From OpenSIGLE to OpenGrey
Changes and Continuity

Christiane Stock and Nathalie Henrot

(INIST-CNRS)

First presented at the GL8[1] conference in New Orleans 2006 as a prototype, OpenSIGLE went life in December 2007. After three years of existence, the results are beyond all expectations. OpenSIGLE has become a reference source for grey literature, and its user community has grown constantly, especially from outside Europe. The integration of the GL conference preprints into the repository from 2008 onwards not only added research papers on the topic of grey literature to its contents, but also permitted OpenSIGLE to be accepted in the "Directory of Open Access Repositories" (OpenDOAR*).

In spite of the success of OpenSIGLE it's not wise to rest on one's laurels. The change of name to "OpenGrey" signifies a shift in the content of the repository as well as in its physical appearance. Besides providing a new look and a more convenient technological environment, OpenGrey closes the gap between the close of the SIGLE* database and today, including recent records and links to the full text.

The paper presents the new website which includes numerous facilities requested by users such as OAI-PMH*, the possibility to export records and an improved access to the document itself. OpenGrey also takes into account a changed user behaviour, where visitors arrive after searching Google or Google Scholar* and want all relevant information at a glance. The paper further explains input procedures and gives other information for the ongoing updates of the repository. Finally we call former SIGLE members and new partners to contribute to OpenGrey.

From SIGLE to 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 accessible through subscription to hosts, e.g. STN International, and available on a CD-ROM produced by Silverplatter/Ovid.

INIST decided to transfer the results of 25 years of work[2] onto an open access platform. As a result OpenSIGLE went live in December 2007 with almost 700 000 bibliographic records, using

---

[1] http://opensigle.inist.fr/handle/10068/697774
[2] the main effort relating to the identification and collection of documents

DSpace* technology and its qualified Dublin Core metadata* format. The operation was to be low cost and easily feasible.

It was very important that key information found in every SIGLE record was preserved during the migration:
- English title or keywords
- SIGLE classification code
- Availability statement, in order to facilitate the order of a paper copy

One of the main goals of SIGLE was to facilitate access to the paper document. Most of today's requests for assistance from users concern the document availability.

OpenSIGLE – its evolution and its usage

Though not perfect[3], with its lack of flexibility for the user, and no possibility of exporting either search results or records, OpenSIGLE found an ever growing audience and new visitors from – nearly - all around 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 is reached through the Google search engine by an increasing number of users, and visits via Google Scholar amount to 30% per month. For 2010 the average number of monthly visits exceeds 35 000. The audience is definitely worldwide, not only European: 10 % of the visitors come from Asia, 20% from North America, continents not involved in the database production. These proportions have remained fairly stable over the past year.

The following statistics (see below) are based on php/MyVisites*, an open source software using a tracker, similar to Google analytics. As a result, figures are lower than with a log analyser[4].

Two years of statistical data collection 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. Simultaneously the average duration of a visit decreased as well as the number of pages viewed in a visit.

We take the data as a sure indication of a change of behaviour in our visitors. Many of them reach the OpenSIGLE after using a search engine, spend a short time looking for relevant information and leave. However some of our users tried to replicate former search strategies in OpenSIGLE and were successful.

---

[3] mainly because it used standard DSpace features without further development
[4] older software based on simple log analysis would yield numbers at least four times higher

Average per month	2009	2010
Visits	12 000	35 000
Pages viewed	35 450	88 580
Length of visit	107 sec	91 sec
Number of pages viewed per visit	3.4	2.8

The most important highlight for 2010 relates to the Greynet collection[5]. As of October 2010 the complete collection of conference preprints from the GL conferences are accessible in full text and in open access. GL5 to GL11 were provided by GreyNet in electronic format and the first conference became available in full text on OpenSIGLE as early as May 2008. GL1 to GL4 were received in paper form, digitized by INIST and uploaded to OpenSIGLE in September 2009.

When OpenSIGLE became OAI-PMH compliant, we decided to be more explicit about the rights and adopted the Creative Commons licence*, in order to inform the users and partners about the different conditions of use of the bibliographic data and GL proceedings.

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 smaller updates. DSpace obviously reached its limits with 700 000 records in the database. Besides, the website layout was rather simple (low key) and today no longer answers current needs for referencing.
The most important reason for thinking of a change was the absence of 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 emails from users request how to obtain the document, a service which was one of the principal goals of SIGLE and OpenSIGLE. This essential information should be seen at first glance.
In fact, the availability statement is « hidden » in the full record display of OpenSIGLE. However, few users take the time to look it up. 55 percent of the visits only display one page in

---

[5] http://opensigle.inist.fr/handle/10068/697753

OpenSIGLE before leaving again, according to our statistics. In addition, in the context of the Google generation we presume that the brief record display becomes more and more the "entry page" to the database, due to the Google Scholar search, and half of the time the unique page-view. 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, INIST-CNRS intended to re-open OpenSIGLE for "new" input. This can be done in two ways:
- add records from 2005 onwards from former EAGLE* members (INIST-CNRS included).
- open the database to new European partners.
However, the technical limits mentioned above would have been a major hindrance, hence the need for a change of software.

