%
Education	7%	7%
Digital preservation/libraries	–	7%
Thesauri initiatives	7%	5%
Indexing/abstracting	29%	4%
Organizing corporate or business information	–	4%
Librarians as knowledge organizers of the Web	–	2%
Cognitive models	29%	1%

The survey illustrates three major trends: (i) the spectrum of research areas has broadened significantly from originally complex and expert driven methodologies and systems to more lightweight, application oriented approaches; (ii) while certain research areas have kept their status over the years (e.g. cataloguing and classification or machine assisted knowledge organization), new areas of research have gained importance (e.g. metadata applications and uses, classifying Web information, interoperability issues) while formerly prevalent topics like cognitive models or indexing have declined in importance or dissolved into other areas; and (iii) the quantity of papers that explicitly and implicitly deal with metadata issues has significantly increased.

These insights coincide with a survey conducted by The Economist (Ref. [7]) that comes to the conclusion that metadata has become a key enabler in the creation of controllable and exploitable information ecosystems under highly networked circumstances.[5] Metadata provide information about data, objects and concepts. This information can be descriptive, structural or administrative. Metadata adds value to data sets by providing structure (i.e. schemas) and increasing the expressivity (i.e. controlled vocabularies) of a data set.

According to Ref. [19], "The association of standardized descriptive metadata with networked objects has the potential for substantially improving resource discovery capabilities by enabling

[5] Results from an econometric survey [18] point to the fact that from a certain point in time the economic value of metadata is rising faster than the actual instance data that is being managed by it. In other words, a controlled metadata strategy is a necessary means to preserve and leverage the value of data for information management purposes like aggregation, processing and bundling of data as well as corresponding business processes like strategic planning, designing and selling of goods and services. The concrete economic impact of an integrated metadata strategy is illustrated by the retail giant Walmart. They announced that improved semantic interoperability in their e-commerce system has increased sales up to 15% since its deployment in 2012 (see http://www.computerworld.com/s/article/9230801/Walmart_rolls_out_semantic_search_engine_sees_business_boost).

field-based (e.g., author, title) searches, permitting indexing of non-textual objects, and allowing access to the surrogate content that is distinct from access to the content of the resource itself."

These trends influence the functional requirements of the next generation's knowledge organization systems as a support infrastructure for knowledge sharing and knowledge creation under conditions of distributed intelligence and competence.

3.1.2. Classifying knowledge organization systems

According to Ref. [20], a knowledge organization system (KOS) "is intended to encompass all types of schemes for organizing information and promoting knowledge management. […] A KOS serves as a bridge between the user's information need and the material in the collection […] the KOS guides the user through a discovery process".

Traditional KOSs include a broad range of system types from term lists to classification systems and thesauri. These organization systems vary in functional purpose and semantic expressivity. Term lists may be simple authority lists. Classification systems put aggregate resources into groups with shared attributes. Thesauri express broader/narrower, synonymous and associative (or related term) relationships. More advanced systems use logic and rule models to enrich the model's expressivity by defining constraints between resources and their relationships. These and other traditional KOSs were developed in a print and library environment. They have been used to control the vocabulary used when indexing and searching a specific product such as a bibliographic database or when organizing a physical collection such as a library [21].

With the proliferation of the World Wide Web, new forms of knowledge organization principles have emerged based on hypertextuality, modularity, decentralization and protocol based machine communication [22]. New forms of KOSs have emerged such as folksonomies, topic maps and knowledge graphs, also commonly and broadly referred to as ontologies.[6]

Gruber notes that "a common ontology defines the vocabulary with which queries and assertions are exchanged among agents" [23] based on "ontological commitments to use the shared vocabulary in a coherent and consistent manner" [24]. From a technological perspective ontologies function as an integration layer for semantically interlinked concepts to improve the machine readability of the underlying knowledge model. Ontologies leverage interoperability from a syntactic to a semantic level for the purpose of knowledge sharing. According to Ref. [21], "semantic tools emphasize the ability of the computer to process the KOS against a body of text, rather than support the human indexer or trained searcher. These tools are intended for use in the broader, more uncontrolled context of the Web to support information discovery by a larger community of interest or by Web users in general."

In other words, ontologies are considered to be valuable in classifying web information in that they aid in enhancing interoperability — bringing together resources from multiple sources [17].

Reference [25] introduces a model to classify ontologies along their scope, acceptance and expressivity, as can be seen in Fig. 1.

According to this model, the design of KOSs has to take into account the **user group** (acceptance model), the **nature and abstraction level of knowledge** to be represented (model scope) and the adequate **formalism** to represent knowledge for specific intellectual purposes (level of expressiveness). Although the proposed classification leaves room for discussion, it can help to distinguish various KOSs from each other and gain a better insight into the architecture of functionally and semantically intertwined KOSs. This is especially important under conditions of interoperability.

---

[6] It must be critically noted that the inflationary usage of the term 'ontology', often in neglect of its philosophical roots, has not necessarily contributed to a clarification of the concept itself. A detailed discussion of this matter is beyond the scope of this publication. Gruber defines ontology as "an explicit specification of a conceptualization", which is commonly being referred to in artificial intelligence research [23].

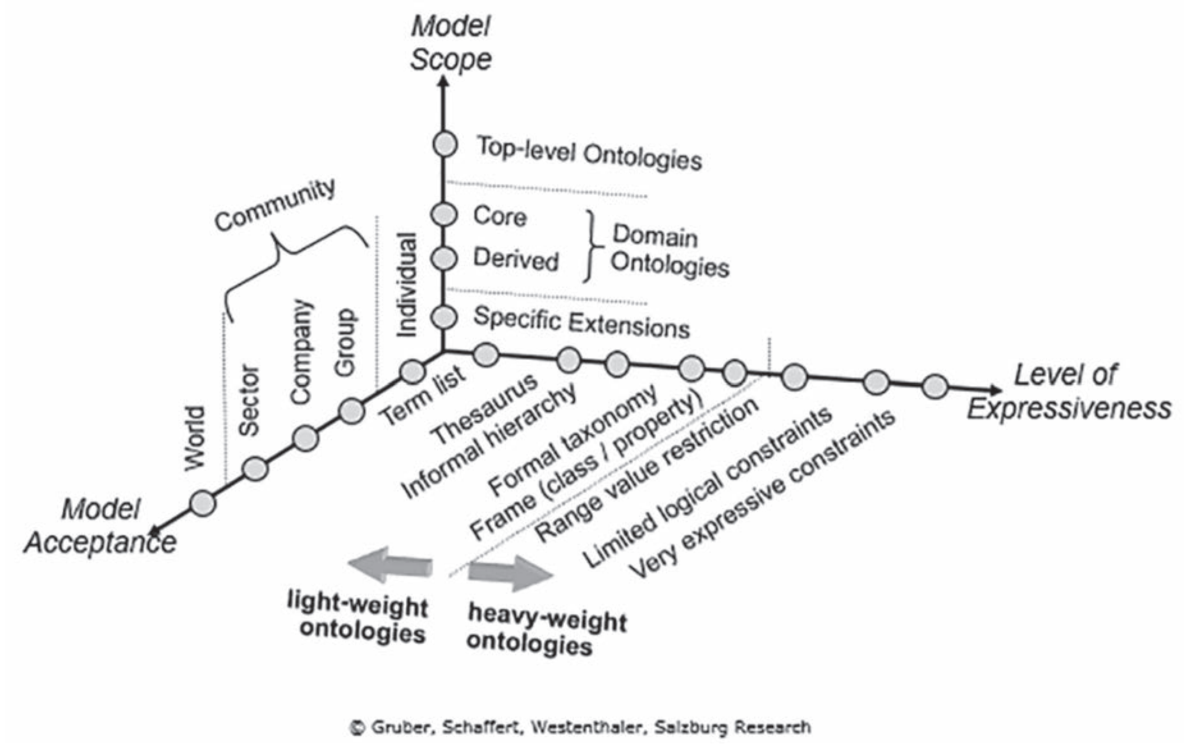


FIG. 1. A classification model for ontologies (adapted from Ref. [25]).

3.1.3. Designing interoperable KOSs

Enabling and managing interoperability at the data and the service level is one of the strategic key issues in networked KOSs and a growing issue in effective data management [26]. Without interoperability, silos of data and systems are being created that prohibit the ecosystem from nurturing network effects and positive feedback loops around its knowledge sources. This is especially relevant for data driven organizations that rely on fast and flexible data analytics from federated sources to generate content for decision support purposes. Here, interoperability is a precondition to enable the cost effective transaction and integration of data and provide adequate services for its consumption [27].

The concept of (data) **interoperability** can best be understood in contrast to (data) **integration**. While integration refers to a process whereby formerly distinct data sources and their representation models are merged into one newly consolidated data source, the concept of interoperability is defined by a structural separation of knowledge sources and their representation models, which allows connectivity and interactivity between these sources by deliberately defined overlaps in the representation model [28]. Under circumstances of interoperability, data sources are being designed to provide interfaces for connectivity to share and integrate data on top of a common data model, while leaving the original principles of data and knowledge representation intact. Thus, interoperability is an efficient means to improve and ease integration of data and knowledge sources.

When designing interoperable KOSs it is important to distinguish between structural, syntactic and semantic interoperability [29]:

Structural interoperability is achieved by representing metadata using a shared data model like the Dublin Core abstraction model or resource description framework (RDF).

Syntactic interoperability is achieved by serializing data in a shared markup language like XML, Turtle or N3.

Semantic interoperability is achieved by using a shared terminology or controlled vocabulary to label and classify metadata terms and relations.

Given the fact that metadata standards carry a lot of intrinsic legacy, it is sometimes very difficult to achieve interoperability at all of the three levels mentioned above. Metadata formats and models are historically grown; they are usually a result of community decision processes, often highly formalized for specific functional purposes and most of the time deliberately rigid and difficult to change. Hence it is important to have a clear understanding and documentation of the application profile of a metadata format as a precondition for enabling interoperability at all of the above mentioned three levels.

In this context, semantic technologies and Semantic Web standards play an important role in creating an infrastructure of interoperable data sources [30]. Semantic Web enabled KOSs are the grounding layer of automated information access and use. According to Ref. [5], "ontologies are the backbone technology for the Semantic Web and — more generally — for the management of formalized knowledge in the context of distributed systems. They provide machine-processable semantics of data and information sources that can be communicated between different agents (software and people). In other words, information is made understandable for the computer, thus assisting people to search, extract, interpret and process information."

3.2. ENABLING INTEROPERABILITY WITH SEMANTIC WEB TECHNOLOGIES

3.2.1. Semantic technologies and the Semantic Web

Semantic technologies are utilized across a wide area of application, including classification systems, knowledge bases, search engines, recommender systems, storage technologies and metadata management. The term itself lacks a clear definition. According to the Internet Corporation for Assigned Names and Numbers (ICANN) and for the purposes of this publication, semantic technology is understood as "an encoding process whereby meaning is stored separately from data and content."[7]

Semantic technologies are usually applied to structure data according to its meaning in context. Typical objects of interest include persons, organizations, points of interest or industry terms relevant to a specific knowledge domain. Machine readable semantics of data can be utilized to enrich information sources with contextually relevant information and can use this information to improve its expressiveness for purposes such as indexing, filtering, mashing and analytics. KOSs such as taxonomies, thesauri or other forms of ontologies can be considered semantic technologies as they explicate the semantic and pragmatic relationships within a knowledge domain. Creating and maintaining KOSs and the applications that make use of them (search engines, recommender systems, tagging systems, etc.) is supported by a range of tools (text mining, data mining, natural language processing, etc.) that automatically process information (indexing, filtering, clustering, etc.) according to implicit (i.e. statistical) or explicit (i.e. metadata) attributes. Semantic Web technologies should be seen as a subset of semantic technologies with the specific purpose of enabling interoperability of semantically structured information sources by providing standards for representing, linking and sharing information. Table 2 illustrates the difference between semantic technologies and the Semantic Web. Both technological strands complement each other and are often symbiotically intertwined.

According to the W3C the Semantic Web "is about common formats for integration and combination of data drawn from diverse sources, where on the original Web mainly concentrated on the interchange of documents. It is also about language for recording how the data relates to real world objects. That allows a person, or a machine, to start off in one database, and then move through an unending set of databases which are connected not by wires but by being about the same thing."[8]

This is achieved by improving interoperability at the system level and portability at the data level, thus creating a wide-ranging infrastructure of semantically linked information that makes it easier to integrate, process and publish data. As Semantic Web standards attempt to change the ecosystem on the

[7] See https://icannwiki.org/Semantic_Technology

[8] See http://www.w3.org/2001/sw/

TABLE 2. DISTINGUISHING SEMANTIC TECHNOLOGIES AND THE SEMANTIC WEB

Semantic technologies	Semantic Web
Tools for representing and processing meaning	Standards and norms for interoperability
Data and text mining, statistical indexing, natural language processing, reasoning, etc	Dereferenceable URIs (uniform resource identifiers)
	Data model: RDF (resource description framework)
Term lists, schemas, vocabularies, taxonomies, thesauri, etc	Ontology language: OWL (Web ontology language)
	Query language: SPARQL (SPARQL protocol and RDF query language)

Web by building network effects around (meta)data, the Semantic Web has also been called Web 3.0, the Web of data, giant global graph and linked data.

3.2.2. Semantic Web design principles

The Semantic Web uses a set of standards to enable interoperability at the data and the system level. These standards and their functional interdependence, commonly referred to as Semantic Web Stack[9] or Semantic Web Layer Cake, are illustrated in Fig. 2.

The core elements of the technology stack according to the Semantic Web's design principles are discussed below.[10]

Use HTTP-URIs to name things: The core idea of the Semantic Web can be boiled down to the handy formulation: **things not strings**. This basically means that on the Semantic Web machines are processing not words (strings) but concepts (things). On the Semantic Web all things or resources have a uniform resource identifier (URI) thus distinguishing ambiguous words from unique concepts (e.g. 'Java' the island from 'Java' the programming language). Hence, the island Java can be uniquely identified by the URI http://dbpedia.org/resource/Java, while the programming language is identified by the URI http://dbpedia.org/page/Java_(programming_language).

The Semantic Web uses HTTP-URIs not to address things (e.g. as URLs do for HTML pages) but to name things via dereferenceable namespaces. Namespaces provide a way to use markups from multiple sources. On the Semantic Web namespaces are based on the domain name system of the Internet, thus allowing one to create unique URIs on the Internet. HTTP-URIs allow machines to dereference things and retrieve associated information, since an HTTP-URI can be linked to other HTTP-URIs according to the principles of hypertext. URIs should be designed to be user-friendly by supporting content negotiation and providing reasonable information within its string patterns.[11]

Link URIs using XML/RDF: The resource description framework (RDF)[12] is the standard model of data interchange on the Web. Together with XML it enables structural and syntactic interoperability of datasets. RDF is used to express descriptions of resources. It allows one to create simple statements about things, such as 'William Shakespeare (subject) is the author of (predicate) King Lear (object)' or 'Java

[9] See http://en.wikipedia.org/wiki/Semantic_Web_Stack

[10] The core stack embraces the following layers: URI, XML, RDF, RDF-S, OWL and SPARQL. The remaining stack layers will be ignored for the time being as they do not contribute to the understanding of the design principles themselves.

[11] See http://www.w3.org/TR/cooluris/

[12] See http://www.w3.org/TR/rdf11-concepts/

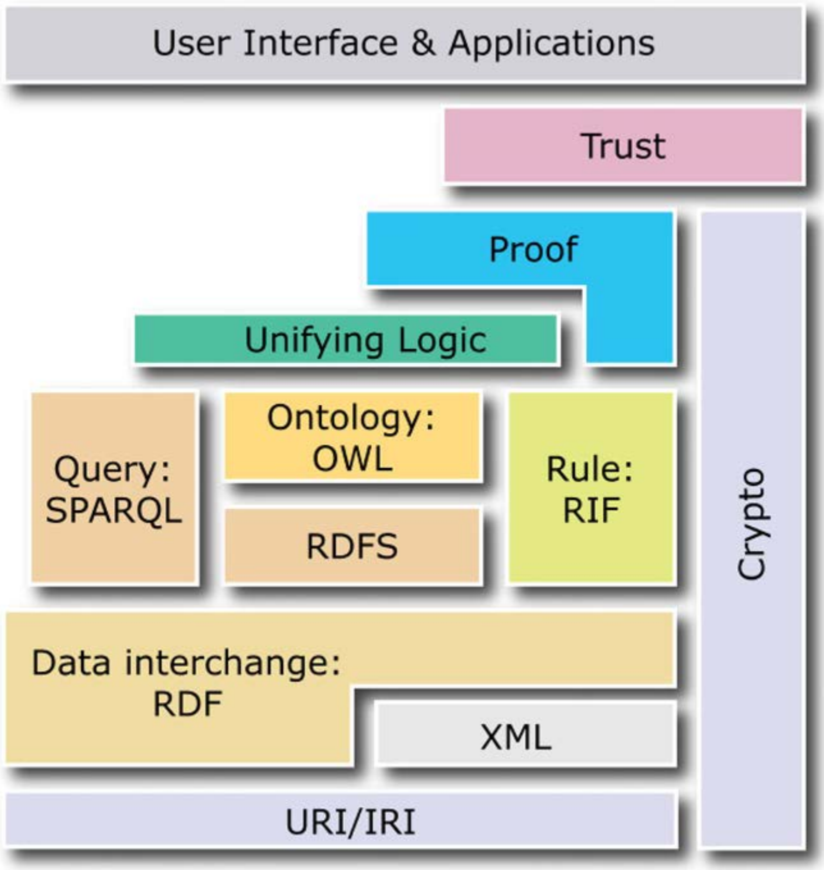


FIG. 2. The Semantic Web Layer Cake. Copyright © W3C®.

(subject) is a geographical entity of type (predicate) island (object)'. RDF links HTTP-URIs into graphs. The tiniest graph is called RDF triple and consists of three components:

(1) The subject, which is a URI or a blank node;
(2) The predicate, which is a URI;
(3) The object, which is a URI, a literal or a blank node.

Use RDF-S and OWL to represent context: An RDF schema (RDF-S) and OWL (Web ontology language) enable semantic interoperability on top of the RDF data model. In other words, they are used to extend the semantic expressivity of an RDF. An RDF schema provides mechanisms for describing groups of related resources and the relationships between these resources. The RDF schema class and property system differs from many such systems in that instead of defining a class in terms of the properties its instances may have, an RDF schema describes properties in terms of the classes of resource to which they apply by allowing a definition of the domain and range of an RDF graph.

For example:

> "[W]e could define the eg:author property to have a domain of eg:Document and a range of eg:Person, whereas a classical object oriented system might typically define a class eg:Book with an attribute called eg:author of type eg:Person. Using the RDF approach, it is easy for others to subsequently define additional properties with a domain of eg:Document or a range of eg:Person. This can be done without the need to re-define the original description of these classes. One benefit of the RDF

property-centric approach is that it allows anyone to extend the description of existing resources, one of the architectural principles of the Web."[13]

Richer vocabularies (like Dublin Core for documents, LOM for Learning Object Metadata, or SCORM for Sharable Content Object Reference Model) or ontology languages such as OWL extend the semantic expressivity of RDF-S, allowing meaningful generalizations about data on the Web. An ontology specifies the meaning of concepts within a certain knowledge domain by typing relationships between concepts and classes and defining logical constraints such as symmetries, transitivity, inverse relationships and many more. Commonly used standards are SKOS[14] or OWL.[15] From a technological perspective RDF-S and OWL lay the basis for a machine to disambiguate, infer and reason over data, thus improving the linking, integration, coherent navigation and discovery of knowledge sources.

Query ontologies with SPARQL: RDF graphs can be queried with the so-called SPARQL Protocol and RDF Query Language, a syntactically structured query language (SQL) for querying RDF graphs via pattern matching. The language's features include basic conjunctive patterns, value filters, optional patterns and pattern disjunction. The SPARQL protocol is a method for remote invocation of SPARQL queries. It specifies a simple interface that can be supported via HTTP or simple object access protocol (SOAP) that a client can use to issue SPARQL queries against some endpoint or federated data sources.[16] Due to its highly standardized syntax, SPARQL provides an application programming interface (API) for the fine granular retrieval of data (allowing experimental use and discovery). Given the fact that a lot of relational databases come with specific query dialects, the uniform interface of SPARQL allows one to access data easily and cost-efficiently compared with other systems. As expressed in a forum comment from a SPARQL expert: "A semi trained user that has experience with one vendors SQL syntax is lost in the next. A semi trained SPARQL user can use any SPARQL endpoint no matter what the implementation is or what the schema looks like."[17]

These design issues are refered to as **linked data** in the following sections in a nod to Tim Berners-Lee [31], one of the key inventors of the Semantic Web, who refers to the linked data approach as the "Semantic Web done right".[18]

3.2.3. Technological benefits of Semantic Web technologies

According to the Semantic Web approach, interoperable data sources can be perceived as "an immense database that can be mined to link to data, rather than document-based resources" [26]. This is especially relevant in environments that are characterized by heterogeneous data formats and sources that are spread across various locations and time zones. By building on Semantic Web standards the linked data approach offers significant benefits compared to conventional data integration approaches. These benefits include the following (see Ref. [32]):

Dereferenceability: International resource identifiers (IRIs) are not just used for identifying entities, but since they can be used in the same way as URLs they also enable the location and retrieval of resources that describe and represent these entities on the Web.

