



**HAL**  
open science

# Degrees of Secrecy in an Open Environment : The Case of Electronic Theses and Dissertations

Joachim Schöpfel, Hélène Prost

► **To cite this version:**

Joachim Schöpfel, Hélène Prost. Degrees of Secrecy in an Open Environment : The Case of Electronic Theses and Dissertations. ESSACHESS – Journal for Communication Studies, 2013, 6 (2(12)), pp.66-85. sic\_01077239

**HAL Id: sic\_01077239**

**[https://archivesic.ccsd.cnrs.fr/sic\\_01077239](https://archivesic.ccsd.cnrs.fr/sic_01077239)**

Submitted on 24 Oct 2014

**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.



Distributed under a Creative Commons Attribution 4.0 International License

# Degrees of Secrecy in an Open Environment The Case of Electronic Theses and Dissertations

Joachim SCHÖPFEL

Senior lecturer in information sciences, director of ANRT,  
GERiiCO, University of Lille 3

FRANCE

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

Hélène PROST

Information professional, INIST CNRS, associated member  
GERiiCO, University of Lille 3

FRANCE

[helene.prost@inist.fr](mailto:helene.prost@inist.fr)

**Abstract:** The open access (OA) principle requires that scientific information be made widely and readily available to society. Defined in 2003 as a “comprehensive source of human knowledge and cultural heritage that has been approved by the scientific community”, open access implies that content be openly accessible and this needs the active commitment of each and every individual producer of scientific knowledge.

Today, the success of the open access initiative cannot be denied. Yet, in spite of the growing success of the open access initiative, a significant part of scientific and technical information remains unavailable on the web or circulates with restrictions. Even in institutional repositories (IR) created to provide access to the scientific output of an academic institution and central vector of the so-called green road to open access, more or less important sectors of the scientific production are missing. This is because of lack of awareness, embargo, deposit of metadata without full text,

*ESSACHESS. Journal for Communication Studies*, vol. , no. () / :

confidential content etc.

This problem concerns in particular electronic theses and dissertations (ETDs) that are disseminated with different statuses – some are freely available, others are under embargo, confidential, restricted to campus access (encrypted or not) or not available at all. While other papers may be available through alternative channels (journals, monographs etc.), ETDs most often are not.

Our paper describes a new and unexpected effect of the development of digital libraries and open access, as a paradoxical practice of hiding information from the scientific community and society, partly while sharing it with a restricted population (campus). We try to explain these different shades of grey literature in terms of different degrees of secrecy related to intellectual property, legitimate interests, expected exploitation and trade secrets, and suggest some ways of increasing availability of ETDs in an open environment (inter-lending loan and document supply, alternative format etc.). The study builds on a review of recent papers on ETDs in institutional repositories and provides empirical evidence on this reality. The study also includes an overview of the thesis mandate policies of IR as mentioned in the ROARMAP directory, and an evaluation of the availability of ETDs in a small panel of European and American academic IR, networks and institutions.

**Keywords:** Electronic theses and dissertations, scientific information, grey literature, institutional repositories, open access, secrecy, embargo

\*\*\*

### *Secret et libre accès - le cas des thèses numériques*

**Résumé (in French) :** Le principe du libre accès exige que l'information scientifique soit largement et facilement disponible pour la société. Défini en 2003 comme une « source de la connaissance humaine et du patrimoine culturel qui a été approuvé par la communauté scientifique », le libre accès implique que les contenus soient librement accessibles, ce qui nécessite l'engagement actif de chaque producteur de la connaissance scientifique.

Aujourd'hui, le succès de l'initiative d'accès ouvert est indéniable. Pourtant, en dépit du succès grandissant de l'initiative du libre accès, une partie importante de l'information scientifique et technique reste indisponible sur le web ou circule avec des restrictions. Même dans les dépôts institutionnels (IR) créés pour fournir un accès à la production scientifique académique et considérés comme le vecteur central de la « voie verte » pour ouvrir l'accès, des pans plus ou moins importants de la production scientifique sont portés disparus. Cette situation s'explique du fait d'un manque de prise de conscience, d'un embargo, d'un contenu confidentiel ou du dépôt de métadonnées sans lien vers le texte intégral, etc.

Ce problème concerne en particulier les thèses et mémoires électroniques (ETD) qui sont diffusés sous différents statuts - certains sont disponibles gratuitement, d'autres sont sous embargo, confidentiels, en accès restreint sur le campus (crypté ou non) ou ne sont pas du tout disponibles. Alors que d'autres documents peuvent être disponibles via d'autres canaux (journaux, monographies, etc.), les thèses et mémoires électroniques le plus souvent ne le sont pas.

Notre article décrit un effet nouveau et inattendu du développement des bibliothèques numériques et du libre accès : nous arrivons à une situation paradoxale où l'information est cachée aux yeux de la communauté scientifique et de la société, alors qu'elle est partiellement disponible pour une population restreinte (campus). Nous essayons d'expliquer ces différents visages de la littérature grise en termes de différents degrés de confidentialité, liés à la propriété intellectuelle, aux intérêts légitimes, aux restrictions d'exploitation commerciales. Nous suggérons quelques pistes pour accroître la disponibilité des thèses et des mémoires électroniques dans un environnement ouvert (prêt entre bibliothèques et fourniture de documents, autre format, etc.) L'étude s'appuie sur une analyse des études récentes relatives aux thèses et mémoires dans des archives institutionnelles et fournit des données empiriques illustrant cette réalité. L'étude comprend également un aperçu des politiques de dépôt des thèses dans les archives institutionnelles, d'après les informations fournies par le répertoire ROARMAP, puis l'article propose une évaluation de la disponibilité des thèses électroniques au sein d'un petit panel rassemblant des archives institutionnelles, des réseaux et des établissements américains et européens.

