

INIS: NUCLEAR GREY LITERATURE REPOSITORY

Dobrica Savić

d.savic@iaea.org

International Atomic Energy Agency (IAEA), United Nations

This paper is licensed under the Creative Commons licence: CC-BY-SA-4.0 (<http://creativecommons.org/licenses/by-sa/4.0/>).

Abstract

As one of the world's largest collections of published information on the peaceful uses of nuclear science and technology, INIS represents an extraordinary example of world cooperation. Currently, as INIS members, 130 countries and 24 international organizations share and allow access to their valuable nuclear information resources, preserving them for future generations and offering a freely available nuclear knowledge repository. Since its creation in 1970, INIS has collected and provided access to more than 3.8 million bibliographic references to publications, documents, technical reports, non-copyrighted documentation, and other grey literature, as well as over a million full texts. Public interest throughout the years in accessing the INIS Collection has been remarkable. This paper deals with the challenges faced by INIS in its endeavour to increase the use, accessibility, usability and expandability of its on-line repository. It also describes document collection, the features and characteristics of implementing a new search engine, as well as the lessons learned.

Keywords

Nuclear Information, Document Repository, Grey Literature, INIS, IAEA

Introduction

The International Nuclear Information System (INIS) hosts one of the world's largest collections of published information on the peaceful uses of nuclear science and technology. It offers on-line access to a unique collection of 3.8 million bibliographic records and over a million full texts of non-conventional (grey) literature. This large repository collection suffered from most of the well-known shortcomings of the classic meta-data catalogue. Searching was complex and complicated, it required training in Boolean logic, full text searching was not an option, and response time was slow. An opportune moment to improve the system came with the retirement of the previous database software and the adoption of Google Search Appliance (GSA) as an organization-wide search engine standard. INIS was quick to realize the potential of using such a well-known application to replace its on-line search engine.

This paper deals with the challenges faced by INIS regarding its endeavour to increase the use, accessibility, usability and expandability of its on-line repository. It also describes document collection, the features and characteristics of implementing a new search engine, as well as the lessons learned. Although based on specific INIS practice and experience, this paper also offers guidelines on ways to improve classic collections comprised of millions of bibliographic and full text documents, while reaping the multiple benefits of a well-organized grey literature repository, such as increased use, accessibility, usability, expandability and improved user search and retrieval experiences.

International Atomic Energy Agency (IAEA)

The IAEA is regarded as the world's centre of cooperation in the field of safe, secure and peaceful uses of nuclear technologies. It was set up in 1957 as the world's "Atoms for Peace"¹ organization within the United Nations system. As of September 2015, the IAEA has 165 Member States.

The IAEA Secretariat is headquartered at the Vienna International Centre in Vienna, Austria. It also operates liaison and regional offices in Geneva, Switzerland; New York, USA; Toronto, Canada; and Tokyo, Japan. The IAEA runs and supports research centres and scientific laboratories in Vienna and Seibersdorf, Austria; Monaco; and Trieste, Italy. The IAEA Secretariat is a team of 2300 multi-disciplinary professional and support staff from more than 100 countries.



The IAEA's mission is guided by the interests and needs of Member States, strategic plans and the vision embodied in the IAEA Statute. Three main pillars – or areas of work – underpin the IAEA's mission: Safety and Security; Science and Technology; and Safeguards and Verification.

The work of the IAEA is carried out through six departments²: Nuclear Energy, Nuclear Safety and Security, Nuclear Science and Applications, Safeguards, Technical Cooperation, and the

¹ <http://www.iaea.org/About/about-iaea.html>

² <http://www.iaea.org/Publications/Reports/Anrep2012/orgchart.pdf>

Department of Management. Although supporting the entire Agency, the Nuclear Information Section (NIS) is organizationally part of the Department of Nuclear Energy. The Department's main tasks are to foster the efficient and safe use of nuclear power by supporting interested Member States in improving the performance of nuclear power plants, the nuclear fuel cycle, and the management of nuclear wastes; catalysing innovation in nuclear power and fuel cycle technologies; development of indigenous capabilities for national energy planning; the deployment of new nuclear power plants; and the advancement of science and industry through improved operation of research reactors.