---

[13] See http://www.w3.org/TR/rdf-schema/#ch_introduction

[14] See http://www.w3.org/2004/02/skos/

[15] See http://www.w3.org/TR/owl-ref/

[16] See http://thefigtrees.net/lee/sw/sparql-faq#what-is

[17] See http://web.archive.org/web/20160609070647/ http://answers.semanticweb.com/questions/19183/advantages-of-rdf-over-relational-databases

[18] See https://www.w3.org/2008/Talks/0617-lod-tbl/#(3)

Coherence: When an RDF triple contains IRIs from different namespaces in subject and object position, this triple basically establishes a link between the entity identified by the subject (and described in the source dataset using namespace A) with the entity identified by the object (described in the target dataset using namespace B). Through these typed RDF links, data items are effectively interlinked.

Integrability: Since all linked data sources share the RDF data model, which is based on a single mechanism for representing information, it is very easy to attain a syntactic and simple semantic integration of different linked data sets. A higher level semantic integration can be achieved by employing schema and instance matching techniques and expressing found matches again as alignments of RDF vocabularies and ontologies in terms of additional triple facts.

Timeliness: Publishing and updating linked data is relatively simple, thus facilitating a timely availability. In addition, once a linked data source is updated it is straightforward to access and use the updated data source, since time-consuming and error-prone extraction, transformation and loading are not required.

On top of these technological principles linked data promises to lower the costs of data integration [27], thus improving the reusability and richness of information (in terms of depth and broadness). According to Ref. [26], linked data provides an efficient means for "mapping and interconnecting, indexing and feeding real-time information from a variety of sources. [It allows one to] infer relationships from big data analysis that might otherwise have been discarded".

3.2.4. Application areas of Semantic Web and linked data

The following application scenarios will give practice related insights into how organizations can benefit from linked vocabularies (see Ref. [33]).

3.2.4.1. Semantic content management

The principle question tackled by this scenario is: How can linked data help to establish a metadata layer across systems to link content from multiple sources? The general aim is to provide a single point of access to numerous content repositories.

Based on the fact that managing content in a content management system (CMS) is cost intensive, the maintenance of metadata as an integral part of professional content management is likely to be neglected in many cases. But using dereferenceable metadata on top of all content is key to increase the value of such cost intensive assets.

Text analytics based upon controlled vocabularies can help to keep the cost of managing metadata in a CMS low. Annotating and categorizing content by using thesauri allows for the creation of highly expressive semantic indices of content repositories. Automatic text analytics in combination with comfortable user dialogues for semi-automatic content tagging can be used to link, categorize and annotate content. A linked data based solution approach aims at establishing a metadata layer outside the actual content management system to make integration with other content repositories as easy as possible.

The results and benefits are:

- Automatic document annotation and categorization (XML documents, plain text) based on dereferenceable linked data resources;
- Semi-automatic tagging dialogues based on tag recommender services;
- Ontology based named entity recognition;
- Connectors to other enterprise linked data repositories;
- Laying the basis for content mashups.

3.2.4.2. *Collaborative knowledge bases*

The fundamental question tackled by this scenario is how semantic technologies help to make collaborative knowledge bases more accessible for employees?

Transforming a simple document server into a collaborative knowledge base that serves as a valuable source of knowledge for workers in their daily work is not as simple as it seems. On the one hand, collaboration platforms like enterprise wikis most often are the right choice to encourage people to collect ideas for new content or to make knowledge about products and services more accessible. On the other hand, knowledge bases tend to become poorly maintained over time.

In order to make specific knowledge about business processes, methods or technologies available to as many employees as possible, a combination of three approaches is advisable: enterprise collaboration software, text mining and controlled vocabularies. This results in solutions that fulfil the demand for highly dynamic and flexible knowledge bases that are still sufficiently stable (technical and content wise) to be used in professional environments. Since the knowledge base is generated around a controlled vocabulary acting as a meta-layer, traditional navigation structures like trees are no longer a rigid corset that make the traversing of graph-like structures impossible. Semi-automatic tools for linking, categorizing and content indexing are key to overcoming this problem. Putting a controlled vocabulary in place that grows in parallel to the content base demands new and more agile patterns of taxonomy or thesaurus management than ‘traditional’ approaches would provide.

The results and benefits are:

— Linked knowledge objects on top of collaboration platforms like Confluence or SharePoint;
— Semantic search and knowledge discovery over federated knowledge bases;
— Automatic content enrichment and population of knowledge models from linked data sources.

3.2.4.3. *Semantic data portals*

The high agility and dynamics of an organization and its environment require more agile and flexible methodologies for data integration.

Putting all the information in one place that describes a business object such as a product, a customer, or a certain technology can ease the life of many knowledge workers significantly. On the other hand, the automatic integration of data from various sources can take tremendous effort. Currently, data in enterprises remain locked up in a database. Knowledge workers are forced to manually collect information from a series of data silos to put those pieces together like a puzzle in order to trigger a decision making process. Data integration projects most often are built upon yet another inflexible data structure. Numerous amendments or additions made to the structure or to the semantics of an information component cannot be reflected properly by the integration layer. The result is a landscape consisting of data silos that are scarcely connected to each other. Intelligent linkage happens only in the course of ad hoc processes that are not readily comprehensible.

Web data as well as data in organizations are characterized by a great structural diversity as well as frequent changes. This poses a great challenge for applications based on that data. Linked data methodologies address this problem by using a flexible data model that supports the integration of heterogeneous and volatile data. Linked data technologies for data integration purposes rely on graph based models. This allows for the incremental extension of the schema by various properties and constraints. Linked data is based on open standards, which makes the effort future-proof by avoiding lock-ins and sunk cost investments.

The results and benefits are:

— 360^{o} views on specific business objects (‘topic pages’ like products, organizations, technologies etc.);

— Reports based on sometimes complex queries that can only be answered if data are used from various sources;
— Mashups of unstructured (e.g. business news, social media) and structured data (e.g. statistics, legacy data);
— Mashups of data from the web (e.g. open government data) and internal data sources.

3.2.4.4. Customer support systems

Customer support systems frequently cause disorientation due to the technical terms used and a lack of transparent and easily comprehensible navigation structures. Providers of products and services from various sectors (telecommunications, public administration, law, etc.) use different languages and differing categories than do consumers (or citizens). Thus, in many cases clients have to deal with frustrating translation work, which leads to misunderstandings and increased costs in the call centre. Hence, managing heterogeneity of sources and complexity of information (and its associated management tasks) is key to satisfactory customer support.

Users can benefit from a guidance system that helps to achieve orientation at any point of the support process. The guidance system consists of semantic search facilities such as search filters (faceted search), search refinements, similarity search and integrated fact boxes that display further details about the search term, which might refer to another linked data source. As a prerequisite for these improvements, a knowledge model consisting of concepts (e.g. products, technologies, services), their relations and differing names (including synonyms) has to be created. In some cases, it is advisable to split up the thesaurus into (at least) two modules. A semantic layer of this kind helps to translate between distinct worlds (i.e. supplier/vendor vs. client/customer). While the supplier's thesaurus still links its concepts to the corresponding parts of the client's thesaurus, the thesauri can be managed separately.

The results and benefits are:

— Semantic index of the knowledge base/content base of the support system;
— User-friendly digital guidance system;
— Facilities to refine search queries to find answers to specific questions more easily;
— Help for users to learn quickly: automatically combine search results with facts from other linked knowledge bases.

3.3. CONSTRUCTING DISTRIBUTED KNOWLEDGE BASES WITH SKOS

3.3.1. SKOS — Simple knowledge organization system

SKOS is an open community project governed by the W3C Semantic Web Deployment Working Group[19] and conforms to the ISO 25964 standard about thesauri.[20] Owing to its relatively strict standardization, several international organizations, governmental bodies, libraries and companies provide resources in SKOS format on the World Wide Web, among them the Library of Congress, UNESCO, the Leibniz Institute and NASA.[21]

According to the SKOS Working Group at W3C: "SKOS is an area of work developing specifications and standards to support the use of knowledge organization systems (KOS) such as thesauri, classification schemes, subject heading systems and taxonomies within the framework of the Semantic Web. SKOS provides a standard way to represent knowledge organization systems using the Resource Description Framework (RDF). Encoding this information in RDF allows it to be passed between computer

[19] See http://www.w3.org/2004/02/skos/development
[20] See https://www.niso.org/schemas/iso25964
[21] A current but incomplete list of SKOS datasets can be found at http://www.w3.org/2001/sw/wiki/SKOS/Datasets

applications in an interoperable way. Using RDF also allows knowledge organization systems to be used in distributed, decentralised metadata applications. Decentralised metadata is becoming a typical scenario, where service providers want to add value to metadata harvested from multiple sources."[22]

Several versions of SKOS are currently available under various RDF vocabularies. These are:

SKOS core vocabulary[23] comprising the core concepts and properties of a thesaurus such as skos:Concept, skos:prefLabel, skos:altLabel, skos:definition, skos:broader, skos:narrower to name but a few.

SKOS eXtension for labels (SKOS-XL) vocabulary[24] extending the core vocabulary by allowing one to attach additional metadata to a particular concept.

SKOS RDF Schema — OWL 1 DL Subset[25] a schema representing a formalization of a subset of the semantic conditions described in the SKOS reference document from 2009.[26]

Figure 3 illustrates the basic principle of a SKOS/RDF graph representing the concept 'economic cooperation'.[27]

The following sections take a hands-on view on how to construct SKOS models and build federated knowledge graphs using SKOS-XL.

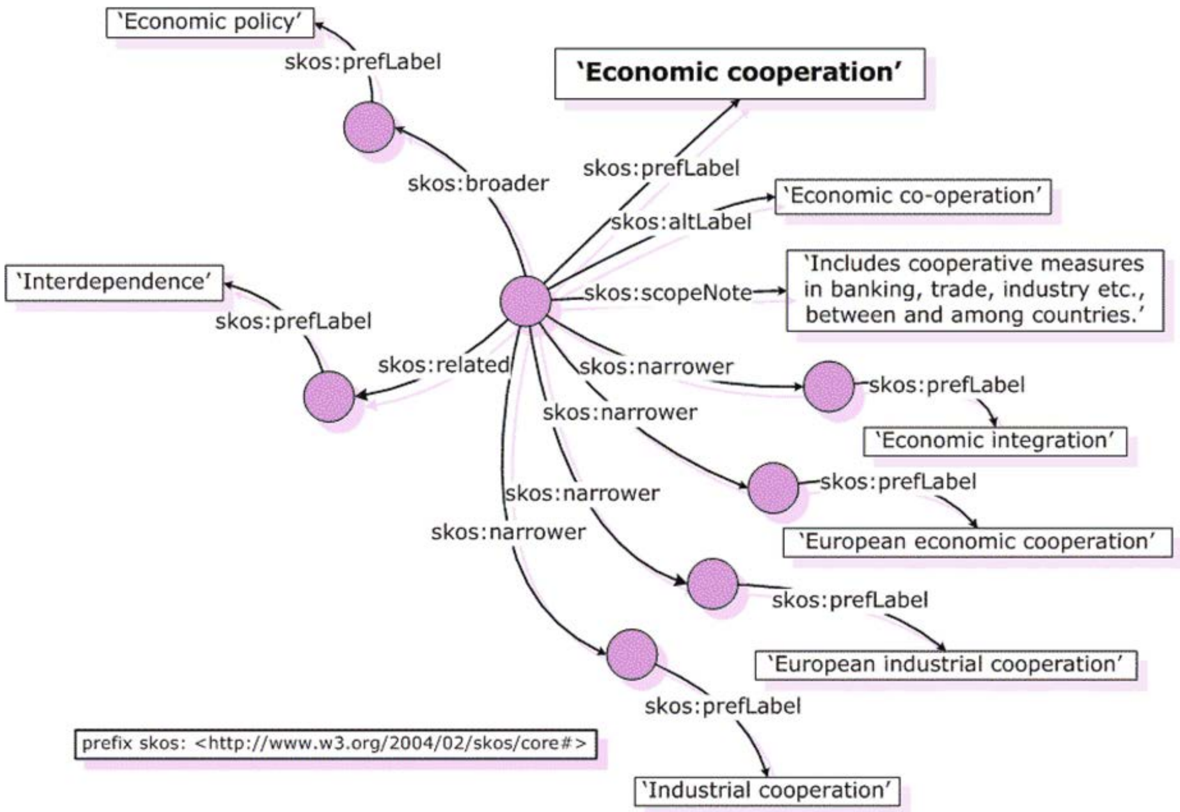


FIG. 3. A typical SKOS graph (Ref. [34]). Copyright ©2005 W3C®.

[22] See http://www.w3.org/2004/02/skos/intro
[23] See http://www.w3.org/2009/08/skos-reference/skos.html
[24] See http://www.w3.org/TR/skos-reference/skos-xl.html
[25] See http://www.w3.org/2009/08/skos-reference/skos-owl1-dl.rdf
[26] See http://www.w3.org/TR/2009/REC-skos-reference-20090818/
[27] See http://www.w3.org/2006/Talks/0428-Beijing-IH/#%2867%29

3.3.2. Using SKOS to build a federated knowledge infrastructure

Building a federated knowledge infrastructure on top of SKOS can be boiled down to four steps: (i) building a SKOS model from the core vocabulary; (ii) adapting the model to the knowledge domain; (iii) extending the expressivity with additional metadata; and (iv) linking to other dereferenceable sources on the knowledge graph. The process is illustrated in Fig. 4 and explained in more detail below.

3.3.2.1. Creating a SKOS model

Organizations do not start from scratch when engaging in a SKOS project. Most of the time some sort of KOS will have been developed in the past such as term lists, controlled vocabularies or more advanced KOSs like taxonomies or thesauri. It is reasonable to reuse these resources by representing them using the SKOS core vocabulary, thus creating a first version of an interoperable KOS entailing the basic vocabulary of an organization. This also includes the machine assisted analysis of text corpora from a set of defined resources to extract additional terms and locate them in the emerging SKOS model. After completing this phase, the model can be populated by importing data from structured sources like Excel sheets, XML files or data harvested from external sources. When the first step is completed an organization has been provided with a basic knowledge model that can be implemented in existing workflows and functions as a dynamic representation of the organization’s knowledge ecosystem.

3.3.2.2. Editing and curating a SKOS model

Under circumstances of networked knowledge management, KOSs should be perceived as living systems prone to inconsistencies and errors. While SKOS is designed to allow a certain degree of tolerance to the formal representation of semantics, it nevertheless needs to be edited and curated on a continuous basis to ensure quality assurance and topicality. The corresponding curation process is usually machine supported comprising a variety of services for purposes such as natural language processing, term extraction and tag recommendations for purposes such as content enrichment, document tagging or data harvesting.

Automatically generated artefacts need to be approved, discarded or populated within the SKOS model, often utilizing a so-called mixed initiative approach, in which machines and humans cooperate

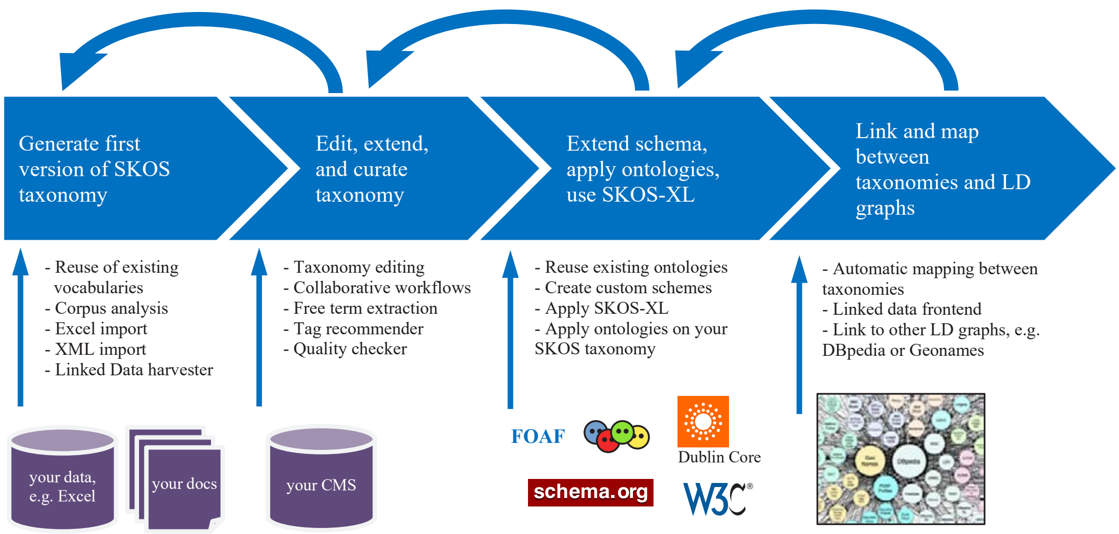


FIG. 4. Building a federated knowledge graph with SKOS.

in feeding, tagging and querying content from the knowledge base, thus continually extending the underlying knowledge model.

The SKOS model's formal consistency is checked via special analytical tools that provide information on quantifiable quality issues, as will be discussed in more detail in Section 3.3.3.

3.3.2.3. Extending a SKOS model

When extending a SKOS model it is reasonable to reuse vocabulary from existing ontologies or custom schemes such as the following:[28]

FOAF: The Friend of a Friend (FOAF) RDF vocabulary can be used for describing persons, organizations, projects and their connections. See: http://www.foaf-project.org/

GEO: The basic geo vocabulary can be used for representing latitude, longitude and altitude information in the WGS84 geodetic reference datum. See: http://www.w3.org/2003/01/geo/

ORG: The core organization ontology is a vocabulary for describing organizational structures. See: http://www.w3.org/TR/vocab-org/

DC Terms: Dublin Core Metadata Initiative (DCMI) namespace for metadata terms. See: http://purl.org/dc/terms/

vCard: Ontology for vCard. See: http://www.w3.org/TR/vcard-rdf/

CUBE: Cube allows multidimensional data to be published in RDF. It is based on the core information model from SDMX. See: http://www.w3.org/TR/vocab-data-cube/

CC REL: The Creative Commons Rights Expression Language (CC REL) can be used for describing copyright licences in RDF. See: http://www.w3.org/Submission/ccREL/

DOAP: Description of a project vocabulary. See: https://github.com/edumbill/doap/wiki

Additional metadata schemes are provided by http://schema.org, an initiative supported by major content aggregators like Bing, Google, Yahoo and Yandex, which rely on this markup to index information on the Web.

3.3.2.4. Linking the SKOS model to other sources

After the SKOS model has reached a sufficient degree of maturity it can be linked to other sources, either within or without one's organizational boundaries. Increasing the degree of interlinking between various KOSs by pointing to other dereferenceable URIs (i.e. by using the skos:related, skos:exactMatch or owl:sameAs properties), SKOS models evolve into knowledge graphs within an ecosystem of interlinked data:

> "Linked Data is a method to publish data on the Web and to interlink data between different data sources. Linked Data can be accessed using Semantic Web browsers, just as traditional Web documents are accessed using HTML browsers. However, instead of following document links between HTML pages, Semantic Web browsers enable surfers to navigate between different data sources by following RDF links. RDF links can also be followed by robots or Semantic Web search engines in order to crawl the Semantic Web."[29]

A prominent ecosystem is the so-called 'linked open data cloud'[30], a federated collection of interlinked RDF datasets allowing access to several billion RDF triples on various domains provided by governments, companies, NGOs, research groups and open-source communities.

---

[28] A non-comprehensive list of available SKOS thesauri can be found in the Appendix.