**Mots-clés (Keywords in French) :** Thèses numériques, information scientifique, littérature grise, archives institutionnelles, libre accès, secret, embargo

\*\*\*

## **1. Introduction**

Openness, i.e. “the relative degree of freedom given to the dissemination of information or knowledge” (Long 2001), is generally considered as a fundamental and necessary condition for science. The belief that “knowledge is most effectively pursued when disseminated without hindrance” began when science first took shape in early Greece (McMullin 1985). Science is supposed to progress through rapid communication of results among scientists (Lawal 2002), which implies “accessibility or lack of restrictiveness with regard to communication” (Long 2001) and requires that scientific information be made widely and readily available to society. This epistemological approach often reflects, too, the conviction that only free debate, open communication and unrestricted exchange can contribute to open society, democracy and human welfare (Popper 1945).

More recently and with regards to Internet and the web, leading research organisations defined the open access principle as a “comprehensive source of human knowledge and cultural heritage that has been approved by the scientific community” (Berlin Declaration on Open Access to Scientific Knowledge 2003). Open access implies that content – original scientific research results, raw data and metadata, source materials, scholarly multimedia material etc. – be openly accessible and this needs the active commitment of each and every individual producer of scientific knowledge. Ten years after the Berlin Declaration, the success of the open access initiative cannot be denied. More than 2,200 repositories and 8,000 OA journals make available several millions of papers, mostly articles but also communications, unpublished papers, theses and dissertations, reports, datasets etc., improving the efficiency of scientific communication by free and unrestricted dissemination of content. The OA principle, supporting and inviting a global readership, appears in accord with the scientific community’s interests (Willinsky 2005). Especially open research repositories (open archives) are considered as an appropriate and promising way; “regardless of how different research areas move into the future (...), on the one- to two-decade time scale it is likely that other research communities will also have moved to some form of global unified archive system without the current familiar partitioning and access restrictions from the paper medium, for the simple reason that it is the best way to communicate knowledge and hence to create new knowledge” (Ginsparg 2001).

Yet, “reality of modern science is more complex” (Long 2001), and in spite of the open access initiative, a significant part of scientific and technical information remains unavailable on the web or circulates with restrictions. The digital revolution complicates the traditional problem of transparency and secrecy (Cohen-Tanugi 2001). Some information are classified for security reasons, others are concealed because of industrial interests or are not available due to price barriers or inefficient dissemination tools. This problem is related to the history of science. McMullin (1985) described how, during the Middle Ages and Renaissance, ownership rights gained importance as science became a means to power and how, especially by patenting innovations, knowledge split up into science and technology, with science being open and technology, secret (Long 2001). The increasing impact of industry and business on American academic research in life sciences and medicine has been blamed by Krinsky (2004) as “corruption” where “secrecy has replaced openness; privatisation of knowledge has replaced communitarian values; and commodification of discovery has replaced the idea that university-generated knowledge is a free good, a part of the social commons”. Krinsky’s critic is a modern version of the former understanding of expressed thoughts as “social property”. Already in 1856, the Italian politician and activist Giuseppe Mazzini stated that “the breath of the

human soul cannot be a monopoly (...) nobody has the right to impede or restrict the circulation of truth.”<sup>1</sup>

With regards to academic publishing as the main vector of scientific communication, Peekhaus (2012) condemns the “capitalistic control” exerted by the commercial academic publishing industry that subverts the dissemination of scholarly research. Yet, even without “capitalistic control”, secrecy, i.e. “intentional concealment” (Bok 1984), accompanies openness as a kind of dark side of science. Even in open institutional repositories created to provide access to the scientific output of an academic institution and central vector of the so-called green road to open access (Harnad et al. 2008, Lynch 2003), more or less important sectors of the scientific production are missing. There are many reasons for this - lack of awareness, publishers’ embargo policy, deposit of metadata without full text, confidential nature of content etc. Especially embargoes - “publisher-imposed delay on the availability of full-text content” (Brooks, 2003) - have been identified as a major problem: “Science is supposed to progress through rapid communication of results among scientists, but the embargo system is a barrier to this free exchange of information” (Lawal 2002). Additionally, Chen (2004) identified no embargo information or incomplete embargo information as a problem that reduces access to online resources.

The following study will take the exploration of openness and secrecy in science in institutional repositories (IRs) a step further. Instead of published articles, we will focus our attention on electronic theses and dissertations (ETDs). In fact, even a superficial search in open repositories reveals that ETDs are disseminated with different status. Some are freely available, others are under embargo, confidential, restricted to on-campus access (encrypted or not) or not available at all. But while articles, journals or monographs may be available through other channels (subscription, interlibrary loan etc.), ETDs most often are not. Our paper will provide an overview on some published studies, communications and posters, present original data from France and will discuss the problem of secrecy with regards to ethics and policies, individual and institutional strategies, and as a problem of workflow management, decision-making and framing.

## **2. Literature overview**

After the launch of several institutional repositories and digital workflows for ETDs, a small number of papers provided data and observations on limited access to electronic theses. The following review covers recent articles published by academic publishers (via Scopus and Google Scholar) and proceedings from the conferences

---

<sup>1</sup> In G. Mazzini, *Scritti editi ed inediti*, vol. IX, Roma, 1877, p. 244.

of the Networked Digital Library of Theses and Dissertations (NDLTD) and of the Grey Literature Network Service (GreyNet), together with conferences on open access, in particular the Berlin Open Access conferences and the conference series on Open Repositories (OR). We included information and comments published on the ETD email discussion list<sup>2</sup> managed by NDLTD and hosted by Virginia Tech. In some cases, we completed the information through direct email contact.

### 2.1. *Typology of access restriction*

How can electronic theses and dissertations be concealed? According to literature, especially to case studies of ETD projects, we can distinguish different types of access restriction beyond open access, i.e. full content available to all.

