Restricted vs open access for electronic theses and dissertations - a challenge for public science

Joachim Schöpfel, Hélène Prost, Marjorie Piotrowski, E. Hilf, Thomas Severiens, Paul Grabbe

Restricted vs. Open Access for Electronic Theses and Dissertations – A Challenge for Public Science

Joachim Schöpfel[1], Hélène Prost[1], Marjorie Piotrowski[1], Eberhard R. Hilf[2], Thomas Severiens[2], Paul Grabbe[2]

1) GERiiCO, University of Lille 3, France
2) Institute for Science Networking Oldenburg, Germany

Introduction

Academic electronic theses and dissertations (ETDs) should be openly accessible to serve best the scientific progress, and inform the scientific community as a whole on the results. This is technically supported at universities by their institutional repositories, by global open archives, and by some publishers. Over the years the percentage of ETDs which are open access (OA) is rising worldwide and in all scientific disciplines.

Open, stated Peter Murray-Rust in his keynote at the ETD2014 conference at Leicester, UK, should be the default, not the exception. However, the reality is different. Even in institutional repositories (IRs) created to provide access to the scientific output of an academic institution and as a central sector of the so-called green road to open access, more or less important parts of the scientific production are missing. This is because of lack of awareness, embargo, deposit of metadata without full text, confidential content etc. ETDs, in particular, are disseminated with different status types – some are freely available, others are under embargo, confidential, restricted to campus access (encrypted or not) or not available at all. Contrary to the OA philosophy, many IRs are not fully OA (Prost & Schöpfel 2014). But while other papers may be available through alternative channels (journals, monographs etc.), ETDs most often are not.

Our paper summarizes empirical evidence from different studies and describes some main results from a small French-German survey. It proposes a conceptual approach designed not only for a better understanding of access restrictions but also for action and policy-making to increase the part of ETDs to be open access. At the end, the paper will present the outline of a European infrastructure project called "Electronic Theses and Dissertations for Open Access" (ETD4OA) to be submitted in the framework of the Horizon 2020 program.

Empirical evidence

A first preliminary study based on a review of recent papers and empirical evidence of a small set of European and American repositories, networks and institutions confirmed that ETDs are disseminated with different status types – some are freely available, others are under embargo, confidential, restricted to on-campus access (encrypted or not) or not available at all (Schöpfel & Prost 2013). The number of hidden theses may amount for some repositories to 40% or more. Figure 1 provides an overview[1].

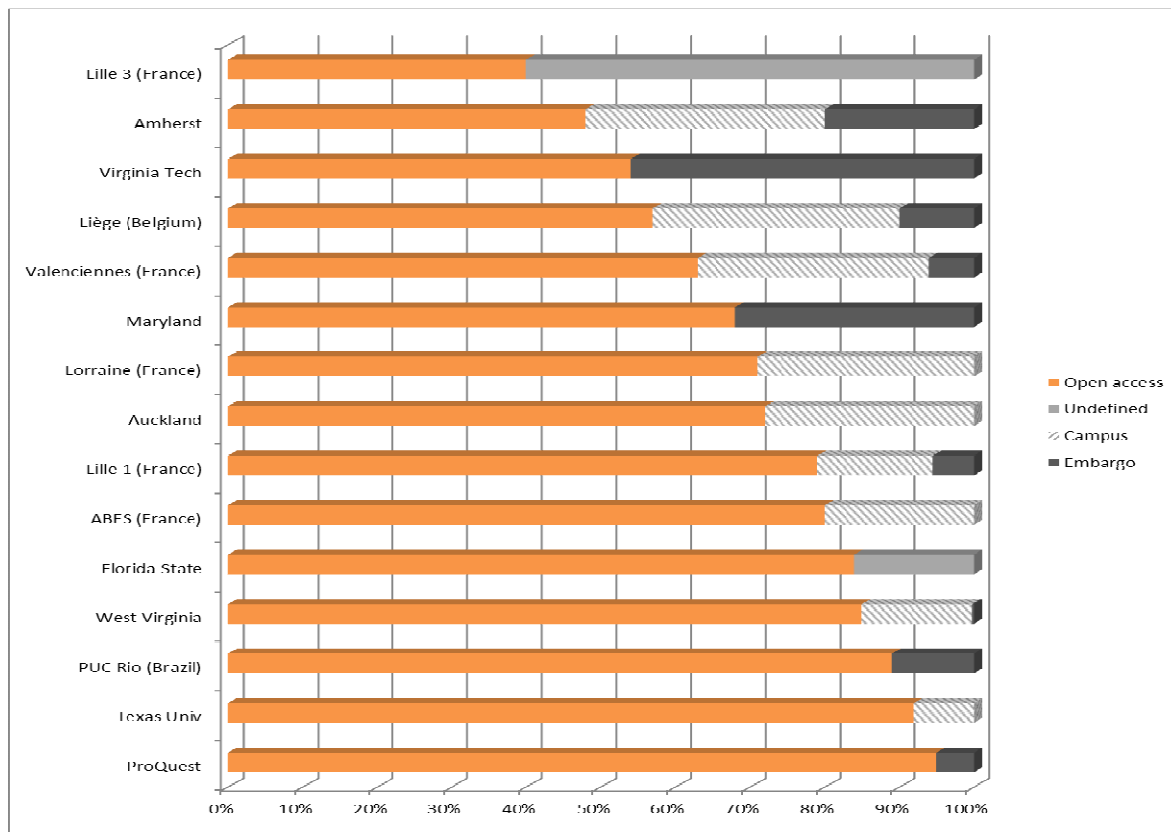[1] For the sources, see Schöpfel & Prost (2013).