[29] See https://wiki.dbpedia.org/services-resources/interlinking

[30] See http://lod-cloud.net/

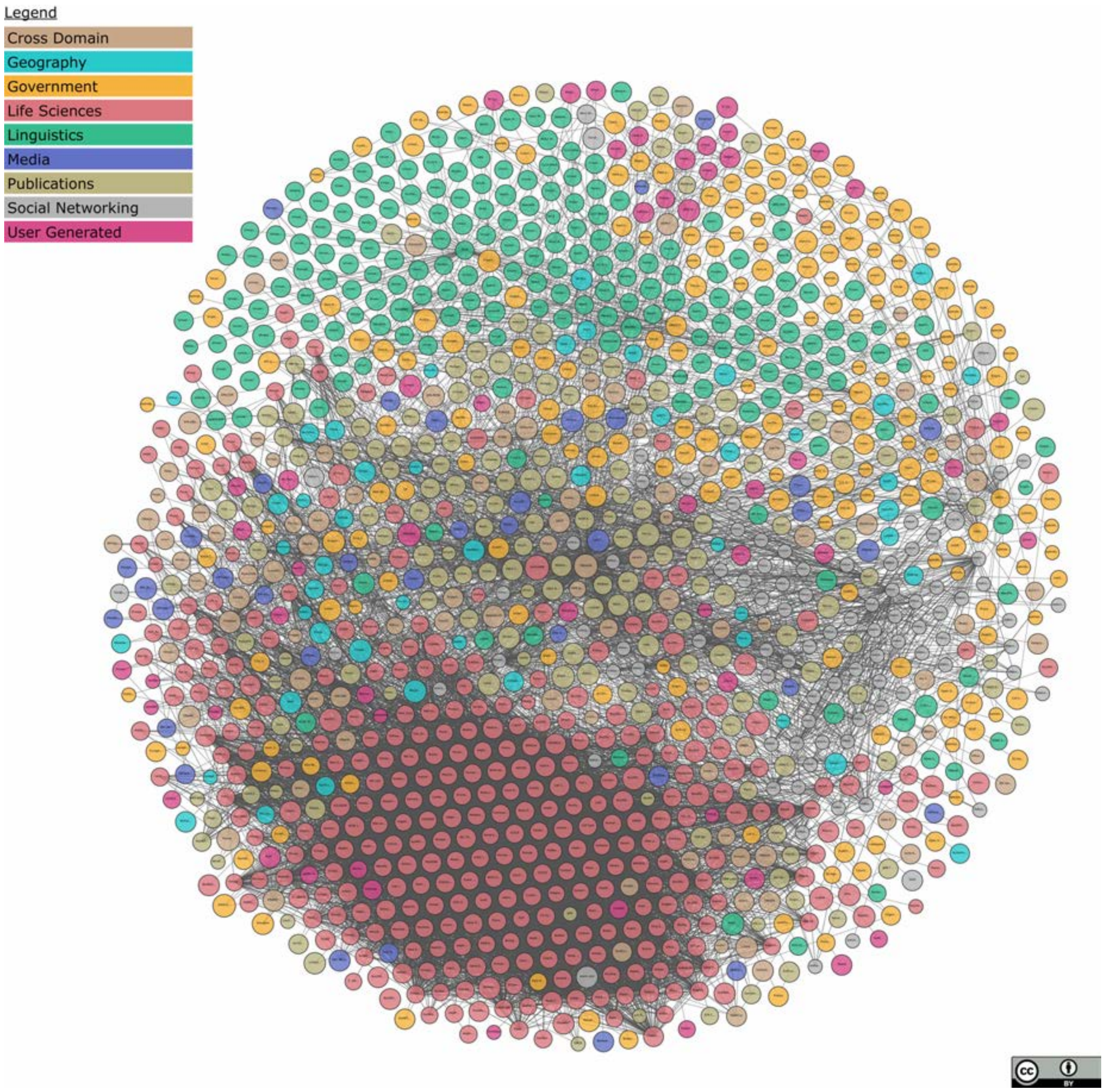


FIG. 5. The linked open data cloud as of May 2020.

DBpedia,[31] is one of the most prominent and popular datasets in the linked open data cloud. The English version of the DBpedia knowledge base currently describes 6.6 million entities, of which 4.9 million have abstracts, 1.9 million have geo coordinates and 1.7 million have depictions. Altogether, the DBpedia 2016–2010 release consists of 13 billion pieces of information (RDF triples), of which 1.7 billion were extracted from the English edition of Wikipedia, 6.6 billion were extracted from other language editions and 4.8 billion from Wikipedia Commons and Wikidata.[32]

Additional datasets are provided by Eurostat, the US Census, The World Bank, the CIA, the BBC, The New York Times, the Association of Computing Machinery and multiple governments from all over the world. Figure 5 illustrates the linked open data cloud as of January 2019.

Linked data clouds are the backbone of the Semantic Web. They provide an infrastructure to create huge knowledge graphs that reach beyond the boundaries of a single organization. If supported by an API,

[31] See http://wiki.dbpedia.org/About

[32] See https://wiki.dbpedia.org/develop/datasets/dbpedia-version-2016-10

the linked source might provide content to be consumed and reused for further machine processing. But most important of all open interlinked data clouds allow publishing information in a standardized format, thus allowing new forms of knowledge discovery. Whether, how and to what extent linked data clouds are made available under an open licence is a matter of policy and not technology.

3.3.3. Limitations and quality issues in modelling SKOS

Reference [34] points to a number of design issues still unsolved by the Semantic Web Deployment Working Group responsible for SKOS specifications. These issues include (i) the extension of SKOS with Boolean operators for concepts; (ii) the indexing of subjects; (iii) provenance information about mappings; (iv) the description of concept schemes; and (v) representation of temporal concept evolution. The working group recognized that these issues are important functional extensions of SKOS, but that they could be resolved by referring to other vocabularies and ontologies (i.e. dct:subject for subject indexing or OWL properties to express Boolean relationships), which provide such mechanisms within their generic specifications. Other issues, such as describing a concepts scheme or representing the evolution of concepts, are deliberately left to be solved by communities of practice due to their context specificities.

The SKOS specification does not mention the notion of quality, but it defines a total of six integrity conditions addressing the formal consistency of a SKOS model [35]. Herein, each condition is a statement that defines under which circumstances data are consistent with the SKOS data model. Apart from these rather formal criteria little research has been done so far addressing further quality issues associated with the creation and utilization of SKOS vocabularies, especially if used to provide and retrieve linked data. The work in Ref. [36] points to critical quality issues associated with SKOS vocabularies resulting from the imminent gap between model level integrity constraints and domain specific quality criteria. Based on their research, common quality issues appear with respect to the following:[33]

Labelling and documentation, including: (i) omitted or invalid language tags, (ii) incomplete language coverage, (iii) no common language, (iv) undocumented concepts, (v) overlapping labels, (vi) missing labels, (vii) unprintable characters in labels, (viii) empty labels and (ix) ambiguous notation references.

Structural consistency, including: (i) orphan concepts, (ii) disconnected concept clusters, (iii) cyclic hierarchical relations, (iv) valueless associative relations, (v) solely transitively related concepts, (vi) unidirectionally related concepts, (vii) omitted top concepts, (viii) top concept having broader concepts, (ix) hierarchical redundancy and (x) reflexive relations.

Linked data specific issues, including: (i) missing in-links, (ii) missing out-links, (iii) broken links, (iv) undefined SKOS resources and (v) HTTP-URI scheme violations.

SKOS semi-formal consistency, including: (i) relation clashes, (ii) mapping clashes, (iii) inconsistent preferred labels, (iv) disjoint labels violation and (v) mapping relations misuse. Without machine support it is practically impossible to detect and repair quality issues within a typical SKOS model, which can comprise several thousand concepts, relations and labels. Tools like qSKOS[34] can be used as a command line tool or API, supporting the identification of the above mentioned quality issues in SKOS vocabularies. Parts of qSKOS's functionality have also been integrated into the GitHub project rsine (Resource Subscription and Notification Service).[35] Rsine enables changes to SKOS vocabularies to be monitored as soon as they occur and sends out immediate notifications (e.g. by email) if some potential quality problems are detected.

---

[33] A detailed discussion of these quality issues can be found at https://github.com/cmader/qSKOS/wiki/Quality-Issues

[34] See https://github.com/cmader/qSKOS/

[35] See https://github.com/rsine/rsine

3.4. ENRICHING KNOWLEDGE BASES WITH LINKED DATA

3.4.1. Connecting information sources with linked data

3.4.1.1. Usage scenarios for linked data

Since 2009 the linked data paradigm has emerged as a lightweight approach to improving data portability among systems. Based on Semantic Web standards, linked data marks a transition from hierarchies to networks as an organizational principle for data and knowledge [6]. Hence, the primary value proposition of linked data is rooted in its modular and simple network characteristics [37]. By sharing RDF as a unified data model, linked data provides the infrastructure for publishing and repurposing data on top of semantic interoperability.

Taking the network characteristics of linked data into account, it is possible to identify three prototypical usage scenarios.

Scenario 1: Internal perspective: From an internal perspective, organizations can make use of linked data principles to organize their information within their closed environments. This is especially relevant for organizations that have to deal with an increasing number of dispersed databases, federated repositories and the legacy issues deriving from them. There is high potential for linked data to consolidate these infrastructures without necessarily disrupting existing systems and workflows.

Scenario 2: Inbound perspective: In the second scenario, organizations use external data sources for purposes such as content pooling or content enrichment. This trend is basically backed by the increasing availability of open data (i.e. provided by governmental bodies, community projects like Wikipedia, Musicbrainz[36] or Geonames[37] and an increasing amount of commercial data providers like Socrata,[38] Factual[39] or Datamarket[40]). Instead of creating these resources on their own, organizations can use this data either free of charge (according to the terms of trade) or as a paid service according to the service levels of an API.

Scenario 3: Outbound perspective: In the third scenario, organizations apply linked data principles to publish data on the Web either as open data or via an API that allows the fine granular retrieval of data according to a user's needs. This process — often called linked data publishing — is basically a diversification of the data distributions strategy of an organization and allows an organization to become part of a linked data cloud. Data publishing strategies often go hand in hand with the diversification of business models and require a good understanding of the licensing issues associated with it. Few enterprises have yet started to engage in such practices. But they are key to leveraging the full potential of linked data for purposes such as e-commerce and e-procurement.

It is important to note that linked data ecosystems are not necessarily open ecosystems. Nevertheless, the degree and extent to which linked data is being provided to the public is an important design issue, if networking effects are intended to unfold around available data. In most cases it will be reasonable to design a differentiated open access policy that makes certain parts of a linked data set available to the public under an appropriate open licence, while providing other parts of a data set under traditional property rights. Choosing the correct set of licences and developing an appropriate open access policy is not a trivial task and is discussed in more detail in Section 3.4.3.

---

[36] See http://classic.musicbrainz.org/

[37] See http://www.geonames.org/

[38] See http://www.socrata.com/

[39] See http://www.factual.com/

[40] See https://www.qlik.com/us/products/qlik-data-market

3.4.2. Linked data in the content value chain

As communication within electronic networks has become increasingly content-centric [38] — in terms of the growing amount of unstructured information that is produced and transmitted electronically[41] — linked data gains importance as a lightweight approach in structuring and interlinking content. The content production process consists of five sequential steps: (i) content acquisition, (ii) content editing, (iii) content bundling, (iv) content distribution and (v) content consumption. As illustrated in Fig. 6, linked data can contribute to each step by supporting the associated intrinsic production function.

Content acquisition is mainly concerned with the collection, storage and integration of relevant information necessary to produce a content item. In the course of this process information is being pooled from internal or external sources for further processing.

Content editing entails all necessary steps that deal with the semantic adaptation, interlinking and enrichment of data. Adaptation can be understood as a process in which acquired data is provided in such a way that it can be reused within editorial processes. Interlinking and enrichment are often performed via processes like annotation and/or referencing to enrich documents either by disambiguating existing concepts or by providing background knowledge for deeper insights.

Content bundling is mainly concerned with the contextualization and personalization of information products. It can be used to provide customized access to information and services (i.e. by using metadata for the device sensitive delivery of content, or to compile thematically relevant material into landing pages or dossiers), thus improving the navigability, findability and reuse of information.

Content distribution mainly deals with the provision of machine readable and semantically interoperable (meta-)data via APIs or SPARQL endpoints. These can be designed either to serve internal purposes so that data can be reused within controlled environments (i.e. within or between organizational units) or for external purposes so that data can be shared between anonymous users (i.e. as open SPARQL endpoints on the Web).

Content consumption is the last step in the content value chain. This entails any means that enable a human user to search for and interact with content items in a user-friendly and purposeful way. This step thus mainly deals with end user applications that make use of linked data to provide access to

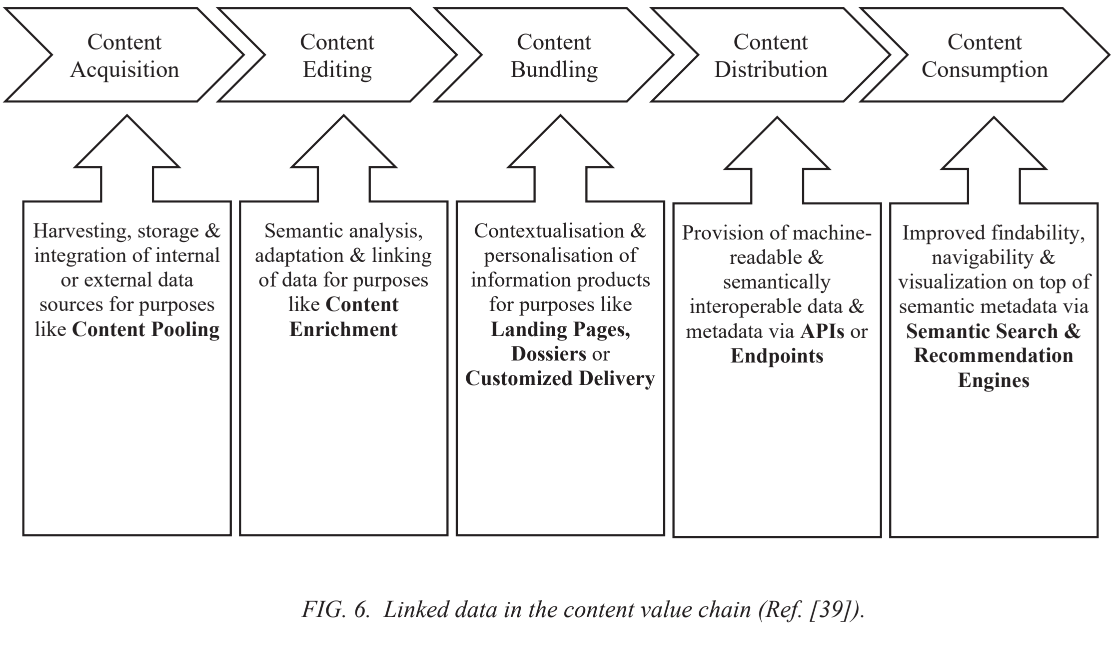


FIG. 6. Linked data in the content value chain (Ref. [39]).

[41] Reference [38] reports for the time period from 2011 to 2016 an increase of 90% of video content, 76% of gaming content, 36% VoIP, 36% file sharing being transmitted electronically.

content items (i.e. via search or recommendation engines) and generate deeper insights (i.e. by providing reasonable visualizations).

Reference [40] proposes a model that describes various stakeholder roles in the creation of linked data assets, as illustrated in Fig. 7.

The model distinguishes between the various stakeholder roles an economic actor can take in the creation of linked data assets and the various types of data and applications that are being created along the data transformation process. Along the value creation process, raw data — which is provided in any kind of non-RDF format (e.g. XML, CSV, PDF, HTML) — is transformed into linked data.

According to the model, raw data is consumed by a linked data provider and transformed into RDF thereby gaining compliance with linked data principles. This step is crucial in enhancing the semantic interoperability of the data. The transformation process can vary significantly in technological complexity, ranging from simple mapping procedures (i.e. with tools like Google Refine) to heavyweight extensible stylesheet language transformations (XSLTs). The complexity depends on the underlying schema's complexity and sometimes requires highly skilled professionals to manage this task properly. In a further step, the linked data is consumed and processed by a linked data application that also provides the interface to the human user. These applications usually do not differ in look and feel from existing ones but use linked data for functional extensions for reasoning and filtering purposes. Finally, the end user consumes the human readable data via functionally extended applications and services.

This view is extended in Ref. [41] with an orthogonal layer called 'support services and consultation', stressing the fact that apart from the value creation process itself, linked data also creates an environment for added value services that transcends the pure transformation and consumption of data. Such services

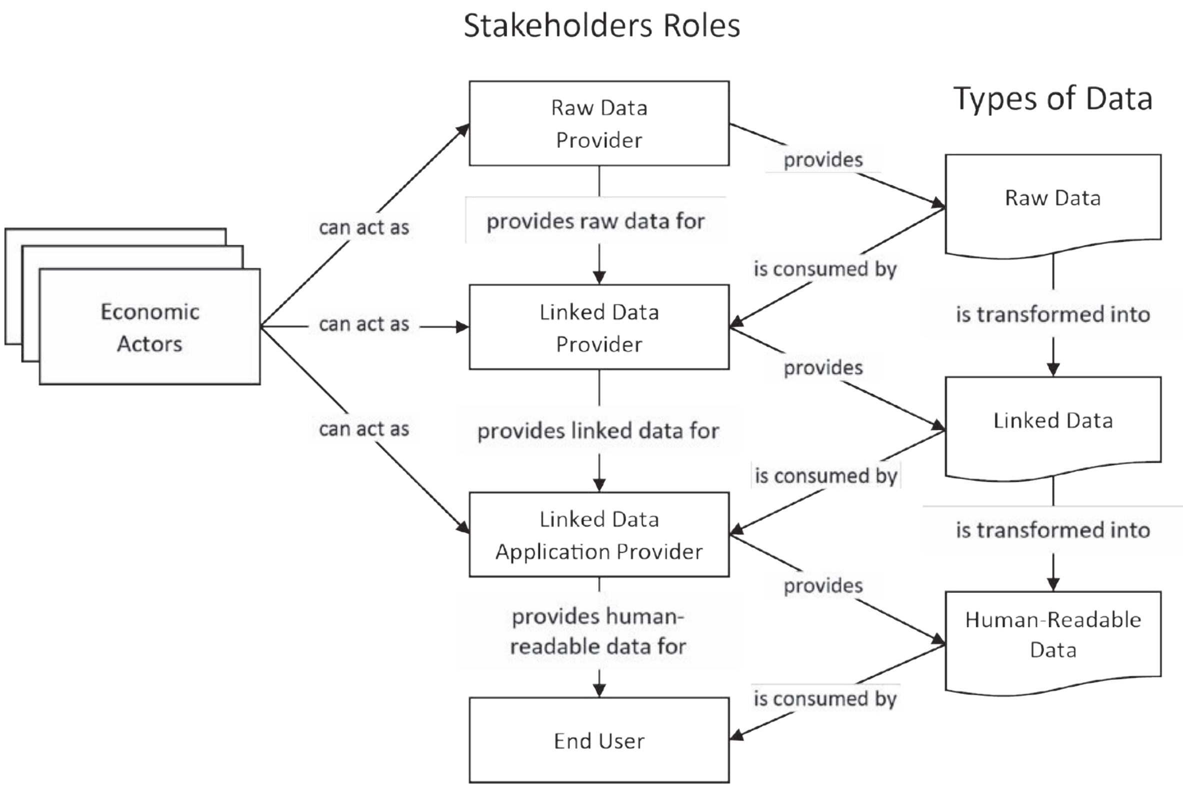


FIG. 7. Linked value chain (Ref. [40]).

are usually provided by data brokers, who collect, clean, visualize and resell available data for further processing and consumption.[42]

As illustrated in Fig. 7, the process of linked data creation can be covered in its entirety by one actor or might require several economic actors depending on the technological complexity of the data transformation process. Existing use cases (e.g. from the BBC or New York Times) reveal that linked data transformations are usually outsourced rather than handled in house. For the time being, it is also difficult to estimate the cost effectiveness of linked data applications, but model based analysis (see Refs [26], [28]) indicates that the savings can be significant depending on the scale and scope of a linked data project. This is due to the network effect linked data generates as an integration layer across various components and workflows in complex IT systems. Herein, linked data can help to reduce technological redundancies thus reducing maintenance costs, improving information access in terms of reduced search and discovery efforts and providing opportunities for business diversification due to the higher granularity and increased connectivity of content.

