Digital Repository as Instrument for Knowledge Management

Zaven N. Hakopov

International Atomic Energy Agency, Vienna, Austria

Abstract. In the modern technologically advanced world, implicit knowledge, but also certain manifestations of tacit knowledge, is accumulated primarily in digital form, increasing the dependence of Knowledge Management (KM) on tools and specifically on digital content management platforms and repositories. The latter, powered by subject classification system such as a thesaurus or an ontology, can form a complete *Knowledge Organization System* (KOS). The purpose of this paper is to describe and (re)define the role of these systems as an integral part of KM, and present an example of such a KOS, including its major role in knowledge preservation.

Keywords. Digital repository, knowledge management, knowledge organization system, semantic technologies

1 Knowledge management challenges and toolset (in)dependence

Generally speaking, *Knowledge Management* (KM) regulates chiefly organizational and infrastructural mechanisms to capture, retain and disseminate knowledge and is, as such, not primarily derived from information technology; however, it profits greatly and is strongly dependent thereon. As common as it isto stress the non-technological components of KM, such as organizational and infrastructural, which are undeniably key, information repositories are just as vital to a common conceptual KM implementation and maintenance. Major and critical amounts of codifiable corporate knowledge comprise information which needs to be archived in an organized and sustainable manner.

Since most of the implicit knowledge which can be codified, as well as major segments of tacit knowledge which are captured on non-conventional media and digitized, is stored in digital libraries or repositories, it is valid to assume that a modern KM strategy cannot be fully complete without thorough planning and implementation of such repositories, which generally leads to increased requirements to the very core of their functionality, supported features, sustainability and compatibility planning.

2 Evolution of repositories and changing roles

As archives evolved with the progress and development of information technology, catalogues were gradually evolving from analog to digital. Their sole purpose, however, did not change that much: to serve as storage (archival) and provide retrieval of information. On the back-end, several additional processes complete an operational repository: information acquisition and cataloguing. This acquisition takes the form of either direct ingestion of information submitted by single entities, or harvesting of – usually large volumes of - information from other sources in an automated manner through gateways that operate according to certain standards.

The organization of repository workflow usually, and not surprisingly, resembles library operation. No wonder – most evolved from conventional library structures.

Despite major technological advancements, many workflows have not changed since analog times, and often a synergy, or a mixture of old and new procedures, is observed in many present-day digital libraries and repositories, especially in the operational structure of the back-end i.e. processing, proofing, formatting and cataloguing, as well as indexing workflows. One should also note that the front-end and retrieval advances and progresses faster, since it is visible to the end-users and is very dependent on technology. More features are being added to the front-end, format and display of retrieved information, flexible search engines are employed for faster search times – which all leads to much higher requirements for the modernization and implementation of cutting-edge technology.

All in all though, repositories have definitely progressed, and nowadays provide highly technological digital information storage possibilities with a vast amount of powerful features to acquire, structure, index, format and retrieve information. One of the most important developments of modern repositories that is relevant for the purpose of their positioning as an instrumental part of KM is, without a doubt, the increasing support for a variety of data objects which can be stored and indexed. If the main emphasis of original digital repositories and libraries was on bibliographic metadata, nowadays it is more and more focused on the storage of media, data, code, accompanying notes and other types of “knowledge objects”. It is, therefore, quite natural to review the changing role of repositories as knowledge object providers in the KM instrumentation and preservation paradigm.

3 Repository as a Knowledge Organization System

A modern digital information repository, apart from serving its direct purpose, may assist the users in a much broader way. Combined with expert classification systems like ontologies and powered by modern semantic technology, it is more than just an electronic archive, but rather forms a complete *Knowledge Organization System* (KOS). Here, one should note that for the purpose of this paper, the interpretation of KOS is re-defined in a more granular manner than it is usually referred to. For example, according to Gail Hodge [1],

> *“The term knowledge organization systems is intended to encompass all types of schemes for organizing information and promoting knowledge management. Knowledge organization systems include classification and categorization schemes that organize materials at a general level, subject headings that provide more detailed access, and authority files that control variant versions of key information such as geographic names and personal names. Knowledge organization systems also include highly structured vocabularies, such as thesauri, and less traditional schemes, such as semantic networks and ontologies.”*

When discussing digital repositories, it is assumed that the schemes for classification and categorization are an intrinsic part of their implementation, just as authority files are. Together with the structured vocabularies which are then used to actually organize the knowledge stored in the repository, these repositories form a system, which we refer to as KOS in its specific form.

