

Prepared for the
Legal Deposit
Advisory Panel

An



Research and Consultancy Report



By **Electronic Publishing Services Ltd, an Outsell, Inc. company**
28 November 2006

Refining the map of the universe of electronic publications potentially eligible for legal deposit

Updated final report

www.epsLtd.com
www.outsellinc.com

CONTENTS

0. Executive summary	1
0.1 Background.....	1
0.2 Work carried out.....	1
0.3 Existing and complementary maps.....	1
0.4 Categorization of e-publications	1
0.5 Key issues	2
0.6 Statistical data.....	2
0.7 Metadata.....	2
0.8 Practicality of phasing regulations.....	2
1. Introduction	3
1.1 Background.....	3
1.2 LDLA definitions	3
1.3 Other terminology.....	3
2. Objectives	4
3. Methodology.....	5
3.1 Stakeholder interviews.....	5
3.2 Desk research.....	5
4. Review and elaboration of original map	6
4.1 General.....	6
4.2 Completeness.....	6
4.3 Affinities	6
4.4 Updated version of the original map.....	7
5. Complementary map	8
5.1 Rationale.....	8
5.2 Terminology used in the complementary map.....	8
5.3 Structure of the complementary map.....	9
5.4 Categories of published e-content.....	10
5.5 Application to legal deposit.....	11
6. Issues arising from the complementary map	13
7. Correlation of original and complementary maps	13
8. Statistical data	14
8.1 Library statistics.....	14
8.2 Publishing output/population statistics	15
9. Metadata issues.....	16
10. Practicality of phasing regulations	17
Charts	18
Chart 1: Original map of the universe of e-publications.....	19
Chart 2: Updated version of the original map	20
Chart 3: Complementary map of the universe of e-publications.....	21
Chart 4: Alternative view of the complementary map.....	22
Chart 5: Matrix showing relationship of original and complementary maps....	23
Appendixes	24
Appendix A: Terms of reference.....	25
Appendix B: List of interviewees.....	30
Appendix C: Updated schedules attached to original map.....	33

0. Executive summary

0.1 Background

Electronic Publishing Services Ltd (EPS), an Outsell, Inc. company, was commissioned in July 2006 by the Department for Culture Media and Sport (DCMS) on behalf of the Legal Deposit Advisory Panel (LDAP) to review, refine, and update an already existing provisional map of the universe of electronic publications potentially eligible for legal deposit under the Legal Deposit Libraries Act 2003 (LDLA), and to expand on information previously gathered relating to categories of publications identified in the map.

0.2 Work carried out

Work carried out fell into four categories:

- ❖ Interviews were conducted with 19 individuals covering a wide spectrum of print and electronic publishing.
- ❖ Data were collected concerning the output and populations of electronic publications in each of the categories of the existing map.
- ❖ Data were collected concerning the electronic publications collected hitherto by the LDLs in each of the categories of the existing map.
- ❖ Considerable effort was put into analysis of data and opinions collected and into the subsequent synthesis of a proposed complementary map.

0.3 Existing and complementary maps

Whilst the already existing map was not comprehensive, it was considered valuable in terms of ensuring that no known type of publication was overlooked.

It was concluded, however, that a complementary map was necessary in order more clearly to demonstrate the characteristics of various types of electronic publication, and the affinities and distinctions which exist between them.

By focusing specifically upon characteristics of structure and of electronic delivery mode rather than upon nature of content and print delivery mode the complementary map was considered more likely to be future-proof.

0.4 Categorization of e-publications

Four questions are used to categorize e-content in the complementary map:

- ❖ Is the published e-content composite (i.e. composed of discrete files) or enquiry-driven (i.e. requiring interrogation by each user to yield results)?
- ❖ Is the e-content available offline or online?
- ❖ If online, is the e-content delivered to or collected by the user?
- ❖ If collected, is the e-content unprotected or protected?

Six theoretically possible categories of e-publication result from these questions, though in practical terms the number can be reduced. The categories are:

- A1. Composite e-content delivered to the user (offline or online).
- A2. Online composite e-content collected by the user without formality.
- A3. Online composite e-content collected by the user following compliance with some formality.
- B1. Enquiry-driven e-content delivered to the user (offline or online).
- B2. Online enquiry-driven e-content collected by the user without formality.

