Preserving and Accessing Nuclear Knowledge

Extending the INIS model

Y. Turgeon, Y. L. Yanev

INIS & Nuclear Knowledge Management Section
IAEA
E-mail: Y.Turgeon@iaea.org

INIS was officially created in February 1969,[11] and the first issue of the INIS Database was published in May 1970 in the form of magnetic tape and printed bulletin. If the contents of the original product and that of the latest release of the Database are compared, it is clear that despite changes (sometimes major) in format and methods used to capture and distribute the information, the core components of the System remain essentially the same: a fully indexed bibliographic database, complemented with a repository collection of non-conventional (NCL) or 'grey' literature.

All systems need to innovate in order to survive and remain relevant, and INIS is no exception. It is clear from the feedback received from INIS Members and users in the last years that they want to see significant changes made to this system. Such changes are expected to make INIS simpler and less expensive to operate; to extend the INIS model to cover more of the accumulated nuclear knowledge and information at the IAEA and in Member States; and to become much more attuned to the System's users needs and expectations. This is not new and, for a number of years, the INIS Secretariat has been enhancing the methods used to maintain the system, while exploring changes that could answer some of the expressed needs. Looking first at recent changes to the System, the most significant are the following:

Data format and processing: A new record format, much less rigid and able to accommodate various levels of input quality, has been introduced. It is designed in XML, which will make it much easier to extend without affecting the underlying processing system. We are in the final phase of redesigning our processing system with off-the-shelf tools (rather than proprietary software) to fully support it.

Acquisition of bibliographic records from publishers: These acquisitions complement rather than replace INIS Members' input and are designed to enhance the comprehensiveness of the System. Started in 2001, they are now expected to constitute nearly 50% of total INIS input.[12] A number of INIS Members are involved in the indexing of these records on a voluntary basis, thus strengthening the co-operative approach of INIS.

Adoption of a computer-assisted indexing (CAI) system: The system became operational in June 2004, and initial results indicate a 80-100% increase in indexing productivity. We are now looking at extending access to the system to selected INIS Centres. In future, we are also interested in using this technology to support automatic categorization and enhanced full text retrieval, features that could be used on the NCL collection and an IAEA Nuclear Knowledge Portal.

Enhancement to the INIS Database on the Internet: INIS will soon release a new version of the INIS Database on the Internet. In addition to an enhanced user interface designed to make

[11] The Board of Governors of the IAEA, Record of GOV/OR Meeting 408, GOV/OR.408, Wednesday, 26 February 1969.
[12] Publishers records input for the period 2002 to August 2004 was about 14,000 (of a total of 71,000) in 2002; 30,000 (87,800) in 2003; and 40,000 (74,000) for Jan. to August 2004.

retrieval easier for end users, the main change is the introduction of direct links to a portion of the NCL collection held at the IAEA. This portion includes all documents held in electronic form from IAEA, and all those NCL whose online access has been authorised by the INIS Liaison Officers. Material on microfiche is being scanned and will be added as it becomes available. We are also activating all hyperlinks to documents not kept at the IAEA, and will be looking at methods to link to documents if no such hyperlink exists.

Redefinition of the role of INIS at the IAEA: The Agency must remain an authoritative source of information and knowledge on the peaceful use of nuclear energy. In order to support this objective, a more efficient relationship is needed between all Agency resources, including INIS. In this context, INIS has been reorganized to make it easier to co-ordinate a range of IAEA nuclear knowledge and preservation activities, and the development of a strong synergy with the IAEA Library has been recommended.

Extensions to the INIS Model

A number of external changes are driving the need for an extension of the INIS model, including:

- A changing nuclear environment, where the earlier focus on government-sponsored basic research has changed to focus on the development of technologies and applications by industry;
- The need to capture and transfer accumulated nuclear knowledge for use by later generations;
- The oft expressed need of the users base to have online access to all nuclear materials found in INIS, including publications only available online (which should then be preserved), and to extend the contents to non-document materials.

While it is agreed that the INIS model needs to be re-evaluated in view of the above, it is also clear that a key to the System's continued relevance lies in retaining and strengthening its core functionality. For this reason, a two-tiered approach is suggested, as follows:

INIS 'Core'

INIS core mission should remain that of preserving and documenting nuclear scientific and technical information in a high quality database, complemented with what is, in fact, an INIS Preservation Collection of 'grey' literature. We need, however, to review what materials should be included in future; firstly, contents no longer has to be limited to traditional document-like materials, as current technology already permits the inclusion of other types of materials requiring access and preservation, e.g. data sets; multimedia objects; knowledge bases, etc. Secondly, INIS has minimal coverage of materials published in the two decades that preceded its creation, a period which saw fundamental progress in nuclear sciences. An extension of the core database to incorporate materials from the period 1945–1970 would greatly enhance the value of the System, and we believe that future generations will want to have access to such information.