IAEA Nuclear Information Goals

One of the IAEA's goals is the collection, preservation and dissemination of nuclear information and knowledge, which in turn is the main responsibility of NIS.

NIS consists of the IAEA Library Unit, the INIS Unit and the Systems Development and Support Group. It fosters the exchange of scientific and technical information on the peaceful use of nuclear science and technology; increases awareness in Member States of the importance of maintaining efficient and effective systems for managing such information; provides information services and support to Member States and to the IAEA; and assists with capacity building and training.

INIS

INIS represents an extraordinary example of world cooperation, where 154 members allow access to their valuable nuclear information resources, including grey literature, in order to preserve world peace and further increase the use of nuclear energy for peaceful purposes. In addition to over a million full texts, more than 3.8 million bibliographic references to publications, documents, technical reports, non-copyrighted documentation, and other grey literature are made available. Overall, there are 800 GB of data in the INIS Collection.

Full text documents available from the INIS Collection almost entirely represent nuclear related non-conventional literature (NCL) or grey literature. This PDF collection contains some very important historic and technical documents collected by INIS during the last 45 years. Much of that documentation originated in paper form only, so digitization of that collection was a huge project which went back and converted millions of microfiche pages to electronic, fully searchable, files. Optical Character Recognition (OCR) was performed on all documents within the collection making it easy to index and search. Besides being a source of information when searching, the availability of full text gives INIS a special role in the area of nuclear information and documentation — being the main custodian of this world information heritage and preserving this codified, specialized, scientific and technical knowledge.

INIS Collection

On average, INIS adds around 120,000 bibliographic records and 13,000 full text PDF documents to its collection annually. The collection is freely accessible from the INIS Search

website³. Only a small portion of the full text documents is regarded as restricted and kept for internal use.

The INIS Collection covers around 50 well defined subject categories which are regularly maintained by INIS and provides the scope descriptions used by national and regional centres to categorize nuclear literature for INIS input. The INIS Joint Reference Series publications are also available on the INIS website⁴.

The INIS Collection covers all aspects of the peaceful uses of nuclear science and technology such as nuclear reactors, reactor safety, nuclear fusion, applications of radiation and radioisotopes in medicine, agriculture, industry and pest control, as well as related fields of nuclear chemistry, nuclear physics and materials science. Special emphasis is placed on the environmental, economic and health effects of nuclear energy. Legal and social aspects associated with nuclear energy are also covered. Fig. 1 lists a complete set of INIS Subject Categories.

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

Figure 1 INIS Subject Categories

³ <http://inis.iaea.org/search/>

⁴ <http://nkp.iaea.org/INISSubjectCategories/>

A break-down of the amount of documents by major category is provided in FIG 2, while FIG. 3 gives a similar break-down according to various document types in the Collection.

