Open Is Not Enough

Grey Literature in Institutional Repositories

Abstract

The paper contributes to the discussion on the place of grey literature in institutional repositories and, vice versa, on the relevance of open archives for grey literature. Even in an open environment, grey literature needs specific attention and curation. Institutional repositories don't automatically provide a solution to all problems of grey literature. Our paper shows some scenarios of what could or should be done. The focus is on academic libraries.

The paper is based on a review of international studies on grey literature in open archives. Empirical evidence is drawn from an audit of the French repository IRIS from the University of Lille 1 and from ongoing work on the development of this site.

The study includes a strategic analysis in a SWOT format with four scenarios. Based on this analysis, the paper provides a set of minimum requirements for grey items in institutional repositories concerning metadata, selection procedure, quality, collection management and deposit policy.

The communication is meant to be helpful for the further development of institutional repositories and for special acquisition and deposit policies of academic libraries.

Note on the authors

Joachim Schöpfel is head of the department of information and library sciences at the Charles de Gaulle University of Lille 3 and researcher at the GERiiCO laboratory. He is interested in scientific information, academic publishing, open repositories, grey literature and usage statistics. He is member of GreyNet and EuroCRIS.
Université Charles-de-Gaulle Lille 3, UFR IDIST, BP 60149, 59653 Villeneuve d'Ascq Cedex, France.
joachim.schopfel@univ-lille3.fr

Isabelle Le Bescond is librarian at the Lille 1 University Central Library, Science and Technology, since 2005. She is currently responsible for developing the digital library IRIS. She has served in different University Libraries (Paris, Strasbourg) since 1994. She studied German language and literature.
University Lille 1, Academic Library, Common Service Documentation, Cité scientifique, Avenue Henri Poincaré, 59653 Villeneuve d'Ascq Cedex, France.
isabelle.le-bescond@univ-lille1.fr

Hélène Prost is responsible for studies at the Institute of Scientific and Technical Information (INIST-CNRS). The different studies concern the evaluation of collections, document delivery, usage analysis, grey literature and open access to information. Expertise in statistical tools and knowledge in library information science allowed her to participate in various research projects and writing of several publications.
INIST-CNRS, 2 Allée du Parc de Brabois, CS 10310, 54519 Vandœuvre-lès-Nancy, France.
helene.prost@inist.fr

1. Introduction

Along with other documents and items, grey literature contributes to the success of institutional repositories. Its non-commercial and alternative nature puts grey literature in close proximity to the community-driven culture of open archives.

But does this mean that "grey literature is at home in open archives" (Luzi, 2010) and that it should be re-defined through this new vector of scientific communication?

After years of debate on open access and grey literature, the international conference GL12 at Prague offered two different perspectives. Marzi et al. (2010) stated that "open access is the key to knowledge" and that "web-base sharing facilities and distributed access to openly available information" are key features of grey literature. For Marzi and her colleagues, institutional repositories became the future of grey literature, and grey literature hardly exists without or beyond open access.

On the other hand, our own communication defined additional attributes for grey literature that are not necessarily linked to open access, such as intellectual property, quality and interest for collections. Institutional repositories are an interesting and important vector for dissemination of grey literature but they are not enough. Based on literature review and survey data, we made a proposal for a new definition of grey literature ("Prague definition") with four new essential attributes: "Grey literature stands for manifold document types produced on all levels of government, academics, business and industry in print and electronic formats that are protected by intellectual property rights, of sufficient quality to be collected and preserved by library holdings or institutional repositories, but not controlled by commercial publishers i.e., where publishing is not the primary activity of the producing body" (Schöpfel, 2010).

Concerning open archives, we added that "institutional repositories have started to take over some of the traditional roles of library holdings. In terms of function, they bear some equivalency with grey literature itself, as their main role consists in dissemination and, to a lesser extent, preservation" (ibid). Institutional repositories are important for grey literature but they are not the only option, and they have to satisfy some minimum requirements in order to offer an adequate home for grey literature.

Institutional repositories and grey literature can become a fertile and profitable encounter for scientific communities. But open is not enough. Here are the reasons.

2. Background: A review of grey literature and institutional repositories

Institutional repositories (IR) became a significant channel of digital scientific communication.[1] Part of the open access movement and alongside with subject-based repositories, research repositories or national repository systems (Armbruster & Romary, 2009), they focus on "serving the interests of faculty – researchers and teachers - by collecting their intellectual outputs for long-term access, preservation and management" (Carr et al., 2008).

They can be seen as "tools (…) for collecting, storing and disseminating scholarly outputs within and without the institution" (Jain, 2011), as "a set of services (…) for the management and dissemination of digital materials created by the institution and its community members" (Lynch, 2003) or as an "organisational commitment to the stewardship of these digital materials" (ibid.).

One of their main characteristics is their great diversity. There is not *one* model but multiple possibilities, not *one* path but a multiplicity of options. Yet it is crucial for their success that the institution clearly defines the objective of its repository, in line with its own strategy and environment. "Each of the reasons for setting up a repository carries implications for the content, design and funding of a repository, and the institution needs to be clear about the implications of different roles for a repository, while being prepared to change or add roles as the scholarly communication environment develops" (Friend, 2011).

