

EVALUATION OF OPTIONS FOR A UK ELECTRONIC THESIS SERVICE

Study report

Prepared by
Key Perspectives Ltd
and
UCL Library Services

CONTENTS

	Page
Executive summary	
1. Introduction and background to the study	1
2. Methodologies	3
2.1 Methodologies employed in the study	3
2.2 Stakeholders and services consulted or examined	4
2.3 Issues examined	5
3. Overview of electronic thesis provision in an international context	6
4. The EThOS model	14
4.1 Technical aspects of the EThOS model	14
4.1.1 User interface	14
4.1.2 Metadata	14
4.1.3 Digitisation and location of theses	15
4.1.4 Preservation of theses	16
4.2 Organisational aspects of the EThOS model	16
4.2.1 The hub	16
4.2.2 Metadata records	17
4.3 Legal aspects of the EThOS model	17
4.3.1 Retrodigitisation	18
4.3.2 Third party rights	18
4.3.3 Deposit licence	18
4.3.4 Embargoes	18
4.3.5 Use and access issues	19
4.4 Business aspects of the EThOS model	19
4.4.1 Participation levels	19
4.4.2 Added-value services for users	20
4.4.3 Usage statistics	20
4.4.4 The EThOS Institutional Toolkit	20
5. Possible models for a sustainable national electronic thesis service	21
5.1 Business model types	21
5.2 Advantages and disadvantages of the business models	22
5.3 Organisational model types	24
5.4 Advantages and disadvantages of the organisational models	24
6. The UK HE community views on a national electronic thesis Service and the EThOS model	26
6.1 Institutional policies and practices	26
6.2 The EThOS business model	28

6.3	The organisational model for a national thesis service	30
6.4	The legal framework	31
6.5	Preservation arrangements	32
6.6	HE community views on associated issues	32
6.6.1	Masters dissertations	32
6.6.2	Rights clearance services	33
6.6.3	Plagiarism detection services	33
6.6.4	Metadata enhancement	33
6.6.5	Complex objects	33
6.7	General summary	34
7.	Risks, benefits and opportunities for a service based on EThOS	37
7.1	Risks	37
7.2	Opportunities	39
7.3	Benefits to UK institutions of EThOS working with other international electronic thesis organisations	39
7.4	Commercial opportunities that might arise from an EThOS Service	40
8.	Recommendations	41
8.1	Project-to-service issues	41
8.2	Other issues, including core business factors	42

EXECUTIVE SUMMARY

This report presents the findings from a study of EThOS (Electronic Theses Online Service) as a model for a UK national electronic theses (e-theses) service. EThOS, under the auspices of the British Library, was chosen by competitive tender to develop a plan for such a service that would operate on the open access principle of providing electronic information free to the user. The current service operated by the British Library (BL) charges users for theses and these are supplied on microfilm.

The main characteristics of national e-theses services in operation elsewhere in the world were identified by desk research and by consultation with managers of those services. A table showing these comparative characteristics is presented in the report. The views of the UK higher education (HE) community were sought on a range of issues relevant to electronic thesis (e-thesis) provision. Interviews, focus groups and an online survey were used for this purpose. The EThOS model was then mapped against the views and requirements expressed by that community.

The essence of the EThOS model consists of a central hub, located at the British Library, where electronic copies of theses would be held. Electronic theses may also be held locally in institutions. Where theses are born-digital they would be harvested from institutional repositories or pointed at in their original location. Digitisation of existing hardcopy theses would take place upon receiving a request for the thesis from a would-be user. Further bulk retrodigitisation of theses selected by institutions forms a major plank of the proposed service. EThOS will provide an Institutional Toolkit to inform and advise institutions on policy, rights and other issues around the collection, curation and provision of electronic theses.

Institutions may opt to participate in the service or not. For those that do there will be a choice of participation modes, including the option of pay EThOS to digitise theses on demand or of paying a fixed sum annually to EThOS to cover the digitisation of a guaranteed number of theses ('Sponsoring' membership). The service will be based on an 'opt-out' principle, whereby theses will be digitised upon request unless specified otherwise.

In broad terms these standards and arrangements fit well with the views and requirements of the UK higher education (HE) community.

The assumptions on which EThOS has based the designing of its model are generally robust. Mostly, institutions do not wish to digitise their own theses and operate a locally-based service but they do want to store locally electronic copies of theses that have been digitised in response to a request from a user, for which EThOS provides. There is a good match between the business model proposed by EThOS, too: institutions like the fact that there are various levels of participation and the sponsorship option, where a fixed annual sum is paid to EThOS, is acceptable in principle, though price points have not been objectively tested in the marketplace. The EThOS guarantee on the expert curation and preservation of digital copies of theses is also well received. Areas where the match between community requirements and the EThOS offering is not so good are those of rights clearance, third party rights issues and plagiarism detection. These issues need further attention if the model is to be fully acceptable to the HE community.

A number of risks have been identified. The main ones are:

- That not enough institutions participate as 'sponsoring' members, paying an annual fee for mass retrodigitisation
- That the 'gap year' between the development of the model and the launch of the service, and the lag time for institutions to formally sign up, will detrimentally affect business planning
- That institutions initially signed up as sponsors switch to associate membership (no annual fee; digitisation on demand) in future years

The main benefits are:

- Hugely increased visibility of UK doctoral research output
- Resulting increased usage and impact of UK doctoral research output
- The opportunities for resulting new research efforts and collaborations

The main opportunities are:

- Being able to provide a world-class electronic theses service to showcase the UK's doctoral research
- Providing an example of good practice and the impetus for other nations to develop electronic theses services of their own
- Possible commercial opportunities for value-added service providers

The report makes a series of sixteen recommendations for stakeholders, the main ones of which are:

- Stakeholders should require further testing of the prospective EThOS pricing regime to assess its acceptance in the marketplace

- Stakeholders should satisfy themselves that planning for the management of the 'gap year' and institutional take-up rates is robust
- Marketing and advocacy activities should be factored into the plans (and costed)
- EThOS should indicate how it will manage institutional expectations with respect to rights issues
- Indications from EThOS of the shape of the long-term business model, when almost all new theses are born digital, should be required
- Short-to-medium term planning for different levels of business should be required: currently EThOS has a business plan based upon a certain level of activity but no accompanying plans that show how the business would proceed if take-up is higher or lower than predicted
- The stakeholders should themselves plan for a programme of active advocacy within the UK HE community at graduate school and registrar level to raise awareness of the benefits of open access e-thesis provision and to advise institutions on policies that will effect this provision

Alma Swan
Key Perspectives Ltd
August 2006

1. INTRODUCTION AND BACKGROUND TO THE STUDY

The British Library (BL), JISC, UK HE institutions and CURL have funded an 18-month project to develop a national framework for the provision, preservation and open access to electronic theses produced in UK HE institutions. The project, called EThOS (Electronic Theses Online Service) was developed in response to a competitive tender invitation released by the JISC and proposes a service set up and run by the British Library.

The British Library's current service, the *British Thesis Service*, offers access to around 180,000 doctoral theses, predominantly from 1970 onwards, though it is estimated that overall some half million theses dating from the 1600s are in existence in the UK. Around 80% of requests are for theses published within the last 13 years and almost all of these exist only in hardcopy. Through this service, theses are acquired 'on demand' and delivered on microfilm at a cost of just over £60 to the user (and at this price the service runs at a loss). Whilst this service, coupled with the *Index to Theses* (Expert Information), enables the location of and access to relatively recent British theses by the determined seeker, no one could argue that the process is optimised. As a result, usage of theses is much lower than it might be and much research is going unnoticed and unused as a result. Conversely, it has been shown that when theses are easy to locate and access, usage is high: at Virginia Tech, a pioneer site in the provision of a formal, systematised ETD (electronic theses and dissertations) service, downloads have been shown to increase over 30-fold when a thesis is available free online and easily located.

A national service for the UK that provides discovery and access to theses in electronic form via the Web will increase the utility of doctoral scholarship. A single interface that directs users to theses wherever they are held, and which addresses the issues of intellectual property, permissions, royalties, preservation, discovery, and other matters associated with the public provision of theses in electronic form, will be of great benefit to the scholarly community in the UK and across the world.

The EThOS project (Electronic Theses Online Service) was commissioned to develop a model for a workable, sustainable and acceptable national service for the provision of open access to electronic doctoral theses. The EThOS project team have completed the task and UCL Library Services in partnership with Key Perspectives Ltd have been asked to undertake a consultative study to assess the acceptability of the proposed model to the UK higher education community in the context of other potential models.