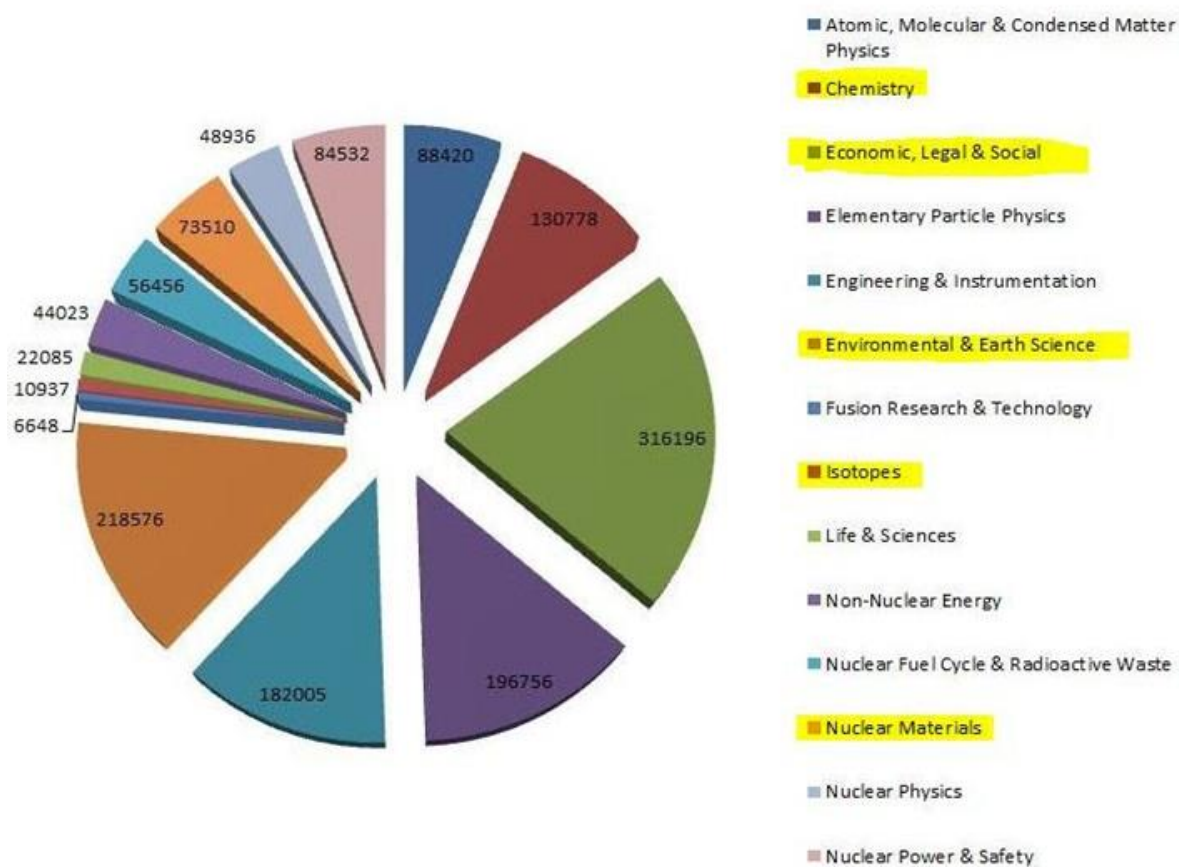


Figure 2 INIS Collection by Subject

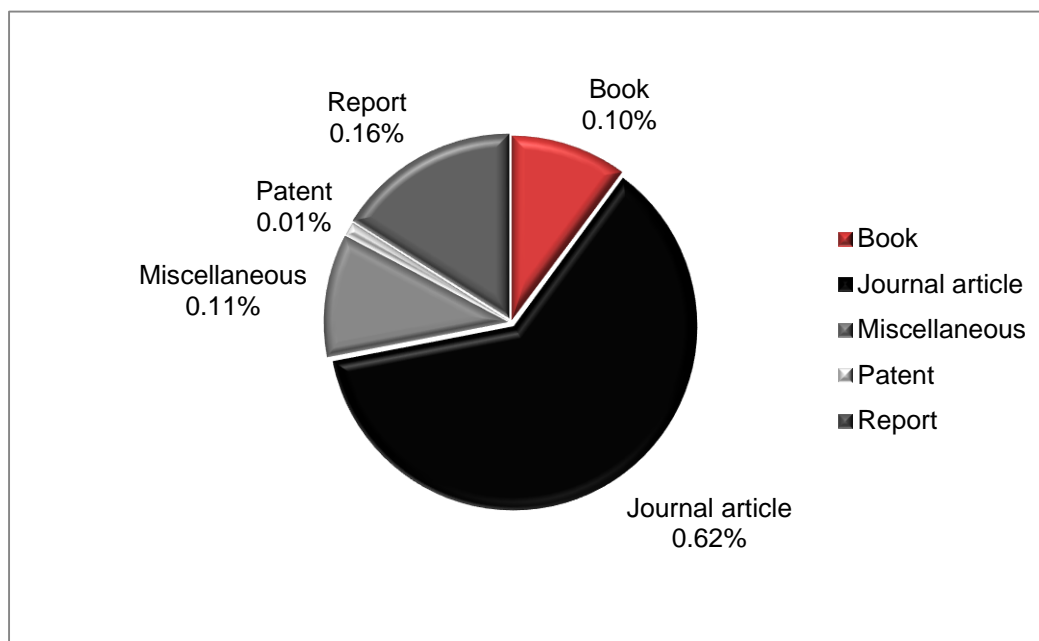


Figure 3 Bibliographic Records by Type (Status on 1 October 2015: 3,846,526 records)

INIS Search engine

Since its inception, the INIS Collection has operated in a controlled environment, requiring users to register through their national INIS centre, as well as the INIS Secretariat headquarters in Vienna, before being given access to the Collection. This changed in April 2009, when INIS became a free, open, and unrestricted information resource for internet users around the world. The opening of the Collection simplified access to reliable nuclear information on the peaceful uses of nuclear science and technology, including non-conventional literature, and made nuclear knowledge readily available worldwide for research, development and other uses. The opening of the Collection resulted in a significant increase of users.

The old search engine, a well-known BASIS Extended Relational Database Management System (ERDMS)⁵, was in operation from almost the beginning of INIS until 2011. Eventually, BASIS merged with an RDBMS system called DM, and became known as BASISplus. Currently, BASISplus is owned and supported by OpenText.