**Campus access:** The access to the ETD is restricted and limited to the campus where the PhD theses have been prepared. For instance, at the Amherst College, campus access means “full content only available to those using a computer on the UMass Amherst campus (or) to those with a valid UMass Amherst user name and password” (Banach 2011). At Amherst, “campus access” is a permanent restriction. Campus access is sometimes opposed to off-campus but this is misleading because authentication tools may allow off-campus access for registered and authorized users. Campus access may be secured by digital rights management techniques, for instance by encryption such as on the West Virginia campus (Hagen 2010). When the ETD is only available on-campus, print copies may be supplied via interlibrary loan (ILL) and document supply services, either as a print copy or as electronic files, free or with ILL fees (Banach 2011, Hagen 2010). At the Florida State University, campus access is the option by default for all retro-digitized theses (Smith 2009), such as at the University of Auckland where it is called “closed access”.

**Embargo:** In connection with ETDs, delays on the availability of full-text content (Brooks 2003) are not imposed by publishers but decided by faculty, the author or both. In most cases, the embargo period is pre-defined and settled by decision, sometimes only after authorisation from supervisor, department or research degree board. At Amherst College, standard embargo on simple demand without any justification lasts six months after degree date; after this period, depending on the author’s choice the full content is made available either on or off-campus, because embargoes can be placed on either campus access theses or open access theses (Banach 2011). At Virginia Tech, authors may choose to temporarily restrict access to their ETDs. Generally, the student can select between different embargo options (six months, one year, two years...). In a survey with American universities, Hawkins et al. (2013) report for those graduate schools that allow an embargo an

---

<sup>2</sup> ETD-L@LISTSERV.VT.EDU

average maximum period of two years. Other universities apply a one to three year extendable restriction on request by student or supervisor (Brown et al. 2010). Generally, too, the embargoed theses are not available via ILL. For instance, the University of Maryland does not currently offer campus access to the electronic copy of the embargoed thesis. Only a non-circulating print copy is available for viewing on campus and it is not available for ILL (Owen et al. 2009 and personal information).

**No access:** Some ETDs are not available at all, or more precisely, only their metadata are publicly available. For instance, Hagen (2010) reports a “No Access” option at the West Virginia University. This option means “ETD file(s) (are) embargoed from access for patent (or) proprietary concern”. Sometimes, this solution may be called “opt-out” (Brown et al. 2010) which means a deliberate and justified decision not to disseminate the digital file.

## *2.2. The part of access restricted ETDs*

A small number of empirical studies on ETDs reveal figures on access restriction. At Amherst College, Massachusetts, 32% of PhD theses cannot be accessed from outside of the campus and 20% are under embargo for at least six months (Banach 2011).

At the University of Maryland, 68% ETDs are available without any restrictions. The other theses are under embargo, 21% up to one year and 11% from one to six years (Owen et al. 2009).

Smith (2009) describes how the Florida State University Graduate School requested campus-community and ETD PDF document security options for FSU ETDs starting in Fall 2008, and he adds that “since retrospective digitized theses and dissertations did not include retrospective digitized access agreement forms, senior leadership recommended IP restriction for all FSU retrospective digitized theses and dissertations in 2009”. Following the published figures, this part of restricted access can be estimated at about 16%. At Auckland, the part of ETDs with “closed access” is 28%.

West Virginia University: Hagen (2010) reports that for the period 1998-2010 85% of the more than 4600 theses are disseminated without any restriction. The part of theses with restricted access decreased from 47% (1998-2000) to 15% in 2010, because the option of encrypted on-campus only access was phased out in 2009 while the part of embargoed ETDs remained stable. At Virginia Tech, the rate of embargoed ETDs is 46%.



University of Texas at Austin: The Digital Repository contains 935 restricted ETDs (8%) that are permanently restricted, accessible only to users who have a University ID (current faculty, students, staff) (Steans 2012, completed by personal information).

In Brazil, Pavani & Mazzeto (2009) describe access restriction for 11% ETDs on the campus of the Pontifícia Universidade Católica at Rio de Janeiro. About 21% of these files are under embargo for five years or longer.

Table 1 summarizes the published results.

**Table 1.** *Published figures on access restrictions of ETDs (\* Amherst: estimation)*

Study	Sample	% open access	% restricted access	% on campus	% embargo
Texas (Austin)	11539	92%	8%	8%	n/a
PUC Rio de Janeiro	2787	89%	11%	n/a	11%
West Virginia	4600+	85%	15%	15%	0.3%
Florida State	3709	84%	16%	n/a	n/a
Auckland	3088	72%	28%	28%	n/a
Maryland	2050	68%	32%	n/a	32%
Virginia Tech	20386	54%	46%	n/a	46%
Amherst	n/a *	48%	52%	32%	20%

According to these surveys, a significant percentage – 10 to 50% - of ETDs are not disseminated, at least not before six months to two years or even longer, or only on the campus. This partial, temporary or total concealment is in conflict with the open access principle and opposite to the need for rapid and unobstructed communication of scientific results. Only very few data on trends have been published. Based on figures from ProQuest (UMI) Hawkins et al. (2013) found an increasing number of embargoed ETDs. The findings by Owen et al. (2009) can be interpreted in the same way, especially for short-term one-year embargoes. On the other hand, the embargo statistics at West Virginia appear to be relatively stable over time (Hagen 2010).

### 2.3. Reasons

In a UK survey on mandates for ETDs, 88% of the universities indicated that they allow authors of theses to impose restrictions on access to their work, i.e. the electronic file, with many different reasons. Students, with the agreement of their supervisor, can request an embargo for the following reasons: commercial contract (for instance, funding by an external organisation), patent pending, ethical confidentiality and/or sensitive material (data protection), publication pending and third party copyright (Brown et al. 2010). The same study reveals that restrictions on grounds of third party copyright, data protection or potential risks to personal safety

were reported only amongst ETDs and that only 60% of the universities allow students to impose restrictions for print theses.