Figure 1: Open and restricted access of ETDs in different OA servers

Figure 1 shows a situation that academic librarians in charge of ETDs know well more or less, the large spread of access-type per repository. Yet, surprisingly only very few studies contain empirical data on access restrictions and allow for comparison and analysis. Also, in some cases, figures are only available for limited periods and relatively small numbers of items. Three points should be emphasized:

1. In the surveyed repositories, the part of openly available ETDs varies largely, from less than 40% to more than 95% (ProQuest OA database).
2. We identified three main types of restricted access: availability limited to authorized users (often, on-campus access or intranet), embargoed ETDs (delayed access) and confidential content or topic. Sometimes, the type of restricted access is not specified.
3. The nature of these types of access restrictions show considerable differences. In some repositories, all non-OA ETDs are under embargo while elsewhere their access is limited to the campus. Also, embargo and on-campus access can take different forms (embargoes for six months to two years or more; access in library, on the campus, via intranet etc.). The part of confidential ETDs appears of be rather low.

Survey data

There is an astounding decrease since 2009 in Germany in the OA theses as archived by the National German Library collected from the official OA institutional repositories of the universities. In spite of the growing success of the OA initiative, a significant part of ETD remains unavailable on the web or circulates with restrictions.

Which are the reasons and determining factors? We started in 2013 a French-German project called EDAR *Electronic Theses: Access and Restrictions*[2] to collect data on the present status of the workflow of ETD and on the local rules for handling restricted access, via a survey of academic libraries and graduate schools. The survey has been published in detail (Schöpfel et al. 2014). Here are some major results.

France

•   The part of OA theses compared to all theses increased from 12% of the PhD theses in 2009 to 24% of the PhD theses in 2012.

•   There is an increase of embargoed PhD theses, from 1% of the ETD in 2009 to 7% in 2012.

•   On-campus access restrictions of ETD increased from 5% in 2009 to 28% in 2012.

•   The part of confidential theses remains small and stable (2% of ETD).

Germany

•   The number of OA theses increased slightly from 41% of all theses in 2009 to 47% in 2012.

•   Furthermore, in Germany, very few libraries reported access restrictions. There are however some individual cases where the author asked for restrictions because of confidential material etc. These cases are in total less than 1% of the total.