Provided necessary infrastructure and guidelines for its population and usage are in place, such a KOS empowers end-users to quickly and efficiently retrieve information needed for KM propagation, and increases the knowledge turnaround by providing dedicated, well-defined and well-indexed collections according to specific needs. Apart from its core functionality, it also has the very interesting and often disregarded effect of promoting KM discipline. In a way, knowledge - manifested in the form of information which needs to be accumulated and disseminated - has to be provided by the knowledge carriers in a structured manner, so the very necessity to submit information already carries an implicit KM disciplining function.

Knowledge preservation is another vital part of the KM strategy, which should be assumed to be a natural or intrinsic part of the repository functionality – after all, it is a general prerequisite that objects stored to be retrieved at a later time, should be preserved. Although no guarantees can exist for longer-term preservation of digital objects, very serious efforts are being made in this area to address issues like sustainability, physical media, access and many others which form the overall concept of a *trustworthy digital object* [2]. Provided proper infrastructure and compliance with the modern digital repository standards, data longevity of several decades can be expected – which, for the purpose of KM covers the typical needs of immediate, mid-term and long-term knowledge transfer and retention. A quick summary of the key features of a digital repository and their relevance to KM:

- Structured information
- Subject and classification and indexing
- Quick and flexible discovery
- Information sharing discipline
- Preservation of explicit knowledge

4 Case example: International Nuclear Information System (INIS)

At the end of the 1960's, the *International Atomic Energy Agency* (**IAEA**) developed and established the *International Nuclear Information System* (**INIS**) as a "co-operative scheme, involving the IAEA and its Member States, for applying computers to the task of disseminating information dealing with nuclear science and its peaceful applications" [3] as a logical continuation of fulfilling the requirements of its Statute to "…take positive steps to encourage the exchange among its members of information relating to the nature and peaceful uses of atomic energy and … serve as an intermediary among its members for this purpose" [3].

The system has since then undergone various modifications and developments, dictated not only by technical progress but by social and political factors; however, as detailed below, the very core of its purpose has not only remained intact, but has evolved into a sustainable structure that operates successfully to date and is on its way to further growth.

Repository structure and scope. INIS hosts bibliographic references of documents of various types: journal publications, books, conference presentations, technical reports, and many others. An extremely important feature of INIS is that, benefitting greatly from its distributed input scheme with the Member States contributing to the contents, it has been able to collect and preserve massive amounts of non-conventional publications from all over the world. These publications contain outcome of research or measurements which have never or only partially been published or made widely available to the public, but often contain unique knowledge which would otherwise be lost or buried in local archives. **Fig. 1** shows the document type distribution of INIS. Another impressive feature of INIS is the colossal amount of information that has been digitized from a microfiche collection. Throughout its years of operation, INIS has collected a substantial amount of microfiche and, at the end of the 1990's, the microfiche-based production system was replaced with an imaging system to process and disseminate the documents in electronic form. This heralded the beginning of a digital preservation project that has, since then, been one of the important tasks of INIS. [4] **Table 1** gives an overview of some of the statistics of the digitization project.