B3. Online enquiry-driven e-content collected by the user following compliance with some formality.

Category B1 is effectively redundant. Category A2 embraces most of the web domain, including large amounts of user-generated content. This is the only category which could easily be harvested, but trials have been hampered by rights clearance procedures, especially non-response by many rights holders. Categories A3, B2 and B3 entail a heavy overhead of liaison to obtain access to content held beyond barriers and to understand software requirements.

This categorization is not considered to conflict in any way with existing priorities – offline; e-journals; web material – since trials in these areas will be valuable in identifying issues in relation to Impact Assessment and metadata.

0.5 Key issues

A 'catch-all' typology will focus attention heavily upon other aspects of deposit, such as territoriality and safeguards for publishers' concerns over access.

A difficulty with the LDLA arises, since important, but integral, parts of the digital content spectrum are excluded. Yet, web site resources often embrace such types of media – i.e. sound and motion video.

0.6 Statistical data

Statistical data from LDLs concerning e-publications present some difficulties. Data received were not always broken down in detail. Significant variation exists between the LDLs in presentation of figures and in the description of such accessions. Break-downs did not usually correspond to the categories of e-publication in the original map of the universe.

Data concerning e-publication output and populations are presented in a table according to the categories of the original map of the universe. Most categories are of 'manageable' proportions, the obvious exception being web sites, of which at least 6.2 million are estimated to exist.

0.7 Metadata

Three main metadata issues are discussed and potential costs highlighted:

- ❖ The essential metadata required adequately to describe specific resources.
- ❖ The practicality of obtaining such data from publishers.
- ❖ The depth of data desired by LDLs to make resources usable/accessible.

Additionally, it is anticipated that most 'publishers' of user-generated content will have no knowledge of metadata issues.

0.8 Practicality of phasing regulations

Attention is drawn to three main points:

- ❖ The requirement to produce an Impact Assessment, which is likely to be simplified by pursuing the complementary approach.
- ❖ The speed with which LDLs will be able to communicate with publishers and to handle the volume of new accessions. Clearly some sort of phasing is likely to be necessary.
- ❖ It may be necessary to assess whether LDLs' objectives can be achieved at all by voluntary codes of practice. Important areas exist where voluntary compliance would be wholly impractical.

1. Introduction

1.1 Background

Electronic Publishing Services Ltd (EPS), an Outsell, Inc. company, was commissioned in July 2006 by the Department for Culture Media and Sport (DCMS) on behalf of the Legal Deposit Advisory Panel (LDAP) to review, refine, and update an already existing provisional map of the universe of electronic publications potentially eligible for legal deposit under the Legal Deposit Libraries Act 2003 (LDLA), and to expand on information previously gathered relating to categories of publications identified in the map. The original map which EPS was asked to examine is reproduced in this report as Chart 1.

This document constitutes the final report on the various activities undertaken in order to fulfil the requirements of the project.

1.2 LDLA definitions

In considering the universe of electronic publications potentially eligible for deposit, it is as well to keep in mind throughout two of the key definitions set out in Section 14 (Interpretation) of LDLA, as follows:

“publication”, in relation to a work –

- (a) means the issue of copies of the work to the public, and*
 - (b) includes making the work available to the public by means of an electronic retrieval system;*
- and related expressions are to be interpreted accordingly.*

“electronic publication” means an on line or off line publication including any publication in electronic form (within the meaning given by section 178 of the 1988 Act¹).

In the context of the present report, the words **‘making the work available to the public’** are particularly pertinent, since (subject to other criteria, such as place of publication) merely offering an e-publication, or posting it in a public place may be sufficient to establish eligibility for deposit.

1.3 Other terminology

Some new terminology is introduced later in this report in order to facilitate the definition of e-publications according to a different view of the universe from the view adopted in the existing provisional map of the universe of electronic publications potentially eligible for legal deposit. This new terminology is defined at the point where it is introduced – specifically in Section 5.2 of this report. It is important that that section is read very carefully.