Institutional repositories have different policies, procedures, functionalities, services and metadata, they have different business models and funding strategies (Swan & Awre, 2006), and their content may include more than current output from faculty. Smith (2008) details a "wide variety of materials in digital form, such as research journal articles, preprints and postprints, digital versions of theses and dissertations, and administrative documents, course notes, or learning objects." Other repositories include datasets, multimedia or cultural and scientific heritage.

Of course, grey literature as unpublished, special or not-for-profit documents is part of the repositories' content. But what is its place in institutional repositories, and what is the relevance of institutional repositories for grey literature?

2.1. The place of grey literature in institutional repositories

Some empirical studies contribute to a realistic vision on grey literature in institutional repositories. Luzi et al. (2008) estimate the part of grey materials eligible for the institutional repository of the Italian National Research Council at about 1/3 of all items, even if not all of these documents are freely available.

[1] See the quantitative study from Mukherjee & Nazim (2011).

In our survey on French repositories, grey literature represents 18% of all documents (Schöpfel & Prost, 2010). Another survey on Spanish repositories reveals that at least 23% of the deposited items with full-text are grey (Melero et al., 2009). Both studies confirm, too, that the number of grey documents in repositories is rapidly growing.

Vernooy-Gerritsen et al. (2009) report results from the EU-sponsored DRIVER project on institutional research repositories. They separate full-text records (33%) from metadata only records and records of non-textual and other materials; 62% of the full-text records are grey literature. This percentage corresponds to 20% of the whole content.

Most of all these grey items are theses, dissertations, proceedings, unpublished papers (working papers) or reports. Up to now, course material is less important.

The part of 20-30% of repository content is somewhat higher than the average percentage of grey literature in citation analyses (see Schöpfel & Farace, 2010).

So far, there is but little evidence on usage of grey items in institutional repositories. Yet, recent studies on access statistics suggest that downloads per item are often higher for unpublished theses or reports than for published articles (Schöpfel et al. 2009, see also Kroth et al. 2010).[2] One reason may be that these items can't be viewed elsewhere.

2.2. The relevance of institutional repositories for grey literature

To which extent are institutional repositories the place for grey literature? According to the information of the OpenDOAR directory of open archives, 82% of all institutional repositories contain grey literature.

Type of documents	Nb IR with these items	% of all IR (n=1,978)
Theses, dissertations	958	48%
Unpublished	616	31%
Proceedings	572	29%
Learning objects	245	12%
Special items	235	12%
total	1,628	82%

Table 1: Grey items in institutional repositories (source: OpenDOAR, June 2011)

The OpenDOAR figures are comparable to results from France and Spain. In France, 94% institutional repositories hold grey documents while their part is significantly lower in subject-based repositories (37%) or national or research repositories (23%) (Schöpfel & Prost, 2010). In

[2] For instance, the 2010 annual report of the French Research Institute for the Exploitation of the Sea shows that the average usage for theses in their IR is 4x higher than for published articles, see http://wwz.ifremer.fr/institut/L-institut/Documents-de-reference/Rapports-Annuels

Spain, more than 80% repositories contain theses, and at least 60% have unpublished working papers and/or proceedings (Melero et al., 2009).

For some of this material, especially for specific types of unpublished items like slides, posters or other, supplementary material, it is surely true that "this is academic output that would not likely be otherwise captured and made freely available were it not for publication in an IR" (Kroth et al., 2010).

Some papers praise the impact of institutional repositories for grey literature. On the word of Luzi (2010), they provide "a natural home for GL" because they amplify its dissemination. Open access makes grey literature "less grey and more white" (Gelfand, 2004); the "distinction between GL and conventional literature is becoming increasingly blurred" (Luzi, 2010; see also Swan 2008 and 2011).

Yet, this "blurring" only applies to potential usage, not to value or quality. Banks (2005) believes that even if the hierarchy between grey and white may shift into a continuum of scholarship, this hierarchy will not completely disappear insofar institution and faculty generally prefer published and peer reviewed documents. A recent study on content recruitment and usage in an institutional repository confirms this belief (Connell, 2011).

2.3. Grey issues

Studies on grey literature in institutional repositories recurrently point out six critical aspects for the success and development of such initiatives.

Community: Describing a conference proceedings repository at Cornell, Rupp & LaFleur (2004) plead for "a specific workflow (…) for the identification and gathering of proceedings" that includes public relations, "one-to-one marketing" and communication with faculty to create awareness and get the documents from the author's desktop into the repository. Without community, no repository.[3]

Quality control: A repository that is "'all things to all people' lacks focus" (Westell, 2006). Specific action from the very beginning of the workflow is required to guarantee a minimum quality of content, data and services. Control procedures and workflow technology should ensure quality of item selection and overall project management (Luzi et al., 2004).

Metadata: Grey literature in institutional repositories has need of specific metadata for identification and bibliographic description. For instance, Ruggieri et al. (2009) propose a table with mandatory and optional metadata fields, including a note field, for conference papers, oral presentations, reports and in-house publications. Jeffery (2007) adds that "the syntax must be formal and precise; the semantics must be present, formal and precise (…); the relationships form a fully-connected graph; (…); the relationships require an annotation richer than the triples of RDF

[3] See also the disillusioning survey from Seaman (2011).