At Brunel University, “while every effort has been made to ensure that embargoing access to theses is not used as ‘a panacea against all ills’, students are offered the option of a 3-year embargo if they have a publication or patent pending” (Brown & Sadler 2010). Academics of the University of Maryland mention future publication, protection of data or work, student request, proprietary data and patent application as primary reasons for approving of embargoes (Owen et al. 2009).

In Italy, Arabito et al. (2008) justify embargo options as indispensable for the same reasons: “(...) the free availability of doctoral theses on the web can be jeopardized by thorny copyright issues, which arise in the following cases: use of third party owned materials (...), third parties involved (possible infringement of privacy), patentable discoveries (...), and ongoing publication of data (according to the publisher policy)”.

The last reason appears by far to be the most important and explains between 1/3 (Owen et al. 2009) and 3/4 (Pavani & Mazzeto 2009) of all embargo decisions. The role of faculty appears to be crucial. At Virginia Tech, nearly half of the students’ embargo decisions were taken on advice by faculty while requests by publishers are insignificant (McMillan et al. 2012). Ramirez et al. (2013) confirm that “scholars continue to doubt the viability of publishing opportunities after a dissertation or thesis becomes available electronically in an open access repository. Perceptions and fear, not data, inform many graduate advisors’ and graduate students’ decisions to restrict access to their ETDs”.

On the other hand, a recent survey with more than 150 American graduate schools show that nearly 30% of all institutions “either don’t allow an embargo at all, or don’t tell students (about it any) where they can find that information readily (...) In their enthusiasm for OA, universities and libraries across the U.S. are cajoling, arm-twisting, or even coercing students into in effect surrendering the copyright to their dissertations and theses, sometimes with the threat that students cannot graduate if they disagree” (Hawkins et al. 2013).

Florida State University Graduate School implemented access restriction - on campus only access – for older, digitized PhD theses: “Since retrospective digitized theses and dissertations did not include retrospective digitized access agreement forms, senior leadership recommended IP restriction for all FSU retrospective digitized theses and dissertations in 2009” (Smith 2009).

## 2.4. Access restrictions and discipline

Only three studies present detailed embargo statistics cut down by scientific disciplines (Owen et al. 2009, Pavani & Mazzeto 2009, ProQuest 2012). Yet, these survey results are not really consistent, as the following table shows (table 2).

**Table 2.** Embargo per discipline

Discipline	Owen et al.	Pavani & Mazzeto	ProQuest
Life Sciences	54%		high
Chemical Sciences	54%	40%	
<i>Sciences</i>			<i>medium</i>
Agriculture, Natural Resources	51%		medium
Business	47%	6%	
Arts & Humanities	41%		
<i>Arts</i>			<i>low</i>
<i>Law</i>		36%	
<i>Languages</i>		19%	<i>medium</i>
<i>Philosophy</i>		10%	<i>low</i>
<i>Theology</i>		24%	<i>low</i>
<i>History</i>		10%	
Engineering	29%		
<i>Applied Sciences</i>			<i>high</i>
<i>Materials Engineering</i>		28%	
<i>Industrial Engineering</i>		11%	
<i>Mechanical Engineering</i>		7%	
<i>Civil Engineering</i>		6%	
<i>Electrical Engineering</i>		4%	
Education	27%	9%	medium
Behavioural & Social Sci	25%	6%	high
<i>International Relations</i>		12%	
<i>Social Work</i>		9%	
<i>Economics</i>		0%	
<i>Psychology</i>			<i>low</i>
Computer, Math & Physical Sci	23%		
<i>Informatics</i>		1%	
<i>Mathematics</i>		8%	
<i>Physics</i>		7%	
Public Health	21%		medium
Public Policy	19%		
Journalism	18%	29%	
<i>Communication</i>			<i>low</i>
Architecture	7%		
<i>Design</i>		14%	
<i>Metrology</i>		6%	

Some disciplines appear to be relatively consistent, such as life and chemical sciences, agriculture and environment, business, some domains of engineering (applied sciences) and public health, all with medium or high rates of embargoes. Yet, we must be careful with these statistics because of more or less small samples.

Another effect by disciplines has been reported by Pavani & Mazzeto (2009). In science and technology, pending publications as a reason for embargo concern mostly articles (73%) while in social sciences students intend above all publishing a book (57%)

Other differences have been found with regards to OA policies, decision-making, communication, advice and embargo options. We'll come back to these differences in the discussion.

### 3. Methodology

In addition to the review of literature, we conducted a survey with a small panel of French and Belgian universities (Liège, Lille 1, Lille 3, Lorraine, Valenciennes), ABES (the host of the French academic union catalogue SUDOC and the gateway to PhD theses, and the operating agent for the national ETD infrastructure STAR), The British Library (EthOS) and ProQuest. The survey was limited to statistics – global figures of processed ETDs in 2012, ETDs in open access, with access restrictions, embargoed or limited to on-campus availability. Yet, we also gathered comments and explanations where available. The objective was to provide some first elements that may be helpful to understand the situation, to allow comparison especially with US universities and to prepare a forthcoming national survey on the availability of PhD theses.

In addition, the study also includes an analytical overview of the thesis mandate policies of institutional repositories as mentioned in the ROARMAP directory.

### 4. Results

#### 4.1. *French and Belgian universities*

Table 3 presents results from a small panel of universities in France and Belgium engaged in open access and electronic theses.

**Table 3.** *Synthesis of figures on access restrictions of ETDs of French and Belgian sample*

Study	Sample	% open access	% restricted access	% on campus	% embargo
Lille 1	833	79%	21%	15.5%	5.5%
Lorraine	52	71%	29%	29%	n/a
Valenciennes	35	63%	37%	31%	6%
Liège	191	57%	43%	33%	10%
Lille 3	124	40%	60%	n/a	n/a

The University of Lille 1 processed 833 ETDs in science and technology from 2008 to 2011. Nearly 80% are in open access on their institutional repositories. 15% are available on the campus only while the other 5% are under unlimited embargo, based on a decision of the faculty to protect intellectual property and innovation (confidentiality). There is no real trend between 2008 and 2011.