Yet, attention should be paid to two limitations (see discussion in Schöpfel et al. 2015). First, because of the small sample size, the survey did produce only preliminary data, interesting for empirical evidence but not necessarily representative. Second, the results reflect also different ways of handling ETDs – in France, all ETDs are deposited in IRs whenever possible, if necessary and requested with embargo or limited availability; in Germany, only OA ETDs are handed in for deposit so that ETDs with restricted access remain virtually invisible, "out of scope". The EDAR survey confirms former empirical evidence on restricted access to ETDs, especially because of embargo periods and on-campus availability. Obviously, there is no need for new facilities and infrastructures on local, national or European levels. Today, the problem of OA ETDs lies upstream in local contexts that facilitate decisions in favor of OA as the regular case, instead of embargoes or restricted access (on-campus access, intranet). To put it in a simple way, pipes exist, but it lacks the fuel and the pressure for ETDs and OA.

Conceptual approach

In a third approach we tried to conceptualize the context of access restrictions – who are the stakeholders, which are the reasons for decision-taking, how can the result be described in terms of different degrees of openness? Here are some results; for more details, see Schöpfel & Prost (2014).

Who are the stakeholders, who takes part in the decision on open or restricted access? At first glance, the situation appears to be rather simple. PhD theses being intellectual work, the student is the only person holding the right to decide about dissemination. Of course, this view is by far much too simplistic. Different actors – people and institutions – can be distinguished who have a more or less impact on the process of decision-making, with different reasons, motivations and objectives.

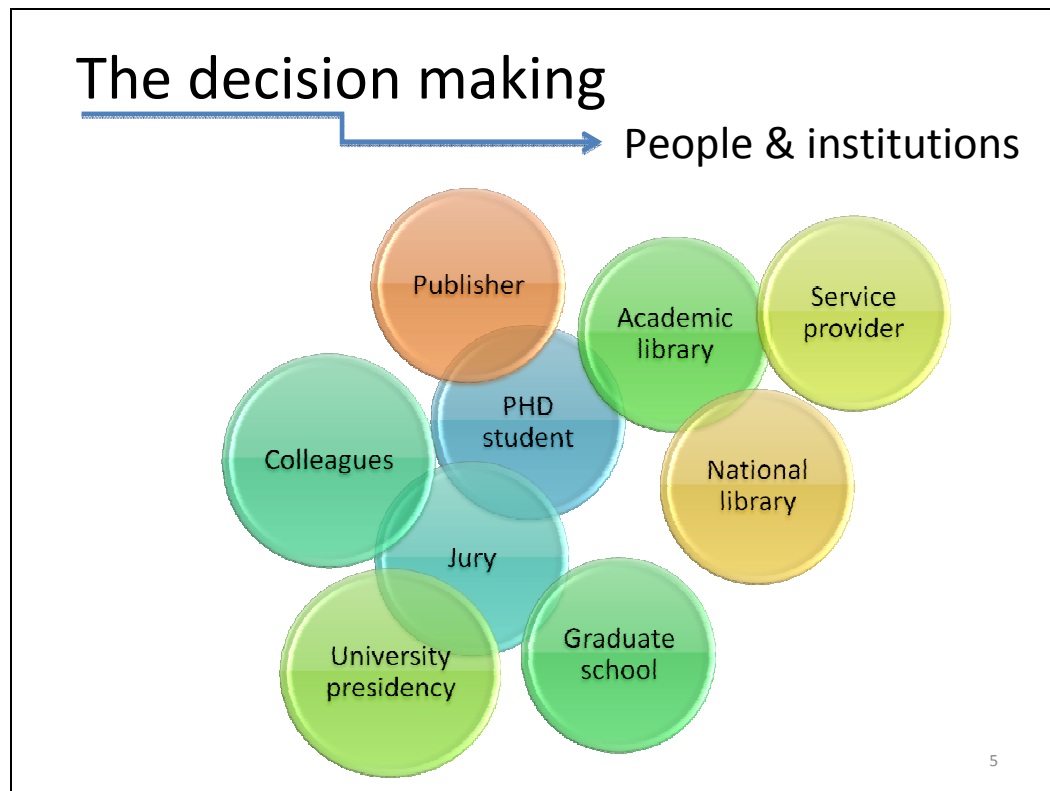[2] See the project website at http://edar.meshs.eu/

Figure 2: People and institutions taking part in decision-making

Figure 2 tries to map these players in a joint sphere of decision-making of dissemination and access to ETDs. Each player sets his own goals, fulfils specific functions, plays his particular role, sometimes consistent with others, sometimes in opposition. For instance, PhD students may deposit their non-reviewed papers in open archives "off-campus", outside of their institution and without any validation or authorization, even when the jury rejects the disclosure.