¹ According to the Copyright, Designs and Patents Act 1988 (CDPA): *“electronic” means actuated by electric, magnetic, electro-magnetic, electro-chemical or electro-mechanical energy, and “in electronic form” means in a form usable only by electronic means.*

2. Objectives

The objectives of this report are set out in the full DCMS specification, which is reproduced in this document as Appendix A. In summary, the objectives are as follows:

- ❖ To refine the current version of the map by testing the categories and the relationships between the categories with members of the LDAP, the JCLD and other relevant stakeholders;
- ❖ To review the supplementary information already collated in the schedules supporting the map;
- ❖ To expand the information within the schedules, taking into consideration relevant work already carried out by the LDAP in this area. Additional information required falls into 2 categories
 - Quantitative:
 - the number of items currently published in each category (or a reasonable estimate where such figures are not available);
 - the number of items collected by the legal deposit libraries in each category at present;
 - Qualitative:
 - Publishing dynamic (static, accumulative, dynamic);
 - Modes of capture (publisher deposit, harvesting, other);
 - Collection priority;
 - Priority for legislation;
 - Special factors (eg: for rights holders, libraries, users);
 - Other issues (such as territoriality).
- ❖ To isolate the attributes of the different categories and to identify their shared affinities (if any);

LDAP also indicated that it would find it useful for the final report to consider the following and provide appropriate comment:

- ❖ The practicality of phasing regulations covering legal deposit of electronic publications by reference to the 16 descriptions identified in the 'universe'.
- ❖ The metadata issues which may mean different approaches to reflect the practicalities of harvesting versus deposit by publishers in certain categories within the 'universe'.

3. Methodology

As envisaged in the final version of the EPS proposal, dated 4 July 2006, the methodology consisted principally of data collection, review, refinement and synthesis using existing secondary materials in this field, supplemented by interviews with appropriate members of LDAP, JCLD and other stakeholders, as well as by EPS' in-house knowledge and expertise in this area.

3.1 Stakeholder interviews

Originally, it had been envisaged that EPS would carry out 17 interviews, mostly by telephone. In the event, 19 were completed. Of these, 12 interviews – of individuals conveniently located in central London – were actually conducted face to face, offering a better opportunity for exchange of views. The full list of individuals interviewed in the course of the project may be found in Appendix B of this report. Overall a good cross-section of stakeholder interests was achieved in the interviews.

In order to facilitate the interviews, a discussion guide was drafted at an early stage (focusing upon the provisional map which had already been produced) and was approved by the Chair of LDAP. In several cases, however, a much wider range of topics and issues was covered. This was especially true of discussions with a number of key experts (notably Michael Jubb, Sally Morris, Christopher Roper, and David Worlock) which, because of diary considerations, had to take place after the interim report had been submitted. These later conversations were a good opportunity to test ideas first mooted in the interim report. Additionally, separate telephone conversations were held some weeks after the interim meeting with both Andrew Green and Andrew Yeates, both members of the project steering group.

The feedback from individual interviews is not reported separately, but is highlighted at various points in the present report, and has been taken fully into account in the discussion and conclusions which follow.

3.2 Desk research

Work to augment the supplementary tables which accompanied the existing map was largely carried out by accessing internet resources, though a number of printed publications were also consulted and expert input sought by telephone and e-mail. The scope of the research was extremely wide, so individual sources are not mentioned here, but are noted in the updated version of the supplementary schedules which are included as Appendix C of this report.

4. Review and elaboration of original map

4.1 General

In its original form, subject to some modifications and a number of additions, the map which EPS was asked to review (see Chart 1, below) is a useful tool for ensuring that no type of publication currently known is overlooked, and for collating characteristics and quantitative data relating to known types of output. At the same time, however, it is immediately clear that the map was compiled from a traditional print publishing industry perspective. Hence, known print publication types have been mapped on to a typology which takes as its principal divisions the book/serial/other categories which have been in use by the publishing and library communities for a very long time.

The above observation reflects a key finding of the interviews. On the one hand, interviewees in book and serial publishing organizations, even when engaged in significant electronic publishing, felt broadly comfortable with the map. Nevertheless, most also acknowledged the print-orientation of the map. Suggestions from this quarter were, however, for the most part fairly minor. On the other hand, those who had a different background or profile found the map unhelpful or even meaningless in an electronic context and very vulnerable to obsolescence.