3.4.3. Licensing linked data

New technology has never been a sufficient precondition for the transformation of business practices but has always been accompanied by complementary modes of cultural change [44]. Although linked data fits neatly into the incremental IT development practices, it comes along with disruptive technological effects that pose significant challenges that change the nature and notion of data as an economic asset [42]. Hence, it is crucial to develop an appropriate licensing strategy for linked data that takes account of the various asset specificities of linked data as intellectual property.[43]

3.4.3.1. Intellectual property of linked data

Semantic metadata is a fairly new kind of intellectual asset that is still subject to debate about the adequate protection instruments [47]. In the European Union the legal framework of property rights related to linked data comprises copyright,[44] database right,[45] competition law[46] and patent law[47]. These appropriative legal regimes are complemented by open access policies and licensing instruments. Table 3 illustrates how linked data assets and intellectual property rights relate to each other.

Copyright protects the creative and original nature of a literary work and gives its holder the exclusive legal right to reproduce, publish, sell or distribute the matter and form of the work. Hence, any literary work that can claim a sufficient degree of originality can be protected by copyright.

Database right protects a collection of independent works, data or other materials that are arranged in a systematic or methodological way and are individually accessible by electronic or other means. Databases are also protected as literary works and need to have a sufficient degree of originality, which requires a substantial amount of investment.

An **Unfair Practices Act** protects rights holders against certain trade practices that are considered unfair in terms of misappropriation, advertising, sales pricing or damages to reputation. Misappropriation

---

[42] See also Archer, et al. (Ref. [42], p. 13), who carried out a study on business models for linked open government data. A discussion of the data broker industry is provided by the US Committee on Commerce, Science and Transportation [43].

[43] A detailed discussion of the licensing issues associated with linked data can be found in Refs [45] or [46].

[44] See Directive 2001/29/EC. See also http://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:32001L0029

[45] See Directive 96/9/EC. See also http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31996L0009:EN:HTML

[46] See consolidated versions of the Treaty on European Union and the Treaty on the Functioning of the European Union – Official Journal C 326, 26.10.2012. pp. 0001–0390. See also http://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:12012E/TXT&from=EN

[47] See http://www.epo.org/law-practice/legal-texts/html/epc/1973/e/ar52.html

TABLE 3: LINKED DATA ASSETS AND RELATED PROPERTY RIGHTS (REF. [39])

	Copyright	DB right	Comp. law	Patents
Instance data	NO	YES	PARTLY	NO
Metadata	NO	YES	YES	NO
Ontology	YES	YES	YES	NO
Content	YES	NO	YES	NO
Service	YES	NO	YES	PARTLY
Technology	YES	NO	YES	PARTLY

especially is relevant to semantic metadata, occuring when data is reused without appropriate compensation (i.e. in terms of attribution or financial return).

Patenting does not directly impact the protection of semantic metadata as — at least in Europe — patents can be acquired soley for hardware related inventions. But as soon as semantic metadata becomes indispensable, is the subject of a methodology that generates physical effects, has a sufficient level of inventiveness and can be exploited commercially, these components can be protected under patent law.

3.4.3.2. Commons' based approaches

The open and non-proprietary nature of Semantic Web design principles allow for the easy sharing and reuse of linked data for collaborative purposes. This offers new opportunities to organizations to diversify their assets and nurture new forms of value creation (i.e. by extending the production environment to open or closed collaborative settings) or unlock new revenue channels (i.e. by establishing highly customizable data syndication services on top of fine granular accounting services based on SPARQL).

Thus, traditional intellectual property rights regimes should be complemented by open access policies and according licensing instruments.[48] Creative Commons[49] allows for the defining of licensing policies for the reuse of work protected by copyright. Open Data Commons[50] does the same thing for assets protected by database right. And open-source licences complement the patent regime as an alternative form of resource allocation and value generation in the production of software and services [44]. Creative Commons and Open Data Commons have gained popularity over the last few years, allowing maximum reusability while at the same time providing a framework for protection against unfair usage practices and rights infringements. Nevertheless, to meet the requirements of the various linked data asset types, a linked data licensing strategy should make a deliberate distinction between the database and the content stored in it. This is necessary as content and databases are distinct subjects of protection in intellectual property law and therefore require different treatment and protection instruments. Additionally, open licences should be provided in machine readable form, using rights expression languages such as CCREL[51] or ODRL[52], so that this information can be used by machines to filter datasets according to their terms of use.

[48] A detailed discussion of licensing issues related to linked open data is provided in Ref. [46].
[49] See http://creativecommons.org/
[50] See http://opendatacommons.org/
[51] See http://www.w3.org/Submission/ccREL/
[52] See https://www.w3.org/community/odrl/

3.4.4. Providing a community norm

In addition to licensing information expressed by Creative Commons and Open Data Commons a so-called community norm is the second component of a linked data licensing policy. Such a norm is basically a human readable recommendation of how the data should be used, managed and structured as intended by the data provider. It should provide administrative information (e.g. creator, publisher, licence, rights), structural information about the dataset (e.g. version number, quantity of attributes, types of relations) and recommendations for interlinking (e.g. preferred vocabulary to secure semantic consistency).

Community norms can differ widely in depth and complexity. Below we present three real world examples that illustrate what community norms look like:

Example 1: A community norm as part of an RDF-S statement by the University of Southampton.

```
36 rdfs:comment "This data is freely available to use and reuse. Please provide an attribution to University of Southampton, if convenient. If you're using this data, we'd love to hear about it at webmaster@ecs.soton.ac.uk. For discussion on our RDF, join http://mailman.ecs.soton.ac.uk/mailman/listinfo/ecsrdf, for announcements of changes, join http://mailman.ecs.soton.ac.uk/mailman/listinfo/ecsrdf-announce."^^xsd:string;
```

Example 2: A community norm provided by http://datahub.io/dataset/uniprot

> Copyright 2007–2012 UniProt Consortium. We have chosen to apply the Creative Commons Attribution-NoDerivs License (http://creativecommons.org/licenses/by-nd/3.0/) to all copyrightable parts (http://sciencecommons.org/) of our databases. This means that you are free to copy, distribute, display and make commercial use of these databases, provided you give us credit. However, if you intend to distribute a modified version of one of our databases, you must ask us for permission first. All databases and documents in the UniProt FTP directory may be copied and redistributed freely, without advance permission, provided that this copyright statement is reproduced with each copy.

Example 3: A community norm provided by the International Press Telecommunications Council (IPTC) for embedding their metadata into media files.[53]

Embedded Metadata Manifesto
How metadata should be embedded and preserved in digital media files
Photographers, film makers, videographers, illustrators, publishers, advertisers, designers, art directors, picture editors, librarians and curators all share the same problem: struggling to track rapidly expanding collections of digital media assets such as photos and video/film clips.
With that in mind we propose five guiding principles as our "Embedded Metadata Manifesto":

- Metadata is essential to describe, identify and track digital media and should be applied to all media items which are exchanged as files or by other means such as data streams.
- Media file formats should provide the means to embed metadata in ways that can be read and handled by different software systems.
- Metadata fields, their semantics (including labels on the user interface) and values, should not be changed across metadata formats.
- Copyright management information metadata must never be removed from the files.
- Other metadata should only be removed from files by agreement with their copyright holders.

More details about these principles:

1: All people handling digital media need to recognise the crucial role of metadata for business. This involves more than just sticking labels on a media item. The knowledge which is required to describe the content comprehensively and concisely and the clear assertion of the intellectual ownership increase the value of the asset. Adding metadata to media items is an imperative for each and every professional workflow.
2: Exchanging media items is still done to a large extent by transmitting files containing the media content and in many cases this is the only (technical) way of communicating between the supplier and the consumer. To support the exchange of metadata with content it is a business requirement that file formats embed metadata within the digital file. Other methods like sidecar files are potentially exposed to metadata loss.
3: The type of content information carried in a metadata field, and the values assigned, should not depend on the technology used to embed metadata into a file. If multiple technologies are available for embedding the same field the software vendors must guarantee that the values are synchronized across the technologies without causing a loss of data or ambiguity.
4: Ownership metadata is the only way to save digital content from being considered orphaned work. Removal of such metadata impacts on the ability to assert ownership rights and is therefore forbidden by law in many countries.
5: Properly selected and applied metadata fields add value to media assets. For most collections of digital media content descriptive metadata is essential for retrieval and for understanding. Removing this valuable information devalues the asset.

[53] See http://www.embeddedmetadata.org/embedded-metatdata-manifesto.php

4. NUCLEAR KNOWLEDGE MANAGEMENT AND SEMANTIC TECHNOLOGIES

4.1. OVERVIEW AND CURRENT STATUS OF NUCLEAR KNOWLEDGE AND INFORMATION MANAGEMENT

Nuclear organizations have adopted many knowledge management methods and techniques and followed well-established good practices that were developed in a generic way without reference to a particular knowledge domain. However, knowledge management in the nuclear field has particular characteristics that have to be considered when adapting knowledge management techniques for nuclear organizations, as described in Ref. [48]. This holds true also for the introduction of semantic technologies, in which the following features should be taken into account when designing new nuclear knowledge management structures or migrating existing ones:

- The nuclear sector typically requires high reliability, validity, actuality and completeness of information.
- There are limitations in the distribution of information such as:
  - Restricted (classified) information on design and operation of the facility;
  - Corporate confidential information (design, financial and technological know-how);
  - Operation/research data, which cannot be distributed outside of the organization.
- Distribution of information to the public domain requires often significant simplifications, which reduce their professional value.
- A wide diversity of information technology is used in different nuclear organizations, sometimes operating aged information technologies on the brink of obsolescence.
- Users in the nuclear field are generally more conservative compared with those in other common business areas, thus the implementation of new information technologies might face some resistance.

These features have played a part in limiting the use of semantic technologies in the field of nuclear knowledge management until now. However, deficiencies in the current knowledge management systems are becoming increasingly apparent, as illustrated by the example of knowledge portals. In many organizations, most of the knowledge management technology is centred on portals, which are constructed on the basis of classical organizational (hierarchical) and thematic structures. One of the principal tasks of such portals is the provision of a central document repository, which may contain thousands of documents, including PDFs and scanned bitmap files from the past with no or a minimum amount of metadata; this impairs the quality of document search and retrieval. Here, the deployment of semantic techniques promises immediate value.

However, the benefits of utilizing new technologies for information and knowledge management go well beyond this single application example. This has been recognized by several nuclear organizations, particularly with respect to modelling knowledge and developing applications on the basis of such models.

This section indicates some of the directions in which semantic technologies might be developing in the nuclear field, with emerging and planned applications. In general, applications of semantic technologies are not specific to a particular domain. Therefore, applications mentioned in the following may hold for many types of organizations; they are, however, of particular importance for organizations operating in the nuclear field. As the application of semantic technologies in the nuclear field is in its early stage, exploration of many other paths beyond the ones indicated in the following is fully under way, and more developments are to be expected.

4.2. SEMANTIC TECHNIQUES PARTICULARLY RELEVANT TO THE NUCLEAR FIELD

4.2.1. Establishing common vocabularies, taxonomies and thesauri

Semantic technologies are particularly effective in providing common vocabularies, taxonomies and thesauri. As many groups of different stakeholders collaborate during the life cycle of a facility, sharing and transferring information is of great importance. By defining the terms used in KOSs, uncertainties with respect to the exact meaning of the term will be removed for all participants in different communities. In addition, language barriers can be removed by translating the definitions into any language.

Efforts to establish a common language within the nuclear community have been undertaken on international, national and organizational levels. Many nuclear glossaries of technical terms, along with their formal definitions, have been produced by the standards development organizations, the IAEA, and many other national and international nuclear organizations. Prominent examples are the IAEA's Safety Glossary and the International Nuclear Information System (INIS) developed by the IAEA and NEA, which provides a large thesaurus of nuclear terms in many languages.

Within a nuclear organization, the effort of structuring knowledge in the form of KOSs is often undertaken in several areas (e.g. management, administration, safety analysis, training, operation, maintenance, engineering, construction, design, supply), and sometimes even on several levels within each area. For instance, one single nuclear power plant (NPP) might operate tens if not hundreds of different databases and information systems, running on various platforms of differing quality. These independently developed KOSs usually contain formal definitions of many terms in different glossaries, leading to overlaps and inconsistencies. This is even more the case when looking at national or international levels, where many examples of different or even contradicting terms may be found in several glossaries. As these technical terms are used in licensing documents, contracts, specifications, design documents and others, problems may arise when interpreting and applying these terms.

Semantic technologies offer solutions to the problem of managing KOSs consistently, and support many additional features. The prerequisite for applying such technologies is the formulation of the KOS in a standardized way. The SKOS standard, in particular, is the best choice for addressing the task of managing vocabularies, glossaries, taxonomies and thesauri. More involved modelling, making use of the extended capabilities of OWL, is usually not required for managing these types of KOSs; the special knowledge required of the OWL language and its features would only be needed when developing ontologies.

In practice, the transfer of an existing vocabulary to SKOS should not present many difficulties, as many vocabularies and thesauri exist in some structured form within MS Word, as PDF files, or in Excel sheets, which usually can be imported by SKOS editors.

Once a vocabulary has been translated into SKOS, the following features are available:

— **Linking between one or more projects:** The concepts of different KOSs may be linked (related to each other) in several ways offered by the SKOS standard: exact matching concepts, near matching concepts, narrower and broader related concepts, related concepts (a general unspecific relationship).
— **Comparing definitions and scopes between different KOSs:** By linking KOSs, their definitions or other attributes such as scope notes may be compared. Variations in definitions of exact matching terms can thereby easily be detected, and homogenized, if necessary (in many cases, differing definitions will reflect the usage of the same term by different user groups and may well be valid within their scope).
— **Integrating concepts defined in other schemas:** Several widely used and accepted ontologies or custom schemes have been developed for the web. These vocabularies should be reused whenever possible in place of recreating existing ones.
— **Attaching linked data sources:** As previously shown, the number of linked data sources is growing at a fast pace. In particular, well known sources such as DBpedia or Geonames may provide additional information to a KOS without much effort on the developer's side. In addition, many organizations

are increasingly publishing their subject matter KOSs either publicly or on private networks. By linking to these KOSs, the original model may be greatly extended to provide a rich interwoven network able to provide related knowledge to the user (‘knowledge discovery’).

— **Multilingualism and synonyms:** In SKOS, a concept is identified by its URI by which it can be retrieved and referred to. For every language considered, one preferred label is attached. In addition, alternative labels may be defined, as well as hidden labels.

— **Ownership and property rights:** Apart from the technical solutions, issues of ownership and property rights have to be observed when interlinking with KOSs from different sources. A reasonable approach would be one in which content owners, who retain the copyright and the control of the content, update and maintain their respective materials, while end users are able to search and retrieve the definition(s) of terms of interest, trace the source of the terms, link to relevant information and publications where the terms were used and access previous revision(s) of terms from the same originator/owner.

Given that KOSs can be linked to produce knowledge descriptions with increasingly larger scopes, the question arises whether, ultimately, a description of the whole nuclear domain could be realized. In theory, it does appear that such an endeavour is possible, provided that the single KOSs adhere to linked data principles, which would enable KOSs developed by different parties to be interlinked. Examples of very large structures have already been established in the medical and biological domain. As producing an in-depth, comprehensive coverage of the whole nuclear domain may require some time, even an effort on a lower scale with less detail would start providing significant benefits: a common (multilingual) vocabulary that can be used in many fields for documentation, communication and development of applications.

4.2.2. Developing knowledge organization systems

Techniques such as mind maps or concept maps are well-established approaches for mapping knowledge (organizing and linking topics and items to form a network). These approaches might be regarded as precursors to modelling knowledge by means of KOSs. KOSs developed according to standards offer significant advantages when describing large, interrelated knowledge domains, since they can be processed by machines, with almost no limitations in extendability and linking.

Describing a complex and extensive knowledge domain, which is usually the case in the nuclear field, requires developing extensive and detailed KOSs. Good practices are available guiding the development of KOSs, usually including the steps of collecting important concepts for the subject under consideration, structuring the concepts in classes and hierarchies, defining attributes for each concept and relations between concepts, and linking concepts to other KOSs.

Still, manually developing a KOS is labour-intensive and costly. However, advances in text analysis and term extraction offer substantial help: a first seed taxonomy of limited complexity and scope might be generated by harvesting available linked data sources and manually refined. In a next step, a body of documents related to the subject at hand can be analysed by text mining techniques, whereby terms significant in this context are extracted and offered to the developer as items to be included in the taxonomy. Such an analysis may be particularly helpful with specialized document corpora which are closely related to a particular, well defined subject, such as a given class of accidents. After some iteration, the number of terms found by text analysis will decrease, indicating that a comprehensive description of the knowledge domain has been achieved.

Today’s search engines make use of methods that allow for (semi-)automatic construction and refinement of knowledge models from information on the Web. For specialized knowledge domains such as the nuclear, such methods are not yet fully applicable. However, the massive effort in developing methods based on artificial intelligence will produce rapid progress in this area.

4.2.3. Integration of heterogeneous knowledge sources

The nuclear domain is particularly well documented, because the organizations in that domain operate in a complex, strongly regulated environment. Every step in the life cycle of a nuclear installation has to be properly documented for purposes of maintaining records on design, the history of the design basis, further developments within phases of the life cycle, and for fulfilling regulatory requirements. A multitude of content management systems and other document repositories are used for this documentation. Even within a single organization, many document repositories exist, distributed among different organizational units, sometimes in different countries and different languages, resulting in a variety of non-interoperable data silos. On a national and international level, this situation is obviously much more prominent.

However, many of the documents and much of the data residing in one repository are related to information in other repositories. Nowadays, the potential of combining data, even very large amounts of data (‘big data’), to extract new information that was not available before, is increasingly recognized.

An example of the value semantic engines could provide by integrating different data sources is in analysing minor incidents at a NPP. Correlating the data from several repositories such as data on the kind of education and training personnel involved in the incident had undergone, the average age of the crew, shift experience with comparable events and other influencing factors could help better determine the root cause of such incidents and the development of appropriate corrective measures.

In view of such potential gains, organizations strive to break the barriers of non-interoperable, heterogeneous data sources. Semantic technologies play a prominent role in integrating heterogeneous sources or making them interoperable. The means of achieving these goals consist in several steps: (i) the data structures of the repositories are analysed (software tools may support this task, e.g. for relational databases), (ii) data from the source system is extracted, (iii) transformed to a format amenable to semantic tools (e.g. RDF), and (iv) loaded into a target system. Software for extraction, transformation and loading (ETL software) is available to perform this task; the RDF triples are stored in No-SQL-databases (triple stores), or even processed on the fly. The consolidated content of the different sources can then be queried by SPARQL, allowing for the issuing of queries on the whole of the triple store, containing all information sources.

4.2.4. Automated indexing, categorization and tagging

Keywords and key phrases consist of a sequence of one or more words that provide a compact representation of a document’s content. Keywords represent in condensed form the essential content (main topics) of a document. Extracting the most relevant words or phrases (keywords) in a document is one of the most essential tasks as they are an important form of metadata. But they also have a variety of indirect uses in performing such tasks as indexing, text classification and categorization, tagging, text summarization, content-based retrieval, topic search, navigation and KOS creation. The effectiveness and relevance depend on discovering a sufficient number of quality key phrases from the text. Though the manual process of providing keywords would be appropriate, the effort and time required with a large collection of data and information resources render it unfeasible. Hence, automatically assigned key phrases provide a highly informative semantic dimension to the documents that would benefit new applications.

Key phrase/keyword extraction involves expressing the main topics using the most prominent words and phrases in a document. Key phrase indexing refers to a more general task where the source of terminology is not restricted as in keyword indexing. Automated indexing is based on natural language processing (NLP) and uses text analytics and mining to extract the entities and list them in the index. It may be done with or without the use of KOS vocabularies (KOS-supported extraction usually produces better results). Key phrase extraction with text analytics can generate terms from text as a source to manually build taxonomies and also to categorize and classify contents against the existing terms in the taxonomy. Automated categorization/classification/tagging performs text analytics or mining to extract concepts from unstructured varied content and makes use of the controlled vocabulary to match to the extracted terms. It is also called term assignment.

The methods of auto-categorization technologies primarily fall into two categories: **machine learning based** and **rules based**. The hybrid method makes use of both. Machine learning based methods automatically categorize on the basis of previous examples. They apply complex mathematical algorithms to learn from multiple representative samples. The system works best if a large collection of pre-indexed records already exists. Rules based auto-categorization works based on rules created for each term in the controlled vocabulary. Rules are generally based on synonyms with more conditions. These systems feature more sophisticated rule writing like advanced Boolean searching, regular expressions and proximity operators.

The relevance for nuclear organizations is obvious: as the number of document and data sources steadily increases, the aforementioned tasks such as classification, retrieval and navigation are increasingly dependent on metadata, which cannot possibly be assigned manually.

