



národní
úložiště
šedé
literatury

Šedá literatura & CRISTin

Asserson, Anne
2012

Dostupný z <http://www.nusl.cz/ntk/nusl-126794>

Dílo je chráněno podle autorského zákona č. 121/2000 Sb.

Tento dokument byl stažen z Národního úložiště šedé literatury (NUŠL).

Datum stažení: 22.05.2024

Další dokumenty můžete najít prostřednictvím vyhledávacího rozhraní nusl.cz.

GREY LITERATURE AND CRISTIN

ANNE ASSERSON

Anne.Asserson@fa.uib.no

University of Bergen, Norway

Abstract

CRISTine is The National documentation system for all research performance institutions there are publicly Funded in Norway. CRISTine is based on CERIF data model which is composed of IContextual metadata where Person, Organization, Output, Projects etc. are integrated. CRISTin holds the contextual metadata to the Institutional Repositories of the research performing institutions. Currently, CRISTine development is focused on tools which will serve for Organizations easier ways to access information within large data sets and to analyze that information in hopes of managing performance and making better decisions.

Keywords

Grey Literature, CRIS, Metadata

Research Process

The rate of acquisition of data, its structuring into information and its interpretation as knowledge is increasing rapidly. There are more active researchers now than ever and the output of white publications per researcher is increasing. The output of grey publications is orders of magnitude greater. Past techniques of experts (librarians) cataloguing manually with metadata each publication do not scale. The problem is to find ways to manage this resource.

The hypothesis is in 4 parts: (a) that the R&D process itself provides some context for managing the information; (b) that linking the records of the process to the publications provides this context; (c) that questions of curation and provenance are addressed automatically in such an environment; (d) that such an environment integrates grey and white literature and other R&D outputs such as software, data, products and patents.

The knowledge society demands that organizations utilise to the full the IP (intellectual property) they generate and acquire. In a context of research, development and innovation, the IP consists of products, patents and publications (in the widest sense – any stored representation of human intellect). While conventional research publications (white literature) provide much of the visible IP, the ‘submerged part of the iceberg’ is the grey

literature. With OA (Open Access) technologies the line between white and grey becomes blurred, because of accessible postprints and preprints. Significantly, white literature (and grey literature) may commonly be hyperlinked to further grey literature in the form of technical reports, procedure manuals, learning materials, software, data (in all media), communications (e.g. e- and voice-mail) etc.

In each case, the relationships need to be annotated with role (such as author or reviewer) and start date/time, end date/time. The relationships can be many-to-many, thus a publication may have many authors each of which may have affiliations with many organizational units and projects. Each instance of organizational unit, project or person may be recursively related to another; this allows correct representation of hierarchic or network (connected graph) relationships.

The underlying ‘magic bullet’ to support the research process is metadata which has formal syntax and declared semantics.

Grey Literature

Typically Grey literature consists of technical research documentation (although many different kinds of material have been classified as grey). The material is usually not peer-reviewed as is white literature. Yet it is not usually the unstructured, unauthorised output from a single source; it may well be commercial in confidence and certainly is likely to contain IP (intellectual property) of value to the organisation. It is between states i.e. between ‘an idea’ and ‘white’.

We wish to demonstrate that grey is the very foundation of the knowledge base both for internal use of an individual organisation (where it may well remain grey) and of world knowledge (where it is transformed to white). The information landscape (which when applied becomes the knowledge landscape) has exposed uplands (white) and hidden valleys (grey).

The dynamics of the landscape concern the way in which an idea, concept or knowledge is generated and transformed: from grey to grey (internal discussion within an organisation with improvement), from grey to white (publication, public relations for an organisation, improved evaluation scores for an organisation), from grey to product or service (wealth creation or improvement in the quality of life within an organisation), from white to product or service (wealth creation or improvement in the quality of life by knowledge or technology transfer).

The prerequisites for these dynamics to work are (1) excellent metadata (to improve discovery and control usage), (2) an institutional document repository of grey, (3) an institutional CRIS for the contextual research information, (4) linkage between the document repository and the CRIS of an institution and thence (in a controlled manner with formal descriptive and restrictive metadata) to other institutions, (5) an e-research repository of research datasets and software, (6) linkage between the e-research repository and the CRIS of an institution and thence (in a controlled manner with formal descriptive and restrictive metadata) to other

institutions, (7) an institutional policy to mandate deposition of the material with appropriate metadata.

From foundation to dynamics, grey is the key ingredient for the knowledge society.

The key results from our research are:

1. conventional grey literature repository metadata (usually based on DC (Dublin Core)) is insufficient;
2. great advantages are achieved when a grey repository is linked to a CERIF-CRIS: contextual metadata, workflow, interoperation and organisational integration;

The adoption of such an architecture ensures:

- a lower effort threshold on input (workflow, formal metadata) and hence increased repository fill;
- improved retrieval (formal metadata, semantic links);
- improved support for workflow and the research process (formal metadata, semantic links);
- improved links to other within-organisation systems (formal metadata, semantic links) including repositories of research datasets and software, library catalog and systems for finance, HR, project management, directories, web-pages;
- improved interoperation with the systems of other organizations (formal metadata, semantic links);

A CERIF-CRIS provides a new way to access and utilise grey information but encourages the user to let the computer system do the tedious work leaving the end-user free to do their research. This is achieved by the use of the formal syntax and defined semantics of CERIF entities, attributes and linkage together with as much or as little intelligence in the system as the user requires. The information space is navigated reliably and reproducibly by the computer, not the user. Presented through a web browser the ease of use remains but with a much more powerful information management capability.