4.2 Completeness

As mentioned above, known print publication types have been translated on to the map. Upon closer scrutiny, however, a key category of electronic publication – one which has never existed in print form – is missing, namely the content eligible for legal deposit delivered by real-time information services². Moreover, some further categories could usefully be added, such as newsletters and sheet music. Arguably the former could perhaps be subsumed into Type 2, 'B2B and B2C magazines', although the means of delivery in a networked environment may well be different. Official publications were also mentioned during the interviews, but do not really constitute a separate category, since government output actually spans a large number of the publication categories already defined.

It is also the case that the eastern third of the map is somewhat less developed, and uses terminology that is sometimes over-general and/or unfamiliar, but in any event not immediately transparent. Some categories are also, according to some of the interviewees, misplaced, whilst others are inappropriate in an online environment.

4.3 Affinities

Although an excellent tool in terms of ensuring that no significant print publishing species has been overlooked and that the attributes of each identified

² Here it is important to remember that it is whatever is made available to the public that is potentially subject to deposit, rather than what is actually delivered to any individual user (see LDLA definitions in Section 1.1, above).

category are properly recorded and analysed, the original map does not readily lend itself to the illustration of affinities which exist between types of electronic publication, and which may be helpful in considering how enabling regulations might be framed. For example, it has been suggested that newspapers, classified as serial publications, could perhaps also be categorized as evolving publications under the sub-category web sites.

After careful thought and discussion with some of the project steering group and some of the interviewees, it has been concluded that product/service structure, together with modes of delivery and access may be more useful for the present purpose. In looking at such characteristics, one key point to be remembered about many e-publications is that they are essentially composed of more separable, free-standing components than traditional books and serials. Thus, historically, the components of a book or the articles of a serial have been inseparable (short of physically vandalizing them). For many, if not most e-publications, this is not the case, since an e-book or e-serial will almost certainly consist of multiple discrete files. This property of e-publications is sometimes referred to as granularity, since in considering issues related to e-publications, attention needs to be given to ensuring that each grain of sand (or sugar) can be collected (if wanted). Since the term granularity may not be immediately understood, however, the term 'composite' is used to describe this property of e-publications in this report and is defined more fully in Section 5.2, below.

4.4 Updated version of the original map

It has proved difficult to update the original map of the e-publication universe so as to take full account of the way in which some types of publication can span more than one of the three major categories in the 'tree'. An updated version is included in this report as Chart 2. In the updated version, three new sub-categories have been added, as mentioned in Section 4.2, above. A number of lines have also been introduced in order to attempt to show how some of the sub-categories link to others and could themselves possibly even migrate into different categories. Comprehensive treatment of such links cannot, however, be achieved without detracting from the clarity of the map.

In addition to the updated version of the original map, the accompanying schedules, which contained descriptions of the original 16 categories, have also been updated. Updating includes, in particular, the addition of a new 'box' in each table, containing information on quantities. Where such information is (all too often) not available, a reasonable estimate has been put forward and, in all cases, references to sources have been recorded. The updated schedules may be found in Appendix C of this report.

5. Complementary map

5.1 Rationale

As has been stated above, a complementary map was considered necessary in order to be able more clearly to demonstrate affinities which exist between various types of electronic publication. Whilst a map rooted principally in the print world is useful to ensure that no type of publication currently known escapes consideration, only a map wholly focused upon the nature of electronic publications can fully demonstrate the characteristics and distinctions which are likely to need to be addressed by appropriate codes of practice or by regulation.

The complementary map is included in this report as Chart 3 and is supplemented by an alternative view of the same universe in Chart 4.

5.2 Terminology used in the complementary map

In the complementary map, the universe of electronic content has been split into two major divisions, namely 'Composite' e-content and 'Enquiry-driven' e-content. In case these terms are not immediately transparent, they are defined and discussed below, as is also the term 'web site'.

Composite e-publications

This concept has already been alluded to above, in Section 4.3. Essentially, the term is used to express the fact that many e-publications – probably the large majority – consist of more than one file, and that each file may have a discrete, independent existence.

Thus, for example, an e-book or a fairly 'ordinary' web site is likely to be composed of text – not necessarily contained in a single file – as well as possibly graphic, photographic and other material (including potentially animated clips, video and sound). Each discrete unit of content is capable of being read, heard, or viewed individually and, importantly, the rights status of each could be different. Conceptually, this is not fundamentally different from a printed work, except that in the print world the components are kept together physically by the print layout and by some form of binding. Conversely, the components of an e-publication, not being held together in the same way, can each have a more independent 'life'. Although, by analogy, the ensemble of components in an e-publication can often be considered to constitute a 'digital edition', the 'edition' is more volatile than is practical in print.