All these people, groups and institutions act in different ways, for different reasons, with different objectives and strategies. The literature review and survey results reveal several components, such as a publishing project, individual attitude toward OA, copyright law or institutional OA policy that may be understood as independent variables of the final decision (figure 3).

Each of these aspects acts in a different way. Some may point to decide on dissemination or non-dissemination, while others are limited to embargo or on/off-campus decisions. Moreover, some are case-by-case decisions while others reflect general attitudes and stable behaviors. This model may need empirical confirmation and perhaps, more details. Yet, its central characteristics are the multi-factorial or multivariate approach to the prediction of decisions on dissemination or concealment of ETDs. Even if individual publishing strategies and attitudes towards open access may play a major role, other variables such as personal advice from the PhD director, easiness of decision and references should not be neglected, in particular when discussing ways of improving accessibility and availability of PhD theses and designing the respective policies.
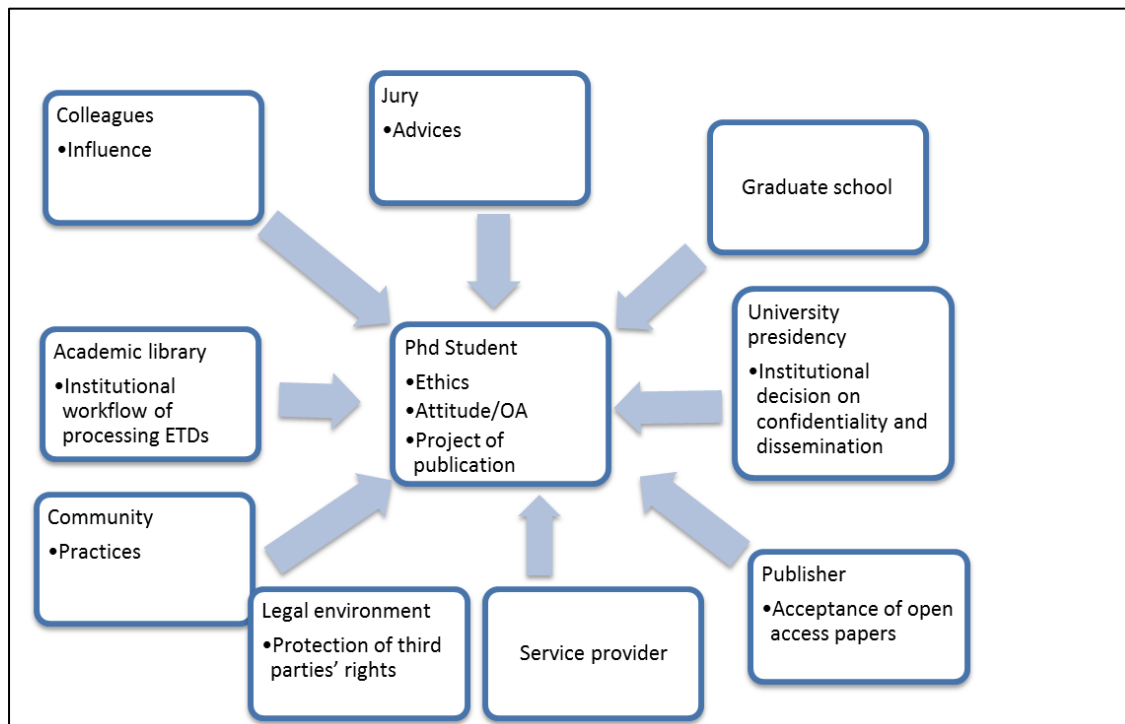
Figure 3: Different aspects of decision-making on OA of ETDs

