národní
úložiště
šedé
literatury

OpenGrey - jak funguje a jak se používá

Stock, Christiane
2011

Dostupný z http://www.nusl.cz/ntk/nusl-82068

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

OPENGREY: HOW IT WORKS AND HOW IT IS USED

CHRISTIANE STOCK

christiane.stock@inist.fr

INIST-CNRS, France

Abstract
OpenGrey is a unique repository providing open access to European grey literature references, the result of 25 years of cooperation. This paper presents the features of the new website, successor to OpenSIGLE, and reports on lessons learned from the transfer of metadata between platforms. OpenGrey usage data and feedback information support our advocacy for more grey literature in repositories.

Keywords
OpenGrey, European repository, repository migration, facetted search, usage data.

SIGLE and OpenSIGLE

SIGLE (System for Information on Grey Literature in Europe) was a unique European database of bibliographic records in grey literature. It was produced between 1980 and 2005 by initially seven and in the end fifteen members of the European Union, represented by major libraries and research organizations. Its contents covered all scientific disciplines (pure and applied science and technology, economics, social sciences and humanities). As a commercial product SIGLE was distributed through subscription to hosts, e.g. STN International, and available on a CD-ROM produced by Silverplatter/Ovid.

SIGLE records had the following mandatory information, essential for it's usage: an English title or keywords allowed search across all countries, the SIGLE classification code for a search by subject and an availability statement (holdings), meant to facilitate the order of a paper copy

INIST-CNRS decided to transfer the results of 25 years of work[1] onto an open access platform. As a result OpenSIGLE went live in December 2007 with almost 700 000 bibliographic records. It was based on DSpace* technology and its qualified Dublin Core metadata* format, which implied a simplification of the metadata compared to the server format. Minimal developments were made for the user interface, which provided search and browse facilities and offered separate information on the partners, the subject classification, and on document supply.

---

[1] the main effort relating to the identification and collection of documents

OpenSIGLE usage

OpenSIGLE had limited features and no new input was added (except for the GL conference collection), but its contents found an ever growing audience and new visitors from all over the world. OpenSIGLE was included in OpenDOAR (Directory of Open Access Repositories) in November 2009, after its integration in the WorldWideScience.org* portal a year earlier. Indexed in Google and Google Scholar since summer 2008, the database was reached through the Google search engine by an increasing number of users, and visits via Google Scholar amounted to 30% per month. Starting in March 2008 the conference preprints from the GL conferences were added gradually with their full text and after the digitization of GL1 to GL4 the collection is now complete.

Two years of statistical data collection[2] reveal a considerable evolution in numbers and the influence of the Google generation. Between 2009 and 2010 the number of visits almost tripled, and the number of page views more than doubled. These are impressive results considering that the documents were published at least 5 years earlier. The average number of pages viewed in a visit is not very high; neither is the duration of a visit. We take the data as a sure indication of the influence of Google. Many visitors reached OpenSIGLE after using a search engine, spent a short time looking for relevant information and left.

Average per month	2009	2010
Visits	14 000	36 000
Pages viewed	40 400	90 600
Length of visit	96 sec	95 sec
Number of pages viewed per visit	3.1	2.8

Time for change

Several reasons lead us to think of changing the system. Right from the beginning of OpenSIGLE we were facing limits in the technical performance of DSpace, especially when uploading and indexing GL collections. The problems were even noticeable for small updates. DSpace technology obviously reached its limits with 700 000 records in the database. Besides, the website layout no longer answered current needs for referencing.

Email contacts made us aware of missing features requested over the years by OpenSIGLE users, such as the possibility to export search results for state of the art studies. An important number of users enquired about how to obtain the document. This service was one of the

---

[2] Numbers are based on php/MyVisites*, an open source software using a tracker, similar to Google analytics

principal goals of SIGLE and OpenSIGLE, but the information was hidden in the full record display of OpenSIGLE, on a separate page. However, few users took the time to look it up.

Usage statistics lead us to presume that the brief record display became more and more the "entry page" to the database, due to the use of search engines. The new website should therefore improve the access to information on document supply through the general layout, improved record display and additional links.