This document reports the results of this consultative study, including a set of recommendations to JISC and other stakeholders for setting up a UK national e-theses service. The stakeholders other than JISC are:

- The British Library
- University administrators (registrars)
- Graduate students and recent PhDs
- Librarians
- Institutional repository managers
- Other e-theses services including:
 - DART-Europe
 - DiVA
 - DissOnline
 - Australasian Digital Theses
 - Theses Canada
 - Networked Digital Library for Theses and Dissertations
- The EThOS team

2. METHODOLOGIES

2.1 Methodologies employed in the study

A set of methodologies was used during the course of this study:

Desk research:

Desk research was the method used to investigate and gather information in the following areas:

- Examination and evaluation of the range of electronic thesis delivery services available worldwide, including EThOS' own information
- Background information on previous studies concerned with e-theses
- Background information on IPR and copyright issues concerned with e-theses
- General context-setting/orienting study

Focus groups:

Two focus groups were held, one in Edinburgh and one in London, attended by librarians responsible for e-thesis stewardship in their institutions or, in the case where there is not yet any formal framework for handling electronic versions of theses, librarians responsible for theses and dissertations in general. The sessions worked to a pre-set agenda, with a formal set of issues for discussion, reviewed and agreed with the JISC's oversight panel prior to the events.

Online survey:

An online survey was developed after the focus groups took place, the latter having helped to distil out some additional issues for examination. The survey was run as a JISC consultative exercise and invitations to respond were disseminated via discussion lists to the following groups of stakeholders:

- Librarians
- Repository managers
- Registrars
- Graduate school administrators
- Graduate students and recent PhDs

Eighty responses were received in total. The majority of these were from librarians (40%) or library directors (41%). Repository managers (some of whom may also have counted themselves as librarians as they play a dual role in many cases) represented 4% of respondents, heads of graduate schools 1% and students 9%. We disaggregated these responses in our

analysis so that the views of the constituent groups could be taken into account separately.

Personal interviews:

Interviews were conducted with some (four) individuals on a one-to-one basis in person or by telephone wherever opportunities arose for this kind of interaction. The views of these people also inform our analysis at the appropriate places.

It is important to note that the large majority of the people whose views we solicited during the course of this study were librarians. Where we refer in this report to the opinions of the HE community we are, therefore, reflecting largely the opinions of this group. Though that caveat should be borne in mind by readers, we can also say that a number of registrars' opinions were also canvassed and that these did not differ significantly from those of the librarians.

2.2 Stakeholders and services consulted or examined

The following stakeholders and services were consulted or studied in the course of the work. The first figure in the brackets is the number of people in each category who responded to the survey; the second figure is the number who attended the EThOS workshop in Glasgow on July 20th):

- EThOS project team
- Academic registrars (0; 1)
- University administrators (graduate schools) (1; 3)
- Librarians (68; 28)
- Repository managers (3; 2)
- Authors (or potential authors) of theses and dissertations (10; 3)
- Networked Digital library for Theses and Dissertations (NDLTD)
- DiVA (Digitala Vetenskapliga Arkivet) Academic Archive Online
- DART-Europe
- DissOnline (Germany)
- Promise of Science (Netherlands)
- Australasian Digital Theses
- Theses Canada
- Charles Oppenheim: response on IPR issues
- ProQuest/UMI

2.3 Issues examined

The issues examined in the focus groups, interviews and online survey for the purpose of comparing services and assessing EThOS' relative strengths and weaknesses were:

- Electronic thesis provision status in the UK and the reasons for its slow development
- Drivers for change in the provision of e-theses
- The administrative and academic contexts in which a national UK e-thesis service would need to operate
- Constraints that might apply, or which have applied until the present
- Architectures and service models for e-thesis provision
- Technical standards
- IPR and other rights issues
- Business models

3. OVERVIEW OF ELECTRONIC THESIS PROVISION IN AN INTERNATIONAL CONTEXT

Several countries or consortia have developed electronic theses services over the last few years. The following list highlights the main elements of the various theses services we have examined. These elements may be dealt with differently by the different theses services:

- The hub: the central focus of the service may offer multiple resources and subservices or, at the other end of the scale, may be a simple resource discovery service: The hub may point at theses located in host institutions or it may contain the full-text of theses itself
- Submission procedure: the simplicity of and requirements for this can vary
- Metadata structure and format: the required metadata formats can vary from very simple Dublin Core with few elements to a deeply descriptive specially developed metadata scheme with many elements
- Metadata dissemination: services vary in the extent to which they disseminate thesis metadata – some only expose it themselves, while others disseminate it via multiple discovery services and routes
- Accepted file formats: some services accept multiple file formats, some few and some just one
- Digitisation: a digitisation service may be part of the offering. If it is, it may be on demand or there may be a mass retrodigitisation programme on offer
- Thesis level: services may offer only doctoral theses or may extend their coverage to masters theses and even undergraduate dissertations
- Copyright and IPR: services may incorporate advice and practical help on rights issues
- Plagiarism: services may offer a plagiarism detection scheme
- Business model: under this heading fall issues such as: whether theses are offered on a pure Open Access basis or whether the access is paid for; whether royalties are paid to authors; how digitisation costs are covered and so forth

The simplest way to present an overview of these services is in tabular form and this we do in Table 1. The main services in operation, which have been surveyed, are:

ADT – Australasian Digital Theses

DiVA – (Digitala Vetenskapliga Arkivet) Academic Archive Online

Theses Canada

NDLTD – Networked Digital Library of Theses and Dissertations
ProQuest/UMI
DissOnline
DART-Europe (in development)

Significant moves have been made in various countries on certain fronts. For example, Canada now has a mandate on the deposition of electronic theses nationwide, resulting in a rapidly-accumulating body of e-theses. In the US, the federal government has mandated ProQuest as the national archive for doctoral works and ProQuest is contracted by the Library of Congress to run the national doctoral theses service, accumulating some 55,000 doctoral theses per year. Students pay ProQuest around \$60 to 'publish' their thesis and each time it is downloaded from *Digital Dissertations* ProQuest pays the author a royalty of around \$6. ProQuest's rights are non-exclusive, however, so many institutions work also with NDLTD in support of open access: for example, OhioLink supports all the institutions in Ohio with a state service for ETDs.

It is worth noting that mandates on submission may not extend to mandatory Open Access to theses. Temporary embargoes, and permanent closed access to parts of electronic theses, are frequently permitted where IPR concerns prevail.

Provider	ADT	DiVA	Theses Canada	NDLTD	ProQuest	DissOnline	DART-Europe	ETHOS
Coverage	Originally 7 CAUL member libraries: now open to all Australian and NZ universities	16 universities in Sweden and Norway; participation in DiVA open to all universities	60 universities in Canada	Any voluntarily participating institutions	Any participating institution	35 voluntarily participating German universities	Pan-European service for voluntarily participating national/European consortia	Any voluntarily participating UK HE institution (geographical limitation defined by the project, not the development)
Hub	ADT	DiVA Portal www.diva-portal.org	Theses Canada Portal	www.theses.org or www.ndltd.org	<i>ProQuest Digital Dissertations</i>	DissOnline is a coordinating body	DART-Europe repository and portal (DEEP) launching to community early summer 2007	ETHOS (at British Library)
Submission procedure	At host institution	By author at host institution in the case of institutional DiVA. Form or template supplied	To ProQuest for digitisation, or universities with ETDs can upload or provide metadata. Also harvests directly from universities with ETDs	At host institutions. NDLTD provides recommendations for interface, and ETD-db software as well as DSpace etc are free	At accredited host institutions via ProQuest form interface	At host institution	At host institution	Born-digital theses submitted at host institutions; digitisation service and central submission by ETHOS for hardcopy & born-digital theses
Metadata format	Details coming soon	99 metadata elements allowing transformation into other standard formats. Metadata enhancement is part of	MARC 21 ETD-ms Dublin Core	ETD-MS (ETD metadata standards). Crosswalking (compatibility) possible between these and MARC	ProQuest standard metadata set	MetaDiss (developed by Die Deutsche Bibliothek)	Developing a basic set of DART-Europe metadata standards consisting of 8 simple DC elements	ETHOS set of c15 qualified DC elements

		service						
--	--	---------	--	--	--	--	--	--

Table 1: National/international electronic theses services: summary characteristics