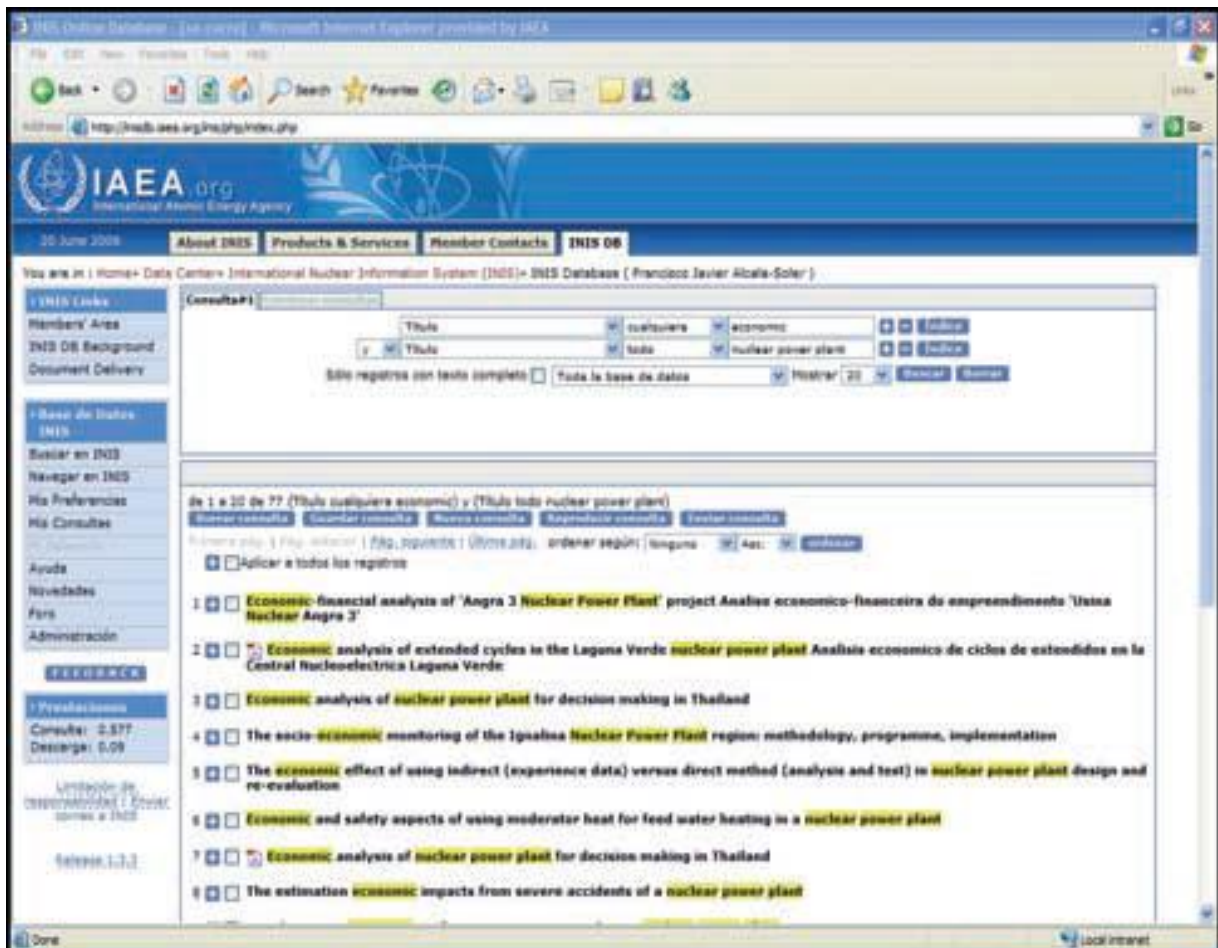


Figure 4 Old INIS BASIS Search Interface

⁵ http://en.wikipedia.org/wiki/Basis_database

The problems INIS experienced with its version of BASISplus included slow response time, a cluttered interface made for librarians, only an advanced search option, no full text indexing, issues with support for a multilingual search interface, and lack of support for the thesaurus and different authorities. The new search engine was installed in 2010, and became fully operation in April 2011. It was based on Google Search Appliance (GSA), which, at that time, became the IAEA-wide search engine standard. Although INIS reviewed a number of different search engine offerings, the decision was made to use GSA as a de facto on-line catalogue and to search the digital collection of INIS records.

This decision was based on the following characteristics offered by GSA: great speed and scalability; uncluttered and easy to use starting interface; the possibility to use advanced options and to broaden or tighten searches; powerful full text indexing; retrieval of more relevant results; and the existence of faceted/filtered search features. The images shown below are actual snapshots of two different versions of the INIS GSA-based standard (simple) search interfaces, both deployed on the INIS website.