(…).” Yet, unfortunately the reality is that “current metadata elements (of electronic theses and dissertations in IR) have a significant level of inconsistency and variation” (Park & Richard, 2011), and often “individual institutions (decide) locally how metadata elements should be defined (ibid).

Interoperability: Institutional repositories are hardly ever stand-alone systems. They should be interoperable or at least three reasons: maybe because their institution is part of a network (Dijk, 2007), maybe because they are connected and exchange data and items, maybe simply because the OAI initiative stipulates interoperability. Pejsova (2011) describes a national system for grey literature that is interoperable with local repositories for documents, metadata and workflow.

Integration: Some authors insist on the integration of institutional repositories and grey literature into current research information system (CRIS) infrastructure. “An institutional repository, being a central point within the organisation for literature and data, is a component of the integration of processes, which promises benefits both to the organisation itself and to the researchers within it” (Lambert et al., 2005).

New item formats: Jeffery (2007) calls for a linkage between CRIS and e-repositories for grey literature on the institutional level, and he suggests that they should be associated to repositories for research datasets and software, via the CRIS. More recently, Doorenbosch & Sierman (2011) focus on the changing nature of scholarly publications, e.g. enhanced publications with both documents and datasets, outline the challenge of these new items for long term preservation in institutional repositories, and suggest the creation of “collaborative virtual research environments are considered to be the new workspaces for researchers”.

3. Case study: The IRIS audit – grey literature at home at Lille

The IRIS repository, hosted by the Lille 1 university, successor to Grisemine, the first French open archive for grey literature. Its development and usage have been presented at the GL5 and GL12 conferences (Claerebout, 2003; Prost et al., 2010). The following case study provides a short overview on the Grisemine/IRIS history and illustrates some conditions that are favourable or not for the deposit and dissemination of grey literature in institutional repositories.

3.1. General remarks

When Grisemine was launched in 2001, it was one of the first open archives in France, a pioneer especially in the academic sector. Its notoriety and popularity among academic librarians was immediate and without doubt superior to its real impact on scientific communication.

Since 2001, Grisemine underwent deep changes. This “Grisemine/IRIS decade” demonstrates the coming out of the hybrid digital library with service marketing rather than collection building. Nearly all has changed –

the name, software, architecture and workflow, content, strategy, policy and institutional positioning.

The story of Grisemine/IRIS is not over. In fact, it just began, again. But which may seem, ex post, logical and necessary often was trial and error, searching for opportunities, benchmarking, exploration and adaptation to a moving context.

3.2. Rise and decline of Grisemine (2001-2005)

Grisemine's purpose was to collect, preserve and disseminate French[4] grey literature, such as theses and dissertations, communications, notes, working papers, preprints, exam topics or educational programs. Grisemine was developed with the CinDoc electronic content management software (Cincom). Its workflow was compliant with the Dublin Core metadata standard and the MARC format.

Even as a prototype, the Grisemine project was technically viable, except for the technical maintenance and development of the CinDoc software. But it had no real institutional recognition, was a "librarians' toy" rather than a labelled, validated and accepted repository for the scientific community. Yet, its content (1,300 documents in late 2005) was widely consulted, in particular from French-speaking countries.

It became obvious, too, that the initial goal – a deposit for all French grey literature – was too ambitious and disproportionate to the allocated resources.

3.3. From Grisemine to IRIS (2006-2010)

In 2006, the French government published a decree on the processing, preservation and dissemination of electronic PhD theses and launched a national network for ETDs called STAR. Grisemine was not able to support the new workflow. For this and other reasons mentioned above (maintenance), the Lille library team considered Grisemine as a technical and documentary dead-end. The next four years were a period of transition.

The most important decision was to migrate from CinDoc to DSpace, and then make the system dialogue with STAR. The migration was operational in 2007. With the migration, Grisemine became IRIS.

Why DSpace? At the time the Lille team took the decision to migrate (2004-2006), DSpace was the most common software for open archives, and it was easy to install. Yet, DSpace is designed for self-deposits, not for an encyclopaedic-like collection (scientific heritage) or an institutional and/or national workflow (theses). Without a dedicated information technology (IT) staff, the Lille library decided to maintain DSpace at best until the new ORI-OAI software became available[5]. "At best" meant

[4] French means: edited in France and/or in French language.

[5] A document management system compliant with OAI-PMH, designed for the publishing, sharing and dissemination of academic digital resources and supported by the French Ministry of Higher Education http://wiki.ori-oai.org

keeping the archive alive, continued uploads but no development. For instance, an early project to separate PhD theses and scientific heritage was put on ice.

The deposit of e-theses became mandatory on the Lille 1 campus in 2008, because of STAR. IRIS was able to provide an operating OAI platform for their dissemination but didn't offer a solution for their management or preservation. The open dissemination of Lille ETDs became the main function of the IRIS repository. In December 2010, IRIS had 625 theses and 711 other documents. Their long-term preservation is supported by the academic data centre CINES at Montpellier[6].