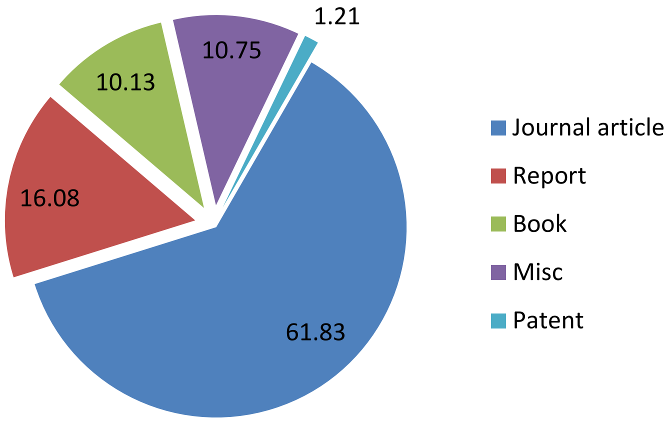

Fig. 1. INIS document type distribution

500,000 bibliographic records	**312,000 Non-conventional reports**
Conversion from paper to microfiche and diazo duplication after database production	**NCL check ensures accuracy of the "eye-readable" microfiche header information**
Over 1 million master microfiches	**~17 million pages**

Table 1. Overview of the INIS digitization project.

S01 - Coal, lignite, and peat	**S36 - Materials science**
S02 - Petroleum	**S37 - Inorganic, organic, physical and analytical chemistry**
S03 - Natural gas	**S38 - Radiation chemistry, radio chemistry and nuclear chemistry**
S04 - Oil shales and tar sands	**S42 - Engineering**
S07 - Isotopes and radiation sources	**S43 - Particle accelerators**
S08 - Hydrogen	**S46 - Instrumentation related to nuclear science and technology**
S09 - Biomass fuels	**S47 - Other instrumentation**
S10 - Synthetic fuels	**S54 - Environmental sciences**
S11 - Nuclear fuel cycle and fuel materials	**S58 - Geosciences**
S12 - Management of radioactive wastes, and non-radioactive wastes	**S60 - Applied life sciences**
S13 - Hydro energy	**S61 - Radiation protection and dosimetry**
S14 - Solar energy	**S62 - Radiology and nuclear medicine**
S15 - Geothermal energy	**S63 - Radiation, thermal, and other environmental pollutant effects on living organisms and biological materials**
S16 - Tidal and wave power	**S70 - Plasma physics and fusion technology**
S17 - Wind energy	**S71 - Classical and quantum mechanics, general physics**
S20 - Fossil fueled power plants	**S72 - Physics of elementary particles and fields**
S21 - Specific nuclear reactors and associated plants	**S73 - Nuclear physics and radiation physics**
S22 - General studies of nuclear reactors	**S74 - Atomic and molecular physics**
S24 - Power transmission and distribution	**S75 - Condensed matter physics, superconductivity and superfluidity**
S25 - Energy storage	**S77 - Nanoscience and nanotechnology**
S29 - Energy planning, policy and economy	**S79 - Astrophysics, cosmology and astronomy**
S30 - Direct energy conversion	**S96 - Knowledge management and preservation**
S32 - Energy conservation, consumption, and utilization	**S97 - Mathematical methods and computing**
S33 - Advanced propulsion systems	**S98 - Nuclear disarmament, safeguards and physical protection**
	S99 - General and miscellaneous

Table 2. INIS Thesaurus subject coverage

Organization and classification. With the growth of INIS, the challenges to standardize contents led to the development of detailed keywords – or descriptors – for precise classification of the literature. This system of indexing the content using keywords was the basis of what later became a controlled vocabulary thesaurus. A substantial amount of effort has been put into further development and maintenance of the INIS Thesaurus, in collaboration with other institutions and Member States. With time, translations have been provided and are regularly maintained, making it a unique multilingual multi-subject thesaurus, available in Arabic, Chinese, English, French, German, Japanese, Russian and Spanish. The system has also evolved into a large scale project which is updated on a regular basis with the input of numerous subject experts world-wide, and integrated with the INIS digital repository [5].

This integration enables the use of INIS as a complex system for knowledge organization and dissemination. Because of its wide subject coverage, listed in **Table 2,** and its enormous amount of non-conventional publications, it is used as the main source of knowledge retrieval in the field of nuclear technology. It also provides an extensive platform for preservation of all types of information, and, coupled with modern semantic technology, can serve as a very powerful tool for information discovery.