Enquiry-driven e-publications

By contrast, an enquiry-driven e-publication does not assemble and make available to users a diverse collection of free-standing files, but allows users to retrieve content from a structured agglomeration of data – i.e. a database (or linked databases) – by means of a process of interrogation in order to satisfy particular information needs. The user does not normally see the ensemble of an e-publication of this sort and is only likely ever to draw out of it a small

proportion of the totality of the data interrogated. Examples include the Royal Mail postcode database or the National Rail Enquiries database.

Web sites

While discussing definitions, it is also worth pointing out that in common parlance (and indeed throughout this report) the term 'web site' is often used with less precision than would be required in regulation. Strictly speaking, the World Wide Web is a subset of the Internet (c.f. the not insignificant number of sites which do not include 'www' in the URL). Similarly, the term 'Internet' is possibly best replaced with a more generic description of a global networked environment. For convenience, however, as stated above, the term 'web site' is used throughout this report with a wide generic meaning.

5.3 Structure of the complementary map

The proposed complementary map has been constructed around four simple questions, which can be applied, for the purpose of categorization, to all published e-content (whether or not the content corresponds to known types of print publication). It is considered that by relying upon fundamental characteristics of structure and e-delivery mode rather than nature of content and print delivery mode, this approach is more likely to be future-proof.

Each of the four questions is discussed fully below, but in summary, the questions are:

1. Is the published e-content composite or enquiry-driven?
2. Is the e-content available online or offline?
3. If online, is the e-content delivered to or collected by the user?
4. If collected, is the e-content unprotected or protected?

When these questions are applied systematically, they lead to six theoretically possible categories of published e-content, which will be discussed, with examples, in the following section. Meanwhile the questions themselves are amplified below.

Question 1: Composite or enquiry-driven?

This question is intended to determine whether published e-content consists of either:

(1a) composite content, that is, pre-existing discrete units (files) of e-content to which users can navigate (e.g. via contents lists, file/folder paths, indexes, dedicated search engines);

or

(1b) enquiry-driven content, that is, e-content held within a database (or linked databases) which will only yield data or information to the user following interrogation.

In case (1a), above, the file(s) accessed by the user already exist within the structure of the e-publication and may be of any type, whether textual, visual or

audio (e.g. Word, Excel, PDF, HTML, JPEG, WMV). In case (1b), each user enquiry will extract or cause to be generated a customized (usually small) subset of the entirety of the data which the database encompasses.

Question 2: Online or offline

This question is intended to determine whether users obtain published e-content either:

(2a) online, that is via some type of telecommunications link, whether fixed-line or wireless;

or

(2b) offline, that is via some tangible 'carrier' such as, for example, a CD, DVD, floppy disc, memory stick, dedicated e-book, or other device, which is delivered physically to users.

Question 3: Delivered or collected

This question is intended to determine whether the means by which users obtain e-content published online is either:

(3a) through delivery (i.e. passively, except where renewal is required), meaning that after initial agreement/request/purchase/subscription, etc., users automatically receive e-content with no further effort on their part until the arrangement expires;

or

(3b) through collection (i.e. actively), meaning that users have to make an effort to retrieve e-content whenever they need it.

Question 4: Unprotected or protected

This question is intended to determine whether access to e-content which is published online is either:

(4a) unprotected, in that the e-content is immediately accessible and is not subject to any sort of barrier;

or

(4b) protected, in that access to the e-content is permitted only after completion of some sort of formality (which may be simple registration, or payment, or may involve higher levels of security).

5.4 Categories of published e-content

When the four questions discussed above are applied to published e-content, six possible theoretical categories may be differentiated (though for practical purposes the number could be smaller). The categories are explained below and are illustrated in graphic form in Charts 3 and 4. It should be noted that although this report does not focus upon offline e-content, which is out of scope, offline forms are referred to for reasons of completeness and context.