With the move from Grisemine to IRIS, the site abandoned its initial strategy as an open repository for French grey literature. The self-deposit of grey items ceased completely. Instead, the library team made another use of the IRIS platform and developed, together with a historical research centre and the academic digitization centre at Lille, a digital library with a collection of copyright cleared documents (articles, papers, books) on the history of sciences. Alongside with the PhD theses, this heritage collection was made freely available on the IRIS platform and is very appreciated by the scientists.

When the university decided the mandatory deposit of e-theses in 2008, it also acknowledged IRIS as the official Lille 1 institutional repository. Yet, this decision was not accompanied or followed by a mandatory policy for the whole scientific production of the faculty. Except some professorial habilitation theses and learning objects, IRIS never received any self-deposits from Lille researchers.

3.4. Rebirth (2010-2011)

At the end of the first decade, the strategic positioning of IRIS was atypical and confusing. The university administration considered IRIS as the official institutional repository. Yet, there was no promotion, communication, incentives or mandate, and the only open archive with a significant number of self-deposits from Lille 1 faculty was (and always is) the French national research repository HAL with 16,143 items.[7]

The library team regarded IRIS as a digital library, more like GALLICA or PERSEE than ArXiv or HAL, yet used the IRIS server for the dissemination of PhD theses, a service usually considered to be a key element of academic institutional repositories, and made some tests with other scientific output from Lille faculty, especially in the context of an emerging learning centre project.

In 2010, with the installation of the ORI-OAI system the Lille 1 repository took a new start. Why ORI-OAI? At least for four reasons: compliance with French metadata standards for theses (TEF) and learning objects (SupLOMFR), interoperability with the nationwide infrastructure for ETDs (STAR) and the national research repository HAL, a French

[6] Preservation of ETDs via STAR, preservation of other deposits on a contractual basis.
[7] But only 13% of these items have full-text, the rest are metadata only records (12 October 2011).

community of software developers and end-users, quality of development and product.

Today Lille 1 hosts a composite repository with two systems accessible through two different interfaces:

- ORI-thèses with theses, habilitations and learning objects.[8]
- IRIS with the collection of history of sciences.[9]

In fact, IRIS became a digital library without input from current scientific production.

A third platform for the self-deposit of scientific production (pre- and post-prints, communications, reports…) is under construction, on the model of the Toulouse OATAO[10] repository or the Luttich ORBi[11] site, and will be launched in 2012 probably with a new name.

3.5. Concluding remarks

As we said above, the story of Grisemine/IRIS is not over and it may be premature to debrief. Yet, we tried to highlight some main characteristics of this project and then to identify the factors in favour of grey literature and success.

The development of the Lille 1 repository was non-linear, dependent on the evolving local and national context, on technology (software) and standards. The library team's quest for legitimacy was complicated by the pluridisciplinarity of their academic community and by the fact that in France, the open archives for scientific information were initially hosted and managed by the public research organisations (CNRS, INRA, IFREMER…).

On the other hand, the national infrastructure for electronic theses (STAR system with TEF metadata standard) and the library's experience with preservation and dissemination of cultural and scientific heritage items – a traditional library function - facilitated the legitimacy and positioning of the project.

So which were the critical key elements for success or failure? Briefly:[12]

- Institutional support and recognition of the project team and the repository.
- Institutional strategy and policy in the domain of open archives and deposit mandate.
- Human resources with sufficient IT and LIS capacities.
- Metadata standard(s) for a careful and precise bibliographic description of the deposited content.
- Software fitting with local needs and IT environment as well as with national infrastructure and standards.

[8] http://ori.univ-lille1.fr
[9] https://iris.univ-lille1.fr
[10] http://oatao.univ-toulouse.fr/
[11] http://orbi.ulg.ac.be/
[12] See also Westell (2006).

- A solution for perennial preservation of deposits (at least for the theses).
- Added value services for legal aspects and usage statistics.
- Knowledge of the scientific community's information needs and behaviours, and integration into the larger academy.

The Grisemine/IRIS case shows also a close link between grey literature typology, IT solutions (software) and workflow features. The repository must cope with specific conditions, such as (for the Lille repository) the national STAR system for theses or the digital university environment (UNT) for the learning objects. The need to align deposit with existing workflows was highlighted by Westell (2006) and Troll Covey (2011). This, together with the different software solutions, argues for a differential approach to grey literature in institutional repositories. Some grey documents may be at home in some open archives, while others in different ones.

4. SWOT analysis: Grey literature in institutional repositories

Based on the review of literature and standards and including the IRIS experience, our evaluative synthesis will take the form of a strategic SWOT diagnostic, keeping apart internal and external factors that are favourable or unfavourable for grey literature in institutional repositories. However, our analysis does not take into account more general aspects that are not directly related to grey literature (for instance, such as Pinto & Fernandes, 2011).

4.1. Strengths

The internal factors in favour of grey literature in institutional repository models are:

1. Grey literature amplifies the content of institutional repositories.
2. Free availability, dissemination, visibility and referencing act as incentives for grey deposits.
3. What's more, relatively high usage of unpublished items may also act as an incentive for grey deposits.
4. Institutional repositories guarantee more security and long-term accessibility of unpublished material than a personal web site.
5. Compared to published articles, there are fewer problems with copyright for grey literature.