Access and usage. The INIS repository is available world-wide and provides unrestricted access to its resources, which comprise bibliographic reference and full-text databases and a multilingual thesaurus. Its powerful search engine provides both simple and advanced search interfaces to construct complex search queries (see **Fig. 2)**. These queries can be used to extract results based on specific metadata in different languages supported by the system. Another part of the INIS KOS is the thesaurus, its interface being integrated in the INIS repository search and providing an advanced interface for the search and discovery of subjects and keywords (**Fig. 3)**. A brief overview of the main INIS statistics is given in **Table 3**.

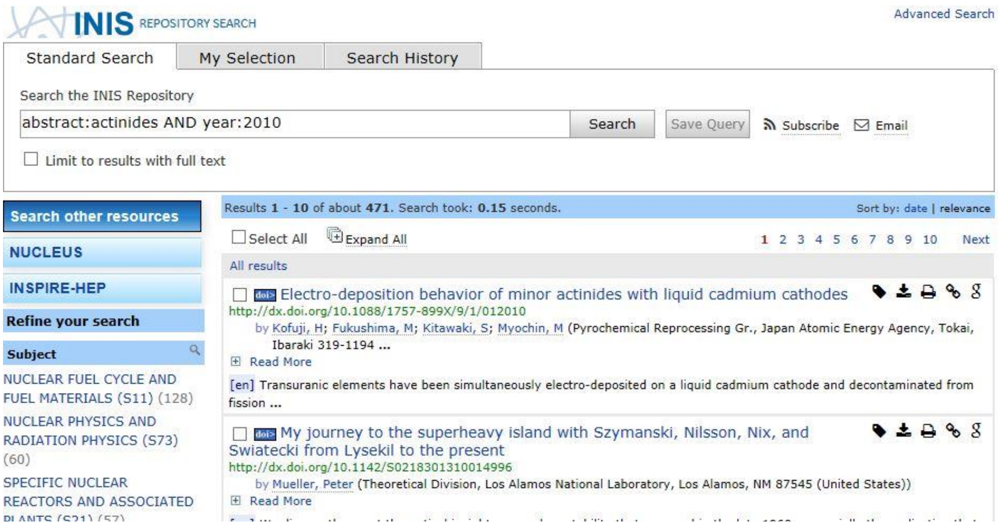

Fig. 2. INIS repository search

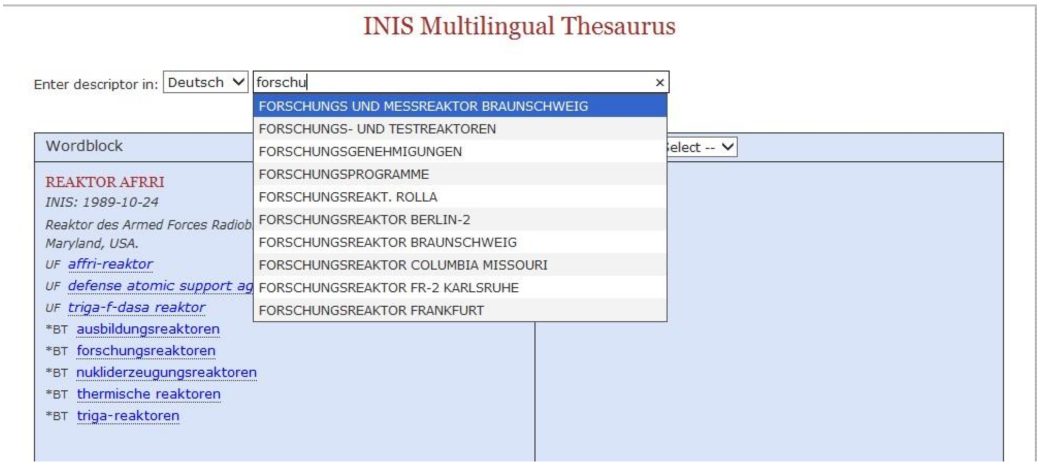

Fig. 3. INIS Thesaurus interface

Repository	Thesaurus	Access (per annum)
Almost 4 million records	Over 31.000 descriptors	Over 2.3 million pageviews
Over half-a-million full-text	Available in 8 languages	Averaging half-a-million document downloads
1.3 million unique documents	Controlled vocabulary	Quick and advanced search functions
Over 100.000 records added every year		Thesaurus search interface

