

## Grey literature in French digital repositories: a survey

Joachim Schöpfel, Christiane Stock

► **To cite this version:**

Joachim Schöpfel, Christiane Stock. Grey literature in French digital repositories: a survey. GL10. Tenth International Conference on Grey Literature: Designing the Grey Grid for Information Society. Amsterdam, 8-9 December 2008, Dec 2008, Netherlands. 2009. <sic\_00379232>

**HAL Id: sic\_00379232**

**[https://archivesic.ccsd.cnrs.fr/sic\\_00379232](https://archivesic.ccsd.cnrs.fr/sic_00379232)**

Submitted on 13 May 2009

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



## Grey literature in French digital repositories: a survey

Joachim Schöpfel

*Charles de Gaulle University of Lille 3*

Christiane Stock

*Institute for Scientific and Technical Information (INIST-CNRS)*

### Abstract

The impact of open archives on the availability and selection of scientific and technical information is growing. Yet, there is little empirical evidence on the deposit and processing of grey literature in digital repositories.

The purpose of this communication is to provide a survey on grey literature in French open archives, e.g. institutional and subject-based digital repositories.

The survey is based on a selection of 56 representative French digital repositories. The different archives are selected through national and international registries of OAI repositories, following a defined set of criteria. The repositories are shortly described (type of repository, scientific domain, software, size, language, institution).

Five aspects are analysed for each digital repository:

1. Typology of grey documents (in particular, theses and dissertations, reports, conference proceedings, working papers, courseware).
2. Part of grey literature in the whole archive (in %).
3. Specific metadata related to grey literature.
4. Quality control and policies (evaluation, validation).

## 5. Conditions of access to the full text.

These information and data are linked to the characteristics of the repositories mentioned above, and specific features of grey literature are discussed.

Furthermore, the question if the New York definition of grey literature applies to the content of digital repositories is discussed.

The communication provides an overview of the preservation and dissemination of grey literature in French digital repositories, contributes to the discovery of French grey literature and open archives, and moves forward the debate on the future of grey literature in the environment of digital repositories.

### **Notes on the authors**

**Joachim Schöpfel** is senior lecturer in information and communication sciences at the Charles de Gaulle University of Lille 3, scientist at the GERiiCO laboratory and associated member of the research group "Document numérique & Usages" (University of Paris 8).

University of Lille 3, UFR IDIST, GERiiCO Laboratory, BP 60149, 59653 Villeneuve d'Ascq Cedex, France.

[joachim.schopfel@univ-lille3.fr](mailto:joachim.schopfel@univ-lille3.fr)

**Christiane Stock** is the head of Monographs and Grey Literature at INIST-CNRS and gives lectures on grey literature.

INIST-CNRS, 2 allée du Parc de Brabois, F-54519 Vandoeuvre-lès-Nancy Cedex, France.

[stockc@inist.fr](mailto:stockc@inist.fr)

## 1. Introduction

“New possibilities of knowledge dissemination (...) through the open access paradigm via the Internet have to be supported. (...) A complete version of the work (...) is deposited (and thus published) in at least one online repository using suitable technical standards (such as the Open Archive definitions) that is supported and maintained by an academic institution, scholarly society, government agency, or other well-established organization.”<sup>1</sup>

On 22 October 2003, five years ago, 19 major European scientific organizations signed this *Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities*. In January 2006 the European Commission published the *Study on the Economic and Technical Evolution of the Scientific Publication Markets of Europe* with policy recommendations in favour of open repositories (“Research funding agencies ... should promote and support the archiving of publications in open repositories”, cf. Dewatripont et al. 2006).

In December 2006, the European Research Advisory Board released a report on scientific publication and policy on open access<sup>2</sup> and recommends, “that the Commission should consider mandating all researchers funded under FP7 to lodge their publications resulting from EC-funded research in an open access repository”. A petition for guaranteed public access to publicly-funded research results launched in early 2007 was signed by more than 27,000 scientists and several hundreds organizations<sup>3</sup>.

In France 17 scientific and academic institutions support the Berlin Declaration. French universities and research organizations signed in July 2006 an agreement on the development of a common infrastructure of open repositories. Central parts of the French “jigsaw puzzle” (André et al. 2007) are the CNRS Center for Direct Scientific Communication<sup>4</sup> at

---

1 <http://oa.mpg.de/openaccess-berlin/berlindeclaration.html>

2 [http://ec.europa.eu/research/eurab/pdf/eurab\\_scipub\\_report\\_recomm\\_dec06\\_en.pdf](http://ec.europa.eu/research/eurab/pdf/eurab_scipub_report_recomm_dec06_en.pdf)

3 <http://www.ec-petition.eu/>

4 <http://ccsd.cnrs.fr>

Lyon and, since November 2008, the institutional repository portal<sup>1</sup> launched by the French academic consortium COUPERIN. -

Last year, the European DRIVER study evaluated France as an advanced country in the open archives landscape (see Van de Graaf & Van Eijndhoven 2007).

Grey literature represents a substantial part of the scientific production (cf. Schöpfel & Farace 2009). Since the 7<sup>th</sup> International Conference on Grey Literature (Farace & Frantzen 2006) at Nancy, the GreyNet community intensified its research activities on the impact of the open access movement on the grey literature. Special attention was paid to institutional repositories, public policies, organisational context and e-infrastructure. Several case studies highlighted the national, cultural and domain-specific differences.<sup>2</sup> All the same, they also confirmed the force and dynamic of this global movement towards unrestricted access to scientific information.

The purpose of our study is to evaluate the integration of grey literature in French open archives. In particular, five aspects are analysed for each digital repository:

1. The typology of grey documents (e.g., theses and dissertations, reports, conference proceedings, working papers, courseware).
2. The relative part of grey literature in the whole archive.
3. The assignment of specific metadata related to grey literature.
4. Information about quality control and policies.
5. The conditions of access to the full text.

Whenever possible, data on development (evolution of deposit) and usage (statistics of access and downloads) are added. These information and data are linked to the characteristics of the repositories mentioned above, and specific features of grey literature are discussed.

The communication provides an overview of the preservation and

---

1 <http://www.couperin.org/archivesouvertes/>

2 See the GL proceedings at the GreyNet website <http://www.greynet.org/greytextarchive.html> and the published articles in *The Grey Journal*, especially the two issues on “Repositories – Home2Grey” (2005, vol. 1, n° 2) and “Grey matters for OAI” (2006, vol. 2, n° 1).

dissemination of grey literature in French digital repositories, contributes to the discovery of French grey literature and open archives, and moves forward the debate on the future of grey literature in the environment of digital repositories.