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 will be *opengrey.eu*.

The OpenGrey homepage

The homepage of the new website OpenGrey aims to meet current needs of the users as well as to facilitate referencing. It is divided into three parts with different groups of information. The upper field holds the name and the logo, - inspired by the lemniscates or infinity symbol -, and provides information on the contents for referencing. Tabs allow the user to access further information, e.g. on the partners and former EAGLE members as well as to choose the language for the user interface.
Besides the "Google" like search field, the centre includes three blocs for short texts: a mini "about", a "focus" on specific subjects and a "news" bloc.
The bottom part, separated into four blocs, provides more information on search and help, a choice of export tools (highly requested by users), legal mentions, and further information (about, tools for partners etc.).

Change and continuity:

The major change from OpenSIGLE to OpenGrey relates to the software: The DSpace platform used for OpenSIGLE is replaced by Exalead® * as the search engine used 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

are identified by a unique identifier, the handle*. The handle system allows to using the URLs as persistent identifiers which remain, even in case of server changes. Many websites linking to OpenSIGLE or a particular collection (e.g. GL11) use the handle. In order to assure continuity the handles of the present OpenSIGLE records will migrate to the new system. A redirection is planned for the handles identifying communities and collections of OpenSIGLE which cease to exist in the same way in OpenGrey.

Another change comes with the Exalead software. The user interface for search provides the Google-like field for full text record search. Exalead also offers the possibility to refine the results through faceted search. The criteria chosen for refinements are similar to the Czech repository NUSL/NRGL[6] (author, subject, date, language…). This feature has a major consequence for the backoffice: metadata which were merged for the simpler DSpace format in OpenSIGLE need to be detailed once more, and even new controlled fields must be added to allow the refinements and to insure consistency.

As mentioned before, we know from usage statistics that the brief record display is the entry page for many visitors and the one and only page viewed by half of them. Therefore the development of features based on this record display has become a necessity. All relevant information must be available at first glance or invite the user to click further. For example, we intend to make an explicit distinction between:

       Access to the paper copy

       Access to the online full text

In addition we must provide easy links to both fields and 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

OpenGrey will provide each partner with a space for information on its institution, a kind of "homepage", accessible from a tab "Partners" on the OpenGrey homepage. OpenSIGLE already offers information on former EAGLE members in the collection page of the "Country Community"[7]. This content will be transferred to the new partner homepage as a starting point, namely:

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

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

Each partner is strongly encouraged to propose any improvement or information about its organization, its policy of document delivery and the fees attached etc… Any information on

---

[6] http://nrgl.techlib.cz/index.php/Main_Page
[7] E.g. NTK : http://opensigle.inist.fr/handle/10068/20

national grey literature initiatives, databases, recent updates, or any information on conferences relating to the subject could be announced in the "News" section of the Homepage.

Contents, input and updates of OpenGrey:

OpenGrey will include all records from the present OpenSIGLE database. In addition, we intend to add new records (created since 2005) from partner organizations, including French grey literature as mentioned before. OpenGrey will not host the documents themselves. The new cooperation will be based on formal agreements with each partner organization. The partner retains the control over the grey documents. An open access policy is encouraged for the full text documents, items which remain under the control of the partner. The bibliographic records as well as the OpenGrey website will be placed under the Creative Commons Licence.

OpenGrey metadata should be as rich as possible. This is why we prefer to receive files in xml format via ftp, although metadata harvesting with OAI-PMH is not excluded. However, harvested material will most likely be in basic Dublin Core format, which is poor material for Exalead®'s refinement feature.
We defined a metadata scheme based on the DSpace scheme (qualified Dublin Core), but taking into account new needs (e.g. for Exalead®'s refinements) and ideas from similar repositories. Suggestions for additions are very welcome.

Updates to the OpenGrey database are done by batch upload of files, excluding direct deposit through the interface, contrary to institutional repositories[8].
Tools for participants will be made available through the website. It might be interesting to open a community space dedicated to partners. On a first meeting it was suggested to create an advisory board to discuss arising questions.

Outlook:

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 might start with the next GL conference.

---

[8] However, a special development might allow partner administrators from partners to directly edit individual records.

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

OpenSIGLE: http://opensigle.inist.fr

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