4.2.5. Semantic search and artificial intelligence techniques

Semantic search is based on the context, substance, intent and concept of the searched phrase. It not only finds keywords, but determines the intent and context meaning of the search words. The focus is on applying the appropriate context for quickly locating the information users are seeking. The context based search will help users to obtain more meaningful results by finding the most relevant documents in the repository.

Controlled vocabularies are designed to support consistent indexing and end user navigation, browsing and searching. A controlled vocabulary gathers synonyms, acronyms, variant spellings, relations and so on. Semantic search engines utilizing controlled vocabularies enable users to search for information by different names or even misspelled terms and retrieve matching concepts, not just words or phrases. Without semantics, mere text string keyword matches may miss the more relevant concepts and retrieve too much irrelevant content.

The advanced semantic search supports all available metadata fields along with terms from controlled vocabularies. It combines concept based search (restricted to controlled vocabularies) and keyword search in metadata or full text to perform a more focused search on specific categories and concepts. Applying synonymy and homonymy, stemming as a part of the searching process, further enhances the accuracy and relevance of the search results. Features such as query extensions (extending the query by synonyms and/or relations to other terms), autocompletion of search terms (offering search suggestions when starting to type), or taxonomy driven facets to quickly refine and drill down search results are typically provided.

For keyword searches without consideration of term interdependencies, information retrieval models like the standard or extended Boolean models, vector space models, probabilistic models or inference networks are generally used. Reaching out further, generalized and enhanced vector space models, latent semantic and neural network based information retrieval models are capable of representing terms with interdependencies. The learning and generalization capabilities of artificial neural networks are used to build up and employ application domain knowledge required for adaptive information retrieval. In addition, evolutionary techniques like the genetic algorithm can be used to increase information retrieval efficiency and optimize the search results.

The first case study presented in the Annex discusses a realization of such systems with advanced search and retrieval features within a nuclear organization, the Indira Gandhi Centre for Atomic Research (IGCAR). It clearly shows the feasibility of such systems today and the benefits they convey.

4.2.6. Visualization

Knowledge models may quickly become large, with many concepts, attributes and relations. Keeping an overview of KOSs for development and maintenance work can be significantly facilitated by

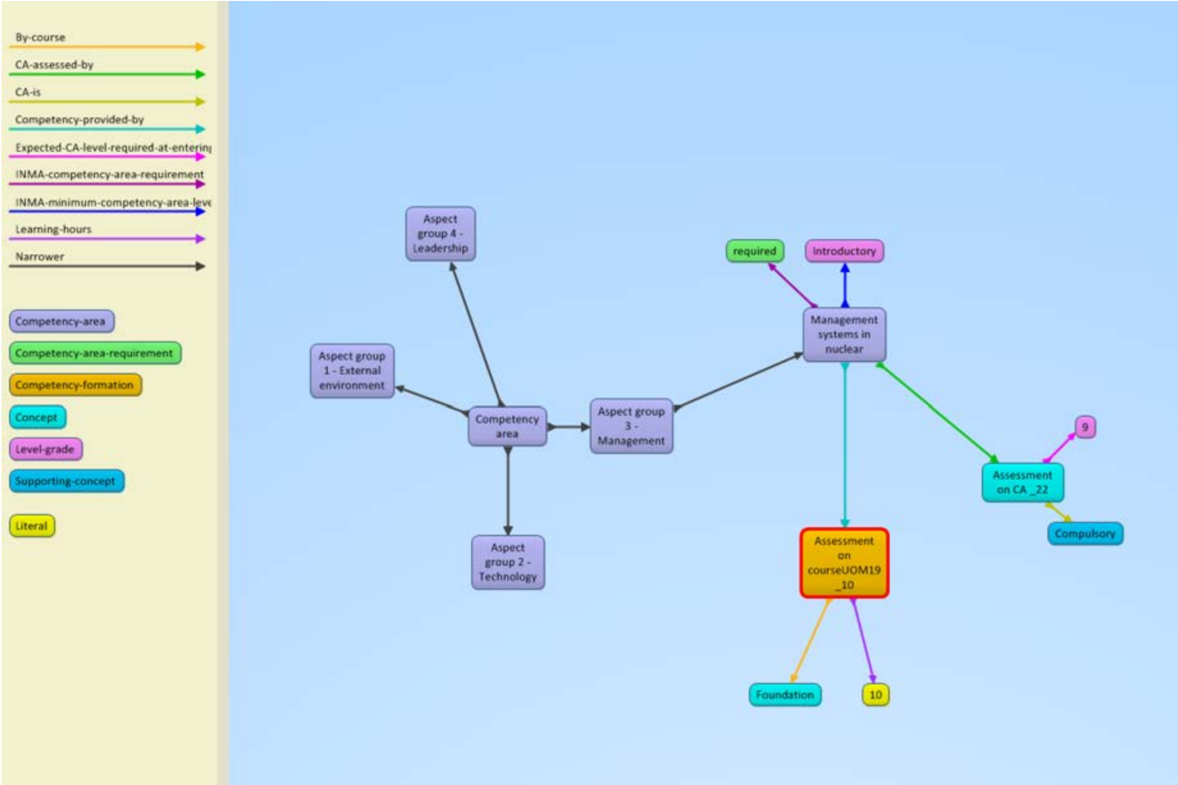


FIG. 8. Graph visualization of part of a knowledge model.

appropriate visualizations. These vizualizations can be further utilized in representing the results derived by querying the knowledge base through SPARQL.

In the case of a simple KOS with few relations, a hierarchical tree visualization may be sufficient. For gaining insights into the structure of more complex ontologies, graph visualizations are usually deployed (see Fig. 8). Graph visualizations may be built into the semantic tools used for developing KOSs or attached to an exported RDF file or to a triple store. Several tools explicitly import RDF formats; for some others a conversion to other graph representation formats is necessary. Graph tools commonly offer several types of layouts, some of them particularly well suited for displaying very large graphs, and provide customizable settings for labels, colours, node and link representations, and many more. Zooming and filtering are indispensable for quickly accessing the requested information.

Visualizing the results of SPARQL queries is mostly equivalent to classical methods of visualizing tabular data (e.g. as line charts, pie charts, bar charts and many other graphical displays). In many instances, the results returned by SPARQL queries have to be transformed into the formats requested by the visualization; however, advanced SPARQL query systems offer direct visualization by choosing the particular graphic display that the user requests.

4.2.7. Text analytics, data mining and knowledge discovery

Text analytics and data mining are defined as the processes of discovering patterns (knowledge) in data. The patterns are valid, understandable, implicit, non-trivial, previously unknown (novel) and potentially useful. The process is generally automatic or semi-automatic as it is difficult to handle huge data manually with traditional techniques. Data mining involves application of more sophisticated computational intelligence algorithms and machine learning techniques in addition to statistical analysis techniques and natural language processing for knowledge discovery.

A knowledge discovery process includes data cleaning, data integration, data selection, transformation, data modelling, pattern evaluation and knowledge presentation. Knowledge discovery techniques perform data analysis and may uncover important data patterns, contributing greatly to knowledge bases and scientific research. This process needs to handle relational and diversified types of data to extract meaningful information. With suitable knowledge representation, real world knowledge can be used for problem solving and reasoning.

The application of hybrid methodologies of soft computing, including neural networks, fuzzy systems and evolutionary computing, provides the power to extract and express knowledge contained in data sets in multiple ways. It improves the performance in terms of accuracy and generalization capability while dealing with highly dimensional, complex regression and classification problems.

The knowledge discovery tools and models help one to discover and extract meaningful knowledge in the form of human interpretable relationships/patterns that are buried in the unstructured or structured document collection, and represent the knowledge components in appropriate ways to facilitate operations like storage, retrieval, inference and reasoning.

The essential aspect for nuclear organizations consists in the capability of these methods to extract new insights and knowledge from the huge amount of data continuously produced by all information sources. Correlating data sources that are currently treated as separate from each other can produce novel and unidentified information (e.g. in the analysis of events or in crisis management), as witnessed by the ever growing big data applications in many scientific and technical domains.

4.3. EXAMPLES OF ONGOING AND PROPOSED APPLICATIONS

4.3.1. Presenting knowledge bases to the users — knowledge portals, wikis, custom applications

As described in previous chapters, knowledge organization systems provide powerful means of modelling knowledge domains. In particular, linked data allows a base KOS to be enriched with other sources: increasingly, many organizations publish their data as linked data. The term ‘publishing’ need not denote making content available to the general public; it may also mean that these data are available to only a limited audience of accredited people. Publicly available sources such as Wikipedia, however, may contribute significantly to enhancing established concepts, leading to rich knowledge bases and enabling knowledge discovery as described in previous sections.

4.3.1.1. Knowledge portals

Knowledge bases provide an efficient backbone for constructing knowledge portals. The knowledge base can be used to structure and navigate the portal according to the given taxonomy, to link portal contents to other knowledge sources and to improve searching with the aid of KOSs.

A KOS describing a knowledge domain is an excellent starting point in constructing a knowledge portal. The taxonomical hierarchy may be immediately translated into pages and subpages. The relations between KOS concepts provide the internal linkage between pages within a portal. External links to linked data from other sources offer direct connections to more information on a particular subject, while attributes of the KOS terms indicate important characteristics of the topic that might be expanded on the term’s page.

The same taxonomy used for auto-tagging provides the backbone for structuring the portal, especially its topical pages. As the taxonomy will often be quite detailed and structured into several levels, the strategy of assigning one page to each entry in the taxonomy might not be practical. However, just addressing the first few levels of the taxonomy will usually provide sufficient structure for the portal’s topical pages.

Another principal task of a portal consists in providing a repository for documents. In organizations using a portal as a centralized organizational memory, the repository (or repositories) will usually grow to very large dimensions. There, the features mentioned in the previous paragraphs on automated indexing, tagging and semantic search will reveal their full potential in improving search, retrieval and knowledge discovery.

Today's portal software products and content management systems support semantic capabilities to some extent. While provisions are usually made for incorporating term management to support manual tagging, features such as auto-tagging, semantic search, inclusion of RDF parsers or linked data connectivity are not yet commonly available. However, third party tools are obtainable to fill these gaps. As an example of such architecture, prototypes of knowledge portals are being developed in the IAEA that link the document libraries of Microsoft SharePoint with the extraction service of the tool used for managing KOSs (PoolParty Server with PowerTagging extension[54]). When a document is uploaded it is sent to the extraction service, which returns the extracted concepts to the library's metadata fields. A semantic search then considers these metadata with high priority, thereby improving the search result and ranking, and providing facets for refining the search based on the concepts in the metadata fields.

4.3.1.2. Wikis

Another way of making knowledge bases provide value to users is to publish them in the form of a wiki. As a guiding principle, a wiki page is assigned to each concept present in the knowledge base. Some items (e.g. the definition of a concept or link related concepts) may be (preferably automatically) derived from the underlying KOS; other page sections will be open to the authors to contribute more information, images and links related to the concept. The hierarchy of pages thus reflects the taxonomic structure of the knowledge base. This proceeding allows for high flexibility in providing content (a wiki is usually used in a collaborative way) and an easy means of maintaining a structure supporting organization of content, navigation and retrieval. Further semantic features may be included by enriching the text with semantic annotations and relations with extensions of the wiki functionality [49].

The wiki of the nuclear knowledge management (NKM) section is an example of utilizing a knowledge base (the NKM knowledge base) to build a wiki.[55] So far, the definitions are automatically synchronized between the KOS management tool (by which the KOS is maintained) and the wiki. Future extensions will deal with synchronizing newly created and deleted pages between the management tool and the wiki.

4.3.1.3. Custom linked data applications

In cases when a knowledge portal or a wiki does not fulfil the needs of the application, a custom linked data application can be created. Several frameworks exist for the development of semantic applications, both open source and commercial. The frameworks offer the freedom to define the required functionality, appearance and user interaction just as any other Web application development platform; in addition, however, they provide an API that allows for accessing repositories compliant with linked data principles and issuing SPARQL queries against them. All data retrieved from (possibly federated) data sources may be processed by the application by standard means of Web programming and presented to the user through the graphic user interface.

4.3.2. Interlinking and comparing KOSs

By publishing KOSs as linked data, they can be straightforwardly compared and interconnected. As KOSs on the same or similar topics are often developed independently, an overlap of the terms is unavoidable. However, the terms might be defined and described differently. In the nuclear domain, this is of particular importance in safety related, authoritative vocabularies, when divergent definitions of

[54] See https://www.poolparty.biz/ and https://www.poolparty.biz/poolparty-powertagging/
[55] See http://wiki-nkm.iaea.org

terms might lead to equivocal interpretation by different parties. This does not necessarily imply that the definitions have to be harmonized: often, the differences in definitions are well justified (e.g. when used in another context).

An initiative has started at the IAEA to collect such authoritative vocabularies, to search for identical or very similar terms and to issue their definitions and descriptions side by side. No attempt is made to harmonize the definitions. As mentioned, there may be good reasons for the diversity; in any case, the decision to modify a definition is left to the authors.

The envisaged availability in the near future of many KOSs covering different areas in the nuclear field opens up the prospect of interlinking these KOSs to produce knowledge models with very broad scopes. As they will be developed independently, these KOSs will undoubtedly overlap. This situation is a common one on the Web, where many KOSs are published on similar subjects. The W3C standards provide the means to deal with this situation. KOSs may be linked by declaring topics in different KOSs as ‘sameAs’, or assigning them relations such as ‘exact matching concepts’, ‘close matching concepts’, ‘broader/narrower matching concepts’ or generic ‘related matching concepts’. This allows for building large, extensible networks of KOSs, which by appropriate harvesting and query can be treated as a single knowledge model.

4.3.3. Competency networks

The potential of semantic technologies lies foremost in their ability to connect disparate information. Thereby, new insight may be gained into systems that interoperate in complex ways. One example is found in competency networks, where different factors are involved: industries looking for qualified staff and advertising their positions by job description; competencies required for given tasks; taxonomies describing required skills and knowledge; training centres and academic institutions providing education and training to gain those skills and knowledge. Combining these aspects may answer questions on existing gaps between job requirements and education, forecast future needs for qualified staff and direct education resources in particular directions.

The means of achieving this connection between areas of interest are provided by semantic technologies. Each of these areas is already well documented in tables, lists and texts: in some instances also by means of KOSs. Correlating these information sources first requires the specification of a common vocabulary. Existing documentation would then be converted to an RDF; the combination with existing knowledge models can be achieved by mappings on the basis of the common vocabulary.

While the need for such systems aligning the requirements of organizations with respect to qualified staff with appropriate education and training is not unique to the nuclear field, it carries particular weight in this domain due to the safety implications present in every stage of design and operation, requiring specific consideration in all nuclear related activities.

4.3.4. Education and training networks

Within the scope of fostering education and training, the IAEA supports the creation and advance of networks such as ANENT (Asian Network for Education in Nuclear Technology), LANENT (Latin American Network for Education in Nuclear Technology), AFRA-NEST (African Network for Education in Science and Technology) and STAR-NET (Regional Network for Education and Training in Nuclear Technology).[56] Their primary objectives are to assist the member countries in building capacity and to develop human and scientific infrastructure through cooperation in education, nuclear knowledge management and related research and training in nuclear technology. The system is highly distributed, offering academic courses, training material and support within the network from partners residing in several locations. To facilitate an overview of all these resources, an information system is needed that combines detailed descriptions of the resources and information on how to access them. For example,

[56] See https://www.iaea.org/topics/nuclear-knowledge-management/nuclear-education-networks

materials should be related to context and subject by topical tags; links to related materials provided; courses denoted by dates, locations, lecturers, required previous certificates and profiles of the people involved (lecturers, students); and more. While a 'classical' database might provide this information, the features realizable by a semantic approach such as a decentralized structure of the information system, the flexibility of changing and enhancing a given common scheme and the interconnection of related information by linked data principles offer significant advantages in providing a comprehensive, up to date overview of the network.

4.3.5. Knowledge and learning objects repositories (K/LORs)

Repositories of learning objects that can be remotely accessed, allowing for the reuse of learning material and its reorganization into new educational formats, are especially important in e-learning environments. Therefore, the IAEA has initiated a project to develop a prototype of a K/LOR intended for the dissemination and preservation of knowledge and open educational resources. The repository has been integrated with a KOS management tool, using the Web service offered by the KOS manager for concept extraction. Thereby, at upload of the document into the repository, the full text of the document is analysed with respect to the keywords present in a given taxonomy. The tagging process has also been extended to include multimedia material, which plays an important role in education, in particular videos of lectures. This is achieved by utilizing tools for transcribing spoken language to text, and then submitting the transcribed text to the Web service in order to obtain the tags in the same way as with textual objects.

This combination of integrated (via service consumption) platforms offers the feature of exposing the metadata in the RDF. Thereby, as demonstrated in the prototype, a SPARQL endpoint could be implemented offering queries not only over the whole of the repository material, but also other external sources of information. The development and implementation of the prototype is described in more detail as a use case in the Annex.

4.3.6. Extracting lessons learned from operational experience and events analysis

Feedback from operating experience (OPEX) is one of the key means of maintaining and improving the safety and reliability of NPP operations and other nuclear organizations. Diverse industries such as aeronautics, chemicals, pharmaceuticals and explosives all depend on OPEX feedback to provide lessons learned about safety that can assist in improving safety performance. An effective OPEX feedback programme helps in improving NPP design, equipment (including mechanical, electrical and instrumentation and control systems) requirements and characteristics, operating and maintenance procedures, and encourages greater proactivity in taking preventive measures to ensure NPPs operation in a safer and more efficient manner. OPEX is also important for improving the methods and tools for analysis, and thus increasing the value and validity of the findings for the analyst. However, effective analysis of incidents and events is knowledge and time intensive, and analysis performed by different parties/stakeholders is often not shared. Moreover, most of the OPEX minor event reports are produced in obsolete and heterogeneous forms. In many cases these reports are not distributed outside a facility (not even at the corporate level). Such practice limits the capabilities to extract useful knowledge using modern achievements in text analysis such as techniques for extracting and searching for concepts as well as finding associations between them. Taking advantage of the capability of semantic technologies to connect different terminologies, which link to ever growing data repositories published on the Web compliant with international Web standards, and to interconnect knowledge models on a multitude of subject matters will give new insights for lessons to be learned. This will also provide broad application scopes, with potentially beneficial impacts in new application areas (e.g. improving the design of systems and components). Semantic engines could provide an analyst working with root cause analysis of events at an NPP with additional correlations; for instance, between the number of events and results of personnel training, average age of the crew and so on. Such unexpected correlations could facilitate better root cause determination and development of the right corrective measures. This is especially valid for minor events analysis since the number of such events is big enough to provide a reliable basis for semantic analysis. Furthermore, the interoperability of minor events repositories will improve analysis and extraction of lessons learned.

4.3.7. Plant information models (PIMs)

In an NPP, multiple information systems and databases from different vendors and for different purposes are used. Most of these systems are not integrated with each other and cannot share plant data throughout their life cycle. This results in redundancies in capturing, handling, transferring, maintaining and preserving a plant's data. Interoperability problems can stem from the fragmented nature of the industry, paper-based business practices, a lack of standardization and inconsistent technology adoption among stakeholders. Recent exponential growth in computer, network and wireless capabilities, coupled with more powerful software applications, has made it possible to apply information technologies in all phases of a facility's life cycle, creating the potential for streamlining historically fragmented operations.

The focal point for consolidating all these diverse data management tasks consists in a plant information model that is comprehensive, detailed, and able to be integrated and interoperable with design requirements, plant design, operations and maintenance processes, as well as databases, document systems and records systems of organizations that own and operate them. These advanced technologies provide an opportunity to radically improve knowledge capture, integration and transfer between stakeholders if industry wide standards are developed and widely used. In consequence, new NPPs are being designed, procured and constructed using modern computer-aided engineering (CAE) and computer-aided design (CAD) systems with three, four and more dimensional modelling along with data, databases and electronic document sources.

Semantic technologies provide the glue to link all the information and documents. As an example, a component of a system, which itself is described within a taxonomy listing all the systems and subsystems in an NPP, may be interlinked with, for example, its design specifications, maintenance protocols, the history of maintenance, operation and failures, or documents pointing to safety relevance and safety analyses. These information sources usually exist in different formats and different repositories, or even reside in outside organizations (e.g. the manufacturer). The role of semantic technologies will be manifold: the taxonomy of systems and subsystems will be particularly useful if based on standards, enabling unequivocal references in the form of URIs, and other information from various systems such as plant data or documents (which by themselves will be identified with unique names) will be linked to it. In the process industry, ISO 15926, a standard for data integration, sharing, exchange and handover between computer systems, defines a reference data library (RDL) containing the terms used within process industry facilities in the form of an extensive ontology, reusable in many applications and for many purposes. The RDL covers many of the artefacts used in engineering, construction and operation of nuclear facilities; the extensions needed for nuclear specific terms are being discussed.

