

Enhanced E-theses Project

Deliverable 8

Legal issues of compound ETDs

Document Description

Project	
Title:	Enhanced E-theses
Start date:	1 st December 2007
Funding Agency	Knowledge Exchange
Document	
Deliverable number:	D8
Deliverable title:	Legal issues of compound ETDs
Actual Date of Delivery:	December 2009
Author:	Chris Awre (University of Hull)
Editor(s):	Keith Russell (Knowledge Exchange)
Workpackage:	WP8
Workpackage title:	Legal issues of compound ETDs
Workpackage leader:	University of Hull
Version/Revision:	1.1
Draft/Final:	Final
Total number of pages: (including cover)	21
File name:	Enhanced Etheses project_deliverable 8_v1.1.doc

This work is made available under a Creative Commons attribution 3.0 licence. For details please see <http://creativecommons.org/licenses/by/3.0/>



Disclaimer

The content of this document is based on desk research. The content represents an overview of the legal situation that emerges from this desk research. The document does not, however, represent any type of legal advice, which should be sought out by those working with e-theses to address any specific needs and/or situations where these arise. The document also does not claim to present a comprehensive picture of the legal situation with regard to e-theses, but an overview of issues and key developments that should be borne in mind when working with e-theses, particularly where there are issues relating to cross-border access, use and management.

Introduction

The work undertaken by the Knowledge Exchange Compound/Enhanced e-Theses Project has sought to identify a model for structuring an e-thesis where this is made up from a number of constituent parts. Other reports and papers produced by the project describe this model and the potential for implementing this within repository systems. The emphasis within the project was on the development of a set of proposals that could be used to stimulate discussion within the thesis management and repository communities on the management of compound e-theses. The focus has largely been on the practical aspects of modelling and implementation.

If the management of compound e-theses is to be taken forward, however, there has to be consideration of the rights held in the thesis and its constituent parts. This paper thus considers these rights and associated legal issues. The development and promotion of e-theses as a concept has always been guided by the desire to make the content and rich research output within the thesis more widely available. This desire has found common ground with the development of open access as a means for the dissemination of research, primarily of journal articles. But as has been all too easily found within open access, when there is a proposal to make content openly available, those with a vested interest in limiting access for whatever reason call on their position as rights owners or guardians to exert control over what can and can't be done. Slowly, increased awareness by researchers of the rights involved is starting to alter the landscape, and allow the exertion of rights to support open access rather than restrict it. And yet this has taken much time and increased knowledge on behalf of all parties involved. By contributing this report to considerations of rights in the context of e-theses it is hoped to help assist in the development of open access to compound e-theses within a framework that is commonly understood and accepted.

What is a compound e-thesis?

In order to understand how we can apply legal rights to compound e-theses, it is necessary to understand what is being described by that term. The model produced by the project, and the exemplars that are used to describe it, help to describe this in detail. However, it is useful here to think of a compound e-thesis on two levels..

- A thesis may have associated with it one or more supplementary files that are in addition to the main text. These may include related video or audio files that either demonstrate the research carried out (e.g., videos of a health care practice being investigated) or act as the basis for the research itself (e.g., excerpts from classical music). They may also include images (e.g., in a study of art history) or associated data generated by the research (e.g., through the use of scientific equipment for chemical experiments). Together, the thesis and the supplementary files can be considered a compound thesis. And since all of these can be, and indeed are possibly more easily produced digitally, they can also together form a compound e-thesis.
- A thesis that comprises a text document only can also be thought of as made up of a set of constituent parts, comprising the different parts of the document. Hence, the thesis can be split up into the title page, table of contents, chapters, appendices, references, bibliography, etc., which together form a compound thesis. And since, when produced digitally, a thesis can easily be broken down or provided as these constituent parts, they can form a compound e-thesis.

A compound e-thesis can thus comprise an aggregation of some or all of the following parts:

Title page
 Colophon
 Table of contents
 Author information (which may be a standard object)
 Awarding institution information (which may be a standard object)
 Abstract
 Chapter (1 – n)
 Appendix (0 – n)
 References
 Bibliography
 Image (0 – n)
 Video (0 – n)
 Audio (0 – n)
 Graphics (incl. figures) (0 – n) (though these may be considered images)
 Dataset (incl. statistics) (0 – n)
 Software (incl. web pages) (0 – n)
 Associated publications (0 – n) (these are themselves compound)
 Journal article / Book chapter / Report / etc.
 Image (0 – n)
 Video (0 – n)
 Audio (0 – n)
 Graphic (incl. figures) (0 – n) (though these may be images)
 Dataset (incl. statistics) (0 – n)

Software (incl. web pages) (0 – n)
Metadata (granularity to be decided, and may vary according to need)
This may include metadata about the above slices or about the thesis as a whole, or about non-digital elements of a thesis.

When produced digitally, the formats of each of the above parts may vary and, thus, each part may be replicated if more than one format is produced.

The discussion of the legal issues around compound e-theses will assume this breakdown within this paper.