The functionality of this core can be greatly expanded by adding hyperlinks for distributed access to most (if not all) materials that is not already in the INIS preservation collection. Until recently, there was a formal distinction between INIS the metadata repository and INIS the document supplier (a function reserved for NCL delivery). Such a distinction will disappear, as INIS metadata become a direct point of access to all types of materials, commercial or not, not located at the IAEA. However, in order to be effective — and not systematically point to documents that cannot be accessed by end-users because of copyright and licensing restrictions

— INIS has a strategic interest in developing partnerships with organizations whose role is to facilitate such access. The IAEA Library proposal for an International Nuclear Electronic Library and a Consortium of Nuclear Libraries would significantly support such a concept and should be of interest to a number of existing INIS Centres.

Extending the INIS 'Core'

The materials and knowledge that cannot be incorporated directly within INIS is much larger than INIS itself. The Nuclear Knowledge Management Unit, the second half of the INIS & Nuclear Knowledge Management Section, is at work building a framework to identify and manage this corpus which, in addition to literature, should include information on experts in the field, products, technologies, methods and services. Without embarking on discussions about what this framework will be, the fact remains that INIS will be asked to index, link to, capture and/or maintain a large quantity of materials in support of activities at the IAEA and in Member States. A common 'glue' to all materials would be a 'Nuclear Knowledge Portal'.

INIS Support for Knowledge Management Initiatives

A key component that distinguishes INIS from generic Internet search engines (e.g. Google) is the subject analysis done using controlled vocabulary.[13] We all know how efficient a service such as Google (or similar Internet search engines) are at automatically identifying highly relevant information; less known, however, are their deficiencies in recall, a shortfall recognized chiefly by information specialists.[14] As we move towards a distributed architecture and evolve our user interface towards a Nuclear Knowledge Portal, it will be essential, in addition to adopting the technologies used by search engines, to retain as much as possible of the INIS strength in subject analysis. For this purpose, and to complement the INIS Thesaurus, INIS intends to develop a nuclear classification for the automatic categorization of nuclear knowledge.

'Multilinguality' and Localization

Pending a sudden renaissance of interest in Western countries, nuclear development is currently most active in developing, non-English speaking states (for lack of a better expression). English is the communication language of science and the core language of INIS. However, as most users are 'passive' consumers of English information (once they have found it), the retrieval process would greatly benefit from the localization of the user interface and the retrieval process. INIS has already developed core elements essential for the internationalization of the System:

- The terms in the Thesaurus are already available in 5 languages: English, French, German, Russian and Spanish, and translations are under way in Chinese and Arabic;
- The new record format has been developed in XML, which uses Unicode as its character set;
- All NCL documents in Western languages (and Cyrillic) are currently 'OCRed' as part of the imaging process at the IAEA;
- The new interface to the INIS Database on the Internet can be localized (a German language version already exists). Future developments will include the choice of language for the descriptors.

[13] The INIS Thesaurus can be considered an ontology. Earlier forms of the INIS Categories were forms of taxonomies.

[14] 'Recall' refers to the amount of relevant information retrieved. Key results that are not on the first (or second) page of results are lost to most users. As the position of a document in a list is as dependent on its 'popularity' (the PageRank system of Google) as its vocabulary contents, coherent result sets for specialised subjects are often impossible to build.

Conclusion

In the recent past, INIS has concentrated on strengthening its core functions, in preparation for an expansion of its role in nuclear knowledge management and preservation. We have already built a solid collection of Internet links to significant resources available online, and will start exploiting these links with automatic classification and indexing.

INIS is an international co-operative agreement between IAEA Member States whose operations are largely guided by Membership Arrangements. The current model of co-operation is static: INIS Members supply input to the IAEA, the Secretariat merges and distributes the output. In order support the approach outlined in this document, INIS will need to be more flexible and able to react more quickly to users' demand. This added flexibility has been added to a proposed Membership Agreement recently circulated to all INIS Members. In this Agreement, the concept of national input centres would remain central, complemented by strategic partnerships to secure access to resources and materials.[15] The function of the INIS Secretariat is also more clearly defined, while the overall responsibilities of Members is strengthened.

[15] For example, a Consortium of Nuclear Libraries (some of which could be INIS Centres) could greatly assist INIS in identifying key journals, conferences and other types of information to add to the INIS Database. It could also be used to negotiate the acquisition of metadata for all this material.