On top of a system based on semantic technology, sophisticated queries will be available to support the generation of reports. The development of applications based on such a PIM will be able to reference all items by their unique identifiers and frame them into a new context.

As yet, efforts in developing PIMs have been centred on utilizing a variety of data sources ('data-centric' PIM). The ultimate goal of the PIM reaches further and is best reflected in the definition of a PIM adopted by the IAEA: "A Knowledge-centric Plant Information Model is a semantically organized set of information describing plant structures, systems and components, incorporating relationships and rules within a knowledge framework that collectively form enriched representations of the plant that provide shared knowledge services and resources over its life cycle" [50]. The transition from a data-centric to a knowledge-centric approach will enable easier, faster, more accurate and sustainable NPP design and design knowledge information sharing, exchange and transfer across the NPP life cycle.

4.3.8. Crisis management, emergency preparedness and Semantic Web

The need to align terminology in the step of processing information during crisis management has been recognized as being of high relevance: "In crisis management, different domain vocabularies are used by different crisis information systems. This presents a challenge to exchanging information efficiently since the semantics of the data can be heterogeneous and not easily assimilated. For example, the word 'Person' can have different meanings — a 'displaced person', 'recipient of aid', or 'victim'. Semantic interoperability is a key challenge to interoperability" [51].

Several publications show that using an ontology helps in sharing and interoperating between several sources of information in crisis management. Still, the question has to be resolved as to which ontologies might be useful in emergencies, as there are no officially registered or recommended ontologies [52]. In the specific case of a nuclear emergency, a KOS would have to be developed by the stakeholders involved. Some additional ontologies in the fields of health care and pathology can then be linked to access additional important information.

5. CONCLUSION

This publication discusses and highlights the benefits of utilizing semantic technologies to manage distributed information and knowledge bases in the context of the nuclear sector. It illustrates how this technology can create new opportunities to derive benefits from large, constantly growing, dispersed data sets and how semantic interoperability and linked data methods can contribute as business enablers and sources of value creation.

It has been shown that, in the specific context of managing nuclear knowledge, many areas exist in which semantic technologies may provide benefits as they can be instrumental in building rich knowledge models, coalescing disparate information sources and supporting applications and services based on them. Such applications and services, with their capacity to capture and synthesize additional meaning from unstructured data and to interlink a multitude of available information sources, will improve the sustainability of managing complex and interdisciplinary nuclear life cycle technology systems, thereby contributing to safety and economics at all phases of the nuclear life cycle.

There is no doubt that linked data is a promising lever to achieve increased organizational efficiency, harmonization of institutional information sources (and systems) and enhanced accessibility and utilization of knowledge resources, all to better empower individuals in their working environment. Semantic and linked data concepts may also touch upon societal issues such as privacy, security and property rights but these go beyond the scope of this publication. The adoption of this technology can greatly improve the efficiency and effectiveness of managing the nuclear knowledge base over the entire life cycle. Member States are encouraged to explore the appropriate application of this technology in their nuclear organizations, to consider utilization of W3C standards and recommendations to improved data integration, sharing and utilization both internally and externally, and to cooperate to pass on their experiences and best practices in doing so.

Appendix

LIST OF AVAILABLE SKOS THESAURI

This is a non-comprehensive list of available SKOS thesauri from various organizations and domain areas.

Agrovoc (UN): Multilingual agricultural thesaurus:
http://aims.fao.org/vest-registry/vocabularies/agrovoc-multilingual-agricultural-thesaurus

Canadian subject headings (CA):
http://www.bac-lac.gc.ca/eng/services/canadian-subject-headings/Pages/canadian-subject-headings.aspx

Common procurement vocabulary (EU):
https://ec.europa.eu/growth/single-market/public-procurement/rules-implementation/common-vocabulary_en

Courts thesaurus (DE):
http://www.thesaurus.com/browse/court

Eurovoc (EU): Multilingual thesaurus of the European Union:
https://op.europa.eu/en/web/eu-vocabularies/
GEMET (EEA): General multilingual environmental thesaurus
https://www.eionet.europa.eu/gemet/

The general Finnish thesaurus (FI)
http://www.nationallibrary.fi/libraries/thesauri/ysa.html

GeoThesaurus (AT): Thesaurus of the geological survey Austria
https://www.geologie.ac.at/en/services/thesaurus/

Getty vocabularies (USA):
http://www.getty.edu/research/tools/vocabularies/

Google product taxonomy (USA):
https://support.google.com/merchants/answer/1705911?hl=en

IPTC NewsCodes (UK):
http://www.iptc.org/site/NewsCodes/

IVOA astronomy vocabularies (UK):
http://www.ivoa.net/documents/WD/Semantics/vocabularies-20080320.html

Jurivoc (CH): Swiss federal court thesaurus:
https://www.bger.ch/index/juridiction/jurisdiction-inherit-template/jurisdiction-jurivoc-home.htm

LCSH (USA): Library of Congress subject headings
http://www.loc.gov/library/libarch-thesauri.html

MeSH (USA): Medical subject headings:
http://www.ncbi.nlm.nih.gov/mesh

NAICS 2012 (USA): North American industry classification system
https://www.census.gov/naics/

NAL Thesaurus (USA): The National Agricultural Library's agricultural thesaurus
http://agclass.nal.usda.gov/

NASA taxonomy (USA):
http://sbir.nasa.gov/topic-taxonomy/52896

ScoT (AU): Schools on-line thesaurus Australia:
http://scot.curriculum.edu.au/

SITC-V4 (UN): The Standard International Trade Classification
http://datahub.io/dataset/sitc-v4

TheSoz (DE): Thesaurus for the social sciences
http://lod.gesis.org/thesoz/de.html

STW Economy (DE): Thesaurus for economics:
http://zbw.eu/stw/versions/latest/about

UKAT UK Archival thesaurus (UK):
http://www.ukat.org.uk/

UNESCO thesaurus (UN):
http://databases.unesco.org/thesaurus/

WAND taxonomies (USA)
http://www.wandinc.com/taxonomies.aspx

World Bank taxonomy (WBG):
http://data.worldbank.org/about/country-and-lending-groups

REFERENCES

[1] BAWDEN, D., Organised complexity, meaning and understanding, Aslib Proc. **59** 4/5 (2007) 307–327.

[2] MAGEE, C.L., DE WECK, O.L., “Complex system classification”, Proc. 14th Ann. Int. Symp. of the Int. Council on Systems Engineering (INCOSE), Toulouse, France, 2004, INCOSE International Symposium **14** 1 (2014) 471–488.

[3] SVENONIUS, E., “Design of controlled vocabularies”, Encyclopedia of Library and Information Science, Marcel Dekker, New York (2003) 822–838.

[4] McHUGH, L., Measuring the Value of Metadata, white paper, Baseline Consulting (2009).

[5] LAUSEN, H., et al., Semantic web portals: State-of-the-art survey, J. Knowl. Manag. **9** 5 (2005) 40–49.

[6] HALFORD, S, POPE, C., WEAL, M., Digital futures? Sociological challenges and opportunities in the emergent semantic web, Sociology **47** 1 (2012) 173–189.

[7] THE ECONOMIST. Data, data everywhere. A special report on managing information (2010), https://www.economist.com/special-report/2010/02/27/data-data-everywhere

[8] RUSSOM, P., Big Data Analytics, TDWI Best Practices Report, The Data Warehousing Institute (TDWI) Research (2011).

[9] RAYFIELD, J., Sports Refresh: Dynamic Semantic Publishing, http://www.bbc.co.uk/blogs/bbcinternet/2012/04/sports_dynamic_semantic.html

[10] DODDS, L., DAVIS, I., MP Data SPARQL Editor,
http://www.guardian.co.uk/open-platform/apps-mp-data-sparql-editor

[11] CARDOSO, J., HEPP, M., LYTRAS, M. (Eds), The Semantic Web: Real-World Applications from Industry (Semantic Web and Beyond), Springer, New York (2008).

[12] WOOD, D. (Ed.), Linking Enterprise Data, Springer, New York (2010).

[13] WOOD, D. (Ed.), Linking Government Data, Springer, New York (2011).

[14] VAFOPOULOS, M., “A framework for linked data business models”, 15th Panhellenic Conf. on Informatics (PCI) (2011),
http://ssrn.com/abstract=1850365

[15] CARACA, J., “Toward a S&T policy for the knowledge-based society”, Science, Technology and Innovation Policy (CONCEICAO, P., GIBSON, D. V., HEITOR, M. V., SHARIQ, S., Eds), Quorum Books, London (2000) 27–34.

[16] VAN DIJK, J., The Network Society, 3rd edn, Sage, London (2012).

[17] SAUMURE, K., SHIRI, A., Knowledge organization trends in library and information studies: A preliminary comparison of pre- and post-web eras, J. Inf. Sci. **34** 5 (2008) 651–666.

[18] HAASE, K., “Context for semantic metadata”, Proc. 12th ACM Int. Conf. on Multimedia, New York, 2004, MM’04, New York (2004) 204–211.

[19] WEIBEL, S.L., LAGOZE, C., An element set to support resource discovery, Int. J. Digit. Libraries **1** 2 (1997) 176–186.

[20] HODGE, G., Systems of Knowledge Organization for Digital Libraries: Beyond Traditional Authority Files, First Digital Library Federation, electronic edn, September 2008. Originally published in trade paperback in the United States by the Digital Library Federation and the Council on Library and Information Resources, Washington, D.C. (2000), https://www.clir.org/pubs/reports/pub91/

[21] HODGE, G.M., ZENG, M.L., SOERGEL, D., “Building a meaningful Web: From traditional knowledge organization systems to new semantic tools”, Proc. 2003 Joint Conf. on Digital Libraries (JCDL’03), Houston, TX, 2003, ACM Press, New York (2003).

[22] BERNERS-LEE, T., Web Architecture from 50,000 Feet (1998), http://www.w3.org/DesignIssues/Architecture.html

[23] GRUBER, T.R., A translation approach to portable ontology specifications, Knowledge Acquisition, **5** 2 (1993) 199–220.

[24] GRUBER, T.R., Toward principles for the design of ontologies used for knowledge sharing?, Int. J Hum.-Comput. Stud. **43** 5–6 (1995) 907–928.

[25] SCHAFFERT, S., GRUBER, A., WESTENTHALER, R., “A semantic wiki for collaborative knowledge formation”, Semantic Content Engineering (REICH, S., GÜNTNER, G., PELLEGRINI, T., WAHLER, A., Eds), Trauner Verlag, Linz (2005) 188–202.

[26] MITCHELL, I., WILSON, M., Linked Data: Connecting and Exploiting Big Data, Fujitsu white paper, (March 2012),
http://www.fujitsu.com/uk/Images/Linked-data-connecting-and-exploiting-big-data-(v1.0).pdf

[27] CRANFORD, S., Spinning a data web, Technol. Forecast (Spring 2009), https://www.pwc.com/cl/es/publicaciones/assets/pronostico-de-tecnologia.pdf
[28] REITBAUER, A., “IT Konsolidierung und Informationsintegration” Semantic Web. Wege zur vernetzten Wissensgesellschaft (PELLEGRINI, T., BLUMAUER, A., Eds), Springer, Berlin (2006) 387–404.
[29] GALINSKI, C., “Wozu Normen? Wozu semantische Interoperabilität?” Semantic Web. Wege zur vernetzten Wissensgesellschaft (PELLEGRINI, T., BLUMAUER, A., Eds), Springer, Berlin (2006) 47–72.
[30] BERNERS-LEE, T., HENDLER, J., LASSILA, O., The semantic web, Sci. Am. **284** 5 (May 2001) 35–43.
[31] BIZER, C., HEATH, T., BERNERS-LEE, T., Linked data – The story so far, Int. J. Semantic Web Inf. Syst., **5** 3 (2009) 1–22.
[32] AUER, S., “Creating knowledge out of interlinked data”, Proc. Int. Conf. on Web Intelligence, Mining Sognal, Norway, 2011, WIMS’11, ACM Press, New York (2011) 1–8.
[33] BLUMAUER, A., “Linked Data in Unternehmen. Methodische Grundlagen und Einsatzszenarien”, Linked Enterprise Data (PELLEGRINI, T., SACK, H., AUER, S. Eds), Springer, Berlin (2014) 3–20.
[34] SKOS Core Guide, https://www.w3.org/TR/swbp-skos-core-guide/
[35] MILES, A., BECHHOFER, S., SKOS Simple Knowledge Organization System Reference, W3C Recommendation (2009), http://www.w3.org/TR/skos-reference/
[36] MADER, C., HASLHOFER, B., ISAAC, A., “Finding quality issues in SKOS vocabularies”, Theory and Practice of Digital Libraries, Lecture Notes in Computer Science 7489 (2012) 222–233.
[37] FREITAS, A., CURRY, E., OLIVEIRA, J.G., O’RIAIN, S., Querying heterogeneous datasets on the linked data Web: Challenges, approaches, and trends, IEEE Internet Comput. **16** 1 (2012) 24–33.
[38] CISCO, Cisco Visual Networking Index: Forecast and Methodology 2011–2016 (2012).
[39] PELLEGRINI, T.; “Semantic metadata in the news production process: Achievements and challenges” (LUGMAYR, A., FRANSSILA, H., PAAVILAINEN, J., KÄRKKÄINEN, H., Eds), MindTrek ’12: Proceeding of the 16th International Academic MindTrek Conference, New York, ACM SIGMM (2012) 125–133.
[40] LATIF, A., US SAEED, A., HÖFLER, P., STOCKER, A., WAGNER, C., “The linked data value chain: A lightweight model for business engineers”, Proc. I-Semantics 2009, 5th Int. Conf. on Semantic Systems, Graz, Austria (2009) 568–577.
[41] KINNARI, T., “Open data business models for media industry – Finnish case study”, Master’s thesis, Department of Information Systems Science, Aalto University (2013), http://epub.lib.aalto.fi/fi/ethesis/pdf/13166/hse_ethesis_13166.pdf
[42] ARCHER, P., DEKKERS, M., GOEDERTIER, S., LOUTAS, N., Study on Business Models for Linked Open Government Data (BM4LOGD - SC6DI06692), prepared for the ISA programme by PwC EU Services (2013), https://ec.europa.eu/isa2/sites/isa/files/study-on-business-models-open-government_en.pdf
[43] COMMITTEE ON COMMERCE, SCIENCE, AND TRANSPORTATION, A Review of the Data Broker Industry: Collection, Use, and Sale of Consumer Data for Marketing Purposes, Staff Report for Chairman Rockefeller (December 18, 2013), https://law.ku.edu/sites/law.ku.edu/files/docs/media_law/2014/Data_Broker_Industry_Senate_Report.PDF
[44] GHOSH, R. A. et al., Economic Impact of Open Source Software on Innovation and the Competitiveness of the Information and Communication Technologies (ICT) Sector in the EU, Final report (November 20, 2006) (Contract ENTR/04/112).
[45] PELLEGRINI, T., ERMILOV, I., Deliverable 7.4: Guide and Best Practices to Licensing Interlinked Data, EU-Project LOD 2, Grant Agreement No: 257943, http://svn.aksw.org/lod2/WP7/D7.4/public.pdf
[46] PELLEGRINI, T., “Linked data licensing — Datenlizenzierung unter netzökonomischen Bedingungen”, Transparenz. Tagungsband des 17. Internationalen Rechtsinformatik Symp. IRIS 2014 (SCHWEIGHOFER, E., KUMMER, F., HÖTZENDORFER, W., Eds), Verlag der Österreichischen Computergesellschaft, Vienna (2014) 159–168.
[47] SONNTAG, M., “Rechtsschutz für Ontologien”, in e-Staat und e-Wirtschaft aus rechtlicher Sicht. Tagungsband des 9. Internationalen Rechtsinformatik Symp. IRIS 2006 (SCHWEIGHOFER, E., LIEBWALD, D., DRACHSLER, M., GEIST, A., Eds), Boorberg Verlag, Stuttgart (2006) 418–425.
[48] INTERNATIONAL ATOMIC ENERGY AGENCY, Knowledge Management and Its Implementation in Nuclear Organizations, IAEA Nuclear Energy Series No. NG-T-6.10, IAEA, Vienna (2016).
[49] SEMANTIC MEDIAWIKI, https://www.semantic-mediawiki.org/

[50] INTERNATIONAL ATOMIC ENERGY AGENCY, Application of Plant Information Models to Manage Design Knowledge through the Nuclear Power Plant Life Cycle, IAEA-TECDOC-1919, IAEA, Vienna (2020).
[51] LIU, S., BREWSTER, C., SHAW, D., "Ontologies for crisis management: A review of state of the art in ontology design and usability", Proc. of the 10th Int. ISCRAM Conf., Baden-Baden, 2013, Karlsruher Institut für Technologie, Baden-Baden (2013) 349–359.
[52] ROLLER, R., ROES, J., VERBREE, E., "Benefits of linked data for interoperability during crisis management", Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci. XL-3/W3 (2015) 211–218.

Annex

USE CASES IN THE NUCLEAR DOMAIN

A-1. USE CASE 1: IT ENABLED NUCLEAR KNOWLEDGE MANAGEMENT SYSTEM AT THE INDIRA GANDHI CENTRE FOR ATOMIC RESEARCH (IGCAR)

A-1.1. Summary/abstract

A knowledge management (KM) system for codification, preservation and utilization of all multidisciplinary knowledge assets accumulated over several decades of nuclear research, development and operation at IGCAR is essential for improved organizational productivity, new insights and high levels of innovation. Realizing the significance of a KM system, an implementation roadmap addressing various challenges related to people, process, technology and resources was formulated and a structured approach was followed for building the system. IGCAR's nuclear KM system, deployed with IT-as-enabler, is built on a federated model consisting of a primary gateway server and a number of secondary level servers having distributed knowledge repositories. A dynamic KM portal with advanced features has been developed and deployed at each organizational entity (group) to acquire, store, share and utilize the organizational knowledge assets in the explicit form of publications, technical design reports, presentations, projects, activities, facilities and so on, along with the tacit knowledge of multimedia modules. The KM portal is designed as a generic, customizable framework and developed fully in house using an open-source platform and application programming interfaces (APIs). The following are the salient features of the portal, which are realized by applying various semantic technologies and standards:

- Taxonomy based controlled vocabulary for knowledge organization;
- Multiformat document upload and conversion facility;
- Automatic extraction of metadata (title, authors, abstract, keywords, etc.);
- Enhanced access control with multilevel rights management;
- Generic keyword based and advanced metadata based search facility;
- Dynamic viewing of metadata and full text of documents;
- On-line analytics and graphical reports generation;
- Customizable menu and contents and interactive user interface.

A-1.2. Organizational context

IGCAR was established in Kalpakkam as a premier research and development (R&D) organization with a mission to conduct a broad-based multidisciplinary programme of scientific research and advanced engineering development, directed towards the establishment of the technology of sodium cooled fast breeder reactors (FBRs) and associated fuel cycle facilities in India. The activities IGCAR carries out include basic and applied research, design, development and applications of materials, techniques, equipment and systems related to fast reactor technology. It is actively engaged in operation and maintenance of the FBRs and design and development of prototype/commercial FBRs to meet the energy needs of the country. At IGCAR, vast knowledge has been accrued from reactor operating experience, engineering experiments, technology development, basic research, modelling and simulations and so on.

The requirement for a robust KM model for leveraging the collective knowledge of the organization to improve its productivity was foreseen some years back at IGCAR. A strategic action plan was devised and an implementation roadmap for a KM system was drafted with a holistic approach to cover people, process, technology and resources, and address the challenges associated with each. A comprehensive

IT enabled KM system has been put in place to cover the various KM processes such as identification, acquisition, creation, storage, retrieval, transfer, sharing, and utilization of knowledge available in explicit and tacit forms.

A-1.3. Main stakeholders/departments

With IGCAR being a multidisciplinary R&D organization, diversified activities are carried out by different organizational entities (groups). The knowledge flow to the repository originates from these groups, which are involved in multidomain activities, as shown in Fig. A–1.

The KM section (falling under the IT group) executes the mandate of devising a strategic action plan and implementing an organization wide IT enabled KM system. The responsibilities of the KM team include promoting a knowledge sharing culture; developing tools and techniques required for effective content management and search; providing a technology platform to facilitate codification, preservation and dissemination of multidisciplinary knowledge assets; and assessing the effectiveness of the KM system.