Legal context

In discussing the legal issues related to compound e-theses, it is important to be conscious of the actual legal context within which the e-theses will be treated: what are the likely laws that will affect how compound e-theses are dealt with if any such dealings do ever end up in court.

Copyright law

Copyright has been an emerging area of law over the past couple of centuries around the world, as more and more countries lay down their own frameworks for how copyright will be dealt with legally. The World Intellectual Property Organisation (WIPO) maintains a directory of laws around the world that address this and related matters. This is not the place to go into the ins and outs of these laws, but suffice to say that copyright is now an almost ubiquitous area of law globally.

The limitation of having national laws covering copyright when materials might be distributed internationally was realised at an early stage. Although many countries had bilateral agreements respecting the rights of authors between them, the Berne Convention of 1886 broadened this through an acceptance that a “work” has copyright in itself by virtue of having been created, without it having to be declared or registered. This principle thus protected works across national boundaries for signatories of the Convention and although not universal this Convention has near global applicability. There have been subsequent international treaties that limit, extend and update the Berne Convention, but this remains at the heart of copyright protection. National laws apply this protection in the context of that country, laying out specific details of what is and what is not considered copyrightable.

Intellectual Property laws

Copyright is one area of intellectual property that should be considered in the management of compound e-theses. Other areas of intellectual property are:

- Patent rights
- Trademark rights
- Moral rights

Theses may well contain content that is patentable: indeed some universities have staff whose role it is to identify this before it reaches the public domain. Hence, it may be the case in a compound e-thesis that certain parts cannot be made openly accessible due to the content within that or those parts, whilst the rest can. Splitting the thesis up into its constituent parts may help with this, so long as the remaining parts still make up a coherent whole.

Whilst patents are quite feasible, it is less likely that compound e-theses would contain trademarks that need to be protected. Two possible exceptions to this are:

- Where the thesis has led to the creation of a spin-off company from the University that is capitalising on the research carried out. This may be to further develop an invention, where patent rights will also play a part, or to deliver a service. In both cases the business created may have trademarks it needs to protect.
- Where the research within the thesis breaches the trademark rights of another. A check of the thesis prior to distribution would be required to ascertain and address this.

Patent and trademark rights constitute rights that can be claimed, negotiated and transferred as agreed. Clarity on the existence of these rights is therefore necessary to be able to best manage them. Moral rights, the right of the author to be recognised as the creator of a work, are largely inalienable, and enshrine the right of the thesis author to be recognised as its writer regardless of how other rights are managed. They may need to be asserted on publication, and can be waived, but they exist per se.

Different types of intellectual property right have been enshrined in either the same or separate laws in different countries. Whilst copyright remains the most relevant legislation that affects compound e-theses, these other forms of IPR law should also be borne in mind.

Legal guidance

As with all areas of law, legal advice and guidance is often beneficial in determining how to act, and also as a means of assessing the risk of carrying out an action that may or may not breach the law. Risk is addressed specifically in a later section. Four key sources of advice of relevance to compound e-theses are listed here to highlight what is available. Additional local sources that give greater heed to national law and its implications should also be sought. None of the toolkits/guidance listed specifically addresses compound e-theses, though the information provided can be equally applied to a compound e-thesis when considered as a work overall.

A. JISC Legal

The JISC Legal service provides guidance on legal matters, though is very clear this does not constitute formal legal advice. It provides a very wide range of briefing papers covering many different types of law, including IPR. One report from 2004 stands out for relevance to compound e-theses: Theo Andrew's paper on Intellectual Property and E-Theses (<http://www.jisclegal.ac.uk/Portals/12/Documents/PDFs/ethesepaper.pdf>). This focuses on issues relating to copyright and licensing (see next section) for e-theses in general.

B. EThOS Toolkit

The EThOS toolkit was produced as part of the development of the EThOS service, a single point of access to UK doctoral theses. It is a community initiative and addresses all aspects of how an institution can get up and running in managing e-theses, including advice on legal issues. The toolkit is an online resource available at <http://ethostoolkit.cranfield.ac.uk>.

C. JISC-SURF copyright toolbox

The copyright toolbox is a tool that allows authors and publishers find a level of agreement between them to suit both parties when contracting to publish a work. As such, it does not address theses per se. It is, however, a useful guide to understanding what should be considered when managing the copyright in a thesis, and also as a way of assisting students in understanding their relationship with publishers if works derived from the thesis, potentially made up from parts of a compound e-thesis, are prepared for publication. Details are at <http://copyrighttoolbox.surf.nl/copyrighttoolbox/>.

D. The Oak Law Project Copyright Guide for Research Students

This guide specifically focuses on what students need to ask themselves when assessing the copyright position of their thesis. Further details can be found at <http://www.oaklaw.qut.edu.au/files/Copyright%20Guide%20for%20Research%20Students.pdf>. Note that this perspective is an Australian one, albeit that the principles can be applied elsewhere.