4.2. Weaknesses

The internal factors unfavourable for grey literature in institutional repositories are:

1. The bibliographic control of grey literature, especially of conferences and reports, remains often mediocre or poor because

of flawed or incomplete metadata format (non qualified Dublin Core).

2. Most often, the hosting institution doesn't provide any solution for the digital curation of metadata.
3. Deposit is time consuming.
4. Deposit of grey literature needs, more than published documents, incentives and support from institution. This support may be missing.
5. Without institutional support or incentives, self-deposits will not have the same quality as a library collection.

4.3. Opportunities

The external factors in favour of grey literature in institutional repository models are:

1. Universities need a solution for the processing, disseminating and archiving of electronic theses and dissertations (ETD). Institutional repositories offer an interesting solution and may at least be an element in the global academic information system for ETD.
2. Institutions want control on research output and content, and this includes unpublished documents.
3. Institutions want to improve presence and impact on the web. Grey literature in repositories adds to both, due to broader dissemination and increased use of grey items, increasing prestige and visibility for the institution.
4. The open access initiative is not limited to published documents.
5. The evolution from "collection development" to "content recruitment" in academic libraries may act in favour of deposit of grey literature in institutional repositories.

4.4. Threats

The external factors unfavourable for grey literature in institutional repositories are:

1. Funding and evaluation agencies put priority on published documents (articles, books) and at least partially neglect grey items. Grey literature is not indexed in the scientometric databases Web of Science and SCOPUS.
2. If institutions introduce self-archiving mandates in order to generate content, researchers may react negatively to any suggestion of compulsion. Most faculties do not respond to the invitation to "add stuff to the IR" (Jain 2011). Another side-effect is the creation of metadata only records, without full-text. This

should be limited to published documents with copyright problems but it isn't.[13]

3. Alternative models, e.g. generating content through deposit by publishers (PEER project) will not impact grey items.
4. Open access through institutional repositories requires funding from particular institutions to set up and maintain a repository (Friend 2011). Poor knowledge on grey literature will make it more difficult to sustain continuous support and commitment from the management and academic staff.
5. A significant part of the scientific community lacks awareness of open access and grey literature.

5. Findings based on four scenarios

Are institutional repositories the future of grey literature? Maybe. But because of the great variety of institutional repositories, we can distinguish at least four different scenarios.

Jain (2011) makes some recommendations for the development of institutional repositories, in particular, promotion and publicity to the faculty, provision of clear policies on ownership, contents, quality and copyright, and an adequate provision of resources. This is in line with the IRIS audit and applies to all scenarios. Therefore, our description is limited to specific criteria for grey literature within this environment.

The differences are with mandatory deposit, strategic vision, services, selection procedure, quality issues, collection management and metadata. Our description is partly based on studies on objectives and business models of institutional repositories (Friend, 2011; Swan & Awre, 2006). We don't describe real cases but potential homes – a kind of ideal archetypes of institutional repositories. The reality will be more complex and composite.

5.1. Scenario 1 – Publishing grey literature

In the first scenario, the institutional repository serves essentially the initial function of open archives, e.g. communication and publishing of scientific papers. Focus is laid on rapid and direct access to full-text, for the scientific community. For grey literature, the strategy is to become less greyish and more white, through institutional digital publishing outside of usual sales channels.

The strategic objective by the institution may be twofold (cf. Friend, 2011):

- "To increase the impact of particular research or teaching programmes through exposure of publications and other outputs on open access.

[13] For instance, only 45% of the deposited working papers, conferences and ETD in the Belgian ORBi repository provide access to the full-text.

- To reduce the cost and increase the benefits from the dissemination of the institution's research and teaching outputs."

The most appropriate business model for repository provision and preservation will be institutionally-supported, perhaps with a contribution by community (learned societies).

Selection procedures for a minimum content and formal quality level (through validation or "labelling") probably will be more important than mandatory issues. Self-deposit of full-text (preprints, postprints but also conference proceedings, unpublished reports and papers…) and institutional workflows for electronic theses, perhaps also for master and habilitation theses, in-house collections of working papers or reports are essential for content recruitment while mandatory deposit policy or incentives are not.

Also, metadata are critical (only) insofar they facilitate content retrieval and access. This means that they are probably of mediocre quality and not very specific for different types of documents, except for ETD.

The primary function of this repository is communication and access to the full-text, via search engines and/or the repository's search and browse interface. The key elements are a high rate of full-text, worthy scientific content, and unrestricted access, followed by a high and representative number of deposits. Other services may be less crucial but would add value to the site:

- usage statistics services,
- preservation services,
- publishing services.

5.2. Scenario 2 – Special items container

In the second scenario, the institutional repository is a container or storehouse for all kind of material produced by faculty, staff and students. In this container, ETD, reports and conference proceedings stand next to images, learning objects, articles, datasets, presentations, posters etc.

The focus is laid on availability and visibility of all kind of materials, "institutional stuff", rather than on selection of scientific relevant results. Quality control through validation or labelling is not an issue.

The strategic objective may be "to collect together all the publications and other research and teaching outputs as a permanent record of the institution's achievements but without any specific use in mind" (Friend, 2011).

Again, the appropriate business model for repository provision and preservation is institutional support. The institution may also decide to establish a mandatory deposit, and/or incentives for self-deposit.