## **2. Methodology**

The survey is based on a selection of 56 representative (e.g. registered either with a dedicated platform or as data provider for harvesting) French digital repositories. The different archives were selected through eight significant international registries of OAI repositories or service providers:

### ***BASE***

Bielefeld Academic Search Engine.  
<http://base.ub.uni-bielefeld.de/index.html>

### ***DSpace***

Repositories using Dspace – Alphabetical.  
[http://www.dspace.org/index.php?option=com\\_content&task=view&id=596&Itemid=182](http://www.dspace.org/index.php?option=com_content&task=view&id=596&Itemid=182)

### ***Eprints***

Sites Powered by Eprints.  
<http://www.eprints.org/software/archives/>

### ***OpenDOAR***

Directory for Open Access Repositories.  
<http://www.opendoar.org/>

### ***ROAR***

Registry of Open Access Repositories.  
<http://roar.eprints.org/>

### ***Scientific Commons***

Register URL  
<http://en.scientificcommons.org/register-repository>

### ***University of Illinois OAI-PMH Data Provider Registry.***

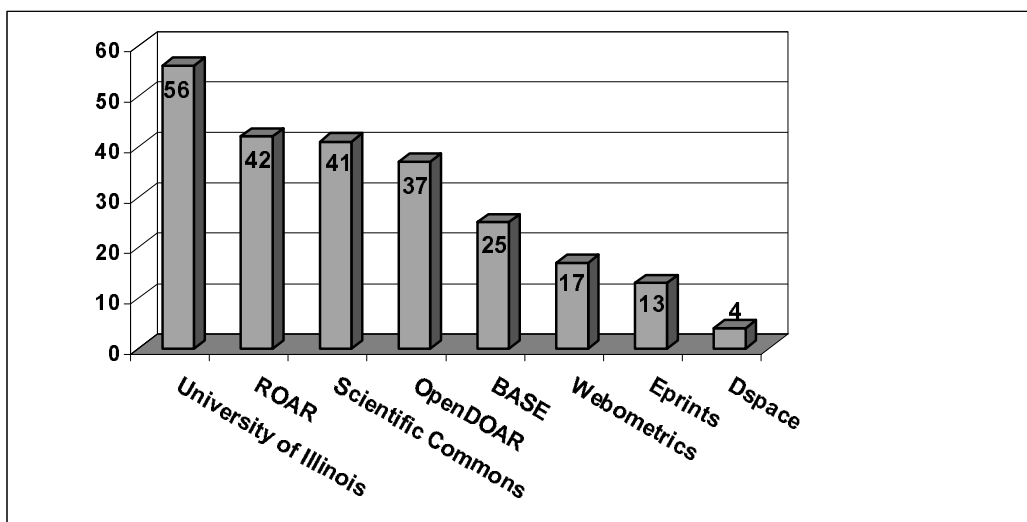
<http://qita.grainger.uiuc.edu/registry/>

## Webometrics

Ranking Web of World Repositories.  
<http://repositories.webometrics.info/>

The selection took place between March and May 2008 and followed a defined set of criteria (located/hosted in France, living archive, size>0).

Figure 1 shows for each registry the number of French archives compliant with the criteria.



**Figure 1: Number of French archives in international registries (March-May 2008)**

Each registered archive (URL) was checked; errors (incorrect URLs etc.) and duplicates were eliminated. Information about the 56 remaining archives were incorporated into a spreadsheet with 37 data columns in 5 categories (see appendix):

1. General (background) information about the archive (10 data elements).
2. Specific information about the archive (6 data elements).
3. Content information (12 data elements).
4. Qualitative data (7 data elements).

## 5. Comments (2 data elements).

If the information for a specific field was unavailable or uncertain, it remained open.

The data were analyzed with basic Excel statistical functions. Qualitative information was added from the spreadsheet if necessary. Several archives had to be excluded, because the URL was no longer valid or no user interface was provided allowing us to obtain data.

## 3. Results

The leading questions for the data analysis are:

- How can the current situation of open archives in France be described?
- Which is their content?
- Which is the importance of grey documents in these archives?
- Which are the main aspects of grey material in French open archives?
- How are grey documents used?

Based on empirical evidence, the following sections try to provide at least partial responses.

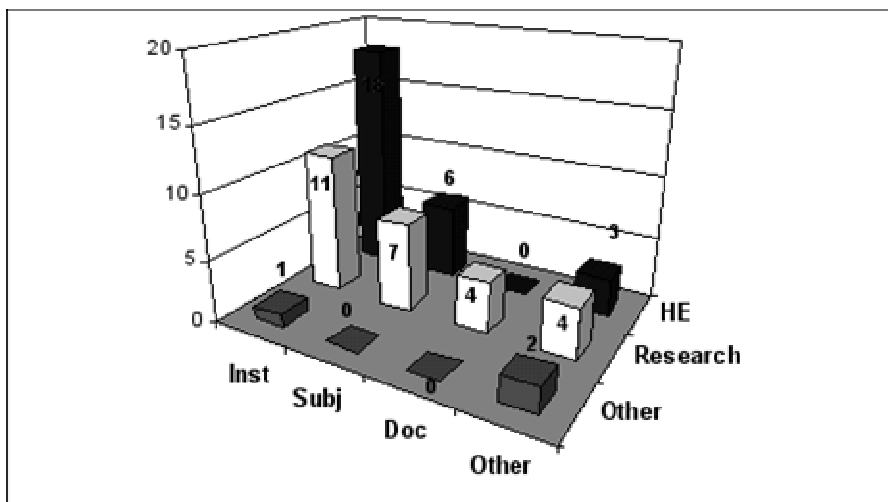
### 3.1. General characteristics of French open repositories

#### 3.1.1. *Institutions and typology of archives*

One half of the French open archives are owned and/or hosted by Higher Education establishments (HE), e.g. universities and engineering schools, with Strasbourg, Lyon and Paris universities in leading positions. The other half is from public research institutes; mostly from the multidisciplinary national research centre CNRS, some other from



INRA (agronomics) or INRIA (applied computer sciences). Only three archives are from other types of organizations.