A-1.4. Organization's KM system

Recognizing the importance of KM, IGCAR has formulated and adopted a policy to create, share, utilize and leverage its organizational knowledge to achieve higher quality, increased productivity and better collaboration through the synergy of knowledge, resources, facilities and employees. The policy states that "IGCAR will consistently endeavour through concerted efforts of all its employees to generate, archive, manage and disseminate the valuable knowledge for improving its productivity and achieve and sustain world class leadership in all its scientific & technological research and development activities."[1] The KM policy is implemented with the complete participation of all employees and with guidance

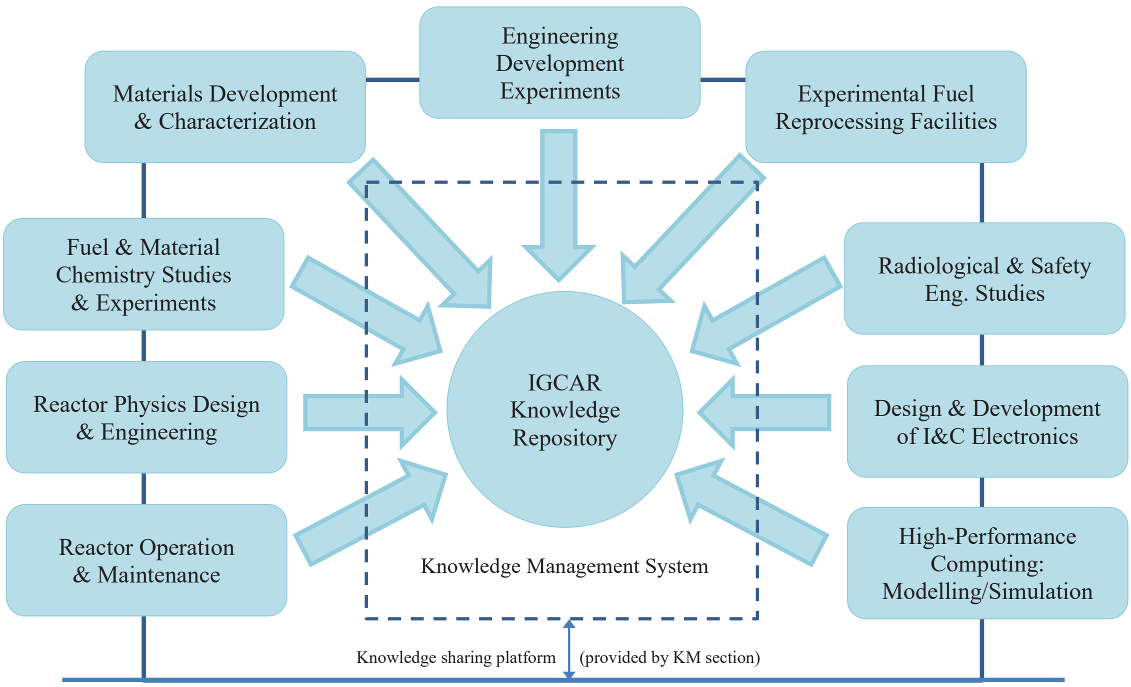


FIG. A–1. Organizational knowledge sources.

[1] http://www.igcar.gov.in/lis/nl92/igc92.pdf

and support from the management at different hierarchies. The KM policy is periodically reviewed to strengthen the activities towards achieving the desired goals.

A knowledge management maturity (KMM) model provides a roadmap for implementing a KM system of an organization in a structured way. A KMM model was developed for IGCAR with six maturity levels, five key areas, twenty key parameters and key maturity indicators. A detailed study was conducted based on the model to assess the current maturity levels for different groups and for the whole organization, identify prominent inhibiting factors and find ways to improve the maturity level of the organization.

A holistic approach for KM implementation covering various dimensions like people, process, technology and resources (knowledge assets) has been adopted with clearly defined deliverables and time frames.

(a) People

KM roles have been defined and responsibilities have been assigned to take care of different aspects like technology development, content creation and management, authentication, monitoring, awareness creation and so on. KM awareness programmes have been conducted to motivate employees and help them to understand the importance of knowledge sharing and thereby build a knowledge sharing organization culture.

A high level committee for information and KM at IGCAR comprising representatives from all groups was constituted to oversee the process of collecting, storing and updating various authentic explicit knowledge assets in the organizational knowledge repository. The committee also monitors the implementation mechanisms and reviews the policy periodically. In addition, group level subcommittees were formed with knowledge officers to manage the acquisition and storing of documents in respective group repositories and keep the contents up to date.

(b) Process

The knowledge life cycle describes the process involved in KM at an individual and organization level. The KM process framework actually defines the different stages of setting up a KM centre. The KM system has been designed to address all micro processes, namely ‘identification’ (determining core competencies and related strategic knowledge domains), ‘acquisition’ (collecting existing knowledge, skills and experiences), ‘selection/store’ (assessment and selection of knowledge for storing), ‘sharing’ (retrieving and making accessible to users), ‘creation’ (uncovering new knowledge) and ‘application’ (using the needed knowledge). Knowledge creation and applied KM revolve around the interplay of tacit knowledge and explicit knowledge and hence the system should support this conversion process.

(c) Resources/knowledge assets

Organization knowledge is the collective sum of infrastructural assets, intellectual capital and personal knowledge. It generally resides in employees’ brains, paper documents, electronic documents and knowledge bases. IGCAR’s explicit knowledge is predominantly available in the form of:

— Technical reports;
— Drawings;
— Manuals/guides;
— Software codes;
— Journal and conference publications;
— Presentations and articles;
— Project progress/review reports;
— Minutes of meetings.

Legacy hardcopy documents were converted into electronic form to enable storing in the knowledge repository. Tacit knowledge is the perceptive/dormant knowledge embedded in contexts and actions. The organization’s implicit knowledge is available in the forms of individual/collaborative experiences, perceptions, bodily skills or mental models. Various avenues are employed to capture individual knowledge and promote creation of organizational knowledge, which include codification and verification of explicit knowledge with domain experts; tacit knowledge elicitation through structured interviews and/or interactive Q&A sessions; discussion forums and technical lectures, colloquiums and/or training.

(d) Technology

The goal of KM technology infrastructure is to facilitate the collection, organization, transfer and sharing of various types of knowledge (explicit and tacit) in secured ways for application or reuse. Information and communication technology is a great enabler in successful implementation of a KM system.

Technology related activities include analysing and evaluating the existing infrastructure, designing the KM architecture, developing the KM system with supportive technologies and tools and deploying the system using results driven incremental methodology. The Web portals provide an easy to use, interactive interface to the users and hence they are commonly used in collecting and disseminating knowledge and information in organizations. Controlled vocabulary like taxonomy is one of the important building blocks of a KM system and it helps to systematically organize the information along with meta knowledge. Application of semantic technologies and standards enhances the features and interoperability of KM portals.

A-1.5. Objectives of the KM initiative

The motivation for the KM initiative at IGCAR comes from a desire to achieve the following objectives:

- Knowledge accumulated over decades of nuclear research, development and operation (organizational memory) has to be preserved and used for the future design, innovations and continued safe operation of nuclear power plants (NPPs).
- Existing skills and competencies have to be retained for a longer period considering the extended time scales of commissioning, service life and decommissioning of NPPs.
- Acquired knowledge has to be transferred to the successors (considering employee attrition) for sustained benefits.
- Organizational learning has to be enabled and knowledge synergized from multidomain R&D activities carried out in the organization to improve its productivity.

IGCAR initiated its KM programme to cater to the knowledge needs of the organization in diverse activities such as research, design, development, project execution and support services by providing the right knowledge to the right person at the right time. The goal is to evolve a KM centre to achieve the organization's mission and vision by effectively utilizing existing knowledge related assets.

A-1.6. Description of the KM initiatives

The IGCAR KM system is designed as a two tier architecture with a federated model of distributed knowledge repositories. It consists of a central gateway server at the primary level and a number of group servers at the secondary level (as shown in Fig. A–2). The structured and unstructured knowledge assets from diversified sources are captured, organized and stored in knowledge repositories distributed across multiple group level servers. The scientific and engineering knowledge repositories maintain explicit knowledge in digital forms of technical reports, journal and conference publications, manuals, drawings, presentations and articles, activities and facilities, project proposals, progress/review reports, software codes, minutes, FAQs and tacit knowledge elicited in the form of audio/video modules. The gateway server maintains the centralized authentication credentials to enable users to access any group level knowledge repository. It also provides a scientific search platform for navigation, efficient retrieval and sharing of knowledge assets distributed across the organization.

A web enabled, taxonomy based KM portal with advanced features was designed, developed and deployed. The KM portal is a generic, customizable framework developed in house fully using open-source technologies and APIs. It is flexible to cater to the requirements of managing content from the diverse knowledge domains of different groups.

The portal provides an interactive and convenient user interface for individual users to upload, organize, list, search, view and share the knowledge assets in different forms and in different file formats with security measures. Also, it enables administrators to carry out tasks associated with content management, user management and rights management with ease of use. Figure A–3 depicts the salient features of the KM portal. The advanced features that are incorporated in the KM portal are highlighted in the following sections.

(a) Taxonomy based KOSs

Hierarchical taxonomy is adopted as a controlled vocabulary in the KM portal and the topics of knowledge assets are organized in an expandable tree structure. Taxonomy provides classification, navigation support, and search and retrieval support. Building taxonomy involves defining the structure for organizing information, specifying individual terms used for classification, and defining the relationships among terms. It is done with due consultation of the domain experts and considering different entities such as scientific processes, products, projects and applications. The taxonomy administration module of the portal facilitates building and maintaining the knowledge taxonomy/map in a consistent way. Presently, the categorization/classification of documents using the taxonomy is done manually and an auto-categorization/tagging module is under development.

(b) Automatic extraction of metadata

An intelligent machine learning algorithm has been developed and implemented in the KM portal to extract the metadata details automatically from the PDF content of publications during the upload. This algorithm performs PDF to text/XML conversion, font analysis, regular expression parsing and pattern identification and filtering to extract metadata details, namely titles, authors, keywords, abstracts, journal/conference names, volumes and dates of publication. The module also applies optical character

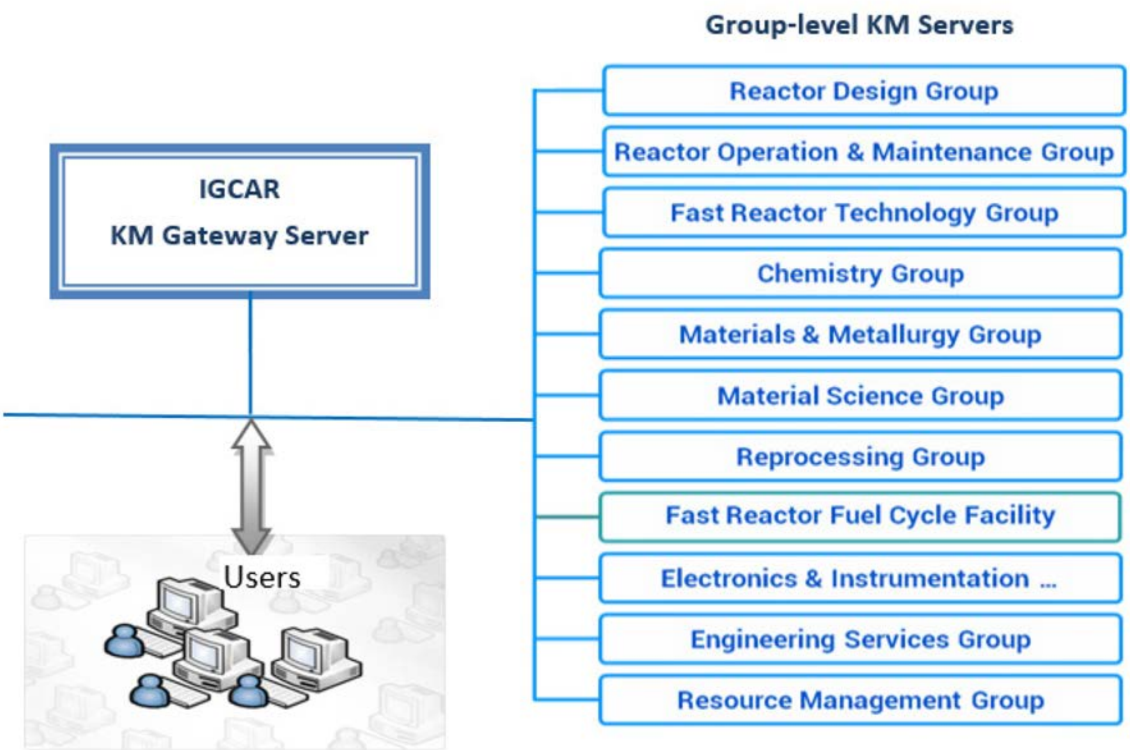


FIG. A–2. Organizational knowledge sources.

FIG. A–3. IGCAR KM portal — salient features.

recognition (OCR) and image processing techniques to extract limited metadata from the scanned PDF content (raster images) of legacy technical reports prepared using predefined templates.

(c) Auto-extraction of index keywords

This feature of the KM portal selects prominent phrases appearing in the document(s) automatically and stores these key phrases as index keywords. Index keywords, in addition to author provided keywords, help improve the accuracy, relevance and speed of the search results. An enhanced rapid automatic keyword extraction algorithm has been used to extract the key phrases, with ranking based on computed weights.

(d) Advanced search engine

An advanced search engine based on an extended Boolean retrieval model has been implemented to perform extensive keyword based searches of the knowledge repository to identify, rank and display the list of relevant documents. It also supports more focused concept based searches based on the subject domains using a taxonomy tree. An advanced option allows one to perform a search based on specific or any combinations of available metadata and produce best or exact matches. Custom filters are provided to narrow down the searches and retrieve the appropriate information quickly.

(e) Enhanced multilevel authorization

The rights management module of the KM portal provides extensive access control over the documents in the knowledge repository. For any document uploaded to the portal, an access control list can be created based on different categories of users (owner/author, group, others) to permit or restrict different operations (list, view, download). The group can be an administrative entity or a customized logical group of members. To protect the confidential information in the repository additional data security and authentication mechanisms are put in place.

(f) On-line copy-safe viewing of documents

The on-line viewing module of the KM portal enables authorized users to view the metadata and full text of the documents. The full text view module converts the PDF pages into lightweight images on-line and displays the page images with scroll, zoom and shrink options. The copy-safe viewing restricts users from saving or printing the documents and also eliminates the need for a PDF viewer (browser plugin).

(g) Dynamic analytics and reporting

This module aids users in performing bibliographic analysis of knowledge assets from different dimensions and generates various textual/graphical reports dynamically. The live statistical reports can be generated based on different parameters such as users, sections, divisions and groups. The analytics module is also provided with drill down features and custom filters to narrow down specific details.

A-1.7. Major achievements and benefits derived

KM is the process of creating value from an organization's tangible and intangible assets. IGCAR has embraced KM as a key strategic initiative to enhance the performance of the organization by leveraging its core competencies. IGCAR has explained its vision and KM strategy in clear terms to all stakeholders. The KM policy was framed and implemented and suitable mechanisms have been established to periodically review the activities to make them more effective.

At IGCAR, knowledge resources are known and needs are well understood. An advanced IT enabled KM system has been successfully put into operation to deal with information flow from distributed sources to a reusable repository. The documented and digitized explicit knowledge assets are acquired, preserved and disseminated using this system. Also, specific solutions have been found to capture tacit knowledge.

The major benefits derived through effective KM include:

— Creating organizational memory and enabling organizational learning with the participation of all members of the organization;
— Converting knowledge into intellectual capital and leveraging it to realize the organization's objectives and mission;
— Improving productivity and innovation capabilities through reuse of knowledge;
— Improving team communication and collective problem solving to enhance work quality;
— Reducing design cycle time and improving the efficiency and safety of NPP operations;
— Reducing the impact of employee attrition and nuclear knowledge loss.

A-1.8. Challenges addressed and knowledge derived from experiences

Developing organization culture: Building knowledge compatible human culture in the organization is one of the foremost challenges in making the KM system successful; this cannot be achieved with technology and tools as it depends on people and their relationships with one another. An organization must build an environment that motivates people to learn, share, change and improve with knowledge; rewards active participants and contributors; and makes people realize that knowledge sharing empowers the organization as well as individuals.

Determining the technology infrastructure requirements: Evaluating the existing infrastructure, and sizing and selecting technology infrastructure in terms of tools, techniques, hardware and software resources is a challenging task. Up to date content, standards, interoperability and support are key concerns in determining technology components.

Data relevancy and accuracy: The available data and information in the knowledge repository should provide relevant and accurate answers to the queries made by the users. The information generated by different groups within the organization may need to be validated before being harvested and distributed. Keeping the contents current by adding new data and eliminating outdated data is a continuous process.

Enabling secured access to knowledge resources: Providing the right level of security for KM is a key issue. Classified information should be protected, and knowledge assets should be accessed with proper credentials. User privileges and configurable access control rules are essential for enforcing the required security in a KM system.

Impact assessment of a KM system: Developing metrics or performance indicators for the assessment of the KM solution is a challenging job as it cannot be easily quantified. A well defined framework with effective measuring and reporting systems based on a set of monitoring criteria and methodology helps in assessing the performance impact of the system qualitatively and making further refinements to the system.

A-2. USE CASE 2: THE IAEA KNOWLEDGE/LEARNING OBJECTS REPOSITORY

A-2.1. Objective

This use case stems from a recent experience at the IAEA creating a repository for the dissemination and preservation of knowledge and open educational resources digital assets. Following current world tendencies and adhering to the Open Access initiative, the agency implemented a repository platform based on open-source software. The project consists in creating a pilot with a set of educational and training material.

A-2.2. Solution design

For implementation, the DSpace software platform was chosen, a well known, mature repository platform with more than 1000 implementations around the world, a strong community of users and developers and the precedent of several international organizations using it as an institutional repository. The implementation followed the model of the JSIC Jorum repository of learning objects.

For the prototype development, the nuclear knowledge management simple knowledge organization system (SKOS) taxonomy was used; it was created by the the IAEA's nuclear knowledge management (NKM) section using the PoolParty[2] semantic suite. In addition to offering the aforementioned functionality for creating taxonomies/ontologies, this tool offers a service via an API for auto-tagging a chunk of text or an entire document. The extracted terms correspond with the taxonomy created using the same platform. One of the challenges of the prototype consisted in integrating DSpace with the extraction service of PoolParty, consuming the tags extracted from the full text of the documents, and offering them to the user for confirmation.

The metadata schema adopted was the IEEE learning object metadata (LOM) schema for educational resources and the extended Dublin Core for description of the resource metadata. A further requirement consisted in making the content of the repository accessible as a SPARQL end point. From the 5.x version on, DSpace includes an RDF module that allows for mapping and exposing the metadata of the items in the RDF. On that basis, it was possible to create a SPARQL end point offering SPARQL queries of all material.

As a starting point, the material to be included in the repository was uploaded to a Web file sharing service. This was consumed via a batch process, converted to an acceptable format to be ingested in DSpace and tagged by the PoolParty service. Using this procedure, nearly 200 items have been uploaded; currently, the resources are manually curated to complete some metadata (i.e. author, issue date, type of resource).

As part of the customization process, the normal submission process of DSpace was modified, beginning with the step of uploading a file (see Fig. A–4). After upload, DSpace extracts the full text of the document (in this pilot only for PDF files) and sends it to PoolParty.

When the upload of the file is completed and the PoolParty API responds, DSpace offers the input form with the pre-added tags obtained to the user for confirmation (see Fig. A–5).