Provider	ADT	DiVA	Theses Canada	NDLTD	ProQuest	DissOnline	DART-Europe	ETHOS
Metadata dissemination	Via website	<ul style="list-style-type: none"> - LIBRIS (national union catalogue – doctoral theses only) - Uppsök (Swedish undergraduate theses only) - OAI-compliance - RSS channels - XML export - Scirus search - NDLTD union catalogue and other services based on OAI-harvest information 	<ul style="list-style-type: none"> - Theses Canada Portal - AMICUS (national union catalogue) - Also harvested for NDLTD union catalogue 	Multiple: web search engines, union catalogues, various national libraries, ProQuest UMI and other third party services	UMI dissertations database via website, OCLC, Dialog, STN and other sites	Via Die Deutsche Bibliothek, ThEO, OPUS and others	OAI-compliance Email alerts XML export	Via discovery service, export to appropriate services, OAI-compliance
Thesis location	At host institution	<ul style="list-style-type: none"> - At DiVA archive - National Library of Sweden archive - At host institutions 	Theses 98-02 and harvested ones are on LAC* server. ProQuest has ETDs from 2002 to present. Some universities retain their	Host institution	At ProQuest	At host institution. Metadata registered with DissOnline	In host institutional repository and, if requested by partner, in DART-Europe repository	At hub harvested born-digital and digitised theses. At host institution if the institution wishes

			own theses					
--	--	--	------------	--	--	--	--	--

Table 1: National/international electronic theses services: summary characteristics

Provider	ADT	DiVA	Theses Canada	NDLTD	ProQuest	DissOnline	DART-Europe	ETHOS
Preservation	Copy at host institution	Copy at National Library 'or other longterm archive'	LAC has microform copies of all theses and a commitment to preservation of electronic ones on its server	Advice available; experiments underway involving LOCKSS	Multiple digital copies. Two microfilm copies kept separately in climate-controlled vaults	Locally	Under development	By BL in keeping with current standards and procedures for emulation and/or migration.
User interface	Simple search form	Portal interface – Swedish, Danish and English - structured searching - free-text searching - full-text searching (using Apache Lucene) - full-text search also in cooperation with Scirus	Theses Canada Portal, AMICUS, ProQuest	www.ndltd.org under search and browse – multiple services	Professional interface	Offers choice of routes into thesis databases: not a simple operation to search	Multi-lingual support Basic search Advanced search	Re-use and development of existing, well-established live systems.

Table 1: National/international electronic theses services: summary characteristics

Provider	ADT	DiVA	Theses Canada	NDLTD	ProQuest	DissOnline	DART-Europe	EThOS
Funded by	Australian Research Council (at least initially)	Universities' consortium (of the participating universities)	LAC funds all costs except ProQuest's publishing fee funded jointly by Theses Canada and participating universities*	Membership dues and membership in-kind contributions	Commercial sales	Die Deutsche Bibliothek	Business model currently being discussed	Initial BL subsidy, contributions from other bodies (if available) and participating institutions, via choice of options
Digitisation provision by service?	At deposit. Also digitisation of requested theses	No	By ProQuest	No	Yes, in return for a publishing fee paid by the author		Yes, probably mass retrodigitisation if funding is secured	Yes – included in 'Funded by'
Number of theses	N/A	11,500 theses, 500 other publications (research reports, monographs)	c50,000 ETDs c250,000 TDs (including ETDs)	Over 250,000	2.3 million (>1 million available in full-text)	c40,000 (includes Masters theses)		Estimated 25,000 digitisations per annum
Masters theses	Yes?	Yes	Yes	Yes	Yes	Yes	Yes?	Limited by project scope – could be included
Open Access	Yes	Yes	1998-2002, and to ETDs acquired through the harvesting programme	Yes	Not via commercial service but OA option now available for publishing fee of \$150, "the lowest OA publishing fee on the market!".	Yes	Yes	Yes

Table 1: National/international electronic theses services: summary characteristics

Provider	ADT	DiVA	Theses Canada	NDLTD	ProQuest	DissOnline	DART-Europe	EThOS
Rights advice	On portal	On portal portal and, in the institutional instances, theses also checked for third party copyright violation by services used at particular member institution. Additionally DiVA is deep indexed by those services.	On Theses Canada Portal	Yes	Yes, and each thesis checked for third party copyright violation by ProQuest	On portal	Yes	Yes, in institutional toolkit and on portal

Table 1: National/international electronic theses services: summary characteristics

Provider	ADT	DiVA	Theses Canada	NDLTD	ProQuest	DissOnline	DART-Europe	ETHOS
Added-value services	Usage statistics	Print on demand service		Usage statistics, online guidance to students / faculty / institutions, annual conference for discussions, union catalogue	Plagiarism checks run against web content, journal articles and other theses and dissertations. Libraries can purchase microfilm copies.	No	Under review for most services. Will provide hosted repository for institutions that do not have their own	Will sell hardcopy (softbound, hardbound, loose leaf) and DVD/CD versions at cost. Usage statistics within core price ('Funded by'). Plagiarism check service under discussion
Royalties to authors	No	No	No	No	Yes. Approx \$6 per sale (download)	No	Yes?	No

Table 1: National/international electronic theses services: summary characteristics

* Library and Archives Canada

** ProQuest sells the thesis but uses royalties to reduce the publishing fee paid by the universities/Theses Canada. Theses Canada pays one-third (currently \$16) and the university pays the rest (\$39) of ProQuest's publishing fee

4. THE ETHOS MODEL

This section describes the model which has been developed over 15 months by the EThOS team for a national e-theses service.

4.1 Technical aspects of the EThOS model

4.1.1 User interface

Users will locate digitised theses via the discovery service provided by the BL, which provides a Google-like interface to enable searching of the metadata and full-text of digitised or electronic theses. Users will be able to limit their search to digitally-available theses if they wish. It is hoped, in addition, to provide the abstracts for theses without digital full-text but this depends upon the establishment of an acceptable arrangement with the Index to Theses.

Searches return title, author, year and institution and users click on the title to obtain access to the full details or the order screen. This gives the full bibliographic data, a scrollable abstract, the source format (options are: paper only at institution; microfilm held at the BL; electronic version available). The user is told how much it will cost to obtain the thesis and how long it will take to supply from the central EThOS hub, i.e. a total price and a turnaround time for delivery of the product. If the thesis is readily available electronically, the user is provided with a link. The user has by this time provided an email address for ID and must now sign a licence agreement and commit to the order. EThOS authorises the credit card payment at this point but takes no payment at the time. When the thesis is ready for delivery EThOS reminds the user of the IPR agreement and takes the payment.

4.1.2 Metadata

The qualified Dublin Core metadata scheme proposed for EThOS is based on the Robert Gordon metadata set. This has a small number of mandatory fields. The EThOS metadata set has been developed in consultation with NDLTD and tested by the Repository Bridge project at the National library of Wales in Aberystwyth. The standards adopted by EThOS for metadata will allow simple cross-searching with other theses services based around similar standards.

Digital theses' metadata residing in institutional repositories will be harvested by EThOS to create a resource discovery service that effectively cross-searches all UK HE institutional holdings. Some centralised services

will be harvested by EThOS too to collect the maximum amount of metadata: those already proposed for harvesting are the Repository Bridge facility for Welsh theses and IRIScotland for theses from Scottish institutions.

Metadata from institutions will be loaded to the BL service and transferred to the central BL system, where they will be enhanced and authority-checked before being returned to EThOS. Given the fact that the abstract and full-text are also searchable, EThOS aims to have “the fullest available description of a thesis anywhere”.

EThOS will need to negotiate with institutions with respect to metadata harvesting. This should be an entirely simple matter, since institutions creating metadata and placing them in open access repositories clearly wish the use of the metadata – including harvesting – by other parties to occur. The formal situation regarding metadata is that IPR rights in metadata, created by University Libraries or a National Library, are governed by the IPR policies and practices at the institution where the library staff are employed. Best practice is for the institution to have an explicit IPR policy covering IPR in materials produced by academic and administrative members of staff (see, for example, the policy from UCL at <http://www.ucl.ac.uk/Library/scholarly-communication/ipr.shtml>). If an institution has no such policy, it is recommended that IPR rights in metadata be addressed as part of any new institutional IPR policy. Unless there is an agreement to the contrary, IPR will reside with the institution at which the member of staff, who creates the metadata, is employed.

4.1.3 Digitisation and location of theses

For theses that are in hardcopy only, EThOS will provide a digitisation facility, scanning textual materials at 300 dpi and photographs and graphics at a higher resolution. Theses that need to be scanned from microfilm will not be as high quality as those scanned from paper hardcopy due to the quality of the source material. Once digitised, the full-text electronic file will be held at EThOS and a digital copy supplied back to the institution if required.