With regards to accessibility and availability of PhD theses, our analysis showed so far that openness is not a simple, binary concept but that the documents can be more or less open, depending on different variables. Some of those variables are similar to articles published in journals or books, but others are specific to PhD theses. In October 2012, the Scholarly Publishing and Academic Resources Coalition (SPARC), "an international alliance of academic and research libraries working to create a more open system of scholarly communication", released a guide called "How Open Is It" that outlines the core components of open access (e.g., reader rights, reuse rights, copyrights, author posting rights, etc.) covering the continuum from "open access" to "restricted access". Compared with our multivariate approach, this Open Access Spectrum[3] (SPARC 2012) helps to get a realistic view on the problems of openness, disclosure and concealment of theses. Figure 4 shows a possible adaption of the SPARC guide to the specific conditions of the dissemination of PhD theses.

---

[3] See SPARC, PLOS, and OASPA (2012). How Open Is It? Open Access Spectrum at http://www.plos.org/wp-content/uploads/2012/10/OAS_English_web.pdf

Access	Reader Rights	Reuse Rights	Copyrights	Institution Rights	Institution Policy	Posting Workflow	Machine Readability
Open	Free readership rights immediately upon posting	Generous reuse & remixing rights (CC-BY license)	No third party claim	Obligation of free dissemination	OA mandat	No embargo option available in workflow, embargo needs special procedure	Full text, metadata, citations & data, incl supplementary data (annex etc), provided in community machine-readable standard formats through a community standard API or protocol
	Free readership rights after 6 months embargo	Reuse, remixing & further building upon the work subject to certain restrictions & conditions (CC BY-NC & other CC licenses)			Institutional repository		Full text, metadata, citations & data, incl supplementary data (annex etc), may be crawled or accessed through a community standard API or protocol
	Free readership rights after 12+ months embargo	Reuse (no remixing or further building upon the work) subject to certain restrictions and conditions	Third party requires embargo	Obligation of dissemination with restrictions (campus)	OA support	Opt-out if claim for embargo	Full text, metadata & citations may be crawled without special permission or registration
	Free readership rights only on-campus (Intranet)				Interest for OA		Full text, metadata & citations may be crawled with permission
Closed	Total embargo (confidentiality)	No reuse rights beyond fair use/limitations & exceptions to copyright (all rights reserved copyright)	Third party claims full IP protection (confidentiality, privacy)	No institution rights	No OA policy	Opt-in for free dissemination	Full text & metadata not available in machine-readable format

Figure 4: Degrees of openness for ETDs in IRs (adapted from SPARC)

This table needs further discussion and comparison with real situations. Also, some openness-scales may be more complicated then those shown in figure 4, as for instance the copyright dimension merging intellectual property, confidentiality and privacy. But it may help to better understand and evaluate the great variety of IRs, also with regards to potential action in favor of OA.

ETD4OA

Following the EDAR project, the project team prepares a proposal for the European Horizon 2020 Framework Programme for Research and Technological Development. H2020 supports the achievement and functioning of the European Research Area in which researchers, as well as scientific knowledge and technology circulate freely, by strengthening cooperation both between the Union and the Member States, and among the Member States, in particular through the application of a coherent set of rules. Also, a set of activities aims at optimizing the use of national facilities by integrating them into networks and opening their doors to all European researchers. The project "Electronic Theses and Dissertations for Open Access" (ETD4OA) will support the coordination of European infrastructures and open access (OA) policies in the field of electronic theses and dissertations. Together with stakeholders and OA initiatives, it will address barriers and access restrictions, and it will take actions (active communication, recommendations, advice) to promote and develop input, openness and impact of ETDs in existing open repositories and portals. The objectives of the ETD4OA project are (see figure 5):

1. to produce and disseminate reliable and consistent empirical data on the processing of electronic theses and dissertations (ETDs) in a wide set of European countries with their different infrastructures and legal environments, in particular to address the problems and variations of access restrictions;
2. to contribute to the coordination of open access policies regarding ETDs in the European research era and to build a sustainable European community focusing on the open access to research theses;