A. Composite e-content

A1. Composite e-content delivered to the user either physically or via telecommunications

Examples:

Online: News services delivered in real-time, as for example by Reuters; e-bulletins or newsletters distributed by some charities and other organizations, such as the Wigmore Hall Trust

Offline: CDs/DVDs containing collections of historical material

A2. Online composite e-content collected by the user without formality

Examples:

Personal web sites; Guardian, BBC web sites

A3. Online composite e-content collected by the user following compliance with some formality

Examples:

Most e-journals (except open access journals); richer resources underlying, for example, some newspapers (e.g. ft.com)

B. Enquiry-driven e-content

B1. Enquiry-driven e-content delivered to the user either physically or via telecommunications

Examples:

Online: No obvious example, but databases (probably small ones) could, at least theoretically, be delivered *in toto* (including all necessary software) via telecommunications

Offline: Thomson Scientific [ISI] journal citations databases used to be (still are?) delivered on disc; formerly, *CD-ROMs in print* was delivered this way; also *OED*.

It should be noted, however, that Category B1 will not only contain few e-publications, but is also obsolescent.

B2. Online enquiry-driven e-content collected by the user without formality

Examples:

British Library catalogue; BT directory enquiries database; AA Route Planner

B3. Online enquiry-driven e-content collected by the user following compliance with formality

Examples:

Royal Mail postcode database; Various B2B and B2C databases (e.g. Glass's Guide); Companies House Direct web site.

5.5 Application to legal deposit

A number of implications (and related questions) emerge from the above consideration of categories of e-content potentially subject to deposit.

Firstly, Category A2 embraces the vast majority of the web domain, including large amounts of user-generated content. Currently, this is the only category which (because the content is open to public access) could readily be harvested without reference to the publisher. In this context, however, it is worth noting that a particularly time-consuming task for the UK Web Archiving Consortium (UKWAC) activities was obtaining adequate licences from rights holders, especially when faced with persistent absence of response.

Related to this point, it is worth noting, also, that there are currently at least two ways of harvesting content. On the one hand, an 'inert' snap-shot can be taken in a manner akin to a screen-grab (obtaining, for example, '.mht' web archive files), thereby preserving the visual impact of the site. On the other hand, in some circumstances it may be more valuable to harvest the discrete capsules of content which make up a web site.

Another consideration is that Categories A3, B2 and B3 are likely to carry a heavy overhead of communication and liaison. Even with the backing of regulations, it will be necessary for the LDLs to embark upon a dialogue with the publishers, both at a strategic or 'diplomatic' level, and at a technical level, in order to communicate the LDLs' objectives and to obtain content which is protected and which may also require software in order to facilitate data interrogation and use.

A further consideration relates to priorities that have already been set, namely: (i) offline; (ii) e-journals; and (iii) web material. Pursuing a revised strategy in terms of the overall definition of e-content potentially eligible for deposit would not, however, conflict with these priorities nor invalidate the schemes already in operation or in a planning phase. On the contrary, these activities will continue to be important in terms of identifying issues relevant to metadata and to impact assessment.

6. Issues arising from the complementary map

Whilst, for the purposes of possible regulation, the categorization discussed above makes the definition of what is potentially subject to deposit easier than would be the case with the much larger number of categories envisaged in the original map of the e-publication universe, a number of questions and issues do nevertheless arise, some of which came to light during the interview exercise.

- ❖ A typology which makes it possible, whether by regulation or otherwise, to define everything in the e-publishing universe could easily become a *bête noire* for publishers. The existence of a catch-all typology is not at all *per se* a disadvantage, but it does mean that there is likely to be much greater focus and pressure upon other aspects of deposit, such as territoriality and safeguards for publishers around the key issue of access.
- ❖ It should be noted that any given web site may embrace types of e-publication which fall into more than one of the categories envisaged in the complementary map. Equally, a web site could embrace the electronic counterpart of more than one of the categories envisaged in the original map.
- ❖ As already mentioned in a previous section, web site resources can and often do embrace types of media – i.e. sound and motion video – which fall outside the LDLA. Here, there is a fundamental difficulty, since important, but integral parts of the digital content spectrum are excluded from the scope of the Act. If only part of certain multi-media web sites were subject to deposit, this would undermine the effectiveness of the Act, quite apart from raising the issue of whether partial harvesting is practically feasible.
- ❖ Whilst it is relatively easy to define all e-publications at the level of the lowest common denominator (LCD), the question of defining ‘digital editions’ is more difficult because of their potential volatility. Nevertheless, the ensemble of a digital edition, which would incorporate ‘look and feel’, can be as important (or even more important, on occasion) as the content. At the same time, however, the LCD approach can provide a solution which may potentially satisfy the objections of some aggregation services.