The underlying idea is to "dig out" hidden material, find a solution for digital dissemination and preservation, together with other published or unpublished documents.

As for quality control or editorship, metadata probably are not an important issue. Most likely, services will be limited to preservation, publishing, resource discovery and perhaps research assessment and monitoring. It is also possible to add social indexing and data mining. There is no clear vision on collection and acquisition. But the most promising perspective may be the linking of the deposits to research data.

5.3. Scenario 3 – Scientific heritage

The third scenario the institutional repository is a showcase for the past and present scientific production, with grey literature alongside with published documents and other material.

Again, the strategic vision will be "to collect together all the publications and other research and teaching outputs as a permanent record of the institution's achievements but without any specific use in mind" (Friend, 2011). The difference with scenario 2 is the heritage character of the collection, the inclusion of older material in the public domain.

But there may (also) be other motivations:

- "To increase the impact of particular research or teaching programmes through exposure of publications and other outputs on open access.
- To make a contribution to the world-wide movement for open access to publicly-funded research" (ibid.).

The definition of an acquisition or content recruitment policy is crucial, together with an institutional strategy for the digitization of older, copyright cleared material (theses, journals, books, papers, images, maps…). This may imply a more thoroughly prepared and pondered indexing and metadata policy. The outcome may be 100% access to full text, as for the IRIS repository.

The appropriate business model is institutional support. But there may be other resources, public funding for scientific heritage or thematic or special collections. For this specific case, it may be possible to experience a subscription-supported model, appropriate for access and authentication, preservation and resource discovery services.

Also, the local presence of a digitisation centre may allow those repositories to populate content more rapidly, especially grey literature, and to attract usage (Westell, 2006).

The underlying idea is digital preservation of heritage collection, together with making these collections available to scientists, students and all interested people. This may be complementary to publishers' backfiles.

This scenario is probably the closest scenario to traditional library collection building, with issues such as quality, indexing, classification etc. Evaluation, scientometrics etc. may be less important, at least not in the heart of the project.

5.4. Scenario 4 – Institutional deposit

The last scenario for grey literature in institutional repositories is mandatory institutional or self-deposit in the way it is promoted by Stevan Harnad: green road (self-deposit) to free online full-text access to peer-reviewed literature, through an explicit and institutional mandatory policy in order to obtain commitment by close to 100% of the authors.

This scenario is meant to demonstrate the value of the institution itself through a kind of quasi-legal deposit showcase, to facilitate control over scientific production and evaluation procedures, and corresponds to one or more institutional strategies, e.g.

- "To report the publications and other research and teaching outputs to funding agencies in support of new grant applications.
- To report the publications and other research and teaching outputs to funding agencies as part of an audit of expenditure.
- To demonstrate to governments or taxpayers the impact of the institution outside its walls (a purpose which will require the compilation of metrics).
- To increase the impact of individual members of the institution's staff through the exposure to potential academic and commercial users of the individual's publications and other outputs on open access" (Friend, 2011).

The business model will surely be institutionally-supported and may include services such as usage statistics, research assessment and monitoring, bridging and mapping, and technology transfer/business advice. Also, a connection to a current research information system (CRIS) should be possible.

The impact on grey literature in this environment is triple:

Peer-reviewed publications will play a major role in this environment, and in comparison, grey literature will be less valued or appreciated. This may have a negative impact on metadata.

The institutional policy of mandatory deposit generate a relatively high rate of metadata only records without access to full-text because of embargo, sensitive content, missing authorization by co-authors etc. Paradoxically, this "collateral damage" also impacts grey literature (see above, footnotes 7 and 12). Only 40% of the HAL grey literature records are with full text.

The number of grey documents will be significant but more or less limited to specific categories evaluated by agencies, such as theses and dissertations, conference proceedings and project reports.

The main interest of these repositories is not collection building but evaluation. Insofar grey literature enters evaluation procedures it will be valued and welcome in this environment.

6. Results and concluding remarks

Our paper started with Luzi's (2010) statement that "grey literature is at home in open archives". This may be right but as we tried to demonstrate, open archives not only offer one but at least four different homes that may be complementary, at least to some extent.

Mapped on two dimensions, policy (evaluation vs. communication) and quality (library vs. container), the four options clearly occupy different positions (figure 1).

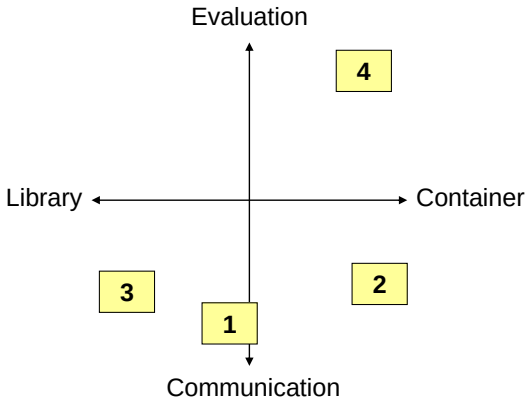

Figure 1: A map of four scenarios for institutional repositories with grey literature

In scenario 1, the political priority is laid on communication of research results, full-text, community, scientific value. Grey literature is part of the content insofar the depositing authors consider it worthy enough for direct communication and preservation. But there is no real control or selection.