Digitisation will operate in two modes:

- on demand, as a thesis is requested by a user
- on selection by host institutions participating in the membership option for EThOS (see section on the EThOS business model below)

The BL currently supplies around 12,500 theses on microfilm annually. The EThOS team estimate that demand will double when electronic theses are

available via the new service. This is not an unreasonable assumption given the increased usage (over 30-fold) of theses in electronic form reported by Virginia Tech, for example. Initially, digitisation will be required for the vast majority of these theses, since they will exist only in hardcopy at the HE institutions. As time passes and universities begin to mandate the submission of digital copies of theses by doctoral students, the proportion of born-digital theses will rise. Eventually the need for a large-scale digitisation programme will diminish (the EThOS team estimates ten years) and digitisation will be required only for fulfilling the occasional request for older theses.

4.1.4 Preservation of theses

EThOS will adopt BL preservation practices and standards for theses in the service. The British Library has ongoing work on preservation standards and procedures for all digital material in collaboration with other similar organizations worldwide. Theses digitized by EThOS or deposited with the service will be backed-up appropriately and they will be actively stored in the British Library's DOM (Digital Object Management) system. This will ensure a preservation watch function, and that migration to new formats or emulation of current formats within future operating systems and applications will be carried out.

4.2 Organisational aspects of the EThOS model

4.2.1 The hub

At the centre of the proposed EThOS service is a hub managed by and located at the British Library (BL). The BL will coordinate the whole service and act as the central facilitator for all HE institutions wishing to participate. Each participating institution will have its own arrangements with EThOS with respect to its level of involvement and the appropriate business terms on which it participates.

Institutions will be offered three options for participation:

- i) Supplying thesis metadata but retaining the full-text of the thesis and associated digital objects locally in the institutional repository.
- ii) Participating by providing theses on demand as users request them and paying for their digitisation in one-off fashion ('associate' membership).
- iii) Becoming a 'sponsoring' participant, paying a lump sum per annum upfront in return for which a certain number of theses will be digitised in that year.

Further details of these options are given in the section on the EThOS business model below.

All electronic theses digitised and then stored by EThOS will reside on a single server at the BL. They will be available to users on Open Access terms, with users paying only for the cost of provision and delivery if they opt for hardcopy or CD/DVD formats.

4.2.2 Metadata records

Negotiations are underway with the *Index to Theses* for the supply of metadata from that database to the BL service. There are currently details of around half a million theses in that index, though this figure includes Masters theses. The BL itself holds metadata for 200,000 theses in its own microfilm holdings. Institutions are to be encouraged to supply metadata for their own hardcopy thesis holdings if possible, too, to provide the best possible locating service through the BL for theses. Any deduplication of these metadata records will be undertaken at the British Library, which has extensive experience in deduplicating MARC 21 and other record formats, for example when rationalising its collections into its integrated library system.

4.3 Legal aspects of the EThOS model

The EThOS team has undertaken a study of the legal issues involved in the provision of an electronic thesis service in the UK. A report was produced by Dr Theo Andrew at Edinburgh University after review by Professor Charles Oppenheim at Loughborough University, recognised as a leading expert on copyright matters in scholarly communication. EThOS has adopted a position on each of the main legal issues, as discussed below.

4.3.1 Retrodigitisation

In general it is accepted in the UK that the ownership of copyright in the whole work remains with the author of the thesis. There are moves now, though, for HE institutions to claim copyright on theses: this is certainly the case at Cranfield University, for example, and at Strathclyde students must actively claim copyright or it remains with the institution. Nevertheless, in most cases the author owns the copyright and permission to copy or disseminate a thesis must be sought from that person. This puts electronic theses services in a dilemma because the chance of contacting the authors of theses to secure this permission is not high even for recent theses and is extremely low for older ones. The solution favoured by EThOS is an 'opt-out' formula, where authors are assumed to have implicitly given permission unless they explicitly opt out of the service and refuse permission for their

thesis to be part of it. It has been recommended that a Risk Assessment be carried out by the EThOS project team to determine the level of risk from authors in this regard. A formula has been provided by Professor Oppenheim for this purpose.

Where institutions own copyright the situation is simpler, since permission may be sought and secured via a service agreement between the institution and the BL.

4.3.2 Third party rights

Where theses contain material whose rights are owned by third parties permissions should be sought from these parties. For born-digital theses clearance of third party rights should become part of the author's normal thesis-preparation activities, but for older theses being scanned retrospectively a more pragmatic solution needs to be found. In the current BL provision of theses on microfilm, third party material is removed (by unspecialised staff) before the thesis is provided to the requester. The risk for a new electronic theses service is considered to be relatively low and the EThOS solution is to proceed without permission but to put in place a rapid take-down policy which would be implemented in the case of a complaint, and to secure insurance cover in case of legal action.

4.3.3 Deposit licence

EThOS intends to use a deposit licence to capture permissions electronically and to display these permissions alongside the theses in the collection.

4.3.4 Embargoes

EThOS offers 2 levels of embargo:

- 1) the institution does not have to supply the thesis at all – it simply makes the thesis unavailable for harvesting. It can be made available for harvesting later, if circumstances change.
- 2) for theses which have a limited embargo period, an embargo date can be entered on the record. In this case, the thesis is harvested as usual but the system does not allow it to become available until after the embargo date.

4.3.5 Use and access issues

Rights metadata will be displayed alongside each thesis. EThOS will provide legal notices on the front page of the website, in pop-up check boxes at download and on coversheets attached to the theses themselves (and other digital objects if appropriate). The institutional toolkit will also provide institutions with information on legal issues concerned with electronic thesis provision and what institutions should do to protect themselves as far as possible from adverse action.

4.4 Business aspects of the EThOS model

4.4.1 Participation levels

EThOS will offer three levels of participation for HE institutions as listed in section 4.2 above.

i) The simplest option is for institutions to hold their own digital theses in their institutional repository, to allow EThOS to harvest the metadata for the EThOS service and for EThOS simply to point prospective users to the institution for access to the thesis required. Presupposing that the institution is willing and able to supply digital theses on a true Open Access basis, no cash changes hands and EThOS's role is as a resource discovery service provider. Institutions may, however, wish to charge for their theses, i.e acting as a non-Open Access supplier, under their own specific arrangements with users.

For institutions that do not have a repository, or that have a repository whose permanence may not be guaranteed, or that have not accumulated a substantial corpus of digitised theses, EThOS offers two further options:

ii) 'Associate' institutions can have theses digitised on a thesis-by-thesis basis as users request the documents. There will be a one-off digitisation fee levied on the institution each time a thesis is digitised. The digital thesis is deposited on the EThOS server and, if the institution requires it, a copy is provided to the institution for its repository.

iii) 'Sponsor' institutions will pay a fixed sum to EThOS once a year for three years and will cover the digitisation of a guaranteed number of theses each year. These will include the theses requested by users; if these do not account for the total number guaranteed to be digitised in any one year, the number will be made up by theses selected by the institution for digitisation. This option will appeal to larger institutions awarding substantial numbers of doctorates annually. The EThOS hub would gain content consisting of the most popular theses and institutions would gain digital copies of these at cost price.

The EThOS development team has calculated that to break even with 50 participating Sponsor institutions each paying a flat rate (this may not be the optimal final formula: there may be a good case for scaling or banding the level of Sponsor payments according to size of institution or other suitable metrics), each would need to pay a sum of around £6,000 per annum. At this level, the service would be able cover its costs. Profit is not factored into this calculation, but any excess revenue would be used to extend the retrodigitisation process to maximise the number of digital theses available. This price point has not been formally tested in the

marketplace, nor has there been any testing of any possible differential pricing, based on the JISC banding scheme for example).

4.4.2 Added-value services for users

As well as providing access to electronic theses, EThOS will provide users with additional service features:

- Print on demand, with the thesis delivered in loose-leaf, softbound or hardbound formats
- Electronic copies on CD or DVD

Users will be charged for these on a cost-recovery basis.

4.4.3 Usage statistics

EThoS will record the entire history of thesis usage from when the thesis is first ordered (or loaded onto the system if born-digital). Items recorded will include the time a thesis is ordered, digitised and supplied. Customers will register with the service so records will be kept of who ordered and when the thesis was delivered to them.

In addition, each year participating institutions will be supplied with a report of the usage of their theses as tracked via the EThOS portal including the number of downloads and when these occurred.

4.4.4 The EThOS Institutional Toolkit

EThOS has prepared a toolkit for institutions. This provides full explanatory, technical and advisory information for participating institutions.