The University of Lille 3 (Social Sciences and Humanities) processed between 2006 and 2007 124 electronic theses. Only 50 are disseminated on the web (40%). The others are available only as print copies or on microfiche.

The University of Valenciennes started to disseminate their PhD theses via a new institutional repository in 2012. So far, 35 ETDs were or are being processed. Two are confidential (permanent embargo), eleven are limited to intranet availability (most of them for a limited period), and the rest is available in open access without restriction. Their experience is that students are often confused by embargo, confidentiality and on-campus options, and that it is not always certain who in reality made the decision. Also, they feel uncomfortable with pressure from industry because of sensitive information, and they prefer to publish a shortened version than to hide away the theses.

In 2012, the multi-campus University of Lorraine (Nancy and Metz) processed 52 ETDs. In 2013, 79% were freely available on their new institutional repository PETALE while the other 21% are only available on the campus (no information about embargoed theses).

The University of Liège (Belgium) document server indicates 191 PhD theses for 2012. 108 are freely available on their institutional repository ORBi (57%). For 33%, the access is limited to the campus; the remaining 10% are embargoed for a non-specified delay.

This small panel is surely not representative and the results should be interpreted with caution. Nevertheless, these figures show very clearly that the problem of restricted access to theses is not limited to the United States or the UK but concerns other countries with ETDs infrastructures and institutional open archives, ranging from 20% to more than 50% with limited access.

#### 4.2. *French ETD infrastructure STAR (ABES)*

Since 2006, French universities have progressively switched from the traditional handling of print PhD theses to the new infrastructure of ETDs called STAR, linked to a national gateway “Theses.fr” (Giloux & Mauger-Perez 2008, Giloux 2012). In July 2012, 108 universities integrated the new network but only 74 were fully

operational while the other 32 were still in a test period. The ETDs are disseminated via the national server for PhD theses TEL hosted by CCSD, via institutional repositories hosted by the universities or by CCSD, or via ABES.

From 2006 to 2012, ABES processed 10,631 ETDs. 8,737 theses were available on the web without any restrictions (80%) while the access to the other 1,894 theses was limited to on-campus availability (20%). ABES did not deliver any information about embargoed ETDs.

#### 4.3. *EThOS (The British Library)*

Up to now, information on embargoed theses is not comprehensive in the UK. Only one part of the institutional repositories record embargo as metadata so that this information cannot be accurately harvested by the central EThOS server. Some EThOS records mention that the item is under embargo or not available at all. Sometimes, only one part (2<sup>nd</sup> volume, appendices) is embargoed. In other cases, the theses cannot be downloaded from the EThOS platform but can be ordered at the university or directly at the author. Recent data from 2013 are more or less brief and not really reliable.

#### 4.4. *ProQuest*

ProQuest Dissertation Publishing, former University Microfilms and part of the corporate information company ProQuest-CSA, has been publishing over two million dissertations and theses since 1938. They have over 700 active university publishing partners, publish more than 70,000 new graduate works each year and provide access to graduate works for thousands of libraries around the world.

In 2012, ProQuest conducted a study on ten years embargo trends (2000-2010) in the ProQuest Dissertations and Theses (PQDT) database. The surveyed corpus of 500,000+ print and electronic theses contained about 25,000 embargoed items (5%). Most of the embargoes are short-term embargoes, for six months to five years, but a small part of theses are under permanent (long term) embargo.

More recently, ProQuest confirmed that the number of embargoed theses is increasing, especially for ETDs and in specific disciplines like social sciences, education, life sciences and engineering. ProQuest pays special attention to embargoes and access restrictions (on/off campus) because each graduate school has its own guidelines and each student can contact ProQuest and ask for embargo, even if the degree-granting institution has required open access.

#### 4.5. *Mandate policies of institutional repositories*

The international Registry of Open Access Repositories Mandatory Archiving Policies (ROARMAP<sup>3</sup>) hosted by the University of Southampton with around 400 institutional and other mandates includes 98 thesis mandates, i.e. specific guidelines for processing and depositing theses and dissertations in institutional repositories. Only 19 mention embargo periods. Some mandates specify the embargo period (one, two or three years), the others don't but it is not always clear if this means "unlimited embargo".

For instance, the University of Pretoria (South Africa) declares that "Since 2004 it has been mandatory for students to submit an electronic copy to the repository before graduation. 2000-2003 ETDs are digitized and added retrospectively by appointed staff. Access is generally open but provision is made for (and clearly indicated) - access on campus only - 2-year embargo for publishing and patent negotiations - indefinite embargo for sensitive information or sponsorship/funding conditions".

At the University of Amsterdam (Netherlands), "At least four weeks prior to the public defence of the thesis, an electronic version of the thesis shall be published by the University library. In cases where the thesis or part of the thesis is subject to an embargo, the electronic publication of the thesis will take place once the longest embargo has expired. In that case the electronic publication of the thesis shall be limited to the cover, the title page and the summary of the thesis in Dutch and in English."

While the University of Ljubljana (Slovenia) asks that "items may not be deposited until any publishers' embargo period has expired", the University of Trieste (Italy) accepts that the "author may place a one-year embargo on access to the deposit (Closed Access) where there is a reason (e.g. patent pending)".

Obviously, there is no standard or pattern that fits for all repositories or that is accepted as a reference. Each institution decides on its own rules and criteria for decision-making.

## 5. Discussion

Our paper is preliminary research, built on a review of recent literature and empirical data from selected academic libraries and networks. Even if our data provide some interesting insight into the phenomenon of (new) barriers to scientific

---

<sup>3</sup> <http://roarmap.eprints.org/>

information, and even if one part of the data is from major institutions like ProQuest, ABES and the British Library, the sample is still too small to produce significant and representative results. Our intention was to add empirical evidence to the small body of research in this domain. More research is needed, especially national or international surveys with large and representative samples of academic libraries.