3. to contribute to the understanding and development of the acceptance, uptake and usage of open access ETD infrastructures, by the research communities, by industry and business and by the Internet community;
4. to develop practical and helpful guidelines and recommendations in order to raise awareness about the challenge, and to support cooperation with developing regions and countries, in order to increase the part of ETDs which are widely and fast disseminated in institutional repositories without any restrictions as the standard solution.
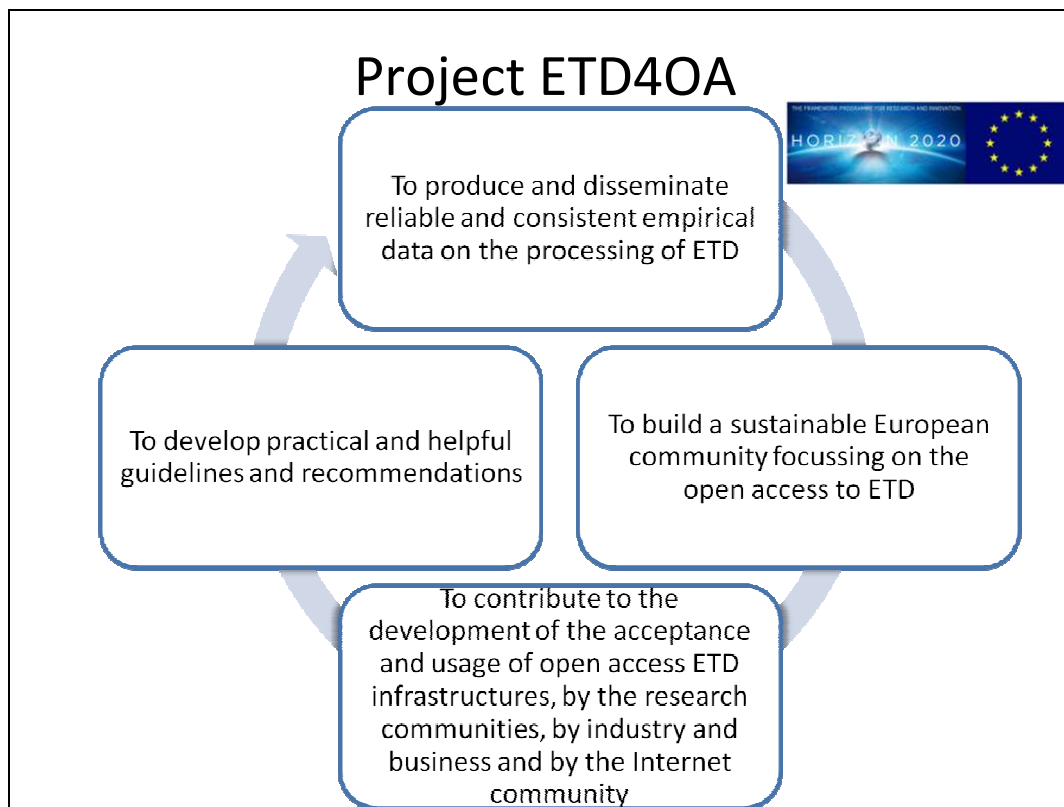

Figure 5: Objectives of the ETD4OA project

The European infrastructure for this already exists, with the interconnection of institutional repositories and other open archives and concurrently registered which can be searched by the DART-Europe portal. The ETD4OA project has not the ambition to develop a new infrastructure or prototype. Instead, today, the problem of open access to ETD lies upstream in local contexts that facilitate decisions in favor of embargoes or restricted access (on-campus access, intranet). To put it in a simple way, pipes exist, but it lacks the fuel and the pressure.

The input to the central portal varies depending on the country and the research field and the institution. To increase the OA-percentage is a problem on many levels, technical (workflow) as well as legal, administrative (regulatory) and tradition and ethics. The answer proposed by the project is awareness raising, lobbying at the national and European policy level, a thorough communication net both for candidates as well as for the responsible local staff, as well as the production of training material for them.

ETD4OA aims at increasing the use of national facilities, i. e. institutional repositories and OA infrastructures, by integrating them into an emerging network of the people, those responsible and involved for European research theses. Its aims are to exploit

synergies between national and Union initiatives by setting up partnerships between relevant policy makers and funding bodies or advisory groups, in order to facilitate the development of global research infrastructures and the cooperation of European infrastructures with their non-European counterparts (in particular US and Russia), ensuring their global interoperability and access.