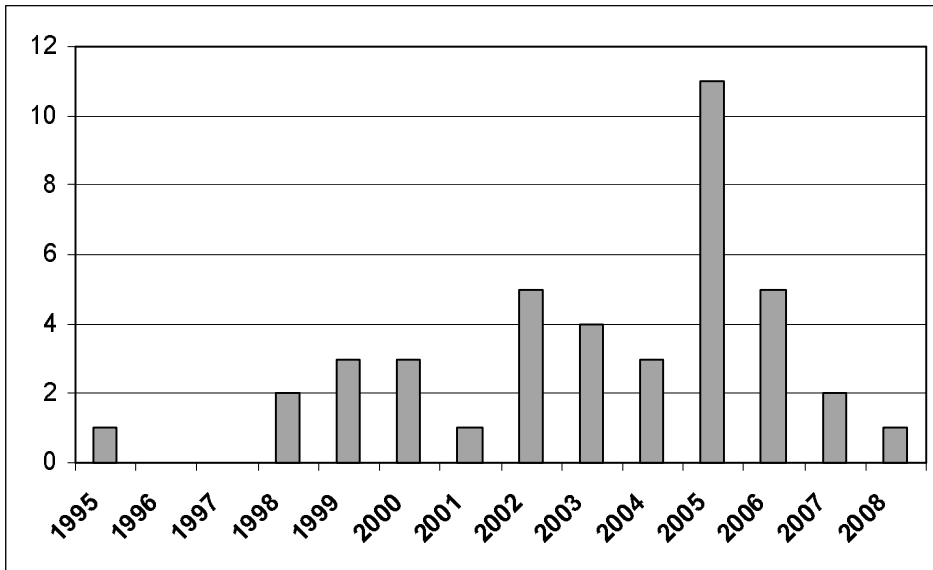


**Figure 2: Institutions and repository typology**

Half of the archives are institutional repositories designed for publications from the scientific authors of the specific institution. In particular, 67% from the HE archives (n=18) are in this category, confirming the academic interest to increase the visibility of scientific production (figure 2).

### **3.1.2. Date of creation**

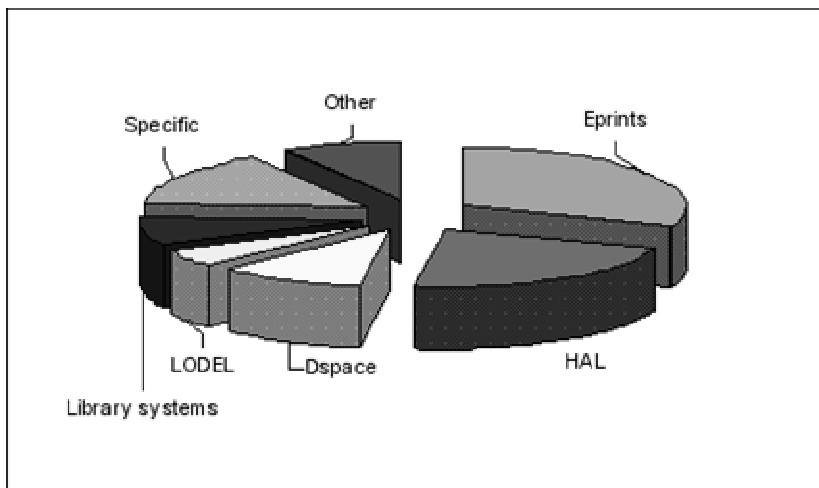
For 16 repositories, we could not determine the exact date of creation. Most of the others were launched in 2005 or later (figure 3). The figures for 2006 onwards would be even higher, had not HAL been agreed upon as a national repository for French research organizations.



**Figure 3: Date of creation of the repositories (updated)**

### 3.1.3. Software

In spite of some early initiatives in favor of national and hegemonic software, the current situation is pluralistic with some major OA-systems and specific (local) solutions.



**Figure 4: Software**

Two-thirds of the repositories were developed with well-established and OAI-PMH-compliant software, namely *Eprints* (CA/UK), *HAL* (F) and *DSpace* (US). This choice offers the opportunity to collaborate with national and international user groups on problem solving and product development.

This landscape will probably change in the next months. The new French open access software *OAI-ORI*, a specific open source solution for HE institutional archives, was launched earlier this year. Nevertheless, during the period of the empirical study (March-May) *OAI-ORI* was only implemented on experimental sites.

### 3.1.4. Language

54 repositories provide French-speaking interfaces, 31 of them exclusively. 25 archives supply at least partial English information for users, two of them also German and Spanish information.

Language (interface)	Number of archives
French	31
English	2
French and English	21
French, English, German, Spanish	2

**Figure 5: Language of interface**

The two fully English-speaking repositories are datasets archives, created for and by international scientific communities (astronomy, crystallography).

## 3.2. Content: scientific domains, types of material and size

### 3.2.1. Scientific domains

The French open repositories and especially the multidisciplinary and often institutional archives cover most of the scientific disciplines. Nevertheless there are some characteristics of the French open access landscape.

Scientific domain	Number of archives
Multidisciplinary	19
Social sciences & humanities	20
<i>Linguistics</i>	4
<i>Library, information &amp; communication sciences</i>	3
<i>Ethnology &amp; cultural studies</i>	3
Applied sciences	11
<i>Engineering</i>	5
<i>Computer sciences</i>	3
<i>Agronomics</i>	3
<i>Telecommunication</i>	1
Sciences, medical sciences	6
<i>Physics</i>	3
<i>Astronomy</i>	2
<i>Chemistry</i>	1
<i>Geochemistry</i>	1

<i>Mathematics</i>	<i>1</i>
<i>Medicine</i>	<i>1</i>
<i>Sport</i>	<i>1</i>

**Figure 6: Scientific disciplines**

Compared to other countries, in particular the US and the United Kingdom, a large important archive for French medical and/or life sciences is missing so far. This probably has two explanations, the importance and force of attraction of the PubMed Central for all international scientists, and the decision of the two French public research organizations with significant research activity in medical and life sciences, the CNRS and INSERM, in favor of a national, multidisciplinary article-based repository (HAL-CCSD).