Table 3. Overview of the INIS features

Summary. With over 45 years of operation, INIS has accumulated an enormous amount of information and developed a knowledge organization system for scientists, engineers and managers in the nuclear industry. Moreover, as technology has further developed, INIS has taken the function of information preservation and successfully stored vast amounts of non-conventional publications, large parts of which would be otherwise lost. Many research facilities around the world are able to benefit from this, especially with diminishing resources or shut-down of operations and when local archival resources, such as libraries, are being disbanded. In such cases, INIS is able to act as an instrument of knowledge preservation.

And albeit, at the time of initiation, the concept of knowledge preservation and exchange was not used, it was, in principle, one of the underlying factors intrinsically accompanying the very definition of INIS, which since then has evolved to become *de facto* the largest scientific and technical knowledge organization system in the field, operated by the IAEA in collaboration with over 150 countries and international organizations.

5 Conclusion

Established as a system to share information on nuclear research and supporting areas and driven by the efforts of the Member States of the *International Atomic Energy Agency* (IAEA) and its mandate to foster peaceful usage of nuclear technology, the *International Nuclear Information System* (INIS) has since then massively evolved, with many factors affecting the shape and development of the system, from political and international communication, to technological and scientific.

However, the very source and the operational definition of the system has not only remained intact but has set an extensive foundation for the continuing vitality and intrinsic usefulness of INIS even now after many decades. And even though the combination of knowledge and management was not the factual agenda back at the end of the 1960's, it was intuitively the right ansatz, and could be the root of knowledge management in the area of nuclear information. The mandate to collect, index and store information, and at the same time create and maintain a list of subjects and related keywords which then formed the most extensive thesaurus in a multi-subject field, was the very foundation of what has later on evolved into a full-scale Knowledge Organization System.

Given the perspective presented in this paper, it would be natural, and almost predictable, to expect the alignment of INIS with the efforts of the knowledge management organization, and integration of the repository and the thesaurus in a global KOS with application of state of the art semantic technologies. Considering the general direction in which digital libraries and repositories have been developing, it is more than natural to assume that this would be the next logical evolution of not only INIS but also many other similar repositories. In the case of INIS, an ongoing project in this direction has been initiated in cooperation with the IAEA's Nuclear Knowledge Management Section.

The case example demonstrates two main points which shall conclude this paper, namely:

- KM aspects as such can be intrinsic to the organization of an information repository. This means that a properly designed information resource with subject categorization, information harvesting, organization and dissemination planned with a high level of sustainability contains many prerequisites for what could be used as an integral part of a KOS which is able to operate and further develop for many decades to come. This has been clearly proven using the example of INIS.
- The natural evolution of digital repositories might lead to their integration or repositioning either as standalone KOS's or as parts of more global KOS's, such as aggregator knowledge systems. In either case, the contribution of a system comprised of a categorized repository and a corresponding organization system such as a thesaurus with semantic relationships is instrumental to KM and should be considered part of a general KM strategy at all stages of its planning and implementation.

1. Systems of Knowledge Organization for Digital Libraries: Beyond Traditional Authority Files, Gail Hodge, First Digital Library Federation electronic edition, September 2008
2. Long-Term Preservation of Digital Records: Trustworthy Digital Objects, H. Gladney, The American Archivist, Sep 2009, Vol. 72, No. 2 (Fall/Winter) pp. 401-435
3. The International Nuclear Information System (INIS), John. Woolston, Unesco Bull. Libr., vol. XXIII, no.3, May-June 1069
4. Digital Preservation at International Nuclear Information System (INIS), G. St-Pierre, D. Savic, The Grey Journal (TGJ): An international journal on grey literature, ISSN 1574-1796, volume 10 (2014), number 3
5. The INIS Thesaurus: historical perspective, B. Negeri, O. Vakula, 45th INIS Anniversary Newsletter, https://www.iaea.org/inis/products-services/newsletter/INIS-Newsletter-2015-17/Duresa-Vakula.html