5. POSSIBLE MODELS FOR A SUSTAINABLE NATIONAL ELECTRONIC THESES SERVICE

The model developed by EThOS is not the only possible model for a service that provides the UK with an electronic thesis service. Any national service has to be viable and sustainable and to take into account the ways in which institutions within that nation work. The costs involved are fairly substantial and lie in these main areas:

- Development and operation of a resource discovery service for a nation's theses
- Infrastructure for providing access to theses electronically online and for providing theses in other formats if required
- Mechanisms for authorising and monitoring accesses, billing customers or suppliers (depending on the model employed) and providing usage figures
- Providing mechanisms to ensure adherence to legal requirements
- Digitisation of theses existing only in hardcopy. The latter is an expensive element if retrodigitisation of theses on any kind of a mass scale is to be provided
- Curation and preservation of electronic theses

5.1 Business model types

There are three main business models:

Commercial profit model

Under this sales-based model the service would sell access to theses. This would not be an open access service, of course. The service would enable users to locate theses of interest, sell access to them, gather the revenue and (probably) pay royalties to the thesis supplier(s). An example is the ProQuest/UMI service which works well in its home base of North America and extends to elsewhere in the world.

Cost-recovery model: supplier recharge

Here the costs of running the operation are recouped by recharging them to a party that values the service for intrinsic reasons. Since cost-recovery models need to operate on a stable, predictable basis, this usually means charging suppliers. Suppliers are willing to pay because they see some sort of non-cash payoff in providing the service. An open access service is operable under this model. The EThOS model falls into this category.

Cost recovery model: subsidised

An alternative to recharging costs to suppliers is to invoke sponsorship to cover the costs. Subsidy may be in the form of donation by a philanthropic third party or it may be in the form of selling some added-value additional services to provide the cash to cover the core service costs. This model could provide an open access service. An example of this model is the proposed DART-Europe service, where ProQuest has been asked to develop plans to sell add-on and added-value services whose revenue will cover the cost of the core open access portal and service.

Sponsored model

The final generic model is the sponsored one, where costs are covered by cash from public funds (because a service is deemed to be for the public good) or from philanthropic sources. Such a service can work perfectly on open access principles so long as the right level of subsidies can be secured. The closest example of this type of service is the Scandinavian DiVA service, which operates entirely on funds contributed by participating universities which view the service as part of their core mission to provide free access to their research output. EThOS will also investigate opportunities for funding from philanthropic sources in association with Supplier Recharge.

5.2 Advantages and disadvantages of the business models

The models have to be assessed for suitability on the basis of their viability, sustainability and acceptability to the UK HE community.

There is a deep level of commitment to the open access model with regard to thesis provision in the UK. In the survey carried out for this study, only one respondent favoured charging for access to theses: all 79 others expressed a preference for free access for users, as did those individuals who participated in the focus groups. The commercial sales model is thus ruled out because of the charge to users that it rests upon. The other three options are compatible with an open access philosophy.

These models are shown in the table below, assessed for viability and sustainability.

	Sales-based model (commercial)	Cost recovery model	Subsidised model	Sponsored model
Match with JISC IE open standards requirements	No	Yes	Yes	Yes
Viability	Probably very good, but discounted due to mission mismatch	Medium. Heavily dependent upon voluntary participation level by HE institutions	Good. Potentially the easiest to plan and run so long as sellable services can be identified and these bring in the required level of revenue. To cover mass retrodigitisation, this is a challenge as the costs are high	Good, because participating sponsors identify attainable goals and can 'cut their cloth according to their pocket'. This may impinge upon the quality of the service, however, if high costs necessitate limiting the offering
Sustainability	Probably very good, but discounted due to mission mismatch	Good in the long term, because with the increase in born-digital theses, digitisation costs will bottom out and the service needs only to cover day-to-day running costs	Good, if the business planners are adept at spotting market opportunities	Good, because participating sponsors identify attainable goals and operate within known constraints
Comments		Success is dependent upon community acceptance and participation but has the potential to deliver a full national service	Success is dependent upon revenue-gaining mechanisms but has the potential to deliver a full national service	Service may be more limited than those based on other models because of budgetary constraints on participants.

Table 2: Assessment of viability and sustainability of business models

5.3 Organisational model types

As for the business models, there is a range of types of organisational model for a national e-thesis service. These are summarised below.

Distributed provision model

Theses remain in full-text at the host institutions. These institutions provide access on an open access basis via an institutional repository. Discovery is through web search engines, OAI search engines and the like. Institutions are fully responsible for collecting, hosting, and preserving e-theses and for all management tasks associated with them (fulfilling legal requirements, etc). The ADT (Australasian Digital Theses) model operates on this distributed database system.

Centralised model

Institutions supply theses to a central service which makes them publicly available. Central services may accept only electronic theses or may offer an additional digitisation service to institutions. Institutions delegate management tasks to the central service. The North American ProQuest service is an example of this type, providing high levels of support to institutions around thesis submission and then delivering a fully managed e-thesis service to the public.

Mixed architecture models

These are where institutions and a central service work in partnership to provide a full service to users. There are numerous permutations and combinations possible, depending on where theses are hosted in full-text, where responsibilities lie for the various management tasks and who, if anyone, is responsible for organising the funding of the service. The EThOS model falls into this category.

5.4 Advantages and disadvantages of the organisational models

Again, it is clearest to present the issues in tabular form.

	Distributed model	Centralised model	Mixed architecture model
Viability	Dependent upon individual institutions' capabilities and resources, which are highly variable	Good, providing service provider selects correct business model and satisfies HEI concerns on rights, liabilities, etc)	Good, providing service provider selects correct business model and satisfies HEI concerns on rights, liabilities, etc)
Disadvantages	Dependent upon individual institutions' capabilities and resources, which are highly variable. This would lead to a service of patchy quality for at least a decade Potentially chaotic with respect to standards and consistency levels	HEIs lose control to an extent and may lose some benefits in terms of PR and other institutional-purpose benefits that accrue with local service provision	Offers potential for inconsistencies unless well-managed by hub provider
Advantages	Self-organising, cheap, simple	HEIs need only to provide access to e-theses: central service provider does the rest: Standards applied across the board: Guaranteed consistent access: Scope for added-value services: One interface; a true national collection as well as a national gateway: Easy to hook up to other national or international services.	Gives the greatest flexibility to HEIs to select the most appropriate options; HEIs can retain control of selected elements: Standards applied across the board: Guaranteed consistent access: Scope for added-value services: One interface (multiple sites of supply): National gateway: Easy to hook up to other national or international services.
HEI community views	Strong feeling against this option	Second most popular option	Highest level of support for this option
Comments	No support in the HEI community	Strong support within HEI community	Very strong support within HEI community

Table 3: Assessment of the organisational models

6. THE UK HE COMMUNITY VIEWS ON A NATIONAL ELECTRONIC THESIS SERVICE AND ON THE ETHOS MODEL

The overall utility and observable success of any electronic theses service will depend to a great extent on HE institutions in the UK changing their existing practices with respect to the stewardship of theses. Until very recently many if not most universities have simply mandated the deposition of a hardbound copy of every thesis in the library, left copyright with the author and have thereafter presided over stacks of under-used volumes. In the last few years a small number of institutions have encouraged – and some have mandated – the deposition of electronic copies of theses alongside the hardcopy. Virtually 100% of theses are now prepared by students in digital format and so cooperating with such conditions is not onerous on the part of the author. The result, however, for the institution issues concerned with the best policies for stewardship, access and rights management.

While the situation is not simple, there are good practice ways to deal with all of these issues. Nonetheless, for most HE institutions implementing an ETD system is new, not particularly well-understood, and presents challenges. For EThOS to construct and operate a successful national electronic theses service, institutional practices must change (or, perhaps, be changed by EThOS) and this is no small task. Encouragingly, it seems that the notion of a national service providing electronic theses is welcomed by the HE community and, moreover, institutions are actively encouraging the idea once they appreciate the *raison d'être* and the principles involved.

6.1 Institutional policies and practices

At the moment, however, over 80% of institutions do not yet collect and store electronic theses at all and only one or two institutions collect them from all doctoral students. More than half have no policies on electronic theses yet either, though 11% have developed one and a further thirty-four percent are planning theirs. On the basis of this information from the community it seems unlikely that the majority of theses will be submitted electronically within the next five years, though there are hopes that born-digital theses will be available ubiquitously beyond this time. In the short to medium term, therefore, EThOS' assumption that digitisation will be necessary in many cases seems reasonable.

The major driving force for the move towards formalising e-thesis collection is the increased accessibility that results, compared to hardcopy theses. HE institutions see electronic access to their theses as contributing to the overall visibility and impact of the institution and in keeping with the move towards open access in general. There is a strong commitment to open access within the HE library community: almost 100% of institutions would intend to make their theses available on an open access basis given the right framework within which to do so. It is JISC's wish, and thus the given framework within which EThOS was designed, that the HE community will want to work on an open access basis for e-thesis provision and these assumptions are borne out by the evidence from that community.