However, our data are consistent and tell the same story, in that the development of digital theses and institutional repositories had an unexpected collateral effect insofar as access to a significant part of PhD theses (5-30% or more) became more difficult, compared not only to other theses but also, to the former dissemination of print and microform copies. In other words, the development of digital infrastructures produces secrecy and concealment. This paradox calls for comments.

#### *5.1. Ethics, law, legitimate interests and policy*

The choice of secrecy is first of all a personal decision of the author. Keeping secrets has been seen as another system of ethics: “To be able to hold back some information about oneself or to channel it (...) gives power” (Bok 1984). Also, Hawkins et al. (2013) prone an ethical approach to embargo decisions and criticize open access because it “allows abuse by allowing illegitimate parties across the world to look for things that they can steal and sell for a quick and easy profit.” In other words, they give individual freedom to disclose or conceal work of the mind priority over other considerations, in particular over institutional strategies (see below).

This ethical approach has law on its side. Theses, written texts “representing the independent research and authorship of a single individual” (Juznic 2010) confirm the scientist’s acquired skills and competencies. Without doubt, they are work of the mind and subject to intellectual property rights that protect the authors’ privilege to decide on disclosure and ways of dissemination of their intellectual work (Schöpfel & Lipinski 2012). Obviously, private ownership is in conflict with public dissemination of science. The “protecting function of secrecy” (Bryon-Portet 2011) primarily protects copyright, not community or institutional needs.

This is not really new. Long (2001) described the complicated relationship between openness, secrecy, authorship and intellectual property in early science. Monier (2000) reviewed the history of secrecy as a history of the limits of public action and policy. Applied to science, the concealment of scientific information can be considered as limits or failure of scientific progress.

On the other hand, individual choice of secrecy is often meant to preserve the authors’ legitimate interests. “Certain seemingly secretive behaviours displayed by



scientists and inventors are expression neither of socio-professional values nor of strategies for the maximization of the economic value of their knowledge. They are, instead, protective responses to unavoidable risks inherent in the process of publication and priority claiming” (Biagioli 2012). In particular, students may fear plagiarism and overall, rejection by journal or book publishers if their PhD thesis is available on the web. “Such a publishing project (with commercial publishers) may incite authors not to allow digital dissemination, especially in social sciences and humanities” (Schöpfel & Lipinski 2012). Yet, different studies show that the great majority of publishers (87%) are not opposed to publish PhD theses already online (Ramirez et al. 2013, McMillan et al. 2012).

Perhaps we can say with Peekhaus (2012) that in an unexpected way, through often unjustified fears and wrong expectations, academic publishing subverts free and open scientific communication. More generally, the exclusive protection of the authors’ freedom of choice denies the research community’s need for information about the results and is in potential conflict with fair use. Thompson (1999) admitted that “secrecy of various kinds is sometimes justified and even desirable in a democracy. But it is justified only under carefully specified conditions, which ensure that the secrecy itself is subject to democratic accountability.” Applied to scientific communication and in particular, to the dissemination of PhD theses, this means that individual choice should be monitored and counterbalanced by institutional reasons.

### 5.2. *Individual and institutional strategies*

Are individual and institutional strategies in opposition? We have already described the main reasons for restricted access and embargo decision, e.g.

- “Patentable rights in the work or other issues in which disclosure may be detrimental to the rights or interests of the author.
- Work based on proprietary research and funding.
- The ethical need to prevent disclosure of sensitive or classified information about persons, institutions, technologies, etc.
- The interest of an academic or commercial press in acquiring the rights to publish a dissertation or thesis as a book.
- Content that is likely to be submitted to a peer-reviewed journal” (ProQuest 2012).

These reasons reflect both individual and institutional strategies. Publishing articles and books not only contributes to scientific career but also to the academic ranking of the institution. Usually also, universities are interested in valorisation of research via patents and try to avoid lawsuits because of intellectual property or privacy infringement. So the question is not whose interests should have priority –

the PhD student's or the institution's – but how the institution deals with inconsistency and conflicting strategies.

Hawkins et al. (2013) were disturbed “in reviewing more than 150 graduate school and library websites (by) their enthusiasm for OA” and their un-ethical “cajoling, arm-twisting, or even coercing (action)” in order to obtain the copyright for open access dissemination. The description and critic may be exaggerated; yet, what they describe is consistent strategy.

Other institutions show more inconsistency. Officially, their goal is to provide open access to the works of their graduate students. But with regards to electronic theses, they implement a strategy of risk avoidance and voluntary concealment. A recent report on UK universities revealed that half of the survey universities impose institutional access restrictions and that “barriers to collecting and exposing electronic theses (...) are overstated in some cases and manageable in all” (Brown et al. 2010). For instance, inconsistency may be produced by different advice and objectives in the same institutions, for instance when the institution (university, department and/or library) promotes open access while faculty suggests protection by embargo or access restriction, or when decisions on embargo are taken or validated on a level not involved in OA.

On the other hand, when the institutional goal is to increase the percentage of open access ETDs, consistency is obtained “through proven strategies involving advocacy, graduate education and measured use of temporary embargo” (Brown et al. 2010) by reducing on campus access, especially encrypted on campus access (Hagen 2010) and by rejecting DRM protected files.

### *5.3. Biased decision-making*

We usually prefer to consider the PhD students as rational agents, making reasoned and conscious decisions about important questions such as the dissemination of their doctoral theses. Speaking of interests and strategies implies logic decisions, based on reality-bound considerations. Also, workflow schemas or tables for the processing of ETDs are built on this “rational-agent theory” where the student makes a series of deliberate choices between different options (see for instance Butler & Taylor 2010, Sefton & Downing 2011, ProQuest 2012).