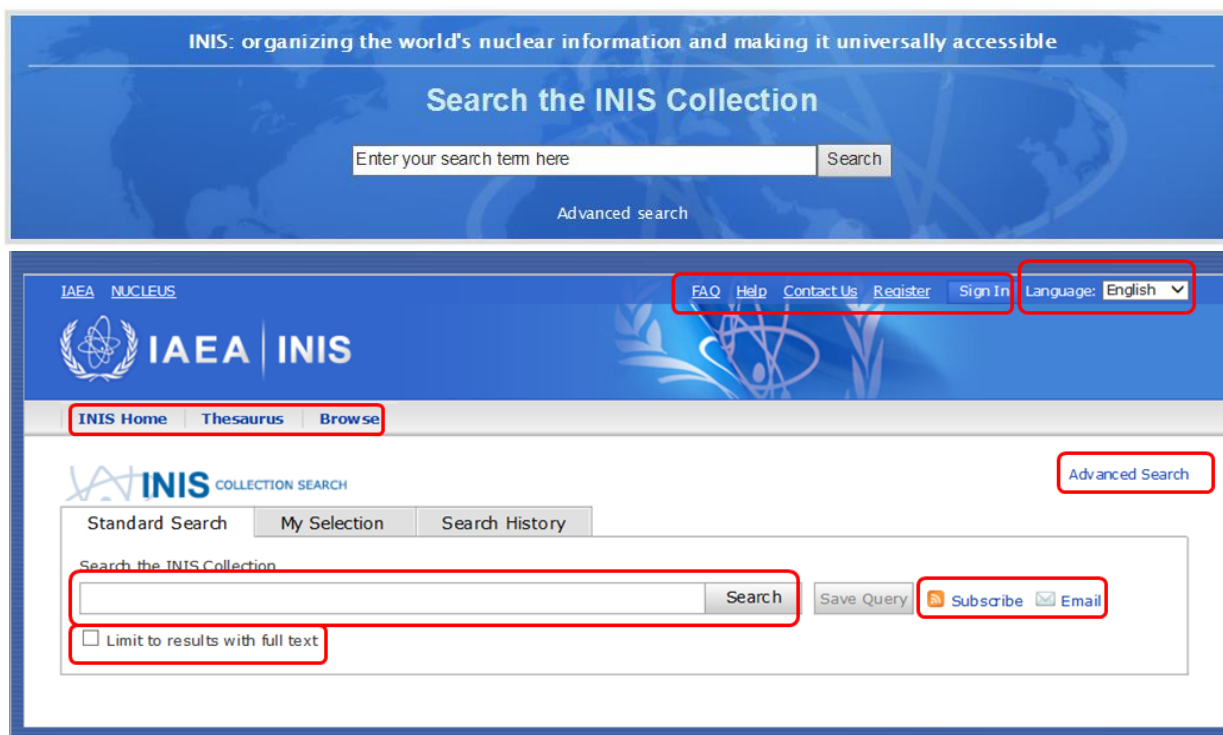


Figure 5 ICS Standard Interfaces

The ICS Standard interface offers a Main Search Box for entering search terms; a Tool Bar linking to INIS Home, Thesaurus, Browse, and Advanced Search, and the Language Selector. Both ICS Standard and Advanced Search interfaces are available in eight languages (Arabic, Chinese, English, French, German, Japanese, Russian, and Spanish). RSS subscribe and email options are also available.

All words put into the query are used; searches are always case insensitive; punctuation is ignored, including @ # \$ % ^ & * () = + [] \ and other special characters; and common articles or determiners, i.e. stop words, such as 'the,' 'a,' and 'for,' are usually ignored. The inclusion of stop words for languages other than English is also possible.