Early on, it was also decided to re-open OpenSIGLE for "new" input and close the gap between SIGLE and today. However, the technical limits mentioned above would have been a major hindrance, hence the need for a change of software.

Moving to OpenGrey

In order to reinforce the changes in the technical environment as well as in the contents and policy for the database, INIST-CNRS decided on a change of name: "OpenGrey". Several new domain names have been acquired for OpenGrey with the extensions .eu, .fr, .net and .org. The new domain name is opengrey.eu.

The major change from OpenSIGLE to OpenGrey related to the software: The DSpace platform used for OpenSIGLE was replaced by Exalead® * as the search engine for the database, completed with in-house developments using php and MySql software for the user interface.

Persistent identifiers are essential to guarantee perennial access to records or documents. In OpenSIGLE each record as well as the communities and collections of the DSpace architecture were identified by a unique URL, the handle*. The handle system is similar to the DOI. In order to assure continuity the handles of the OpenSIGLE records had to be migrated to the new system. Redirections were made for the handles identifying communities, collections and other elements which ceased to exist in the same way in OpenGrey. We also tracked links to individual records (e.g. in bibliographic references) for redirection.

The OpenGrey homepage

OpenGrey went live on June 17th, 2011. The homepage of the new website aims to meet current needs of the users as well as to facilitate referencing. It is divided into three parts with separate groups of information. The logo is inspired by the lemniscates or infinity symbol. Tabs in permanent display allow the user to access records through search or browsing by subject, to gather information on the partners or on export facilities. Besides the "Google" like search field, the centre includes three blocs for short texts: a mini "about", a "focus" on figures on the content and a "news" column to announce conferences or important additions of records. The bottom part offers detailed help on search facilities and document delivery, legal mentions, and further information on OpenGrey, grey literature and GreyNet.

New features

Exalead brought major improvements both for the administration and for the user interface. OpenGrey now provides a more convenient technical environment, besides a new look. On the user side faceted search is the most prominent feature, to be detailed hereafter. Export facilities include an RSS feed for search results. Individual records can be displayed in XML and included as a citation with Zotero*. OpenGrey is also OAI-PMH compliant.

At present we are in contact with several organizations (NTK) for the integration of new records. We have started to add link fields in the existing records pointing to the full text (or its description) in a distant repository. In cooperation with GreyNet we also intend to enhance records from the GL conferences in different ways.

Search and refine

Basic search in OpenGrey is done in a Google-like field. Exalead offers the possibility to refine the results through faceted search. The new criteria are added to the query field and the search strategy can be copied and saved. The indexes chosen for refinements (such as author, subject, date, language…) are similar those of the Czech repository NUSL/NRGL3. This feature had a major consequence for the back office: metadata which had been merged for the simpler DSpace format in OpenSIGLE needed to be detailed once more, and even new controlled fields had to be added to allow the refinements and to insure consistency. Hundreds of records had to be "cleaned" in a semiautomatic way or manually.

As mentioned before, "one page visitors" come directly to the bibliographic record through Google. All relevant information must be available immediately or the graphic user interface must invite the user to click further. The bibliographic record in OpenGrey shows all fields, permits to launch new searches and includes, if applicable, links to partner repositories. Separate fields inform about the location of and access to the paper copy and about the access to the online full text. In addition OpenGrey provides an easy link to further information, e.g. on how to obtain a copy of the paper document or on the terms and conditions of the partner organization.

The partner homepage

This and other useful information is found on a kind of "partner homepage", accessible from a tab called "Partners". As a starting point, information on former EAGLE members in OpenSIGLE was transferred, namely:

- General information of the organization, with a logo, if available.

- Information on how to obtain a copy from old SIGLE records.

- Links to grey literature repositories or projects in the country

---

[3] http://nrgl.techlib.cz/index.php/Main_Page

An icon permits to start a search on the records from this partner

NTK has since enhanced information with a short presentation and a text about NRGL. In a similar way any information on conferences relating to the subject can be announced in the "News" section of the homepage.

First usage statistics

After about 4 months of running OpenGrey, the returns are positive, considering that it opened just before summer break. Frequentation of the website always goes down during summer months, but the figures for September are comparable to September 2010 (30 000 visits and 70 000 page views). 320 visits during that month came from a mobile device, 50 % from an iPad, with the iPhone in the second place. At present, OpenGrey doesn't provide any apps for these devices.

