From nuclear information to knowledge: The role of INIS

Zaven N. Hakopov

International Atomic Energy Agency, Vienna, Austria

Abstract. The International Nuclear Information System (INIS) once created to facilitate information exchange in the broad range of scientific and technical fields related to peaceful applications of nuclear power has evolved into a Knowledge Organization System (KOS) consisting of a digital repository and an advanced multi-language thesaurus. This paper gives a general description of the establishment, evolution and features of INIS not only as a digital repository but as one of the key resources of knowledge organization and preservation in the nuclear field.

Keywords. Nuclear information, digital repository, knowledge management, knowledge organization system, semantic technologies

1 Introduction

With the progress and development of information technology, the bibliographic archives 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, due to the fact that most of them 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.

2 Repository as a Knowledge Organization System

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.

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* [1]. 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.

3 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" [2] 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" [2].

The system has since then undergone various modifications and developments, dictated not only by technical progress but by social and economic 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: serial 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 experimental 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 holdings. 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. [3] **Table 1** gives an overview of some of the statistics of the digitization project.

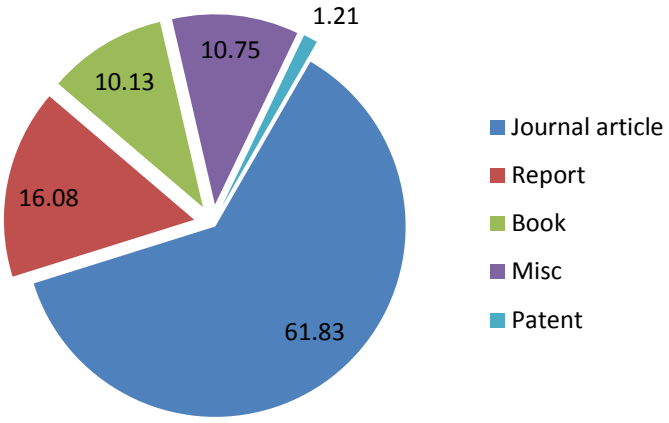

Fig. 1. INIS document type distribution

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 microfiches	**~17 million pages**

Table 1. Overview of the INIS digitization project.

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 in a controlled vocabulary was the basis of what later became a thesaurus, as *"…a controlled and dynamic vocabulary of semantically and generically related terms which comprehensively covers a specific domain of knowledge. This vocabulary is a systematical and/or alphabetical collection of descriptors, non-descriptors as well as indicators of their relationships."*[1] 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 for all areas of science and technology related to nuclear and 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 [4].

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.

[1] "Guidelines for the Establishment and Development of Monolingual Scientific and Technical Thesauri for Information Retrieval" UNESCO, SC/MD/20, Paris, July 1970.

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

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, being integrated in the INIS repository site 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**.

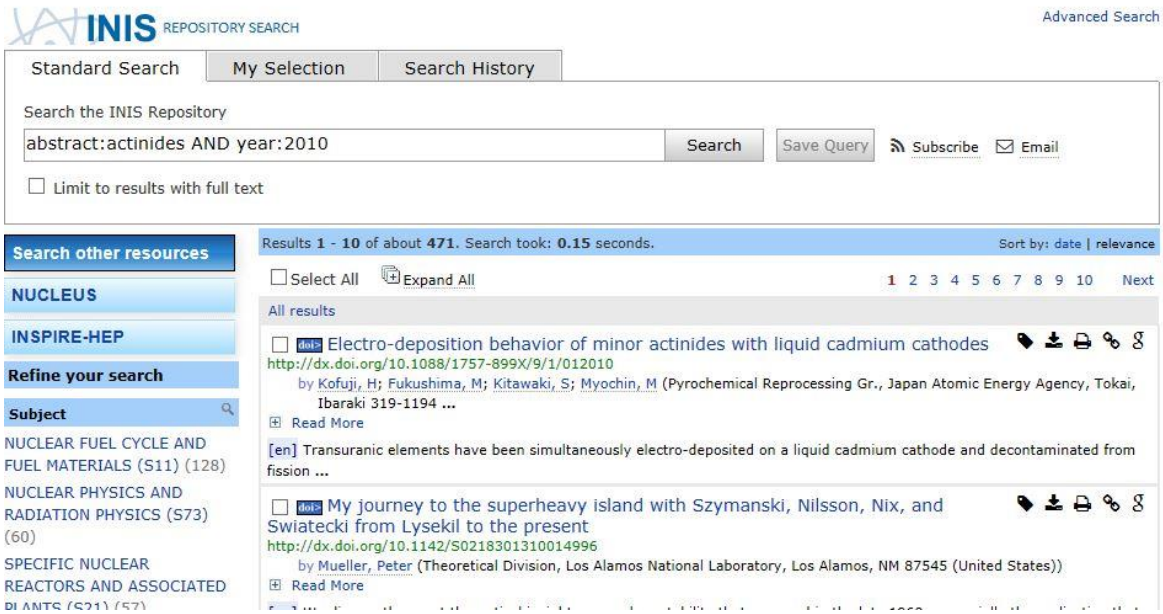

Fig. 2. INIS repository search

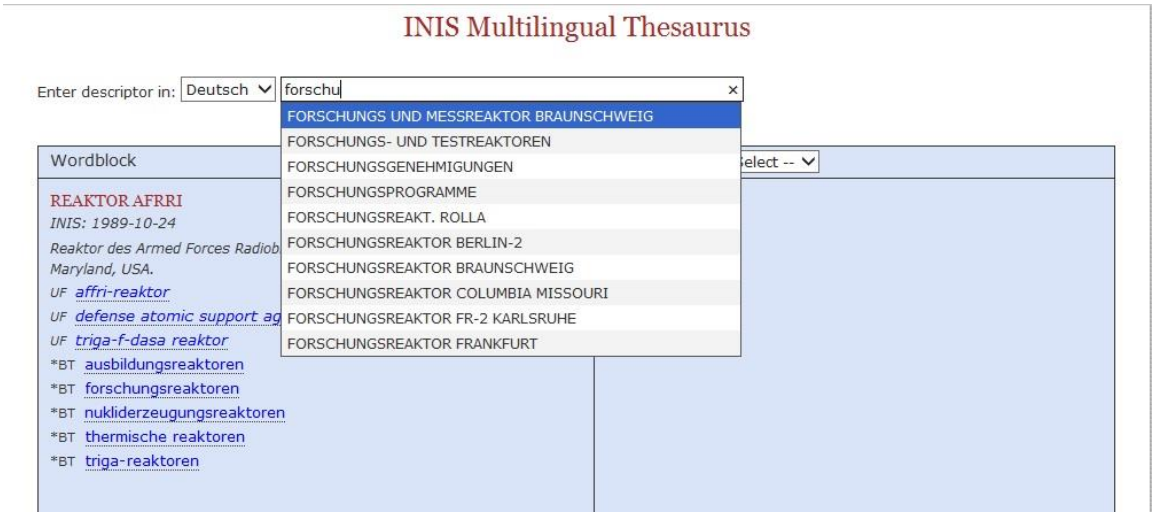

Fig. 3. INIS Thesaurus interface

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 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 organization 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.

4 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* and its mandate to foster peaceful usage of nuclear technology, the *International Nuclear Information System* 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 preservation 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.

1. 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
2. The International Nuclear Information System (INIS), John. Woolston, Unesco Bull. Libr., vol. XXIII, no.3, May-June 1069
3. 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
4. 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