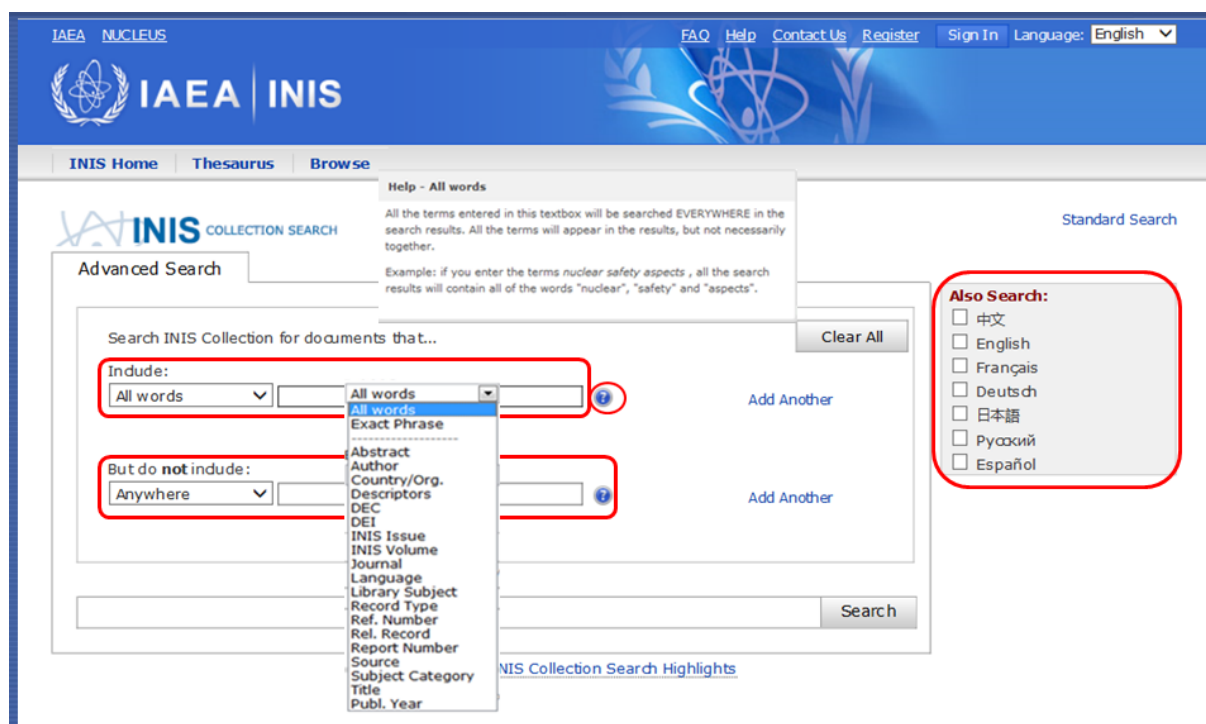


Figure 6 Features of the ICS Advanced Interface

The Standard Search interface provides excellent results for most search requirements. Our statistics show that the overwhelming majority of visitors (almost 99%) use the ICS Standard Search interface, or other channels, such as access through Google Scholar. Many other studies demonstrate that users are not inclined to become expert searchers (Novotni, 2004; Wallace, 1993; Valentine, 1993). However, if a more precise search is needed, a query can be constructed using the query builder form on the Advanced Search page. Query builder also generates a query syntax which appears in the Advanced Search query box. Alternatively, very experienced users can type the query directly into the Advanced Search query box. In this case, the query builder form becomes disabled.

The Advanced ICS Search (Fig. 6), offers the possibility to search all words or an exact phrase; to select (include or exclude) specific metadata fields; and to select the language of publications to be covered by the search. It also supports 'range queries' which enable the user to search for results where field values are between the lower and upper range specified by the query. Range queries are specified using the Range operator, written as two consecutive dots (e.g. year: 2007..2009). The dropdown menu of the Advanced Search offers the list of fields (metadata elements) in which one can search. This includes abstract, author, country/organization, descriptors, title, subject category, publication year, report number, etc.

Main Features of the INIS Collection Search (ICS)

The main features of the INIS Collection Search include:

Platform: Project, started in 2010, using Google Search Appliance® technology - renowned, simple, fast, flexible; Current version is 4.4 (May 2014); Virtual servers are within the IAEA, offering 24/7 availability.

Accessibility: Free and open web-based application; No installation required; Mobile Apps available for iPad, iPhone and Android; Widget prepared allowing searches from other websites, with or without filter.

Ease of use: Intuitive, self-explanatory initial screen; No previous knowledge of the INIS Collection is required; A single search box communicates confidence to users that the search tool can meet their information needs from a single point of entry.

Advanced Search: Offers more precise searches and metadata, as well as Boolean searches; Query builder has a built-in syntax generator; Ability to search for all words and exact phrase; Users can include or exclude metadata, select language; Range queries (2007..2009) and dropdown menus are available.

Faceted search: Dynamic navigation through country, language, publication year, and INIS Volume.

Expandability: ICS can be expanded to new collections, databases, repositories, new formats and type of documents.

Multilingualism: ICS interface and INIS Thesaurus are available in 8 languages, as well as a cross-language search. Automatic Google translation into 8 languages is also present.