### ***3.2.2. Types of material***

The content of the repositories is widespread and of great diversity. A non-exhaustive inventory based on the repositories descriptions gives evidence for more than 20 different types of materials:

*books*  
*manuscripts*  
*speech samples with transcriptions*  
*maps*  
*images*  
*datasets (astronomical observation, crystallography)*  
*journals (backfiles, current issues)*  
*articles*  
*proceedings*  
*courseware*  
*posters*  
*videos*  
*patents*  
*dissertations and theses*  
*bibliographical records*  
*reports*

*preprints*  
*other unpublished materials*  
*cultural heritage materials (rare books)*  
*websites*  
*software*

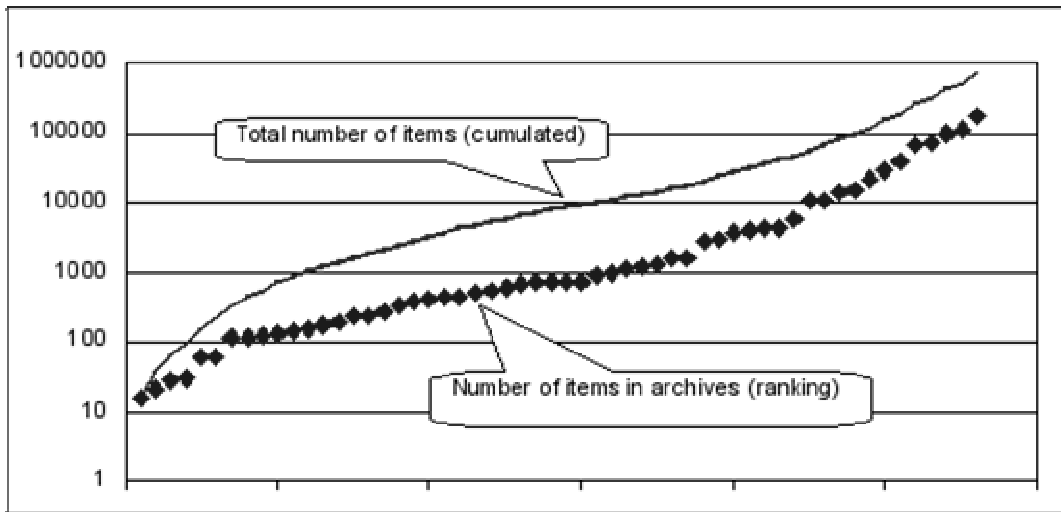
53 archives contain textual material (written documents), seven of them together with other items (datasets, images, maps etc.). Only three repositories don't contain any written document (oral documents and other datasets).

Four archives - all of them produced by the national research organization CNRS - are document-specific, e.g. designed for one specific category of documents and not limited to one institution. The CNRS created a national site for open access journals especially in social sciences and humanities (*revues.org* hosted by CLEO). The other three repositories are dedicated to grey literature: a site for French scientific and technical reports (*LARA* hosted by INIST) and two archives for French electronic theses and dissertations (*TEL* for PhD theses and *MemSIC* for Master theses, both hosted by CCSD).

### **3.2.3. Size of repositories**

The size of the repositories varies largely, between a minimum of 16 items and a maximum of 172,215 items (average size 12,500 items, median size 713 items). Together they total 704,578 deposited items.

32 archives contain less than 1,000 items. Together, they represent 57% of the total number of archives but only 2% of the overall number of items (documents, datasets etc.).



**Figure 7: Size of repositories (number of deposited items)**

On the other side, 12 archives (21%) contain more than 10,000 items each or 94% of the overall number of items in French archives. These most important archives are the following:

<b>Name</b>	<b>Host</b>	<b>Organisation</b>
TEL	CCSD	CNRS
<i>HAL INRIA</i>	CCSD	CNRS
<i>I-Revues</i>	INIST	CNRS
<i>HAL SHS</i>	CCSD	CNRS
<i>Revues.org</i>	CLEO	CNRS
<i>NumDam</i>	MathDoc	CNRS
<i>GALLICA</i>		BNF
<i>Horizon Pleins Textes</i>		IRD

<i>Crystallography Database</i>	<i>Open</i>		University of Maine
<i>ProdINRA</i>		INRA	INRA
<i>HAL</i>		CCSD	CNRS
<i>PERSEE</i>		Lyon 2	University of Lyon 2

**Figure 8: The 12 most important repositories (size)**

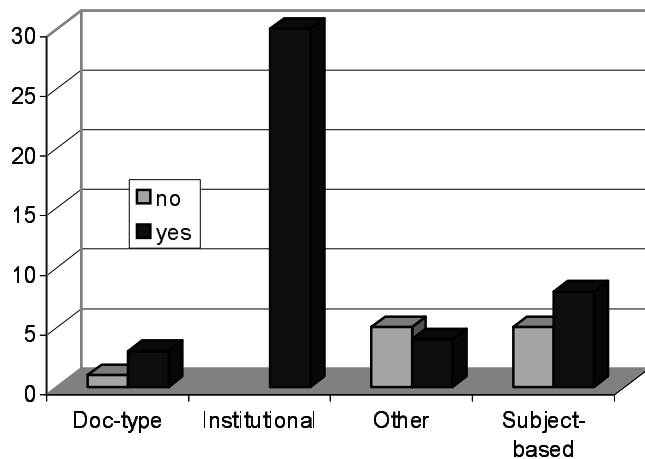
The role of the CNRS is significant; the organization hosts and/or produces more than 30% of the total number of items. However, three of the cited archives provide a mixture of bibliographic records and full text documents (HAL, Horizon Pleins Textes, ProdINRA).

### 3.3. Grey content

According to OpenDOAR data, 50% of the French repositories contain theses and dissertations, 35% conference or workshop papers and 32% unpublished reports or working papers (October 2008). Reports are frequently associated with journal articles and conference papers, whereas 50 % of repositories containing ETD's (10 out of 20 sites) are dedicated exclusively to this type.

Our own survey shows that a significant part of French repositories (79%) includes at least one category of "traditional" grey literature (theses or dissertations, reports, conferences, working papers, courseware etc.). Even more interesting is the fact that 100% of the institutional archives give access to grey material (figure 9).