A search of the web highlights that there are many institutional pages offering guidance that can be followed. They are practical in nature and seek to offer a straightforward means of encouraging students to participate in providing open access to their e-theses. Prior to addressing specific legal aspects of compound e-theses, it is useful first to understand in more detail the options that many of these pages make use of.

Managing the legal issues of digital content

Licences

The creation and projection of digital content policies, including those for e-theses, requires the policies to be encapsulated in something that can be transmitted and understood by those viewing them. Licences fulfil this role. In a paper prepared for a JISC Legal Briefing Day, Richard McCracken talks about the primacy of the licence and licensing in managing access to digital content¹. This derives from the selling of digital products. Unlike when selling a book, the subscriber/purchaser does not own the digital product; they simply purchase a licence to access it. As such, the licence essentially becomes the product. The principle of assigning licences in this way applies as much to institutional digital repositories as to commercial operations. McCracken emphasises the need to be very clear about the scope of the licences developed for disseminating digital content emanating from the University so that the licences are effective and clearly understood.

Digital repositories can use licences to help address the legal needs of the institution and/or the author(s) concerned. Whilst many repositories have focused on open access, those repositories that store learning materials and images, for example, have a need to be sure about how they wish to disseminate these materials and under what terms. Much of the need for this detail derives from the authors/creators wishing to know who is using them, plus wishing to exert a degree of control. E-theses sit somewhere between open and controlled access: the desire is to make them available on open access, but there is still an equal desire to manage how the e-thesis and its associated components are subsequently used. Whilst open access does not have such a need for licensing per se, it is recognised that licensing is important to enable openly accessible material to be used with confidence and protect against plagiarism or improper changing of the content by others².

There have been calls in the past for common licences and templates that can be used by repositories, based on wide consultation³. There is, however, a dilemma in developing a generic licence, in its ability to meet the needs of working with different digital content types. An increasingly popular approach is the use of Creative Commons licences, which have been adapted to work with a wide range of legal jurisdictions, including all Knowledge Exchange partners and many others across Europe. These licences allow individuals and organisations to state what can be done with resources in the areas of attribution, commercial use and derivation.

¹ McCracken, R (2004). *Copyright, performance rights, moral rights and your digital materials*. A paper for a JISC Legal Information Service Briefing Day, held 15th September 2004, available at www.jisclegal.ac.uk/events/ppts/CopyrightNewcastle/RMcCrackenpaper.doc

² Poynder, R (2005). The role of digital rights management in open access. *INDICARE Monitor*, 2 (2), available at http://www.indicare.org/tiki-read_article.php?articleId=93

³ Charlesworth, A (2005). *Rights in digital environments*. A report on two workshops run by the JISC, 3 and 22 March 2005, available at http://www.jisc.ac.uk/uploaded_documents/JISC%20Rights%20in%20Digital%20Environment%20Report.doc

University adoption of Creative Commons licensing has been sporadic thus far, but it is individuals within them that have often led the way.

Recognising that the basic Creative Commons licences may not suffice for specific areas of education and research, companion initiatives have emerged to meet these needs. The Science Commons and ccLearn initiatives, both based at MIT, are looking to create open platforms for the exchange of digital content within scientific disciplines and for learning & teaching, respectively. It should also be recognised that Creative Commons licences are not a panacea, and careful consideration needs to be taken to ensure their appropriateness⁴.

Copyright and Creative Commons

Although, as highlighted in the previous section, licences can be used in a powerful way to manage digital content, when stating the rights held in a thesis or e-thesis up front there are two approaches that can be taken. The rights can state what restrictions there are in place affecting the thesis, or they can state what permissions are allowed for the thesis. Creative Commons epitomises the latter, but the use of copyright is also worth further consideration.

A. Copyright

Copyright is a form of intellectual property right that gives the author of an original work the exclusive right to the work. This is intended to allow the author control over how the work is used and, as the name suggests, copied. It is time-limited, after which the work enters the public domain. The original purpose of copyright was to prevent publishers creating their own copies and benefitting financially whilst the author lost out. By conferring copyright, the author could receive payment in lieu of published copies. Copyright can thus be considered a form of financial protection, and whilst it isn't intended to be restrictive in determining how content is copied and used per se, the net effect has often been to prevent the distribution of content for valid purposes.

B. Creative Commons

The development of Creative Commons (CC) emerged from a frustration with copyright and the way this legislation was sometimes being used, which it was felt hindered academic and cultural sharing. Instead of being an indication of what you can't do without permission, CC takes as its starting point the principle of stating what you can do with content. An appropriate licence is associated with the content that offers a range of alternative permission sets, depending on whether the author or creator of the content wishes to allow commercial use, whether derivatives of the work can be made, or whether to insist that all subsequent uses of the content should be made available using the same licence. All licences insist on the author being attributed whenever the content is used.