Authority files: Journal titles, CODEN, ISSN, Subject category, Descriptor, Country/Organization, Author, and Report number authorities are incorporated in the ICS. Report number authority files are not frequently encountered in other collections and data bases, but INIS maintains this unusual feature as a sign of its dedication to quality document management and in order to accommodate INIS users who know the report series.

Usability: Users can print or export results in different formats (PDF, HTML, Excel, XML), download citations in plain text, RIS format, RefWorks, EndNote, create RSS feeds, e-mail search results as a link; Stop words for languages other than English, and translation of bibliographic records into other languages using Google Translator.

User profiling: User registration offers cross-Agency Single Sign-On, interface language set up, number of displayed results per page, saving queries and search updates, as well as emailing query results; Workspace concept also offers to save documents and translate into other languages.

Help: FAQ (on INIS and on ICS), on-line help, pop-up hints (examples on how to build a query using metadata), and e-training course are available.

Conclusion

The decision to replace the previous, almost 40 year old, INIS classic bibliographic on-line database and search engine was neither easy nor cheap. Finding a best fit replacement was even more difficult. A number of available search engines were examined and evaluated by a team of various information and computer specialists. The decision was to go for a market leader, namely the Google Search Appliance (GSA). Although it might have seemed risky at the time, the years following its implementation have shown that it was the right decision. It has brought considerable benefits, primarily to INIS Collection end-users.

The overall goal was to increase use, accessibility, usability, and expandability of the INIS on-line repository with 3.8 million bibliographic records and over a million full text nuclear related documents. This was achieved in the following manner:

Increased use - By making the repository open and freely accessible to the public and by replacing a legacy database search with a top-of-the-line one. While working on this goal, it became obvious that provision of full text documents was the users' most sought-after feature, making the new interface particularly beneficial to nuclear grey literature.

Improved usability - Meeting user needs was a tricky requirement due to the users' diverse experiences and backgrounds. Some users desired advanced search capabilities, while the majority wanted a simple and straightforward search. Statistics have proven that users mostly use a simplified, basic search interface. Because only by a small number of users (less than a few percent of all users) use the advanced search, this feature was improved upon but made discrete.

Expandability (scalability) – The most efficient way of providing for expandability is by getting a top performance technical solution, while making sure that different record formats are part of the overall solution and that they can be used from the beginning or added at a later stage.

Increased accessibility - Locating and accessing a relevant repository requires preparation and substantial ground work. Incorporating the repository and the document collection with Google.com, Google Scholar and other search engines (e.g. worldwidescience.org), can dramatically increase both the number of searches and users looking for relevant information. Together with incorporating the repository and making it available through other search engines, promotion is also essential. Conferences, articles, presentations, and poster sessions, all provide a great channel to promote the repository to the relevant user groups.

The success of the upgraded INIS repository and user satisfaction were evident from the feedback received. The number of visits dramatically increased and most importantly, the number of full text document downloads went up. The number of INIS Collection users dramatically increased with the introduction of GSA, even more so by connecting INIS to WorldWideScience.org, and, finally, by making the INIS Collection available on Google.com and Google Scholar. The number of searches went up fivefold and the number of downloads more than tenfold within the first few months.

In conclusion, this example of the INIS grey literature repository clearly demonstrates the options and tools that are available to attract more users and successfully boost the use of available information resources.

References

NOVOTNI, Eric (2004). I Don't Think I Click: A Protocol Analysis Study of Use of a Library Online Catalog in the Internet Age. *College & Research Libraries* [online]. 2004, **65**(6), p. 525-537. Available from: <http://crl.acrl.org/content/65/6/525.full.pdf>.

VALENTINE, Barbara (1993). Undergraduate Research Behavior: Using Focus Groups to Generate Theory. *Journal of Academic Librarianship*. 1993, **19**(5), p. 300-304.