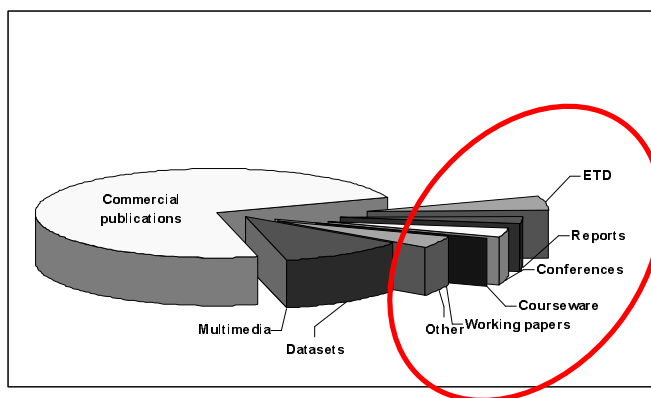




**Figure 9: Type of archive and presence or grey content (nb of OA)**

18 open archives are 100% grey, e.g. their content is set up by theses and dissertations (14), conference papers (2), reports (1) and courseware (1). Nevertheless, their importance is limited. Together, these "grey OA" contain but 2,5% of all publications in French OA.

The overall part of grey documents (items) in the global French OA content is 16%, e.g. one out of six deposited publications in French archives is grey literature. The other material is commercial (mainly journal articles), multimedia and datasets (figure 10).



**Figure 10: Document types (nb of items in OA)**

One third of the deposited grey documents are electronic theses and dissertations, followed by conference papers (22%) and reports (16%). Surprisingly low - below 1% - are the indexed deposits of courseware and working papers. On the other hand, the part of undefined grey documents is relative high - 28% (especially in three archives from IRD

and CCSD).

<b>Grey material</b>	<b>Relative part</b>
ETD	34%
Conference papers	22%
Reports	16%
Courseware	0,3%
Working papers	0,2%
Other	28%

**Figure 10: Typology of grey documents**

Related to the size of the archive and the number of grey items, we can distinguish five types of repositories (figure 11):

- (1) Important archive, no grey material: PERSEE (only journal articles).
- (2) Important archive, relative high number of grey items: IRD, HAL.
- (3) Important archive, average number of grey items: INRA.
- (4) Medium-sized archives, average number of grey items: TEL, HAL-SHS, INRIA.
- (5) Smaller archives, no grey content or low number of grey documents.

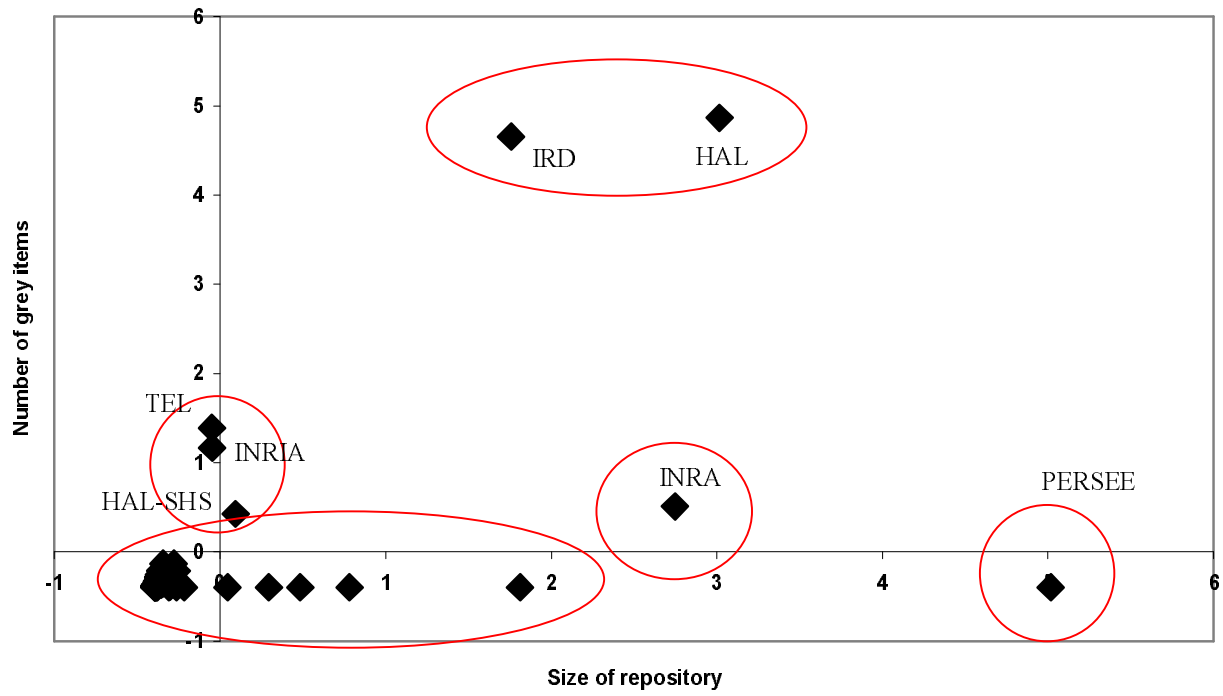


Figure 11: Size of repository and number of grey items (standard scores)

### 3.4. Qualitative aspects of grey content in French repositories

#### 3.4.1. Policy statements

Keith Jeffery in his paper on "Greyscale" (Jeffery 2007) asked whether a repository mentions an "institutional policy to mandate deposition of material". The OpenDOAR registry provides information about policy statements of archives and distinguishes 5 aspects:

- Metadata re-use policy
- Full data item
- Content
- Submission
- Preservation

For the 38 French archives registered with OpenDOAR at the time of our

survey 21 sites give no policy statement at all. For the remaining 17 sites we find the following statements (figure 11).