Some of the research questions the project will address: Which are the (apparently) increasing demands by the authors for keeping the document not open to the public for fear of copyright infringement, e. g. of figures used? Is there increasing pressure of commercial publishers to not allow OA distribution if a book is planned? Are there more recent other ways of posting an ETD in OA which are not seen and thus not counted by the institutional repository network, such as for instance institutional servers, personal websites, gold OA journal publications? Language barriers may lead to publish a thesis as a book to make use of the publisher's editorial office. Figure 6 resumes the five work packages.
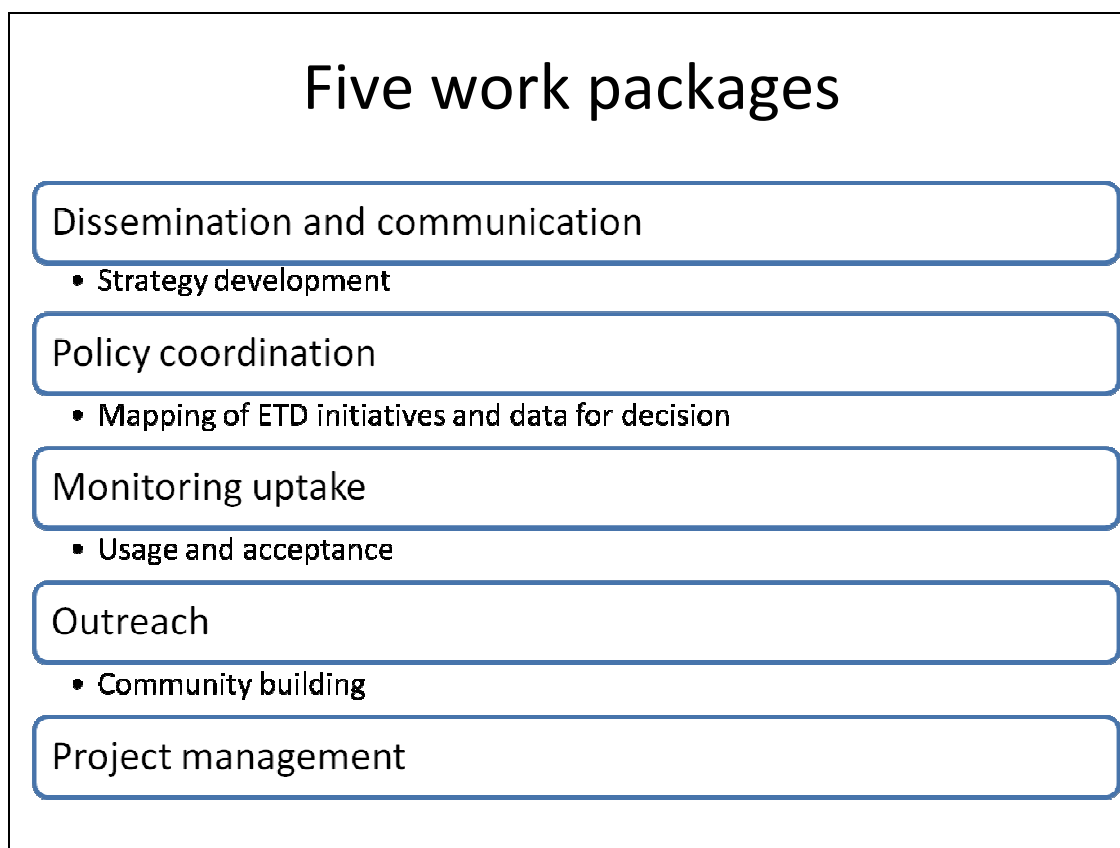

Figure 6: The five work packages of ETD4OA

The project is coordinated by the GERiiCO laboratory at the University of Lille 3 (France) and the ISN Oldenburg (Germany). Actually more than ten member states take part in the project, represented by public or private institutions, experts and networks (see figure 6).