In scenario 2, the main objective is the container function, the deposit of all materials produced by faculty, students and staff. Again, the institutional policy is communication-centred but without selection or validation criteria. Grey literature has its home here – in a (too) large sense and together with a lot of other stuff.

Selection or validation criteria are introduced by the 3rd scenario. Here the institution applies a policy of showcase and scientific heritage, most likely accompanied by digitization programs. The place of grey literature depends on the institution's acquisition policy and digitization program.

The 4th scenario reflects the institutional policy in favour of evaluation and ranking. Full-text and communication are secondary goals while metadata and a minimum quality control are necessary. Deposit of grey literature will be welcome insofar it enters evaluation.

Now, which is the most adequate option for grey literature? The response depends on institutional policy, library goals and professional viewpoint. For the scientific community, end-user and consumer of scientific information, perennial open access to validated items in full-text format is priority. This priority implies at least five minimum requirements:

Access to full-text. Open archives with metadata only records are like libraries with empty shelves.

Quality through selection, validation and/or labelling. Even without peer-review or other, web-based reviewing procedures, grey deposits should meet with some basic quality criteria. Incite deposit of all kind of uninteresting stuff is like keeping waste paper on the desktop. Self-deposit is not collection building.

Openness without restriction and/or embargo. Confidential, classified or non-copyright cleared material should not be part of open archives but should be managed via catalogues, databases or other systems.

Metadata quality. Repositories should guarantee a minimum level of metadata quality, e.g. compliance with standards and curation. This requirement is necessary for information retrieval, interoperability and the semantic web.

Long-term conservation. Institutional repositories should offer a solution for the ephemeral nature of grey literature, via a clear statement on and investment in perennial content preservation, if necessary also via outsourcing or "in the clouds".

For the scientific community, the best option for grey literature may be a mix of scenarios 1 "publishing grey literature" and 3 "scientific heritage". Other elements will add value (standard format and metadata, usage statistics, discovery functions, scientometrics) or increase sustainability (institutional support, integration in research community, promotion and communication, interoperability). But they are not specific to grey literature.

We didn't speak about format and legal matters; yet, they may be critical matters for the future of repositories. With the words of Swan (2011), "we (can't) relax (and) watch repositories fill with articles and datasets". Or as Anderson (2011) put it, "accessibility is not access."

The IRIS case should raise awareness that the same solution may not be appropriate to all kind of grey literature and disciplines and that the system should be evolutionary and flexible enough to easily adapt to and keep up with new conditions and opportunities.

A last and rather paradox remark: the success of institutional repositories may become a problem for grey literature, especially when the institution implements a mandatory deposit policy that gives priority to evaluation and control and not to publishing and communication. Anna Clements, a data manager from St Andrews University, described the problem some time ago on the JISC-Repositories listserv: libraries create institutional repositories with full-text or full objects as the main content, and they are then asked by the institution to look at hosting citations without full-text as well.

A library with empty book shelves may be interesting to research managers but not for scientists. In this case, grey literature would definitively not be at home in institutional repositories. Open is not enough.

6. References

K. Anderson (2011). `Does Access Create New Types of Scarcity?'. *The Scholarly Kitchen* Aug 31.

M. A. Banks (2005). `Towards a Continuum of Scholarship: The Eventual Collapse of the Distinction Between Grey and non-Grey Literature?'. In *Seventh International Conference on Grey Literature: Open Access to Grey Resources, Nancy, 5-6 December 2005*.

L. Carr, et al. (2008). `Institutional Repository Checklist for Serving Institutional Management'. In *Third International Conference on Open Repositories 2008, 1-4 April 2008, Southampton, United Kingdom*.

M.-F. Claerebout (2003). `Grisemine, a digital library of grey university literature'. In *Fifth International Conference on Grey Literature: Grey Matters in the World of Networked Information, 4-5 December 2003*.

T. H. Connell (2011). `The Use of Institutional Repositories: The Ohio State University Experience'. *College & Research Libraries* **72**(3):253-275.

E. Dijk (2007). `Accessing grey literature in an integrated environment of scientific research information'. In *Ninth International Conference on Grey Literature: Grey Foundations in Information Landscape, Antwerp, 10-11 December 2007*.

P. Doorenbosch & B. Sierman (2011). `Institutional Repositories, Long Term Preservation and the changing nature of Scholarly Publications'. *Journal of Digital Information* **12**(2).

F. Friend (2011). `Open Access Business Models for Research Funders and Universities'. Tech. rep., Knowledge Exchange, Copenhagen.

J. Gelfand (2004). `"Knock, Knock:" Are Institutional Repositories a Home for Grey Literature?'. In *Sixth International Conference on Grey Literature: Work on Grey in Progress, New York, 6-7 December 2004*.

P. Jain (2011). `New trends and future applications/directions of institutional repositories in academic institutions'. *Library Review* **60**(2):125-141.

K. G. Jeffery (2007). `Greyscape'. In *Ninth International Conference on Grey Literature: Grey Foundations in Information Landscape, Antwerp, 10-11 December 2007*.