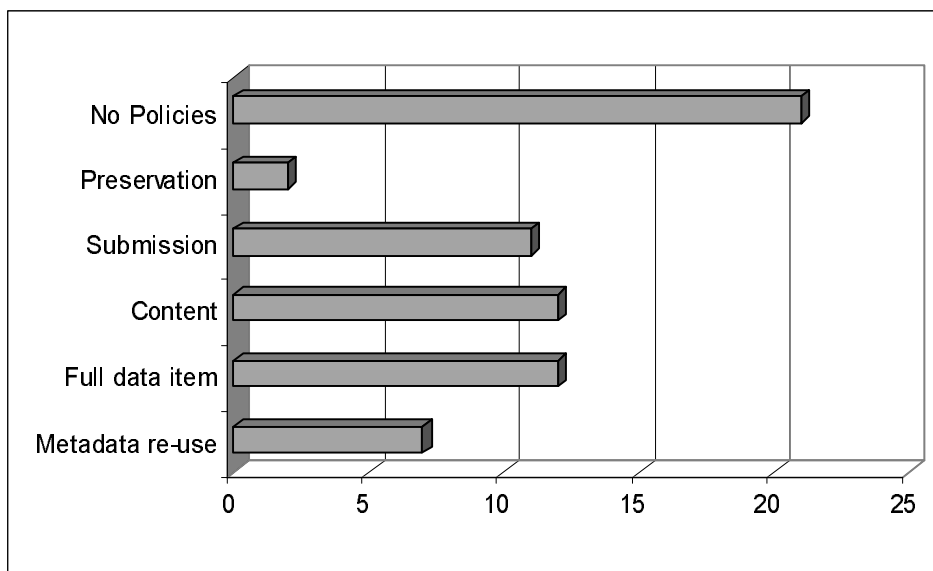


Figure 11: Policies defined by repositories (Source: OpenDOAR)

Policies are expressed in comparable proportions with regards to full data item reuse, content and submission. Metadata re-use is probably implicit for many in the OAI-PMH context, whereas preservation policies are mentioned only twice. Although the majority of the 17 sites make more than one statement, only one repository (OATAO, created in 2007) give information on all 5 issues.

L'Hostis (2006, 23) provides additional information about the mandatory deposit for publications within the major French research organizations. One of the most successful institutes with regards to the submission policy for its research output (effective since 1992, and attaining almost 100%) is the Cemagref institute (agricultural and environmental engineering research). Strangely enough this organization has no visibility as institutional archive whatsoever and doesn't appear in the registries used for our study. Cemagref publications may be accessed through a database (Cemadoc), whenever the full text is available.

CNRS, among the first organizations to sign the Berlin declaration on open access, and operator of HAL, still doesn't oblige it's researchers to

submit their documents to HAL.

### 3.4.2. Metadata

Three main grey document types occur in our survey: theses, reports and conference papers. We consider that specific metadata are added when at least one of the following elements is given.

**Report:** report number, funding organization, project name.

**Doctoral thesis or dissertation:** defense date, university, degree, discipline, and thesis advisor.

**Conference:** name, date, and place (town).

Although 45 repositories contain grey documents, only 37 of them add specific metadata. For the remaining 8 archives either the part of grey documents is very low, or they hold particular documents and fall into the category "other" document type.

Among those who add specific metadata, the number and quality of information vary: from adding the name of the university or defense date for a doctoral thesis to the members of the jury or very detailed information for reports on sponsors or projects.

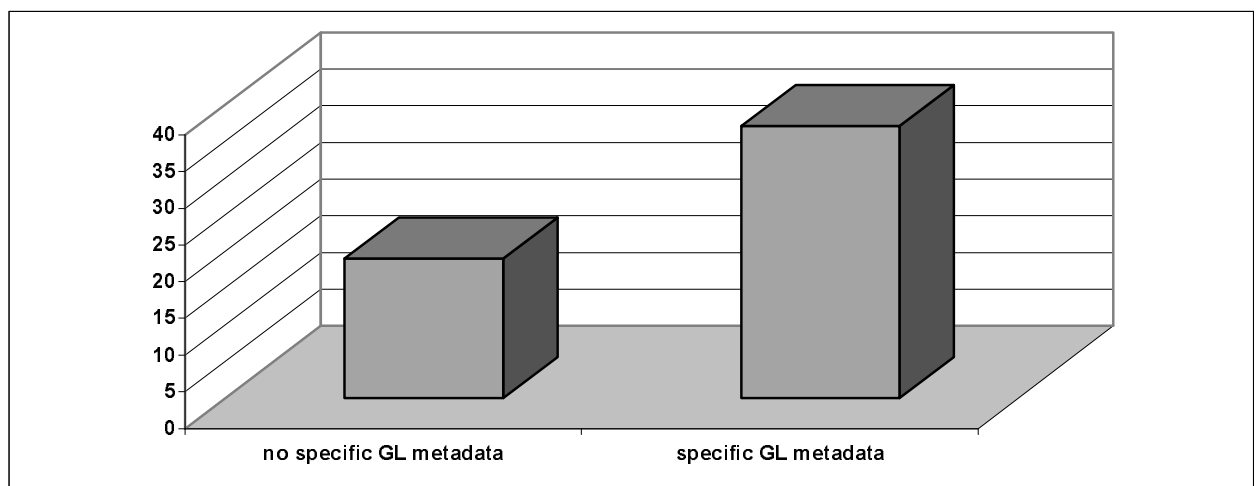


Figure 12: Repositories with specific metadata for grey documents

### 3.4.3. Access to the full text

Both ROAR and OpenDOAR registries deal with the distinction between metadata records and access to full text. Data supplied by OpenDOAR refer to full text items only, whereas ROAR gives an estimate on the availability of full text.

In our survey 71% of the repositories in France provide access to full text, and for 48% of the sites the entirety of the documents is available in open access. We distinguish two categories with restrictions:

- Part of the archive is accessible through an intranet or limited to a community.
- A moving wall for commercial repositories. The goal of e-publishing platforms such as I-Revues is to provide access to the full text, but for some titles a temporary embargo is applied.

16% of the archives contain a mixture between bibliographic records and full text. A part of them (e.g. IRD - Research Institute for Development) provide access through a library catalogue, which necessarily includes bibliographic records. ProdINRA currently enhances a bibliographic database to add full text documents. HAL included a publication database of CNRS researchers in its archive. However, it's possible to search for full text records only.