[2] See https://www.poolparty.biz/

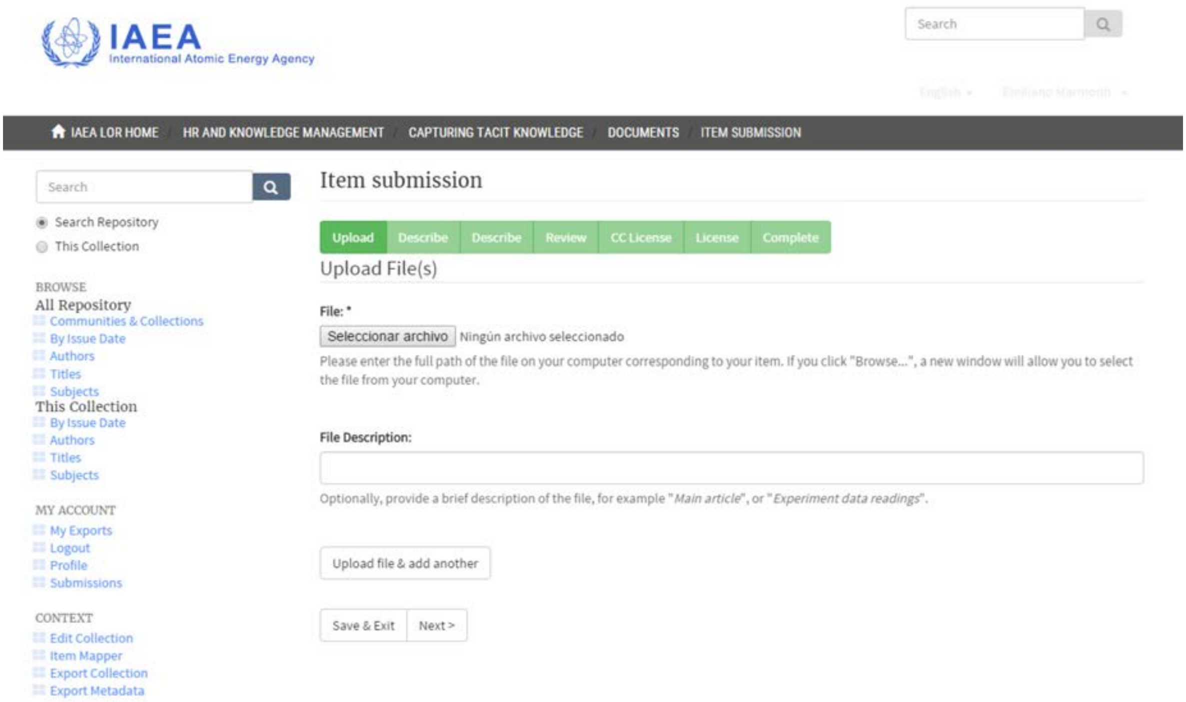


FIG. A–4. The submission process starts with uploading the document, before the metadata input.

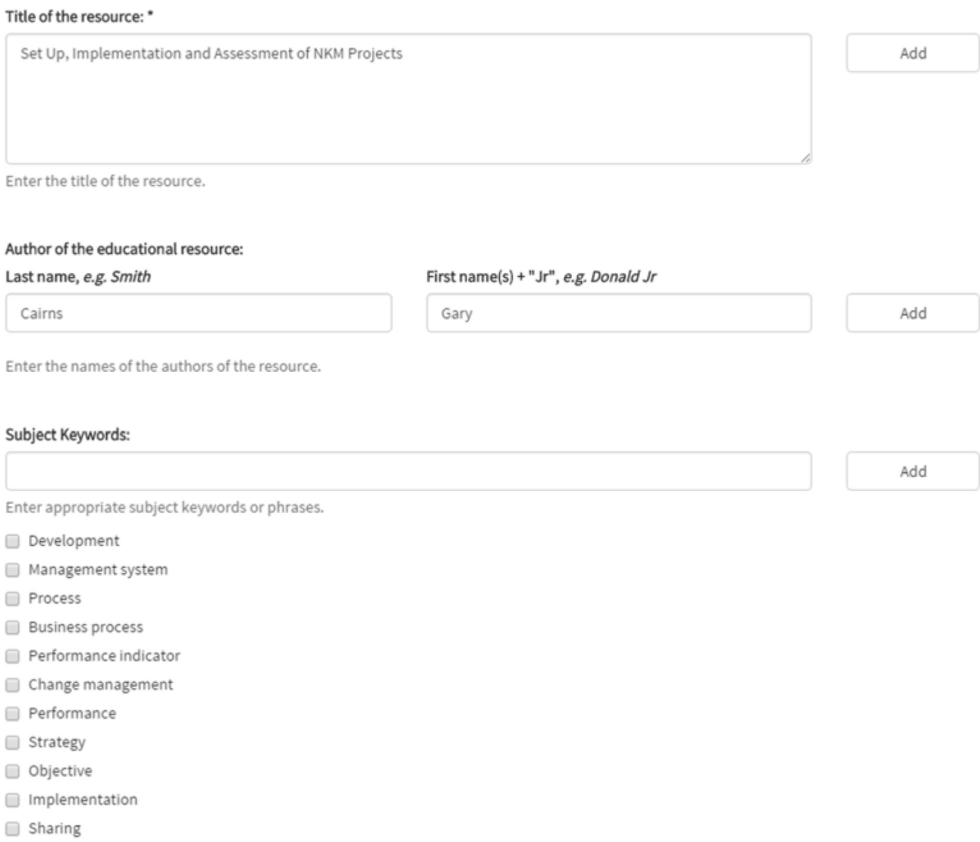


FIG. A–5. Form to confirm the extracted keywords.

If the user wants to add new tags, a new subject can be optionally inserted, or the NKM taxonomy used. To this purpose, the NKM taxonomy was processed and converted into a common controlled vocabulary of DSpace (see Fig. A–6).

By means of the Java libraries included in DSpace it is possible to enhance the search features and offer related terms based on the NKM taxonomy to the user. For this prototype, the decision was made to include skos:related and skos:altLabel for searching related concepts and synonyms.

A-2.3. Special case: Multimedia material

Multimedia material plays an important role in education and training. A requirement for developing the prototype consisted in including videos to be classified and searched for with the same facilities as textual resources. Therefore, the extraction of audio information was included in the batch process, which provides the input for an open-source tool to obtain the text transcription of the video. In a next step, the text transcription is sent to PoolParty in order to obtain the tags in the same way as with text/PDF files. In this prototype stage, all of the intermediate objects are currently stored in the repository to be validated and enhanced; it will not be necessary to keep these objects in a production version. For the time being, this ingestion process is still a batch process due to the several processes applied to the material and the size of the files.

Figure A–7 shows the appearance in DSpace of a multimedia item included in the repository.

A-2.4. Linked data principles: RDFization of the repository

Using the RDF module that comes with the DSpace platform, each item will have its own RDF export which conforms to the linked data principles: Every label is converted to a uniform resource identifier (URI); every URI is an HTTP resource that can be explored by the user.

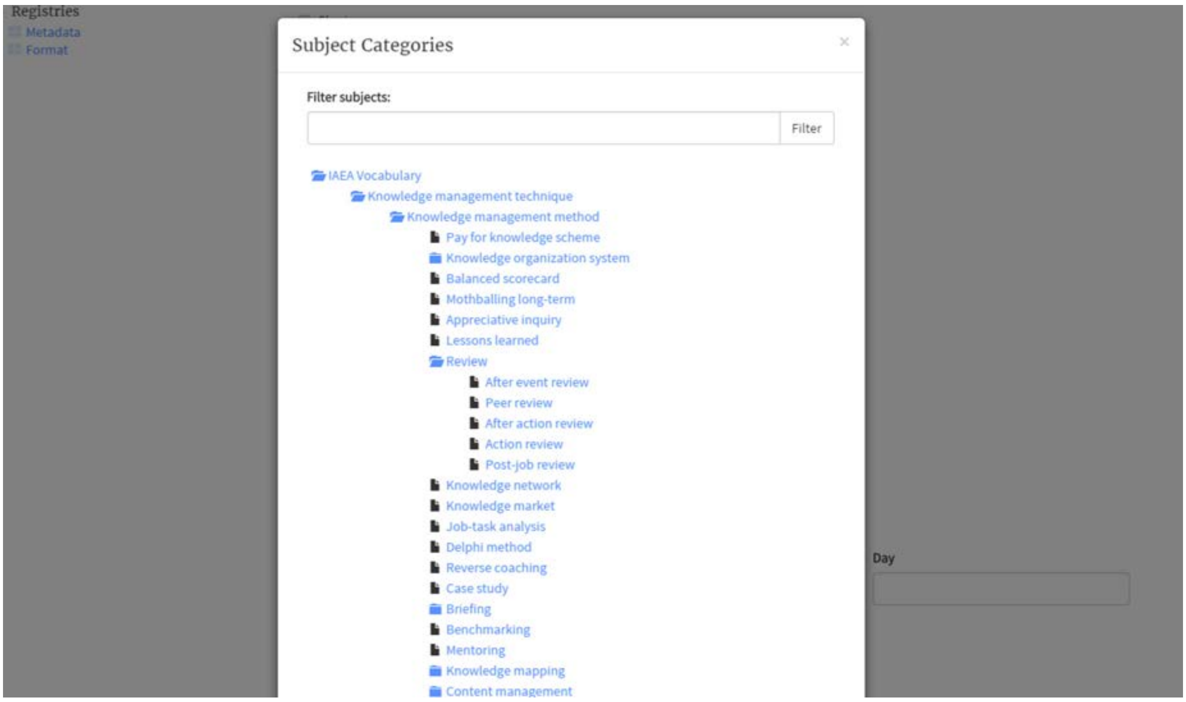


FIG. A–6. The SKOS taxonomy is offered to the user for selecting appropriate terms for indexing.

```
@prefix void:  <http://rdfs.org/ns/void#> .
@prefix rdf:   <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix xsd:   <http://www.w3.org/2001/XMLSchema#> .
@prefix skos:  <http://www.w3.org/2004/02/skos/core#> .
@prefix dcterms: <http://purl.org/dc/terms/> .
@prefix rdfs:  <http://www.w3.org/2000/01/rdf-schema#> .
@prefix bibo:  <http://purl.org/ontology/bibo/> .
@prefix lom:   <http://ltsc.ieee.org/rdf/lomv1p0/lom#> .
@prefix iaeankm: <http://www.iaea.org/rdf/metadata#> .
@prefix foaf:  <http://xmlns.com/foaf/0.1/> .
@prefix dspace: <http://digital-repositories.org/ontologies/dspace/0.1.0#> .
@prefix dc:    <http://purl.org/dc/elements/1.1/> .

<http://             /rdf/resource/123456789/340>
        rdfs:type                "Presentation" ;
        dspace:hasBitstream      <http://            /iaea/bitstream/123456789/340/1/D5.NKM%20
        dspace:isPartOfCollection  <http://          /rdf/resource/123456789/110> ;
        dc:contributor           "Kosilov, Andrei" ;
        dc:date                  "2016-01-25T22:35:27Z"^^xsd:dateTime ;
        dcterms:available        "2016-01-25T22:35:27Z"^^xsd:dateTime ;
        dcterms:creator          "Kosilov, Andrei" ;
        dcterms:hasPart          <http://            /iaea/bitstream/123456789/340/1/D5.NKM%20
        dcterms:isPartOf         <http://            /rdf/resource/123456789/110> ;
        dcterms:subject          "Implementation" , "Database" , "Process" , "Change management
, "Capability" , "Objective" , "Benchmarking" , "Behaviour" , "Norm" , "Collaboration" , "Continu
learned" , "Nuclear knowledge" , "Business process" , "Sharing" , "Performance indicator" , "Stra
        dcterms:title            "Set Up, Implementation and Assessment of NKM Projects" ;
        bibo:uri                 <http://            /iaea/handle/123456789/340> ;
        void:sparqlEndpoint      <http://            /fuseki/dspace/sparql> ;
        skos:Concept             <http://nkmpp.cloudapp.net/NKM-Model/235> , <http://nkmpp.clou
<http://nkmpp.cloudapp.net/NKM-Model/192> , <http://nkmpp.cloudapp.net/NKM-Model/339> , <http://n
<http://nkmpp.cloudapp.net/NKM-Model/200> , <http://nkmpp.cloudapp.net/NKM-Model/131> , <http://n
<http://nkmpp.cloudapp.net/NKM-Model/248> , <http://nkmpp.cloudapp.net/NKM-Model/198> , <http://n
<http://nkmpp.cloudapp.net/NKM-Model/302> , <http://nkmpp.cloudapp.net/NKM-Model/247> , <http://n
<http://nkmpp.cloudapp.net/NKM-Model/265> , <http://nkmpp.cloudapp.net/NKM-Model/189> , <http://n
<http://nkmpp.cloudapp.net/NKM-Model/292> , <http://nkmpp.cloudapp.net/NKM-Model/294> , <http://n
        foaf:homepage            <http://            /iaea> .
```

FIG. A–7. RDF expression of a repository item.

A-2.5. Key benefits of using Semantic Web technologies

The main benefits in deploying semantic technologies are seen in the following:

— To accelerate the submission and annotation of a resource, resources are immediately pre-indexed at upload, with the help of the PoolParty service, based on a curated and maintained taxonomy.

— The repository implements a SPARQL end point and is integrated into the linked data cloud. Users may initiate semantic queries from it; additionally, the end point may contribute answers to queries initiated by a remote SPARQL end point. In the future, as educational organizations publish additional information as linked data, complex semantic queries might be posed; for instance: “Give me all the IAEA’s KM resources produced by Canadian authors that have experience in nuclear fuel treatment and that cover tacit knowledge capture, where the licence allows me to modify the content” or “Give me all the open educational resources for students under the age of 25 with examples from NPPs in Europe that started operation after the year 2000”.

— The end users take advantage of the IAEA’s work in creating taxonomies. These tools are used to expand queries and discover resources that are not commonly shown as the results of a normal query.

— Using a common, established vocabulary increases the quality of the metadata.

ABBREVIATIONS

API	application programming interface
DSP	dynamic semantic publishing
HTML	hypertext markup language
HTTP	hypertext transfer protocol
IGCAR	Indira Gandhi Centre for Atomic Research
KM	knowledge management
KOS	knowledge organization system
NKM	nuclear knowledge management
NPP	nuclear power plant
OWL	web ontology language
PIM	plant information model
RDF	resource description framework
RDF-S	resource description framework schema
SKOS	simple knowledge organization system
SPARQL	recursive acronym for ‘SPARQL protocol and RDF query language’
SQL	structured query language
TTL	time-to-live
URI	uniform resource identifier
W3C	World Wide Web Consortium
XML	extensible markup language

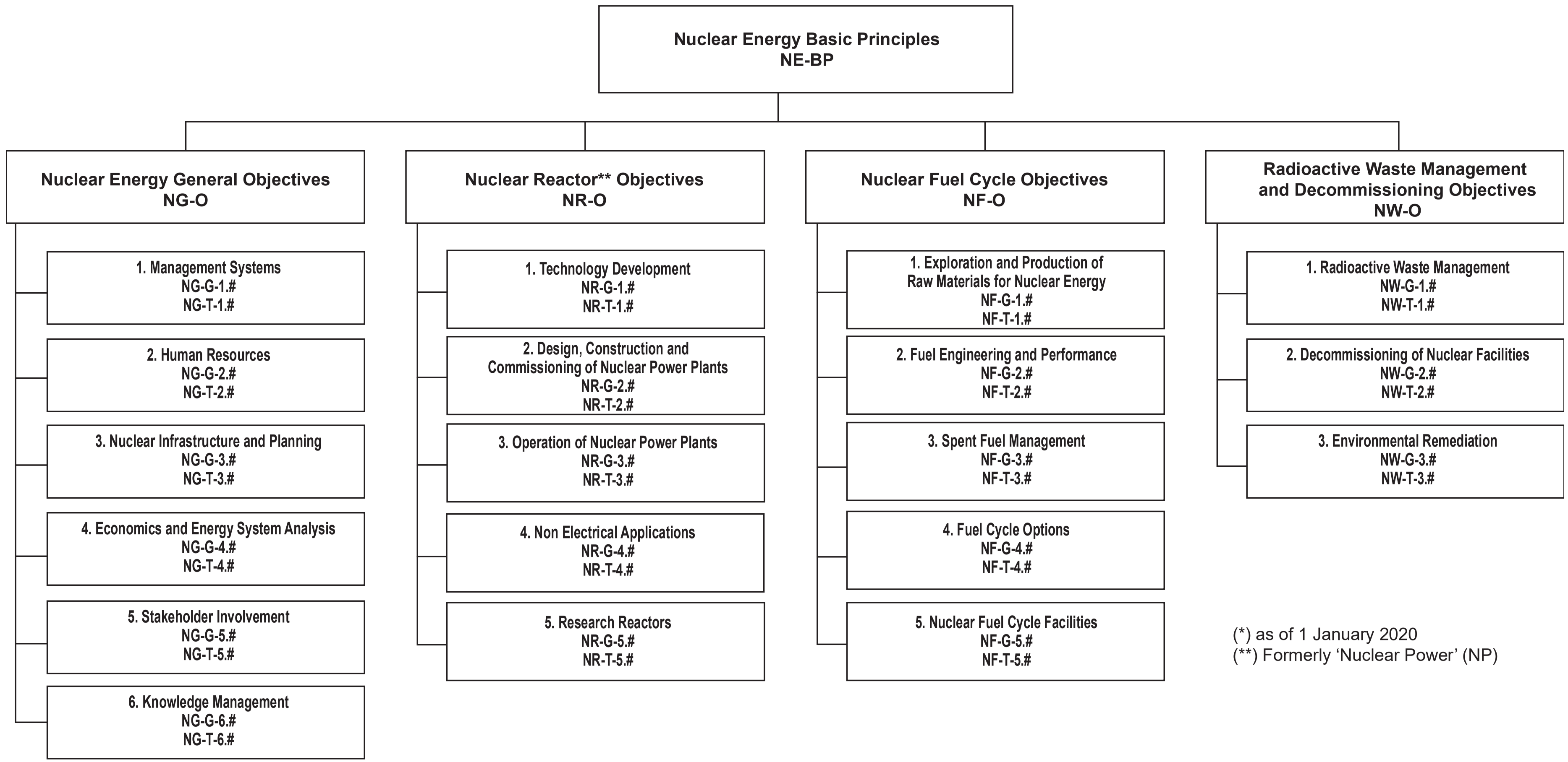
CONTRIBUTORS TO DRAFTING AND REVIEW

Beraha, D.	International Atomic Energy Agency
Gladyshev, M.	International Atomic Energy Agency
de Grosbois, J.	International Atomic Energy Agency
Hakopov, Z.	International Atomic Energy Agency
Jehadeesan, R.	Indira Gandhi Centre for Atomic Research, India
Markov, A.	Institute of NPP Operation Support, Ukraine
Marmonti, E.	Private Consultant, Argentina
Nenadic, G.	University of Manchester, United Kingdom
Pellegrini, T.	St. Pölten University of Applied Sciences, Austria

Consultants Meetings

Vienna, Austria: 11–13 March 2015, 1–5 February 2016

Structure of the IAEA Nuclear Energy Series*



(*) as of 1 January 2020
(**) Formerly 'Nuclear Power' (NP)

Key

BP: Basic Principles
O: Objectives
G: Guides and Methodologies
T: Technical Reports
Nos 1–6: Topic designations
#: Guide or Report number

Examples

NG-G-3.1: Nuclear Energy General (**NG**), Guides and Methodologies (**G**), Nuclear Infrastructure and Planning (topic **3**), **#1**
NR-T-5.4: Nuclear Reactors (**NR**)*, Technical Report (**T**), Research Reactors (topic **5**), **#4**
NF-T-3.6: Nuclear Fuel (**NF**), Technical Report (**T**), Spent Fuel Management (topic **3**), **#6**
NW-G-1.1: Radioactive Waste Management and Decommissioning (**NW**), Guides and Methodologies (**G**), Radioactive Waste Management (topic **1**) **#1**

IAEA
International Atomic Energy Agency

No. 26

ORDERING LOCALLY

IAEA priced publications may be purchased from the sources listed below or from major local booksellers.

Orders for unpriced publications should be made directly to the IAEA. The contact details are given at the end of this list.

NORTH AMERICA

Bernan / Rowman & Littlefield

15250 NBN Way, Blue Ridge Summit, PA 17214, USA
Telephone: +1 800 462 6420 • Fax: +1 800 338 4550
Email: orders@rowman.com • Web site: www.rowman.com/bernan

REST OF WORLD

Please contact your preferred local supplier, or our lead distributor:

Eurospan Group

Gray's Inn House
127 Clerkenwell Road
London EC1R 5DB
United Kingdom

Trade orders and enquiries:

Telephone: +44 (0)176 760 4972 • Fax: +44 (0)176 760 1640
Email: eurospan@turpin-distribution.com

Individual orders:

www.eurospanbookstore.com/iaea

For further information:

Telephone: +44 (0)207 240 0856 • Fax: +44 (0)207 379 0609
Email: info@eurospangroup.com • Web site: www.eurospangroup.com

Orders for both priced and unpriced publications may be addressed directly to:

Marketing and Sales Unit
International Atomic Energy Agency
Vienna International Centre, PO Box 100, 1400 Vienna, Austria
Telephone: +43 1 2600 22529 or 22530 • Fax: +43 1 26007 22529
Email: sales.publications@iaea.org • Web site: www.iaea.org/publications

INTERNATIONAL ATOMIC ENERGY AGENCY
VIENNA