⁴ Korn, N and Oppenheim, C (2006). Creative Commons licences in higher and further education: do we care? *Ariadne*, October, Issue 49, available at <http://www.ariadne.ac.uk/issue49/korn-oppenheim/>

Although at first view the copyright and CC approaches can appear mutually exclusive they can be used in tandem with each other. Claiming copyright is a clear statement that a particular person, persons or organisation is the copyright owner of a work. Whilst this means that permissions should be sought for use of the work, adding a CC licence alongside helps smooth this process by also stating up front what permissions are granted. This can help to address one of the main practical issues with copyright, that of the time it can take to get permission to use a work. There will be occasion where the intended use of a work goes beyond what the CC licence grants permission for, and separate negotiation with the copyright holder is called for, but the use of the CC licence does provide a very useful starting point.

This combination works well so long as copyright and the CC licence used do not contradict each other. This is most likely to occur where the author does not hold copyright in all the content within the thesis – the issue of third-party copyright, which will be addressed later. But it can also work the other way round. One example highlighted on the Creative Commons ccLearn FAQ website demonstrates this⁵. Making a translation of a work is creating a derivative under copyright law, and so allowed. There is therefore little point in assigning a licence that does not allow derivatives unless there is a very specific reason why that should be the case.

Copyright and CC are just two possible ways of managing the intellectual property implicit in a thesis, but they appear to be predominant ways in which digital repositories are tackling the issue. There are other types of intellectual property right that may apply to content within the thesis (e.g., patents deriving from the work), and there are other types of licences that can be applied (the basis of many DRM systems and processes revolve around these). These two examples will be explored further in their role with regard to compound e-theses and the open distribution of such theses later in this paper.

Digital rights expression languages

Having a way to state the rights associated with an e-thesis, via copyright, CC or other means, is one matter: making this available, the expression of those rights, is another. Where the rights are limited to physical items, the location of the item and its storage is an effective means of achieving this. For digital content the rights need to be closely associated with the digital content referred to. For this to happen there needs to be a way in which the licence can be expressed electronically. Digital rights expression languages (DREs) have been designed to address this issue.

A JISC Techwatch study on DREs⁶ reviewed existing approaches and the options available currently.

⁵ See <http://learn.creativecommons.org/resources/faq#2.6>

⁶ Barlas, C (2006). *Digital Rights Expression Languages (DREs)*. JISC TechWatch study TSW0603, available at http://www.jisc.ac.uk/whatwedo/services/services_techwatch/techwatch/techwatch_ic_reports2005_publish.d.aspx

- Open Digital Rights Language (ODRL) was developed by Renato Iannella in 2000 and is backed by a consortium of over 200 companies. It is a generic rights expression language based around an extensive vocabulary of terms describing the terms under which a piece of digital content can be used, with the aim of being flexible enough to encompass many different licence approaches.
- eXtensible rights Markup Language (XrML) came out of work by Mark Stefik in the late 1990s. It is considered to be a meta-language, insofar as it is more abstract than ODRL and can be used as the basis for the development of DREs for more specific purposes.

These DREL developments have used XML as the basis for their development. The Creative Commons licences have their own XML and other encodings (as well as plain language versions), which can be considered a specific instantiation of a DREL around these licences, and an ODRL version of Creative Commons exists.

Though trends and initiatives are emerging, there is, thus far, no hard and fast agreement or consensus on which DREL is most useful for different circumstances and it is the requirements at hand that determine which will be most effective. A Library of Congress white paper lays out a framework to help guide this decision-making. Institutions and their repositories must be guided by their needs and local interest in constructively managing and enabling access to digital content⁷.

Factors that require attention in the distribution of e-theses

Relationships

A compound e-thesis on its own does not need to have rights or its legal status addressed. It is when that compound e-thesis is made accessible, used or managed, in other words moved around in some way, that being clear about rights is of value, and recommended. The relationships involved are then important to understand so that rights can be assigned and managed appropriately, and need to be understood to enable the constructive management of digital content institutionally.

1. Author to University relationship

Theses are generated by students (authors) through work carried out under the auspices and guidance of a particular University. Hence, when the thesis is created, the first rights issue at hand is what the relationship is between the University and the author and how the University manages the thesis. Who owns the thesis? What rights does the author/University have in affecting how the thesis is managed? How and what rights are transferred between them? Establishing a clear picture here

⁷ Coyle, K (2004a). *Rights expression languages*. A report for the Library of Congress, available at <http://www.kcoyle.net/topiclist.html#copyright>

can bring dividends if more external bodies become involved following the completion of the thesis.

Case study: University of Hull

Within the UK it is generally agreed that the student/author owns their thesis. This would be different if the author was an employee of the University, where the rights would default to the institution, but students retain the rights to their own work (a fact reinforced where students are paying customers). However, the University retains the right to hold and manage the thesis as part of the institutional output. This applies to both print and electronic versions.

Taking this as a starting point, both parties then make requests of the other. Where access to a thesis needs to be restricted for whatever reason, the author asks the University to agree to this embargo. The University, in turn, asks the student for permission to load it into a national plagiarism detection service (to protect the thesis from future plagiarism, not to detect plagiarism), to load it into a national discovery service for thesis (EThOS), and to alter the format of the e-thesis as required for long-term digital preservation.