Now, how can this approach explain the important differences of the part of ETDs available in OA? Why, while on one campus only 10% choose access restrictions elsewhere they are 20%, 30% or more to prefer at least temporary concealment of their work? Are the students really different from one campus to another? And how for instance, do we explain the fact that concerns about

publishers' rejection because of open access are obviously often overstated and unrealistic?

Perhaps we should change our mind about PhD students and take into account other effects, like risk avoidance, loss prevention, framing and experience-related bias. "Making things public is an inherently risky business and it is impossible, I argue, to ensure that priority may not be lost in the very process that is supposed to establish it" (Biagioli 2012). What Biagioli describes for articles and peer review, applies also to the open access publishing. Making theses available on institutional repositories is a "risky business", with regards to subsequent publishing, plagiarism etc. And "risky" decisions are not always taken in a thoughtful way (Kahneman 2011).

We have to be conscious about the potential effect of formulation and environment on the choice of secrecy or not. The way an institution asks questions, the way it proposes options, the way it explains the decisions will impact the decision itself. Are options presented as opt-ins or opt-outs? "Large changes of preferences (...) are sometimes caused by inconsequential variations in the wording of a choice question" (ibid.).

Also, as we have already said, the personal advice by tutors, staff or other students may be different from the official institutional strategy. When the accent is laid on the risk of plagiarism and of not being published, the PhD students will probably prefer options that will reduce the risk and prevent copyright infringement, especially when there are many different options, most of them designed to reduce potential risk without any information about the probability of that risk. Are other options different from the risk-reducing options clearly presented and explained, or offered as "default" option?

We have to be conscious, too, of the impact of different levels of decision and validation, for instance if an authorisation is needed from the head of department or a service provider like ProQuest who may modify certain demands made by students. In some cases, short-term embargoes (full content made available six months after degree date) do not require any justification; a period longer than six months needs approval by the academic department and Graduate Dean.

So when the role of context and framing is acknowledged, "a policy question arises: Which formulation should be adopted?" (ibid.). This takes us back to the question of consistent strategy and open access policy.

## 6. Conclusion

While openness remains a fundamental goal and condition of science, our paper provides a different image of conflicting interests, inconsistent strategies and biased decision-making. Openness itself is called into question, by legal or ethical considerations, or both. The result is a non-optimal dissemination of electronic theses and dissertations. In the Gutenberg past, most PhD theses were disseminated in small numbers, via the institutions, through interlibrary loan and often in microform. Yet, the percentage of concealment was very low and confidentiality was temporary, usually up to five years. Today, the percentage of concealment has increased in a significant manner, and sometimes the concealment is permanent, not justified by confidential content and without any alternative option.

Technology is not the cause of new barriers but it facilitates the choice and management of new forms of secrecy and it raises concerns about plagiarism (priority), about refusal by the editors and publishers, and about financial loss. Together with the copyright and academic evaluation it contributes to a secrecy-prone situation. The research community is struggling to find an appropriate answer. In the environment of PhD theses in institutional repositories, secrecy and concealment are not a consistent concept. It is more appropriate to distinguish between different degrees of secrecy, such as restricted access (on-campus access), temporary or permanent embargo etc. Each option is designed to meet specific concerns. “No two dissertations are alike; (...) a one-size-fits-all approach is inappropriate” (Hawkins et al. 2013).

There may be three ways to go on further. First of all, we need consistent and exhaustive data on the access restricted ETDs, on their distribution and on the reasons of concealment, from representative national and international surveys with institutions and networks. The lack of reliable empirical evidence reduces the understanding of secrecy in openness and impairs the development of appropriate and sustainable solutions.

Second, we need a more formal description of restricted openness in institutional repositories, especially with regards to grey literature. In 2012, the Scholarly Publishing and Academic Research Coalition (SPARC) together with the Public Library of Science (PLoS) published a description of components that define open access and that make a journal more or less open<sup>4</sup>. The same approach should be applied to repositories and ETDs and it should be accompanied by a model of explanatory variables related to the choice of secrecy.

Finally, the existing ETD workflows and management systems should be evaluated against inconsistency and bias, and new infrastructures should be developed that not only consider the concerns and fears but that, instead of privilege

---

<sup>4</sup> <http://www.arl.org/sparc/media/HowOpenIsIt.shtml>

risk avoidance, what is more are able to carry out institutional policies in favour of open access. If openness is necessary for science, and if “knowledge is most effectively pursued when disseminated without hindrance” (McMullin 1985), the scientific community and its institutions must adjust their infrastructures and functioning according to this goal. Just complaining about new barriers and secrecy in an open environment is no option.

## References

- Arabito, S., Cermesoni, D., Galimberti, P., and Vignocchi, M. L. (2008). Time to harvest: Electronic doctoral theses in Italy. In *ETD 2008 11th International Symposium on Electronic Theses and Dissertations, 4 - 7 June 2008, The Robert Gordon University, Aberdeen, UK*.
- Banach, M. (2011). The benefits of managing and publishing ETDs "in house" using an open access repository. In *USETDA 2011: The Magic of ETDs...Where Creative Minds Meet. May 18-20, Orlando, Florida*.
- Biagioli, M. (2012). From ciphers to confidentiality: secrecy, openness and priority in science. *The British Journal for the History of Science*, 45, 213-233.
- Bok, S. (1984). *Secrets: On the Ethics of Concealment and Revelation*. Oxford University Press, Oxford.
- Brooks, S. (2003). Academic journal embargoes and full text databases. *The Library Quarterly*, 73(3), 243-260.
- Brown, J. and Sadler, K. (2010). Vision, impact, success: mandating electronic theses. Case studies of e-theses mandates in practice in the UK higher education sector. UCL University College of London.
- Brown, J., Sadler, K., and Moyle, M. (2010). Influencing the deposit of electronic theses in UK HE: report on a sector-wide survey into thesis deposit and open access. UCL University College of London, London.
- Bryon-Portet, C. (2011). La culture du secret et ses enjeux dans la « Société de communication ». *Quaderni*, 75, 95-103.
- Butler, P. R. and Taylor, S. (2010). Guidance on making your thesis available to others. Manchester eScholar Services, University of Manchester, Manchester.
- Chen, X. (2004). Assessment of full-text sources used by serials management systems, OpenURL link resolvers, and imported e-journal MARC records. *Online Information Review*, 28(6), 428-434.
- Cohen-Tanugi, L. (2001). Le clair-obscur d'Internet. *Pouvoirs*, 2(97), 85-90.
- Giloux, M. and Mauger-Perez, I. (2008). A new French circuit for the electronic theses. In *ETD 2008 11th International Symposium on Electronic Theses and Dissertations, 4 - 7 June 2008, The Robert Gordon University, Aberdeen, UK*.
- Giloux, M. (2012). Theses.fr - access to French PhD. In *ETD 2012. 15th International Symposium on Electronic Theses and Dissertations. Lima, September 12-14, 2012*.