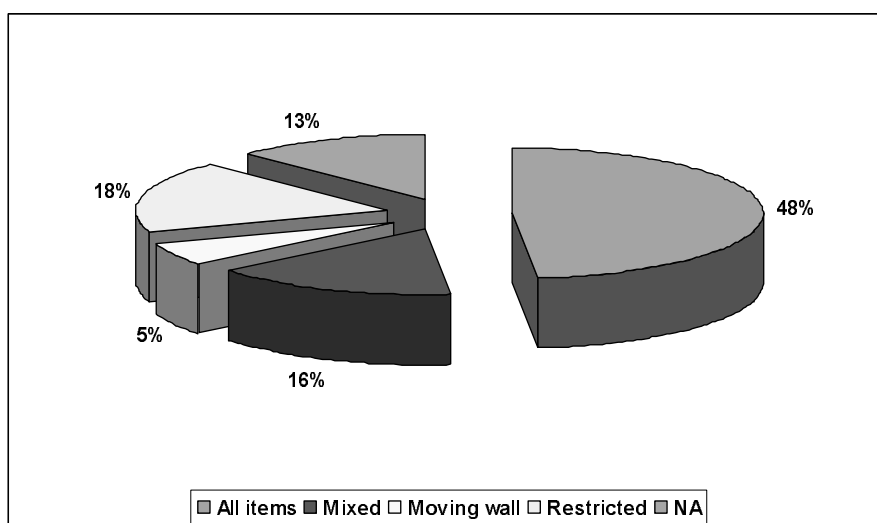


Figure 13: Access to full text

### 3.4.4. Quality control

Along with the explicit information on the surveyed web sites, about 41% of the repositories mention some kind of quality control and/or evaluation of the archived grey documents.

Above all, this quality control concerns electronic theses and dissertations that have been evaluated before their deposit. A few number of archives mention "archive administrators" who obviously act as a kind of scientific editor for the selection (but not for the revision, as far as one can see) of materials. Others only accept peer reviewed or published documents (mostly not grey, however), "outsourcing" by the way the quality control.

### 3.4.5. Usage statistics

During the period of the survey (March-May 2008), no reliable statistics or other usage related data or information were found on the repositories' web sites.

Shortly before the GL10 conference, we discovered the IFREMER report on the functioning and the usage of the IFREMER institutional archive (Merceur 2008). This report not only publishes the usage statistics (cumulated downloads of archived items) since the creation of the repository (April 2004) but also compares the usage of different types of documents.

The result is rather interesting (figure 14).

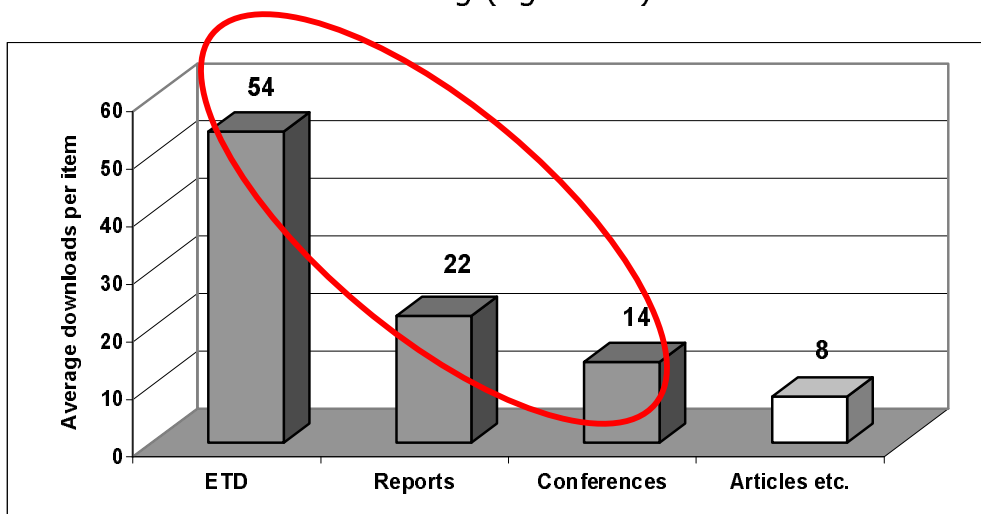


Figure 14: Usage of different document types in the IFREMER archive (source Merceur 2008)

Even if the IFREMER archive contains two times more white material than grey (e.g. articles, books), the average download per item is up to seven times higher for grey documents, especially for theses and dissertations but also for reports and conferences.

## 4. Discussion

Difficulties we came across for this survey were numerous, making it sometimes necessary to go into details such as counting items, reading records or even the documents themselves to obtain information. In the following we shall discuss some significant problems.

### 4.1. Counting items

The overall number of items in a given archive is difficult to define. Data obtained may differ from one source to another depending on what is taken into account: all items, the automatic item count, only full text items, only open access items, items open to harvesting, etc.

On the national level we face the problem of double or triple entries. This situation is similar to other countries; the one and same document may be counted two to four times, because it is included in different repositories. For example a PhD thesis may be submitted to PASTEL and to TEL, then integrated into HAL.

Another confusing situation exists in Toulouse. Several technical universities maintain ETD repositories (INP and INSA and UPS). All their documents can be harvested through a specific website "Toulouse theses", with addition of the items from the veterinary school. OATAO (Open Archive Toulouse Archive Ouverte) is an institutional repository of the recently founded PRES (group of universities and engineering schools), mainly for articles, eprints, but including some ETDS as well. It



is planned to have a unique repository for Toulouse in the future.

#### **4.2. Where to find reliable information**

Certain information are difficult to obtain. Policies can be expressed anywhere: on the homepage or the "about" page to an article which is deposited in the archive. Administration, validation and quality control of submitted items are often enough part of the back office and difficult to assess from the outside.

#### **4.3. How to identify grey material**

Identifying grey content and obtaining quality information, especially reliable numbers, is a real challenge. As mentioned above, it sometimes became necessary to open all items of a given category to control if they are grey. Unfortunately this was not possible for the bigger archives where we know that inconsistencies exist. The earlier archives such as ArchiveSIC and HAL in particular have constantly refined categories (document type, discipline) since their beginnings without always updating existing metadata. Therefore an unknown number of grey documents like reports or conference papers can be found in categories such as "Miscellaneous" or "other".

#### **4.4. The HAL case**

One of the oldest archive in France, HAL became the national archive for scientific and technical organizations in France in 2006. At the same moment, other independent archives such as TEL were integrated into the global HAL, and customized views or portals were created.

As shown in the list below only 25% of the "sub-portals" are referenced in international registries. Most of them led an independent life before 2006. This "fusion" explains the low figures for repository creations from 2006 onwards.

HAL Portals

- Generic

-HAL, TEL, CEL

- Thematic

-*Archive-EduTice, HAL-SHS, Artxiker, @rchiveSIC, HAL-CSS, HAL-SDE, hprints.org, (MemSIC)*

- Institutional

-*HAL-IN2P3, INRIA, Institut Nicod, INSERM, UJM, PRUNEL, LIRMM, Académie des Sciences, EMSE, IRD, CIRAD, PASTEUR, OBSPM, Bioemco, INERIS, CEA, INSU, SSA, IRSN, METEO, UNICE, Paris Descartes, Univ-Paris1, MNHN, SUPELEC, UNIV-BREST, UNILIM*

This situation may change in the next years with the evolution of independent institutional archives hosted and maintained by the universities themselves.