Other rights might be included in such agreements at different institutions according to need, and form a rudimentary licence.

2. Author to reader relationship

Authors and readers of theses may never meet. And yet, for a reader to get the most out of a thesis it is valuable for them to be aware of what rights they have for interpreting the material, using the material, disseminating the material, etc. The situation is akin to the reader of a book, where the publisher makes it clear who owns the rights and what those rights allow a reader to do (most usually evidenced by the publisher stating who owns the copyright and the inferred permissions/restrictions that associated copyright law brings). Hence, when a thesis is made available to a reader it is valuable for the rights involved to be clearly stated and made apparent up front. The author themselves may not have a hand in this, but may delegate this role to the University holding the thesis (or online service for e-theses), which is obviously akin to a publisher role.

3. Author to third-party relationship

A reader is a third-party, but one with a specific interest in reading the thesis and gaining benefit from this. There may be other third parties, such as publishers, websites, search services, etc. who have an interest in the thesis for other primary reasons, usually to support a service and/or business model that seeks to make use of research reported through the thesis. As for the reader, stating the rights up front can make it easier to manage such relationships, both in terms of subsequent negotiation for any use of the material, or to redress any misuse of the material that was carried out without permission.

4. University to third-party relationship

Whilst the key relationships are between the author as holder of the rights in their thesis, the e-thesis will most likely be managed digitally by the University. It may thus be the case that the relationships an author has with third parties are delegated or passed to the University for management, as for the case of requesting permissions to pass the e-thesis to search and plagiarism services, as described above.

Rights

Another element that needs to be considered when considering the rights in an e-thesis is exactly what we are assigning rights to, as this may affect the type of rights and permissions involved. The same applies to print theses, though the issue is exacerbated in the digital environment where there is greater flexibility about how content is managed. There are four areas of rights that require consideration:

- Rights in the content of the e-thesis (or constituent part of it)
- Rights in the metadata of the thesis (an issue pertinent to digital repositories holding the e-theses)
- Rights in the aggregation of the compound e-thesis
- Rights in a collection of e-theses

Rights in the content

The content of the e-thesis is the core of what is being created and what will be distributed. The role of protecting this content has been at the heart of content in this paper thus far. And where the e-thesis is a single piece of content, the statement of copyright and assignment of an appropriate licence is a relatively straightforward path that can be adopted. Where the thesis is compound, however, there is scope for recognising the rights in each constituent part of the thesis. In such cases, the e-thesis can be treated as a single whole, or each part can have its own copyright statement and associated licence (and the licences may vary according to the part involved, e.g. the text of the thesis may not allow derivatives, but an associated dataset might, to facilitate ongoing research). This issue is addressed further later.

Clearly, treating each constituent part on its own, and treating the whole e-thesis as a complex rather than a compound object, has workload and management implications. Potential benefits of following this approach, though, are:

- It facilitates re-use of the constituent parts where this is permitted
- It allows clear recognition of any third-party rights that need to be stated
- It permits more granular management of the e-thesis within digital repositories
- It allows the assignment of the most appropriate licence (CC or other) according to the type of content

- It allows for any uncertainty in ownership or rights to be clearly flagged against the constituent part concerned

Rights in the metadata

When e-theses are made available, predominantly through digital repositories, metadata is created to describe the e-thesis, support its management and organisation, and facilitate access to it. This metadata can be created through a variety of means: the author can create it as part of an online submission system, librarians might create it or amend it as part of the processing of the e-thesis, or it might be automatically generated by appropriate tools. Dealing with rights in the metadata can be guided by considering the relationships described earlier. There needs to be agreement between the author and the University about who owns the rights to the metadata, which is likely to be affected by who creates it or how it is created. There then needs to be a clear understanding between the author and/or University and third parties about how the metadata can be used. Metadata can be used to promote the existence of content, and it can therefore be in the interests of the author/University to release the metadata for others to use accordingly. Where there is considered to be inherent value in the metadata itself, though, any licence associated with it may wish to place limits on what can be done.

An alternative to expressing rights explicitly through copyright or CC is the use of policies that lay down practices the repository undertakes. The policy development tool provided by the OpenDOAR service is a starting point for generating such policies, which can also cover data, content, submission and preservation⁸. Whilst the aim of this tool is to generate repository-wide policies, collection specific policies may also be applicable, and offer another way of stating what can and can't be done up front for clarity.

Rights in the aggregation

Where a compound e-thesis is considered in its constituent parts, with each part having rights declared for it, there is also the issue of whether the overall aggregation of the constituent parts, or indeed aggregations of only some of them, need to have rights declared as well. In the model proposed by this project, having such rights is clearly of benefit, and allows the aggregation to be dealt with overall. This is, in effect, the same as if the compound e-thesis is treated as a single entity, though it is important that rights in the aggregation are not overlooked whilst the constituent parts are managed separately.