CRIS (Current Research Information System)

CRIS Development

The historical development of CRISs (Current Research Information Systems) has emerged from the world of IR (Information Retrieval). The motivating reasons were that the recorded information was usually analogous to the {title/ author / date / keywords / address of source} form of library card catalogues.

Used stand-alone such systems were adequate until requirements for statistical analyses, integration with data in other DBMSs, flexible reporting (including integration with the client office environment) and handling of multimedia (or hypermedia) became important. These requirements emerged from the mid 1970s onwards. However, the emergence of the requirement to access information in multiple heterogeneous CRISs (and other DB systems) distributed geographically exposed harshly the inadequacies of these systems: they lacked any underlying theoretical model (to allow structural matching); they lacked common standards of data recording (to allow interworking) and they lacked common external interfaces (to allow integration with client office environments both for input/update and for retrieval/reporting). Worse, they lacked the metadata to allow such interworking to be provided 'on the fly' when required.

CERIF provides a comprehensive data model for R&D Information agreed by representatives of European countries (both European Union and associated States). The original CERIF1991 recommendation was used as the basis for the ERGO Pilot Project demonstrating interoperation of 100,000 metadata records from 10 countries. The new CERIF data model has been produced within the CERIF Revision Working Group. CERIF2000 has redefined CRIS in a formal way so providing a stable platform for CRIS builders, for inter-CRIS data exchange and for provision of metadata to describe CRIS contents succinctly. The key is formal syntax and declared semantics.

CRIS and Grey Literature

The products of R&D are listed below. CRISs should support them. Items in bold are regarded as 'grey literature': there is a need to design to manage all of these kinds but publications are clearly of great importance:

- (a) publications: journal / conference papers, technical papers, theses, dissertations, reports
- (b) patents
- (c) products: prototypes, fully engineered products
- (d) results: data and its presentation / visualisation
- (e) know-how and IPR: reports, procedure instructions
- (f) education and training: documentation, courses
- (g) publicity: press releases, product or organisational 'flyers'

CERIF2000 assumes that external databases are available for patents, products and publications. Experience of the authors has shown that such databases are incomplete. Therefore, for accurate recording of CRIS products it has been found necessary to record the metadata even on refereed publications in the same CRIS as the metadata recording the grey literature.

The recording of the data provides the context including the workflow of the R&D process, history and provenance. Grey documents produced as early ideas are captured in a temporal and organisational context, just as well as white publications, via the linked repository. CERIF allows, in a multidimensional framework, deduction or induction of relationships between documents – for example between a grey internal report and a white published paper - and with other research outputs. Furthermore, relationships between documents can be expressed explicitly: references and / or citations can be recorded. In this way a rich context for understanding the R&D output is provided, including versions, history and provenance.

Recording facts once in a structured R&D process environment and then re-using them in many ways reduces - by automated provision assistance - the need for user input of metadata to describe research outputs (especially grey literature) and thus addresses the scalability problem.

CRISTin

CRISTin (CRIS in Norway) is a national information system used by all research-performing institutions that are receiving public support for their research. CRISTin is used by the ministry to assess the institutions' research performance based on research output in the form of peer-reviewed publications that includes articles, books, book chapters and review articles. Grey literature is not assessed but CRISTin makes it available.

The key feature of a CERIF-CRIS used in CRISTin is that publications are linked in context with records of the researchers, their organisational units, and related CRIS (Current Research Information System) information.

euroCRIS

euroCRIS is a professional association of CRIS experts (but with strong links to other communities like OA). It is – by declaration of the EC - the custodian of CERIF: Common European Research Information Format an EU Recommendation to member states. It is the International recognized point of reference relating to all CRIS matters and provides stimulation of development and linking of CERIF-compliant CRIS. euroCRIS membership stands at 97 institutional, 38 personal and 18 affiliate, with 274 named delegates (out of a potential delegate count of 541) from 40 countries.

euroCRIS works strategically with many partners. Of interest to the GL community is the Rome Declaration on CRIS and OAR :

We, representing the communities concerned with CRIS and OAR, hereby declare:

That high quality research information is critically important to research institutions, research funders, policymakers and society at large;

That information on publicly-funded research should be available, shareable and integrated seamlessly;

therefore we commit:

To develop, support and promote an architecture (including data model and services) suitable to pursue these principles;

Jointly to adopt, develop and promote applicable open standards;

To advocate these principles among all stakeholders.

Acknowledgement

Much of the material in this paper is based on work done jointly over the last 15 years with Professor Keith Jeffery of STFC (Science and Technology Facilities Council) UK.