In the view of EPS, none of the above-mentioned issues alters the validity of the approach implied in the complementary map.

7. Correlation of original and complementary maps

It has been made clear throughout this report that both maps have validity in considering issues of legal deposit – hence the term ‘complementary’, rather than ‘alternative’. Accordingly, an attempt has been made to show how the two maps relate to each other by means of a two-dimensional matrix setting the categories of one map against those of the other. This correlation of the two maps is presented in this report as Chart 5.

8. Statistical data

8.1 Library statistics

One of the tasks mentioned in the DCMS brief for this study (see Section 2, above, and Appendix A) was to obtain information on 'the number of items collected by the legal deposit libraries in each category at present'. Accordingly, EPS contacted each of the LDLs and requested such statistics as they could reasonably supply. Five of the LDLs provided data, which are summarized in the table below. The figures in the table do not take account of any UKWAC activities.

	BL	Bodley	CUL	NLW	TCD
Latest year	2005/06	2005/06	2005/06	2005/06	2005/06
Serial parts**†					
Online	32,878	1,037		7,080	101
Offline	986				312
Monographs					
Online	2,149	298		443	
Offline	922				
Other		119		20	1
TOTALS	36,935	1,454	1,389	7,543	414

* These figures need to be divided by a factor of possibly 8-10 to gain an idea of how many titles may be involved

† Except TCD, for which the figures represent titles.

Overall, these figures offer no more than a very superficial impression of relative activity and of coverage relative to the universe. Some of the apparent disparities arise because of different collection priorities (e.g. NLW collects a lot of newspapers), or different approaches to the collection of statistics (e.g. Cambridge does not differentiate between serials and monographs, and provided no online figures). It should be noted also that some materials are sent to the Agency for the Legal Deposit Libraries which forwards them (usually monthly) to the five libraries it represents. Such items (perhaps 150-200 items per month on average) may not be fully reflected in the above figures and, in any event, there may be a degree of overlap.

A number of conclusions may be drawn from this exercise. Firstly, the data received were, quite obviously, not always broken down in very much detail (there being significant variation between the LDLs on this matter). Secondly, in most cases, the break-downs did not correspond, except at the very top level (i.e. monographs, serials, other), to the categories of e-publication envisaged in the original map of the universe. Thirdly, the ways in which such accessions are described vary amongst the LDLs.

As regards UKWAC activity (in which the BL, NLS and NLW participate), access was being provided to 1,100 web sites as of March 2006.

8.2 Publishing output/population statistics

EPS has sought statistics relating to the output and/or populations of all 16 of the categories of e-publication included in the original map. In almost all categories, solid and reliable figures are lacking. Quite simply, no bibliographical control exists for e-publications. The *British National Bibliography (BNB)* does not record e-publications comprehensively, generally doing so only when there is no printed counterpart, and is in any case based upon BL acquisitions (whether by deposit, purchase or donation), so does not extend to many online resources. Nielsen BookData similarly is nowhere near comprehensive in its coverage.

In consequence, EPS has been obliged most of the time to use meagre data to derive figures which are at least plausible, but which can often be justified only on the basis of reasonable probability rather than hard evidence. The figures and estimates are, as already mentioned above, included in a new 'box' in each of the tables which make up the schedules in Appendix C of this report, and are accompanied by information on sources and by an explanation of EPS conclusions.

Generally, the figures are of 'manageable' proportions, except in the area of web sites for which the figures, however derived or calculated, are always in multiple millions. A summary of order of magnitude figures for each of the 16 categories of e-publication included in the original map is given in the table below.