Rights in a collection of e-theses

Individual e-theses collected together within a University form a collection of that institution's postgraduate research output. They are collated together within repositories for that reason, and also to facilitate access by content type. As the

⁸ OpenDOAR policy tool, <http://www.opendoar.org/tools/en/policies.php>

creators of the collection, the repository manager(s), and by default the University as the employer, hold the copyright in this collection. In the European Union the collection may also claim database rights as specific protection, though in the US copyright depends entirely on the demonstrable originality of the collection: whilst there are many collections of e-theses in repositories, each collection will, though, most probably be unique due to the e-theses within it.

Managing rights and legal issues for compound objects

Constituent files within an e-thesis

As recognised in the breakdown of a compound e-thesis at the start of this paper, the issues around rights in an e-thesis may have implications for the individual rights of the constituent parts, and also how these parts are aggregated together. Do the rights of the constituent parts need to be stated through copyright and/or licensing?

If an e-thesis comprises a text document made up from a number of parts, as described earlier and re-iterated below, then the purpose of assigning different rights to each of these parts will relate to how much the author/University wishes to disseminate the e-thesis on this basis, allowing for the different parts to be used as separate digital objects.

- Title page
- Colophon
- Table of contents
- Author information (which may be a standard object)
- Awarding institution information (which may be a standard object)
- Abstract
- Chapter (1 – n)
- Appendix (0 – n)
- References
- Bibliography

A single rights statement can be made for the whole document, though, which can also state, if using an appropriate licence, how the constituent parts can or cannot be used. This latter approach is less work, though potentially reduces the flexibility available.

If an e-thesis comprises a text document and a number of other items, textual or non-textual, then the same approaches can be taken.

- Image (0 – n)
- Video (0 – n)
- Audio (0 – n)

Graphics (incl. figures) (0 – n) (though these may be considered images)
Dataset (incl. statistics) (0 – n)
Software (incl. web pages) (0 – n)
Associated publications (0 – n)

The case for assigning separate rights to the constituent parts, however, is potentially greater, and especially so where the rights address aspects of the format of the part. For example, an image or video may require rights that clarify how excerpts from them might be used, that would not be applicable or appropriate for a wider rights statement. Some of the issues arising for particular content types are addressed in the next section.

Individual content types

A. Journal articles / e-prints (associated publications)

There is value in understanding the benefits of rights in regard of open access journal articles and e-prints, so that potential misuse of the openly accessible copies is addressed before it happens. Having said that, the aim of open access is to make such articles openly and freely available, and their inclusion as part of an e-thesis is a part of this. The level of rights required, therefore, is likely to be at the level of an appropriate licence, the terms of which will need to be agreed by the institution for use in the repository. Differences may occur as a result of disciplinary requirements, though anything too specific may result in over-restriction where it wasn't intended.

B. Audio and video files

Audio and video files can be generated as part of the research that is submitted for a doctoral degree. The approach taken to rights for such files is thus dependent on the specific purpose of the files, and the extent to which the files can or should be shared with others, both within and outside the institution. Audio and video recordings can involve people other than the author of the thesis, and therefore the rights of those individuals, and their permission, need to be taken into account and incorporated into the licence or other rights statement used.

C. Datasets

Datasets can be produced as a result of much postgraduate research activity. There has been a great deal of interest in the possibility of exposing datasets alongside the journal articles / e-prints that describe them, to allow full exposure of the research carried out and to allow subsequent investigation and collaboration: the same principle applies to associating a dataset with an e-thesis. There are accompanying licence issues when doing so, to ensure the rights of the dataset are maintained and that it is not misused in any way. It is also important to maintain the connection between the dataset and the e-thesis referring to it through the use of appropriate identifiers that can also help link the licence to both. Datasets are a good example of how constituent parts of an e-thesis can be usefully shared in their own right, even if the link to the thesis should remain in place wherever possible as a record of

provenance. They are, thus, a good case for where separate licensing may be appropriate in addition to that used for the e-thesis as a whole.

Notwithstanding the issues of ownership described earlier, and the prominence of the author as owner of the thesis, there is still some uncertainty around ownership of data. Are doctoral candidates allowed to give access to their data, even though they have made use of facilities of the University to create such data? In other words, are they really the only rightsholders? Data used for thesis research can also be part of a much larger research project that is ongoing, which has a lot of datacreators / researchers. In this case it is unclear who the real rightsholders are of the overall research data outputs. Data generated by one doctoral candidate might subsequently be used by another student for their research: who owns the data in this instance? Clarity from the start in generating the data is advisable to clarify what can and can't be done further down the line.

As for audio and video files, the recording of details about other individuals is also a factor that has data protection issues if the dataset is to be made available.

D. Software

In certain disciplines software will be one of the predominant outputs from a doctoral degree, particularly in Computing Science and Software Engineering. The software outputs will be described in the text document of the thesis. The extent to which the software will be made available separately from the text will need to be decided. Some Universities and individuals may wish to capitalise on the software output commercially. For those that are happy for others to use their output there is a range of software licences available for open source dissemination. These can sit alongside rights for the thesis as a whole.