P. J. Kroth, et al. (2010). `Institutional Repository Access Patterns of Nontraditionally Published Academic Content: What Types of Content Are Accessed the Most?'. *Journal of Electronic Resources in Medical Libraries* **7**(3):189-195.

S. Lambert, et al. (2005). `Grey literature, institutional repositories and the organisational context'. In *Seventh International Conference on Grey Literature: Open Access to Grey Resources, Nancy, 5-6 December 2005*.

D. Luzi, et al. (2004). `The integration of GL documents with a research information system on occupation safety and health'. In *Sixth International Conference on Grey Literature: Work on Grey in Progress, New York, 6-7 December 2004*.

D. Luzi, et al. (2008). `Towards an Institutional Repository of the Italian National Research Council: A survey on Open Access experiences'. In *Tenth International Conference on Grey Literature: Designing the Grey Grid for Information Society, 8-9 December 2008*.

D. Luzi (2010). `Grey Documents in Open Archives'. *The Grey Journal* **6**(3):137-144.

C. A. Lynch (2003). `Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age'. Tech. Rep. 226, ARL Association of Research Libraries.

C. Marzi, et al. (2010). `A terminology-based re-definition of grey literature'. In *Twelfth International Conference on Grey Literature: Transparency in Grey Literature. Grey Tech Approaches to High Tech Issues. Prague, 6-7 December 2010*.

R. Melero, et al. (2009). `The situation of open access Institutional repositories in Spain: 2009 report'. *Information Research* **14**(4).

B. Mukherjee & M. Nazim (2011). `Open access institutional archives: a quantitative study (2006-2010)'. *DESIDOC Journal of Library & Information Technology* **31**(4):317-324.

E. G. Park & M. Richard (2011). `Metadata assessment in e-theses and dissertations of Canadian institutional repositories'. *The Electronic Library* **29**(3):394-407.

P. Pejsova (2011). `Czech National Repository of Grey Literature'. *The Grey Journal* **6**(3):117-122.

M. J. Pinto & S. Fernandes (2011). `Gaining a sustainable IR: thinking SWOT'. In *QQML2011: 3rd International Conference on Qualitative and Quantitative Methods in Libraries (DLMC 2011: Digital Library Multi-Conference), Athens, Greece, 24-27 May, 2011*.

H. Prost, et al. (2010). `Usage assessment of an institutional repository: a case study'. *Twelfth International Conference on Grey Literature: Transparency in Grey Literature. Grey Tech Approaches to High Tech Issues. Prague, 6-7 December 2010*.

R. Ruggieri, et al. (2009). `From CNR Annual report to an Institutional repository: Results from a survey'. In *Eleventh International Conference on Grey Literature: The Grey Mosaic, Piecing It All Together, Washington DC, 14-15 December 2009*.

N. Rupp & L. J. LaFleur (2004). `Making Grey Literature Available through Institutional Repositories'. In *Sixth International Conference on Grey Literature: Work on Grey in Progress, New York, 6-7 December 2004*.

J. Schöpfel, et al. (2009). `Usage of grey literature in open archives'. In *Eleventh International Conference on Grey Literature: The Grey Mosaic: Piecing It All Together. Washington D.C., 14-15 December 2009*.

J. Schöpfel (2010). `Towards a Prague Definition of Grey Literature'. In *Twelfth International Conference on Grey Literature: Transparency in Grey Literature. Grey Tech Approaches to High Tech Issues. Prague, 6-7 December 2010*.

J. Schöpfel & D. J. Farace (2010). `Grey Literature'. In M. J. Bates & M. N. Maack (eds.), *Encyclopedia of Library and Information Sciences, Third Edition*, pp. 2029-2039. CRC Press, London.

J. Schöpfel & H. Prost (2010). `Développement et Usage des Archives Ouvertes en France. Rapport. 1e partie: Développement'. Tech. rep., Université Charles-de-Gaulle Lille 3.

J. Schöpfel & H. Prost (2011). `IRIS. Etat des lieux et perspectives. Rapport d'audit'. Tech. rep., University of Lille, Villeneuve d'Ascq.

K. Smith (2008). `Institutional Repositories and E-Journal Archiving: What Are We Learning?'. *Journal of Electronic Publishing* **11**(1).

A. Swan & C. Awre (2006). `Linking UK Repositories: Technical & Organisational Models to Support User-Oriented Services Across Institutional & Other Digital Repositories'. Tech. rep., JISC, London.

A. Swan (2008). `Study on the availability of UK academic "grey literature" to UK SMEs: Report to the JISC Scholarly Communications Group'. Tech. rep., JISC, London.

A. Swan (2011). `Institutional repositories - now and next'. In P. Dale, J. Beard, & M. Holland (eds.), *University Libraries and Digital Learning Environments*, pp. 119-134. Ashgate Publishing, Farnham.

D. Troll Covey (2011). `Recruiting Content for the Institutional Repository: The Barriers Exceed the Benefits'. *Journal of Digital Information* **12**(3).

M. Vernooy-Gerritsen, et al. (2009). `Three Perspectives on the Evolving Infrastructure of Institutional Research Repositories in Europe'. *Ariadne* (59).

M. Westell (2006). `Institutional repositories: proposed indicators of success'. *Library Hi Tech* **24**(2):211-226.

Websites visited between September-November 2011.