ETD4OA will be submitted within the H2020 project call H2020-INFRASUPP-2014-2, "Support to innovation, human resources, policy and international cooperation", a sub call of the H2020-INFRASUPP-2014-2015 program. The deadline of this call is September 2, 2014.
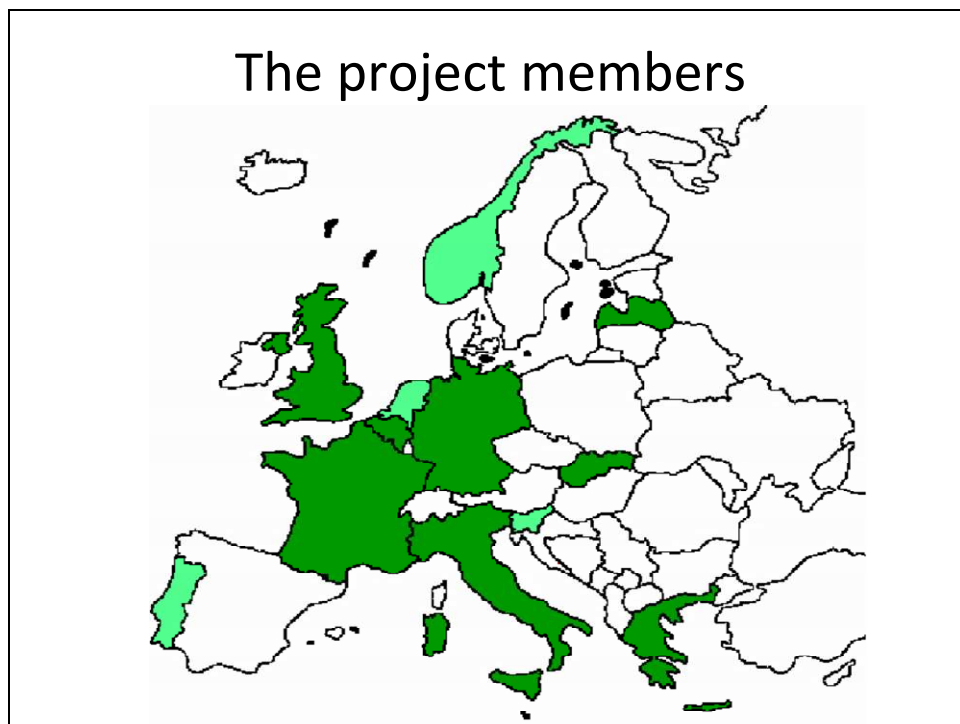
Figure 6: ETD4OA project members (draft July 2014)

References

Schöpfel, J. and Prost, H. (2013). Degrees of secrecy in an open environment. The case of electronic theses and dissertations. *ESSACHESS - Journal for Communication Studies*, 6 (2).
http://www.essachess.com/index.php/jcs/article/view/214

Prost, H. and Schöpfel, J. (2014). Degrees of openness. Access restrictions in institutional repositories. *D-Lib Magazine*, 20 (7/8).
http://www.dlib.org/dlib/july14/prost/07prost.html

Schöpfel, J. and Prost, H. (2014). Back to grey. Disclosure and concealment of electronic theses and dissertations. In *GL15 Fifteenth International Conference on Grey Literature. The Grey Audit: A Field Assessment in Grey Literature. CVTI SR, Bratislava, Slovak Republic, 2-3 December 2013*, pages 110-118, Amsterdam. TextRelease.
http://archivesic.ccsd.cnrs.fr/sic_00944662

Schöpfel, J., Prost, H., Piotrowski, M., Hilf, E. R., Severiens, T., and Grabbe, P. (2014). E-Dissertations: Access and Restrictions (EDAR) - survey 2014. Report, MESHS, Lille.
http://archivesic.ccsd.cnrs.fr/sic_01045115

Schöpfel, J., Prost, H., Piotrowski, M., Hilf, E. R., Severiens, T., and Grabbe, P. (2015 (forthcoming)). Electronic theses and dissertations: access and restrictions. *D-Lib Magazine*, 21 (3/4).