E. Images

Images have many of the same aspects as audio and video files described above. However, whereas audio and video submitted as part of a thesis is most often generated as part of the research, images can be frequently sourced from elsewhere, and are not the primary creation of the author. Licensing for images is thus impacted by the need to consider origin, who actually owns the rights, and whether permission to use them in the thesis and disseminate them as part of this has been granted. Where the image itself has an associated CC licence, for example, then this can be clearly interpreted, but many images are protected by copyright and cannot be distributed further. Such third party copyright is dealt with in the next section.

Third party legal issues

A compound e-thesis is, as we have seen, made up from a number of constituent parts. The majority of these parts will be the creation of the student undertaking the research, who is then the copyright owner. However, as many areas of research involve a study of materials sourced or located elsewhere or owned by other, third, parties, there is a clear possibility that the thesis will include materials for which

copyright is owned elsewhere. Where this arises, action is required to manage these third party rights to ensure the thesis can be submitted for a degree and/or made available for wider consumption. Three possible scenarios arise:

- The third party material is referenced and/or quoted in an acknowledged way. In this scenario, where the third party material is clearly being identified as belonging to someone else, and it is being acknowledged in its own right, there is no requirement to carry out any further action. Work by others is regularly referenced in research, and is an integral part of this. Where the work is not academic in its origin, it may be appropriate to contact the copyright owner to be sure, but this will depend on the nature of the material.
- The third party material is included in the thesis in whole form or as a substantial extract, but in an embedded form (e.g., as part of the text). This may arise where the material forms the basis of the research. In such circumstances, action will depend on how the thesis will be treated: is permission required to use the material for submission?; is this different from the permission required to disseminate the thesis? As the level of third party content increases, it is, though, increasingly advisable to confirm that the author has permission to use the material.
- The third party material is a distinguishable component of the compound e-thesis, and exists as a whole in its own right. This might be the case where the material is a video or image. In such cases, it will almost always be necessary to gain permission from the third party copyright owner for use within the thesis, and also specifically for the different uses the thesis will be used for (both marking and dissemination).

As the creator of the e-thesis, the author is primarily, and best, placed to take forward any seeking of permissions from third parties. They may have an existing relationship with the copyright owner through the research undertaken, or they may at least have contact details that can be followed up. Nevertheless, following up on third party copyright is a task many authors are not familiar with and omit to carry out. If they are then mandated to submit their e-thesis to a repository, does the repository have the responsibility to carry out the check for them? There is a wide spectrum of practices across different institutions and repository owners in this, ranging from detailed checks of the thesis content through to no checks and information/education for the authors and their supervisors. The repository has an interest in supporting the author in ensuring they are only making available what they have permission to, although this make take time and resource that is not available. And yet, because of the specialised subjects of theses, the author is best placed to judge what does and what doesn't need checking. A balance needs to be struck to manage the disclosure of material that contains third party copyright. This is an area where there is a need for clear education amongst the parties involved to better understand how this balance is best managed.

In the current climate, and regardless of the level of copyright checking that does take place, it is necessary to have an effective and well publicised take down policy so that if any complaints are made the e-thesis can be withdrawn from public access as soon as possible whilst the matter is dealt with. Many repositories now have this type of policy. It is clear that in making e-theses and their components parts openly available that copies with potential breaches of third party copyright could circulate after a complaint has been made: they cannot be called back like a faulty product (although a request can be made to search engines like Google to withdraw the thesis from their cached index). But if the repository, and author where available, is responsive to copyright owners' requests then content can be managed effectively. Wider understanding and education of the issues involved will reduce any such occurrences of copyright breaches, but repositories can work with the owners to find a mutually acceptable way forward.

Risk management

Risk analysis

The previous section highlighted the need to carry out a degree of risk analysis when establishing the balance between author and repository checking of third party copyright. This is one area of risk analysis that might be considered when planning the management of digital content generally. A consideration of how systems and approaches might be applied to different types of digital content emerging from a University, including e-theses, reveals two important factors.

- That there is little need for strong protection (such as DRM) in the same way it has been applied in other sectors. The cost of implementing such protection systems can be prohibitive and would need to be balanced against the perceived value of what is being protected.
- That there is little hard and fast guidance about how to address the legal issues relating to many materials. Like much legal advice, the exact details need to be agreed internally by the organisation holding the rights to it (or claiming them). There is a lack of tests of the licences employed in a court of law where there has been a breach, or it may be because institutions are still establishing their own licence arrangements and are not yet in a position to make such licences public or share them within the community.

In both cases there is an element of risk analysis and management involved in assessing the level of protection required for any particular digital content type. The availability of the DRAMBORA toolkit to facilitate a risk assessment of repositories can assist in addressing the relative risks of storing and making accessible digital content on a case-by-case basis, and can be applied to repositories holding e-theses as much as other repositories.

Proposed best practice for rights management as part of repository management

The approach recommended for use by a digital repository is summarised below for reference.