The usage of records from a given partner is monitored with an application based on a log analysis for the handle (Record identifier). Since the opening of the new website Czech records (total 5778 items) were accessed more than 22 000 times.

European grey literature, even 5 to 30 year old documents, still interests scholars all over the world. In September 2011 visitors came from over 160 countries, United Kingdom, United States and Germany being in the lead. Visits from Australia, China and India are more and more frequent.

Further developments and outlook

Adding new content to OpenGrey and enhancing existing records are the priorities for the coming months. Nevertheless, some technical issues still exist; and referencing of the website can be improved.

Indexing of OpenGrey by Google and Google Scholar started soon after the migration. It is done by packages and still ongoing. OpenGrey is OAI-PMH compliant, but since it is organized in a different way from OpenSIGLE, a set structure must still be implemented. The sets will be organised to allow harvesting on a subject basis. Look and features for several web pages (list of results, record display and partner page) can still be improved. In the back office, tools for partner organizations (such as controlled access for editing) and for the administrators (with regards to import, export, and updates) are on the waiting list.

Knowing than many documents referenced in the former SIGLE database have since been digitized or became available in electronic format, leads to another point for improvement. Our plans for future developments include the addition – if possible through batch upload – of links to the now available full text in the partner repository.

Another project planned with GreyNet is to add persistent links to datasets in the bibliographic records, when they are available in a repository. The initiative starts with the next GL conference.

Glossary:

Creative Commons: The Creative Commons copyright licenses and tools forge a balance inside the traditional "all rights reserved" setting that copyright law creates: http://creativecommons.org/licenses/

DSpace: DSpace is a software for academic, non-profit, and commercial organizations building open digital repositories: http://www.dspace.org/

Dublin Core Metadata: The Dublin Core Metadata Initiative (DCMI) is an open organization, incorporated in Singapore as a public, not-for-profit Company engaged in the development of interoperable metadata standards that support a broad range of purposes and business models: http://dublincore.org/specifications/

EAGLE: European Association for Grey Literature Exploitation, producer of the SIGLE database:
http://en.wikipedia.org/wiki/European_Association_for_Grey_Literature_Exploitation

Exalead® is a global software provider in the enterprise and Web search markets: http://www.exalead.com/software/

Google Scholar: Google Scholar provides a simple way to broadly search for scholarly literature ...: articles, theses, books, abstracts and court opinions, from academic publishers, professional societies, online repositories, universities and other web sites: http://scholar.google.com/

Handle: The Handle System provides resolution services for unique and persistent identifiers of digital objects, and is a component of CNRI's Digital object (Corporation for National Research Initiatives® is a not-for-profit organization formed to undertake, foster, and promote research in the public interest) : http://www.handle.net/

OAI-PMH: The Open Archives Initiative Protocol for Metadata Harvesting (referred to as the OAI-PMH in the remainder of this document) provides an application-independent interoperability framework based on metadata harvesting: http://www.openarchives.org/OAI/openarchivesprotocol.htm

OpenDOAR: Directory of Open Access Repositories : The OpenDOAR service provides a quality-assured listing of open access repositories around the world: http://www.opendoar.org/find.php

OpenGrey: http://www.opengrey.eu

SIGLE: System for Information on Grey Literature in Europe: SIGLE was an online, pan-European electronic bibliographic database and document delivery system for grey literature. http://en.wikipedia.org/wiki/SIGLE

php/MyVisites: is a free and open source (GNU/GPL) software for websites statistics and audience measurements: http://www.phpmyvisites.us/

WorldWideScience.org: a global science gateway comprised of national and international scientific databases and portals: http://worldwidescience.org

Zotero: is a free bibliographic manager, and a Firefox add-on. http://www.zotero.org

Literatura

Schöpfel, Joachim et al. From SIGLE to OpenSIGLE: An in-depth look on resource migration in the European context. Amsterdam, Textrelease, 2007.

Farace, Dominic, et al. Linking full-text GL to underlying research and post-publication data: An Enhanced Publications Project. (Forthcoming)