Coupled with this is a general increase in interest on the part of institutions in preserving their research output electronically and this is reflected in the fact that a growing number of UK HE institutions now have a digital repository that is suitable for housing e-theses (as well as other digital research output of the institution). At the same time, many institutions feel that they lack the knowledge, expertise or resources to provide their own e-thesis service and thus look favourably upon the notion of a workable service that can coordinate and provide access on a national basis. They do welcome the idea that EThOS would provide this coordination, expert advice and the infrastructural arrangements within which they can participate in a national thesis service.

There is another issue that was raised in the focus groups with respect to institutional practices and that is the relationship between the registry, academic departments/schools, supervisors and doctoral students. In some cases the relationships work well and all parties share a vision for thesis submission and dissemination. It seems that in rather more cases the relationships can be obstructive to this vision. Registrars may be the instigators of an open access system for theses, or they may act as a bottleneck. Supervisors may facilitate the speedy dissemination of the student's work, or may be ultimately responsible for a long embargo in the interests of future publications or patents.

On this topic, the information we have collected during the course of this project indicates that embargoes on e-theses are likely to be quite common and typically will last for up to one year after examination. The main reason for these is the understanding (mostly by supervisors, who may be poorly informed on this yet hold considerable power over the procedure) that publishers are reluctant to publish articles written using material from a thesis that is already in the public domain in electronic form. Librarians do not in general consider this to be a necessary reason for an embargo. Reasons that gain more support from librarians are, in order of importance,

to protect patents, where the funder of the student claims the IPR, and to protect the first book. It is worth noting here that more librarians consider that embargoes are specifically NOT necessary to protect future journal articles and books than consider them necessary for those reasons.

Embargoes can be extremely simple to obtain, the process amounting to no more than a simple formality. At one institution, for example, there is a simple tick-box on the submission form to request an embargo, which is rarely refused. At other institutions obtaining an agreement for an embargo is more formal and more difficult to achieve. In other words, there is a whole spectrum of conditions and ways of working and this varied landscape may always be a feature of the UK HEI community. Meanwhile, the national service will need to make provision for embargoes and how to manage them.

Institutions remain eager for advice and help on the implementation of best practice in the electronic thesis arena. To mitigate concern in this direction EThOS has developed its Institutional Toolkit, working closely with the NDLTD during the process, to provide institutions with the information they might need. This concept is welcomed by the community.

6.2 The EThOS business model

The EThOS business model is predicated on the assumption that retrodigitisation of theses is required and that digitisation of current theses will be required for some time to come. Based on the figures (from the BL's current thesis service) that 80% of requests are for theses published in the last 13 years, the aim is to work towards digitising as much of this part of the older-thesis corpus as possible, thus providing a reasonable size corpus of digital theses as the backbone of the service. The EThOS plan is for 25,000 theses to be digitised per annum. Institutions will be gradually moving towards mandating the deposition of electronic versions of theses, even if at least for the foreseeable future these are additional, and not alternative, to hardcopy versions. The business model thus accommodates these digitisation needs for a period which is as yet undefined but which is assumed to be finite.

The HEIs understand that there are considerable costs involved in this and accept that new costs will fall onto institutions if theses are to be effectively provided in electronic form through a national service, though there are a number of ways in which costs might be covered: the EThOS model is not the only way, as other services around the world have demonstrated.

The EThOS model would levy charges on institutions for e-thesis provision, a solution that would turn on its head the current system where it is the user who pays for the retrieval and supply of a thesis of interest. This relocation of costs is acceptable to the HE community, though institutions would appreciate a choice of ways in which they might participate and EThOS offers them this. In an age where theses are produced and stored electronically, institutions do not see any savings in cataloguing and management procedures with respect to electronic theses over and above the same costs for hardcopy theses. They do, however, identify where savings can be made and these are particularly in shelving/storage space and in the cost of retrieving theses when they have been requested by users.

More than half of UK institutions would view the EThOS costs as a part of their institutional repository provision; this tallies exactly with the proportion of institutions that have suitable repositories at the moment and indicates that where a repository has been established there needs to be a flexible budget and scope for inclusion of new material and new uses as the repository matures: electronic thesis provision should lie alongside research article output at the heart of a repository's function. Large numbers of institutions also see the new costs that would be incurred by participating in EThOS as part of their normal library technology expenditure and over 60% would use existing library budgets to cover them. Nearly 40% of institutions would also view the EThOS costs as a legitimate part of the expenditure made by the institution to ensure that students and staff are productive scholars. Indeed, according to the survey carried out for this study it can be expected that almost 20% of graduate schools would contribute towards a national e-theses service and 30% of registries/central administration departments would do so too. Some institutions seem prepared to ask graduate students themselves to pay for the digitisation of their thesis if necessary and appropriate. Finally, institutions may see savings associated with reduced use of inter-library loan services if theses are available electronically on open access.

On the issue of the level of charges that EThOS might levy, our survey specifically avoided mentioning the actual charges explicitly. In the focus groups, however, it was possible to explore this issue without revealing EThOS planned figures and here, when asked, institutional representatives volunteered that 'a few thousand pounds' would be an acceptable figure to pay per annum for sponsoring membership of the EThOS service. The EThOS team held a well-attended workshop in Glasgow recently where the business model was presented in detail to delegates, including hard figures: the report from that meeting was that the audience found the sums involved unalarming. It should be noted, though, that formal testing of

potential pricing structures has not taken place. Informal feedback from a small number of libraries suggests that more than half would subscribe at the suggested price but that, of those who would not, some wish to concentrate on obtaining born-digital theses and do not wish to spend resources on older ones and others take the view that they can digitise themselves at a lower cost than EThOS is proposing.

There seems little doubt, then, that if the EThOS model is acceptable to HE institutions in the sense that they see it as a workable, fair and sensible way to provide a national e-thesis service, finding the funds to enable them to participate will be seen as part of the library budget process at an institutional level. In principle, then, the business model that EThOS has developed, in terms of who shoulders the costs and how the product is delivered in return for covering those costs, is acceptable to the UK HE community.

6.3 The organisational model for a national thesis service

With respect to the structure of the service, there is more a dichotomy of opinion. Equal numbers of people favour *in general* a national service that (a) as EThOS proposes, houses the metadata, full-text and associated digital objects of e-theses and provides access to these from a central hub and (b) acts on behalf of participating institutions by harvesting and holding metadata, providing a resource discovery service and pointing users to e-theses held locally in institutional repositories, like the Australian model. There is no significant support for any other model, such as locally-based provision, subject-based provision, or regional services by consortia of institutions.

There is some support for a service that offers a combination of (a) and (b). In the survey several people appended comments to this effect and there was some discussion of this at the focus group sessions. And the EThOS model does offer this: it houses and provides the full-text of theses it has harvested and/or digitised but for those where only metadata has been harvested from institutional repositories (or from other sources, like Repository Bridge) it simply points users to the required thesis in its original location.

There was also some low level of support for a European perspective on theses. DART-Europe would provide this, of course, since it aims to give searchers a single entry point into all European thesis content, wherever it resides. There may be scope for collaborative approaches between the UK service and international services like DART-Europe.

6.4 The legal framework

As reported earlier, there is considerable concern at institutional level about the legal issues around the provision of e-theses. Overall, just over half of the people who responded to the survey considered themselves to be familiar with IPR issues in relation to e-theses. A further 11% considered themselves to be very familiar with them, though 30% said they were not very familiar and 8% said they know very little about the topic. Institutions find the legal situation with respect to information something of a minefield now, and welcome expert help and advice. It was generally agreed by the focus group participants, for example, that the Freedom of Information Act (FOI; in Scotland, FOISA) is poorly understood with respect to what must and what need not be made publicly available upon request.

The majority of people consulted for this study concurred with the view that understanding of IPR issues – by both students and supervisors – could be better. This is the case even where their own rights are concerned and is exaggerated where the rights of the institution or third parties are involved. In many cases institutions leave the clearance of third party rights to the student, yet in the light of the views of the people we consulted this is not an altogether satisfactory way to proceed. It is considered, however, good practice for students to tackle this issue early in their academic careers because it arises throughout their research life. Institutions may provide students with advice on this early in their doctoral training programme (for example, some Australian universities do this particularly well and the NDLTD has a full advisory service on this topic, and the same sort of advice is provided within some UK institutions, such as by UCL Library Services) and this helps, but may not be enough. The OAK (Open Access to Knowledge) Law Project based at Queensland University of Technology in Brisbane¹ has carried out studies on this and a useful paper has been published setting out some of the issues and practical ways to resolve them².