#### **4.5. Usage statistics**

The missing information on usage and access to open archives in France confirms the statement of the 2007 DRIVER study that 70% of the repositories do log the statistical data on access but analysis and interpretation are "in development" or "problematic".

In some cases (CCSD), the depositing authors have dynamic access to «their» statistics, e.g. the figures on hits and downloads of records and full texts. Nevertheless, no global figures are provided.

There may be different explanations: technical problems with software development, conceptual problems with standards, problems related to project planning and priorities, missing capacities for data capture and interpretation, low usage data. Even so, the main problem seems to be the silence, the general lack of any explanation why data are missing or not provided. In a competitive environment where commercial publishers and other vendors provide detailed and standard statistics on usage of journals, databases and e-books, and where the significant public investment in open access creates new business and communication models, public structures can't justify missing information about usage of their repositories.

## **5. Conclusion**

Our survey, even if the dataset remained incomplete for reasons we indicated above, describes a landscape in movement. Pushed by the information market and fostered by new technologies, information services, communication channels and behaviors of scientific communities are undergoing rapid change. The situation of French open archives is changing, and we already mentioned the most important factors of change, e.g. the development of independent institutional archives by the French universities, supported by the academic consortium COUPERIN.

The survey shows how the grey literature takes its place in this environment. The impact of grey material – theses, reports, conferences etc. – in open archives is real and will stay. In the future, the link to new items, multimedia, datasets etc., will need attention and exploration.

On the other hand, the survey reveals three main problems of French open archives, especially in relationship with grey literature:

(1) Policy statements need improvement. Often, the strategy and positioning of repositories are not explicit or simply missing.

(2) Especially grey items in open archives need improved bibliographic control. Compared to traditional cataloguing standards, metadata for grey material are less specific or again, simply missing. This is a problem for referencing, efficient search strategies and evaluation.

(3) Mostly wanted are detailed usage statistics on access and download of documents and other items in open archives.

The survey didn't gather data on the development of the archives (evolution of deposit). This, together with a deeper investigation of usages data, will be the object of a follow-up study in 2009/2010.

## **6. Bibliography**

André F. et al. "Institutional repositories. The repository jigsaw". *Research Information* 2007, April/May, p. 27.

Baruch P. "Open Access Developments in France: the HAL Open Archives System". *Learned Publishing* 2007, vol. 20, p. 267-282.

<http://hal.archives-ouvertes.fr/hal-00176428/fr/>

DOI : [10.1087/095315107X239636](https://doi.org/10.1087/095315107X239636)

Bruley C.; Huet N.; Kalfon J.; Thirionet G. "Bilan d'une enquête sur les archives ouvertes dans les établissements d'enseignement supérieur et de recherche". *AMETIST* 2007, n° 2.

<http://ametist.inist.fr/public/pdf/N2P2A1.pdf>

Correia A.M.R.; Neto M.D. "The role of eprint archives in the access to, and dissemination of, scientific grey literature: LIZA - a case study by the National Library of Portugal". *Journal of Information Science* 2002, vol. 28, n° 3, p. 231-241.

Davis P.M.; Connolly M.J.L. "Institutional Repositories". *D-Lib Magazine* 2007, vol. 13, n° 3/4.

<http://www.dlib.org/dlib/march07/davis/03davis.html>

Dewatripont M. et al. *Study on the Economic and Technical Evolution of the Scientific Publication Markets of Europe*. Final Report. European Commission, Brussels, 2006.

[http://ec.europa.eu/research/science-society/pdf/scientific-publication-study\\_en.pdf](http://ec.europa.eu/research/science-society/pdf/scientific-publication-study_en.pdf)

Farace D.J.; Frantzen J. (ed.) *Seventh International Conference on Grey Literature: Open Access to Grey Resources, 5-6 December 2005*. GreyNet, Grey Literature Network Service. TextRelease, Amsterdam, 2006.

Jeffery K.; Asserson, A. "Greyscape". In: *GL9 Conference Proceedings. Ninth International Conference on Grey Literature: Grey Foundations in Information Landscape*. Antwerp, 10-11 December 2007.

L'Hostis D.; Aventurier P. *Archives ouvertes - Vers une obligation de dépôt ? Synthèse sur les réalisations existantes, les pratiques des chercheurs et le rôle des institutions*. Report. 2006.

[http://archivesic.ccsd.cnrs.fr/sic\\_00115513/fr/](http://archivesic.ccsd.cnrs.fr/sic_00115513/fr/)

Lynch C.A. "Institutional repositories: Essential infrastructure for scholarship in the digital age". *ARL Bimonthly Report* 2003, 226, 1-7.

<http://www.arl.org/newsltr/226/ir.html>

Merceur F. *Fonctionnement et usages d'une archive institutionnelle*. IFREMER, October 2008.

<http://www.ifremer.fr/docelec/doc/2008/rapport-4632.pdf>

Schöpfel J.; Farace, D.J. "Grey Literature". In Bates, M.J. & Maack, M.N. (ed): *Encyclopedia of Library and Information Sciences*. 3rd edition. Taylor & Francis 2009 (forthcoming).

Van de Graaf M.; Van Eindhoven K. *The European Repository Landscape: Inventory Study into the Present Type and Level of OAI-Compliant Digital Repository Activities in the EU*. Amsterdam, AUP 2007. <http://dare.uva.nl/document/93725>

## 7. Appendix

### Format of spreadsheet

#### General (background) information about archive (10 fields)

Name  
Acronym  
URL  
URL alternative  
Type institution  
Institution  
Host  
Description  
Source description  
Creation (yr)

#### Specific information about archive (6 fields)

Repository Type  
Content  
Subjects  
Software  
Language  
Size (items)

#### Content information (12 fields)

Presence GL  
ETD  
Reports  
C-Paper  
Proceedings  
Working papers  
Courseware  
Other  
Datasets  
Multimedia

Total nb GL

% GL

**Qualitative data (7 fields)**

Policies

Specific metadata GL

Quality control

Evaluation

Validation

Limited access fulltext

Other

**Comments (2 fields)**

Comments

Date