- Ginsparg, P. (2001). Creating a global knowledge network. In *Second Joint ICSU Press - UNESCO Expert Conference on Electronic Publishing in Science, UNESCO, Paris, 19-23 Feb 2001*.
- Hagen, J. H. (2010). Managing ETDs and evolution of the repository: Twelve years of case studies at West Virginia University. In *ETD 2010 13th International Symposium on Electronic Theses and Dissertations. June 16-18, 2010, Austin, Texas*.
- Harnad, S., Brody, T., Vallieres, F., Carr, L., Hitchcock, S., Gingras, Y., Oppenheim, C., Hajjem, C., and Hilf, E. (2008). The Access/Impact problem and the green and gold roads to open access: An update. *Serials Review*, 34(1), 36-40.
- Hawkins, A. R., Kimball, M. A., and Ives, M. (2013). Closing the deal: Coercion, ethics, and the enthusiasms for open access ETDs. *The Journal of Academic Librarianship*.
- Juznic, P. (2010). Grey literature produced and published by universities: A case for ETDs. In Farace, D. and Schöpfel, J., editors, *Grey Literature in Library and Information Studies*, pages 39-51. De Gruyter Saur.
- Kahneman, D. (2011). *Thinking, fast and slow*. Allen Lane, London.
- Krimsky, S. (2004). *Science in the private interest: has the lure of profits corrupted biomedical research?* Rowman & Littlefield Publishers, Lanham.
- Lawal, I. (2002). Scholarly communication: The use and non-use of e-print archives for the dissemination of scientific information. *Issues in Science and Technology Librarianship*, (36).
- Long, P. O. (2001). *Openness, secrecy, authorship: technical arts and the culture of knowledge from antiquity to the Renaissance*. Johns Hopkins University Press, Baltimore, Maryland.
- Lynch, C. A. (2003). Institutional repositories: Essential infrastructure for scholarship in the digital age. Report 226, ARL Association of Research Libraries.
- McMillan, G., Ramirez, M. L., Dalton, J. T., Hanlon, A., Kern, C., and Smith, H. (2012). Do open access ETDs affect publishing opportunities in the sciences? In *CNI Fall Membership Meeting, December 12, 2012*.
- McMullin, E. (1985). Openness and secrecy in science: Some notes on early history. *Science, Technology, and Human Values*, 10(2), 14-22.
- Monier, F. (2000). Le secret en politique, une histoire à écrire. *Matériaux pour l'histoire de notre temps*, 58(1), 3-8.
- Owen, T. M., Hackman, T., and Harrod, T. (2009). ETDs in Lock-Down: Trends, analyses and faculty perspectives on ETD embargoes. In *ETD 2009 12th International Symposium on Electronic Theses and Dissertations. June 10-13, 2009, Pittsburgh, Pennsylvania*.
- Pavani, A. and Mazzeto, A. (2009). Restricted ETDs and open access. In *ETD 2009 12th International Symposium on Electronic Theses and Dissertations. June 10-13, 2009, Pittsburgh, Pennsylvania*.
- Peekhaus, W. (2012). The enclosure and alienation of academic publishing: Lessons for the professoriate. *tripleC*, 10(2).
- Popper, K. R. (1945). *The open society and its enemies*. Routledge, London.

- ProQuest (2012). Ten year embargo trends. In *CGS Summer Workshop and New Deans Institute, Boston, Massachusetts, July 7 – July 11, 2012*.
- Ramirez, M. L., Dalton, J. T., McMillan, G., Read, M., and Seamans, N. H. (2013 forthcoming). Do open access electronic theses and dissertations diminish publishing opportunities in the social sciences and humanities? *College & Research Libraries*.
- Schöpfel, J. and Lipinski, T. A. (2012). Legal aspects of grey literature. *The Grey Journal*, 8(3), 137-153.
- Smith, P. (2009). ETDs, IRs, Collaborations, and Access at Florida State University. In *ETD 2009 12th International Symposium on Electronic Theses and Dissertations. June 10-13, 2009, Pittsburgh, Pennsylvania*.
- Thompson, D. F. (1999). Democratic secrecy. *Political Science Quarterly*, 114(2), 181-193.
- Willinsky, J. (2005). *The Access Principle: The Case for Open Access to Research and Scholarship (Digital Libraries and Electronic Publishing)*. The MIT Press, Cambridge MA.

The paper was prepared as a contribution to the research seminar on “Secrecy, Publicity, Publishing – A Question for Information and Communication Sciences” organized at the University of Lille 3 by the GERiiCO laboratory in 2012-2013. Acknowledgement to Julien Baudry, Sandrine Berthier, Sara Gould, Isabelle Le Bescond, Marta Lee-Perriard and Jean-François Lutz for their help and advice.