Nonetheless, additional support and reassurance that would come from the EThOS service was considered to be advantageous. EThOS' position on the digitisation of older theses is to go ahead if the author's permission cannot be obtained, when reasonable attempts to secure this have failed. Some people expressed nervousness at this position but the majority accepted that it is a pragmatic solution to an intractable problem. EThOS has carried out a risk assessment, as have other such entities, and found the risk to be

¹ <http://www.oaklaw.qut.edu.au/>

² <http://eprints.qut.edu.au/archive/00004344/>

low: this risk is to be managed by a combination of rapid take-down procedure and insurance cover. These measures satisfy the majority of HEIs.

6.5 Preservation arrangements

The preservation of e-theses is a concern for institutions, just as is the preservation challenge for other types of digital material. Some institutions have taken the decision to develop their own digital archive (e.g. UCL, Cambridge University) but to date most institutions are not far advanced in this context. There is an appreciation within the library community of HEIs of the importance of digital preservation but only a low level of real understanding and experience of the issues involved. Such an admixture has led to real concern amongst many individuals about how to act in the best interests of their institution. The alternatives to institutional approaches are (a) national-level initiatives such as PADI in Australia³, Germany's KOPAL project⁴ and the Royal Dutch Library's e-depot in the Netherlands⁵ or (b) third party offerings such as the SHERPA-DP project in the UK⁶ which is focused on research eprints.

In the context of e-theses, the community's view is supportive of a national solution to the issue of preservation and the British Library was voluntarily mentioned by many as the 'obvious' provider. Had EThOS not been associated with the BL, this would have meant that the community would favour EThOS contracting this task to a trusted third-party (the BL). In the event, subcontracting is unnecessary since the service would be an integral part of BL's operations and would benefit from the BL's customary preservation practices.

6.6 HE community views on associated issues

6.6.1 Masters dissertations

In the survey a little over half of the respondents thought the possibility of covering Masters theses would be useful. In discussion in focus groups there was not quite so much support for this idea, though some institutions already deposit Masters theses in their repositories, especially for masters-by-research degrees. Undergraduate dissertations were not considered

³ www.nla.gov.au/padi/

⁴ <http://kopal.langzeitarchivierung.de/index.php/en>

⁵ <http://www.kb.nl/dnp/e-depot/e-depot-en.html>

⁶ <http://www.sherpa.ac.uk/advice/preservation.html>

suitable for a national service: there are far too many and the standard of work is highly varied and inconsistent.

6.6.2 Rights clearance services

Over 80% of survey respondents view IPR checks of deposited theses as important. A few (8%) are neutral on this issue and 5% don't have an opinion. The number of people for whom it would not be useful is insignificant (4%).

6.6.3 Plagiarism detection services

Almost 80% of respondents to the survey feel that plagiarism checks against deposited theses would be useful. Only 5% of respondents view this as not useful.

6.6.4 Metadata enhancement

Librarians who might be responsible for creating metadata for theses in a UK national service agree that they are happy to provide metadata to a reasonable level of complexity as required by the service, provided the requirements are clear. If the national service were further to enhance the metadata provided by institutions the great majority of librarians would find that a useful service.

6.6.5 Complex objects

The increasing use of digital technology by doctoral students, resulting in more multimedia digital objects being incorporated in theses, also contributes to the imperative to move to the electronic provision of theses. Concomitant with the increasing use of multimedia digital objects is an acknowledgment on behalf of the community that such things require special skills and knowledge to manage their storage and preservation. This is an issue that has arisen with respect to open access repositories for research output: the management of thesis output produces the same requirements. Librarians responsible for this expressed the opinion that a national electronic thesis service would seem to solve this problem for them. It would also act as a persuasive factor for institutions which like to encourage the use of complex objects but fear the consequences. Almost 90% of the survey respondents said they would find the ability of a national thesis service to cater for digital multimedia objects (and similar) to be useful. EThOS can do this.

6.7 General summary

For clarity and conciseness, the views of the UK HE community on the EThOS model are summarised in a table below. In comment, we can add that, in principle, three quarters of institutions say they would participate in a national service based on the EThOS model.

Almost 90% want digital copies to house locally. No universities we consulted have any plans for their own retrodigitisation of thesis holdings. This is simply not a high priority for their budgets. Nevertheless, they all welcomed the chance of obtaining digital versions of older theses and considered the EThOS proposal to digitise on demand a good one. It would ensure that digital copies of theses that were requested by a user were eventually obtained. Institutions that have a high number of doctorates awarded each year have concomitantly high demand for access to theses. The formula that enables them to pay a fixed sum annually for a guaranteed number of digitisations is an acceptable way to proceed, since it provides digital copies of the most popular theses and increases the digital holdings of those institutions.

Issue	Community requirements/views	EThOS model	Level of match
Institutional policies and practices	Open access requirement	Provides for fully open access model	Good
	Objective to maximise visibility of research output, including theses	Provides national gateway and discovery service to UK thesis corpus	Good
	Institutions lack resources/means to offer e-thesis services individually	Provides national service inclusive of all institutions that wish to participate	Good
	Wish to be able to facilitate embargoes (of various lengths)	Provides for embargoes: mechanism formally included in service plans	Good
	Desire for support and advice on best practices	Provides detailed Institutional Toolkit	Good
Business model	Choice of options for participation	Offers choice of options for participation	Good
	Cost must reflect real value gained	Price paid reflects actual numbers of theses digitised and added to the open access corpus	Good
Organisational model	Overall preference is for a mixed architecture, allowing institutions to retain full-text theses locally but provide metadata for the EThOS discovery service.	Offers this choice	Good

Table 4: Level of match between community requirements and views and the EThOS offering

Issue	Community requirements/views	ETHOS model	Level of match/mismatch
Legal framework	Anxiety about digitisation without author permission	Provides risk-assessed action, backed up by insurance cover	Fair
	Concerns about third party rights clearance	Not included in service	Low
Preservation arrangements	Reticence on the part of most institutions in tackling proper preservation of digital objects: most would prefer a national trusted-party solution.	Offers a national trusted-party solution	Good
<i>Associated issues:</i>			
Inclusion of Masters dissertations	Medium-to-low level of support for inclusion of Masters dissertations	Not planning to include Masters dissertations initially but could easily be accommodated at a later date	Fair
Rights clearance service	High level of demand	Not explicitly included in service plans	Low
Plagiarism detection service	High level of demand	Not explicitly included in service plans	Low
Metadata enhancement	Willing to enhance to reasonable level locally: appreciate the usefulness of additional enhancement by a national service	Guidelines to institutions will provide for a reasonable, functional level of metadata	Good
Handling complex objects	Envisage continuing increase in the complexity of digital theses	BL provides expertise to handle and store complex objects	Good

Table 4: Level of match between community requirements and views and the ETHOS offering

7. RISKS, BENEFITS AND OPPORTUNITIES FOR A SERVICE BASED ON ETHOS

Despite the high level of community acceptance of the EThOS model, and the robustness of the EThOS plan, it is possible to identify risks as well as benefits if the UK national electronic theses service were to be provided by EThOS. We present these in concise form below.

7.1 Risks

The main risks to the service, to JISC and other stakeholders if a national electronic theses service based on EThOS were put in place are that:

7.1.1 Not enough institutions will buy into the service as sponsoring members. The EThOS model rests upon a certain number of institutions participating as 'sponsoring' (in EThOS' own terminology) members, and thus committing to pay a sum of around £6000 annually for at least three years. Since there has been no testing of pricing on the market the willingness of institutions to buy into the service in large enough numbers is unknown.

7.1.2 Institutions that signed up as sponsors for the first three years opt subsequently to be associates (paying for digitisation on demand). This will affect the cashflow and viability of the service, whose business plan is predicated upon a certain level of take-up by institutions. Moreover, a reasonable corpus of electronic theses will not be attained if insufficient numbers of institutions sign up as sponsors.

7.1.3 It is likely that a period of some twelve months will elapse before the service comes into full operation: if approved, EThOS expects to launch towards the end of 2007. EThOS has done an effective job in advocacy and marketing of its concept to the community but the 'gap year' between the end of the project phase (autumn 2006) and the launch of the new service (end of 2007) presents a challenge in keeping up the momentum and retaining the interest of the community. It is possible that some institutions in the vanguard of thinking on thesis provision may well make their own arrangements during this time.

7.1.4 According to our survey, whilst firmly intending to join an EThOS service, fewer than 20% of institutions would join it within one year.

Around 50% would take between 12 and 24 months to join and 10% would take longer. This, too, may affect EThOS' plans.