- Open up dialogue between repository management and institutional legal advisers
- Carry out a risk analysis for each area of content to be stored within the repository to assess the level of protection required when making that content accessible, either internally, externally, or both.
- Agree on the level of protection to be enacted for each area of content
- Develop or adapt appropriate licences for use with the content types involved. Where appropriate and where effort is available, develop both human and machine-readable versions of these
- Embed the assignment of licences within repository workflow
- Agree on and enact appropriate click-throughs and other measures to be put in place for specific materials requiring them
- Monitor protection measures at regular intervals to test their appropriateness and applicability over time

Incorporating rights

It is one thing to state rights in an appropriate form like an e-thesis, it is another to state them in a way that can then enable them to be interpreted as part of the overall compound e-thesis.

Use of OAI-ORE

This project has highlighted the potential capability of using OAI-ORE as the basis for sharing compound e-theses, and proposed a model for structuring them for this purpose. As part of this rights statements can be incorporated at different levels:

- In the abstract model for OAI-ORE section 4.2 states that rights metadata may be added as one of a number of properties metadata types about a Resource Map. These rights can then be included as part of the serialisation of the Resource Map as appropriate. Ideally, the rights should have an available machine/web-readable version to link to for consistency.
- Rights can also be assigned to the aggregations themselves. It is harder to appreciate the value of assigning rights to a concept rather than to the serialisation, the actual aggregation in practice, but there is value in ensuring when an aggregation is effectively managed by a particular party. Aggregations can describe different collections of components and may or may not represent the whole e-thesis. As such, rights can be applied in a granular fashion according to the aggregation breakdown.
- Rights can be assigned to the individual objects, by associating metadata with them. The OAI-ORE specification does not address this specifically as its focus is on managing aggregations, but so long as the rights statement has a

machine/web-addressable location then OAI-ORE can link to it and incorporate it as part of the overall aggregation.

As such, the use of OAI-ORE offers flexible options for incorporating rights in compound e-theses, according to need and taking account of the factors described in this document.

Comparison with OAI-rights

Prior to the development of OAI-ORE, of course, many of the same team had created the OAI-PMH specification. In the light of adopting OAI-ORE to model compound e-theses, it is valuable to consider how the development of OAI-rights might guide, or have guided, adoption and presentation of rights within OAI-ORE. OAI-rights focuses, as does the Protocol for Metadata Harvesting, on the metadata that is being shared, rather than the content that OAI-ORE also addresses. However, some of the principles of application are the same. OAI-rights focuses its attention on rights statements associated at the repository, set and record levels. The first two are aggregations, and it is emphasised that the rights statement is therefore about the aggregation, not the constituent parts. This is also adhered to within OAI-ORE. But it is also equally valid to apply rights at all these levels, according to need, again as is enabled within the OAI-ORE model.

Conclusion

This document has attempted to outline the legal considerations that merit attention when dealing with compound e-theses. There is much that will be learned from working with compound e-theses in practice that will enable a fuller, more rounded view of which legal aspects require the most attention, but the areas raised here should all be borne in mind to prevent unnecessary breaches of someone else's rights. Reading a document like this could, as is so often the case with documents describing legal aspects, be disheartening. And yet, where we can protect the rights of those concerned it is important that we do so in the best and most reasonable ways we can so we can be open with the content and disseminate it in good faith. This document concludes that awareness and planning are both vital in enabling compound e-theses to be managed effectively, whilst recognising that all rights management has a degree of risk management that needs to be assessed at individual institutions.

In summary:

- Compound e-theses need to take account of the legal standing of the constituent parts and cannot be treated only as an entity in their own right

- Specifically, there is inherent copyright in each of the constituent parts of a compound e-thesis, hence there needs to be awareness of who owns this copyright and what permissions for use exist
- Be conscious of other types of IPR law as they might apply to an e-thesis, particularly in respect of patents or trademarks. This is of direct relevance in the context of the affiliated institution where the thesis was awarded so mutual agreement can be reached on exploitation of the thesis findings
- Sourcing appropriate local advice within the jurisdiction the compound e-thesis is being managed within is highly recommended
- Licensing can provide great flexibility in making e-theses openly available, but must be carefully assigned so as not to lead to misunderstanding later
- Understanding the relationships between the parties involved in the dissemination of a compound e-thesis are important to ensuring clarity of its legal status for dissemination
- It is valuable to ask the question, what do the rights need managing for? This can clarify what protection an institution and/or individual is seeking
- Third party copyright issues have the potential to be like icebergs, largely hidden as the Internet makes it almost too easy to use content without clarifying rights on its re-use. Education in the need to request rights is vital to prevent backlash from copyright owners who discover this re-use after dissemination
- Risk analysis as an approach to assess and manage rights issues relating to compound e-theses is a constructive way of managing rights management in a realistic way
- As has been promoted by this study through its models, OAI-ORE is advocated as a mechanism for recording all information related to a compound e-thesis, and this encompasses its use as a mechanism for recording rights associated with compound e-theses