7.1.5 Planning the long-term future of the service once most theses are born-digital is a challenge. The EThOS service business plan focuses on the early years, when institutions are expected to require digitisation as a core part of the service offering. Once theses are deposited in digital form at institutions, retrodigitisation will be something that is required much more rarely, and only in the case of much older theses that are still available only in hardcopy.

7.1.6 EThOS predicts digitising 25,000 theses per annum in the early years and has based its business model around this figure. This may turn out to be correct, but contingency planning for lower demand or, perhaps realistically, *higher* demand should be undertaken. Evidence from other services shows that when theses are easily locatable, demand is high. EThOS should be prepared for various demand and activity levels in its business planning.

7.1.7 There is a risk that the service may face pressures to adopt a commercial model over time. There is no support within the HE community for a non- open access approach to e-theses delivery and any moves towards selling e-thesis content would be poorly received. To gain maximum levels of support and participation at early stages there must be an assurance to the community that the open access model will be retained indefinitely.

7.1.8 Legally-based risks are present, and must be flagged up here, but we are satisfied that EThOS has examined these thoroughly and plans to put in place measures that mitigate against problems arising. Nonetheless, the institutions remain anxious on the issue of third party rights in theses to be digitised and unless proper reassurance can be provided on this issue sign-ups may be reduced. The onus, under the 1988 Copyright Designs and Patents Act, is on the end-user to comply with requirements and so this means that doctoral students must take ultimate responsibility for dealing with rights in their theses. For older, print-based theses, however, institutional worries are not unfounded and so measures that EThOS can take to allay fears in this respect will be worthwhile.

7.1.9 Plagiarism and third party rights checks are currently not an integral part of the EThOS offering and as a result of (v) above may need to be costed in as part of the service. It may be possible to add these as optional extras, but if they must be integrated across-the-board then this may affect the overall cost-revenue equation.

7.1.10 Marketing/promotion activities do not appear to take high priority in the EThOS plans at present but these will be important if the service is to enjoy a high level of support and adoption within the HE community.

7.2 Opportunities

The main opportunities for the EThOS service would be:

7.2.1 To provide a world-class electronic theses service

7.2.2 To increase the visibility of UK doctoral research output

7.2.3 To provide an impetus for the production and collection of electronic theses in the UK by establishing a centrally managed service with guidelines for standards and protocols

7.2.4 To provide a template for other would-be electronic thesis services

7.2.5 To encourage the adoption of such services in other countries

7.2.6 To link or collaborate with other national or international electronic theses services

The model proposed by EThOS as it currently stands will facilitate all these opportunities.

7.3 Benefits to UK institutions of EThOS working with other international electronic thesis organisations

UK institutions would benefit if EThOS were to develop international collaborations with other electronic theses services by:

7.3.1 Increased visibility for UK doctoral research output due to dissemination of details of UK theses through meta-search engines and collaborative discovery services

7.3.2 Increased usage of UK theses as a result

7.3.3 Probability (i.e. more than a possibility) of new scientific collaborations and associations being instigated as a result of new visibility

and accessibility of UK research, with all the concomitant advantages that arise (new funding opportunities, new ideas, new lines of investigation, new questions to answer)

7.4 Commercial opportunities that might arise from an EThOS service

Commercial opportunities are certainly possible from an EThOS service. Opening up UK thesis metadata provides entrepreneurial organisations with the opportunity to develop services based upon thesis provision. An example of a small publishing company in The Netherlands is a case in point: it locates theses that describe the synthesis of organic compounds – information that chemical companies find very useful because it saves them effort and time – pays for the digitisation (via the DARE service) and sells the thesis on to a chemical company interested in the compound. The thesis is digitised at the expense of the publisher, who is effectively fulfilling a resource discovery role, and the university gets a digital copy of the thesis in exchange. It is not stretching the imagination to see this sort of service extending to other industries, the pharmaceutical and biotechnology industries being prime candidates.

There are also opportunities for data-mining and text-mining by informatics companies. This is likely to become more important and represent more of an opportunity as time goes on.

Such commercial openings need to be properly managed and exploited. There is an opportunity here for the sponsors of the EThOS service to develop – or permit the development by third parties of – entrepreneurial exploitation of the service. JISC and the other stakeholders need to be aware of such possibilities and, if the opportunities are deemed attractive, to put in place mechanisms to utilise them.

8. RECOMMENDATIONS

Our recommendations fall into two categories – on immediate project-to-service issues and on more fundamental aspects of the proposed service.

8.1 Project-to-service issues

8.1.1 Testing of proposed price points has been carried out in the context of EThOS workshops, where it is reported that delegates were curious as to what the level of contribution might be but expressed no dissatisfaction with what was suggested. In the focus groups there was positive discussion around a figure of 'a few thousand pounds' per annum. Early signs are, then, that an average subscription fee of £6000 will be acceptable to the HE community but stakeholders should note that other details, such as the possibility of tiered pricing based on the JISC banding structure, have not been explored. It is recommended that EThOS carries out more work on developing the detail of the proposed price points and that these are carefully tested in the marketplace.

8.1.2 Given that most institutions will take between 12 and 24 months to decide to sign up to the service, stakeholders should expect a lag period of three years from now during which activity may be slow to build. Business planning should take this into account.

8.1.3 Marketing of the service will need to be ongoing to fill the 'gap year', retain institutional interest and commitment and ensure rapid take up of the service once in operation.

8.1.4 There will need to be a Marketing and Communications Campaign to promote a new service to the community; there will be costs in doing this which need to be met by a budget (which currently does not exist)

8.1.5 Stakeholders should require firm proposals from EThOS on the management of third party rights in theses to be digitised. This is an issue that concerns institutions and may reduce sign-ups if these concerns are not adequately addressed and reduced. The EThOS team has taken the appropriate steps to cover foreseeable risks but these may not completely satisfy the institutions. We suggest that EThOS investigates mechanisms for clearing third party rights or, if this is not a viable way forward, of managing institutional concerns to reduce the likelihood of resistance to

participate. This may be as much a case of laying urban myths as something more elaborate.

8.1.6 Since plagiarism detection was something the HE library community found desirable, EThOS should begin early discussions with the JISC Plagiarism Advisory Service about the best way to proceed on this.

8.2 Other issues, including core business factors

8.2.1 Business planning for the era when most theses are born digital is important. It is a long term issue, of course, but stakeholders should require EThOS to articulate how the business might operate once the digitisation service is no longer required. Digitisation is a fundamental part of the start-up service and plans for its phase-out should begin to be developed now.

8.2.2 The stakeholders should require EThOS to present plans for lower or higher levels of business than those predicted in the current business plans. Neither situation need necessarily be critical in its effect but planning for such contingencies ensures that the business does not need to cope with surprises at short notice. The management of unexpected success is as difficult as the management of disappointment if not planned for.

8.2.3 The stakeholders, particularly the JISC, should take steps to actively engage the HE community at levels other than the libraries. The latter have been fairly effectively brought onside during the EThOS advocacy and marketing programme and their views are well understood and reliable. There is an element of disconnectedness between the intra-institutional stakeholders regarding e-thesis policy and provision, and some misinformation which may damage the proposed service (in small ways). We recommend, therefore, that the JISC invests some effort in bringing registries and academic staff into the discussions with the aim of informing and enthusing these groups of the merits of a national electronic thesis service.

8.2.4 Following on from the point above, there needs to be continuing effort made to promote to registries the advantages of having an institutional policy to encourage the submission of electronic versions of theses at the time of examination. Virtually all theses are now prepared in digital form but the level of capture by institutions of the digital copy is poor and needs to be improved. The JISC and BL can make a difference here.

8.2.5 There is potential for additional add-on products and services that could be charged for, and the potential for new business opportunities. The

stakeholders should develop plans for exploiting these avenues as the service expands.

8.2.6 The stakeholders should explore the potential for fruitful collaborations with other national and international electronic thesis services. Users welcome one-stop-shops and there is much potential for developments in this direction. There are many initiatives being taken at international level that the service may engage with – the other e-theses services and also programmes such as the DRIVER project – to explore the potential for synergies and collaborations.

8.2.7 There will be a need for members of the library community, who are the most informed and knowledgeable about the issues involved in the provision of an e-theses service, to act as 'champions' within their institutions - and beyond – to promote the concept to other stakeholders, particularly the graduate schools, registrars and students.

8.2.8 The service should make proper provision for ongoing development costs, although we recognise that as time proceeds the overall shape of the service should simplify rather than become more complex. The most likely area in which R&D costs would be significant is curation and preservation of all the digital objects that may be encompassed by modern theses.

8.2.9 It will be important that the service conforms to the JISC Information Environment open standards to enable synergies to be exploited in the future.