Type of e-publication	Order of magnitude output/population figures
1. Scholarly journals	Between 5,000 and 7,000
2. Magazines	B2B/Professional c. 5,000; B2C c. 3,500
3. Newspapers	c. 1,500 including regional and local
4. Databases	Probably not higher than tens of thousands, possibly fewer.
5. 'Flat' e-books	c. 100,000 'institutional'; c. 70,000 retail; but many replicate titles published in print.
6. Interactive e-books	1.7 million; currently increasing at a rate of c. 400,000 p.a.
7. e-prints	2.5 million
8. Blogs, discussion lists	Conservative estimate 6.2 million (could be >10 million)
9. 'Flat' web sites	Probably not higher than tens of thousands, possibly fewer.
10. Database driven sites	4-5,000 conference proceedings; c. 70,000 other 'grey lit'
11. Conference proceedings	15-16,000 p.a. (doctoral)
12. Theses/dissertations	No figure found
13. Digital maps	48-52,000 p.a.
14. Patent specifications	c. 700 image banks in UK; no figure for annual increase
15. Image banks	Probably no higher than 1,000
16. Looseleaf	

9. Metadata issues

There is scope for a whole report in its own right on the metadata issues surrounding electronic deposit. Discussions with Book Industry Communication (BIC) have been very helpful in this connexion (and no doubt will be so in the future). For the present report, however, there are a number of fairly general points which can be categorized, not solved, fairly simply. Here, it may be helpful to refer to the complementary map and the matrix correlating it to the original map.

In outline, there are initially three major considerations, *viz*:

- ❖ The essential metadata required adequately to describe specific resources
- ❖ The practicality of obtaining such data from publishers
- ❖ The depth of data desired by the LDLs to make resources usable/accessible

Firstly, as regards the essential metadata required to provide a reasonable description of resources, there is a significant difference between the metadata which would describe Category A e-publications and the metadata which would describe Category B e-publications. For Category A, at least down to A2, existing schemata such as the Online Information Exchange (ONIX) standard should be entirely adequate, whilst for Category B, it seems likely that a different approach will be necessary, possibly on an *ad hoc* basis.

Secondly, whether such metadata are readily available from publishers depends to a considerable extent upon their workflow, and whether, for the purpose of compliance with deposit requirements, they would willingly make adjustments to workflow to accommodate a standard such as ONIX. Potentially there are significant cost considerations here for some publishers, especially those which, for whatever reason, are reluctant to adopt current standards in this area.

Thirdly, the metadata essential for an 'adequate' description of a resource may not be sufficient to meet all LDL needs, particularly where the requirements of technology and preservation are concerned. It seems likely that a 'full' set of metadata would go well beyond what publishers can readily supply from routine workflow. Undoubtedly, there could be important cost considerations here for the LDLs if they had to supplement metadata obtained from publishers.

Beyond the three points mentioned above, however, it has to be borne in mind that whilst some sort of accommodation is likely to be achievable between LDLs and 'formal publishers' (corresponding to the core of publishers which produce 80-90 per cent of print output in the UK), there are many, many more 'informal publishers' in the electronic universe making available content which is to a considerable extent user-generated. Probably the vast majority of the latter will not even understand the basics of metadata. In such circumstances – normally where some form of harvesting will be necessary – not even basic metadata are likely to be available.

10. Practicality of phasing regulations

To a considerable extent, this is a debate for LDAP after the present report has been completed and digested by the Panel. Three points relating to the practicality of phasing are immediately apparent, however, *viz*:

- ❖ One major consideration in moving towards regulation will be the requirement to produce an Impact Assessment, which will need *inter alia* to address questions of cost on both sides of the 'equation' – that is, for publishers and for the LDLs. Unless the LDLs are able completely to exclude several categories from the original map, on the face of it, there is merit in looking at the possibility of pursuing the complementary approach as the basis for regulations, proceeding in possibly just two or three steps. Either way, the task of impact assessment is substantial.
- ❖ A second major consideration in moving towards regulation will be the speed with which the LDLs will be able to communicate with publishers and to handle the volume of new accessions. Clearly some sort of phasing is likely to be necessary, though that could be handled within any regulations by staggering the implementation date for different categories.
- ❖ Thirdly, though perhaps this is implicit in the first bullet point above, it may be necessary to assess whether the LDLs' objectives can at all be achievable by voluntary codes of practice, including those already in operation or under consideration. There are, however, important areas where voluntary compliance, such as is being trialled with offline publications, would be wholly impractical (as, for example, with Category A2, where the numbers are huge), or liable to result in unacceptably slow progress (as, for example, in Categories A3, B2, and B3). Here again, it is worth referring to the difficulties experienced by UKWAC in seeking permissions for a voluntary scheme (non-responses reached almost 75 per cent in the case of one of the consortium members).

Charts

Chart 1: Original map of the universe of e-publications

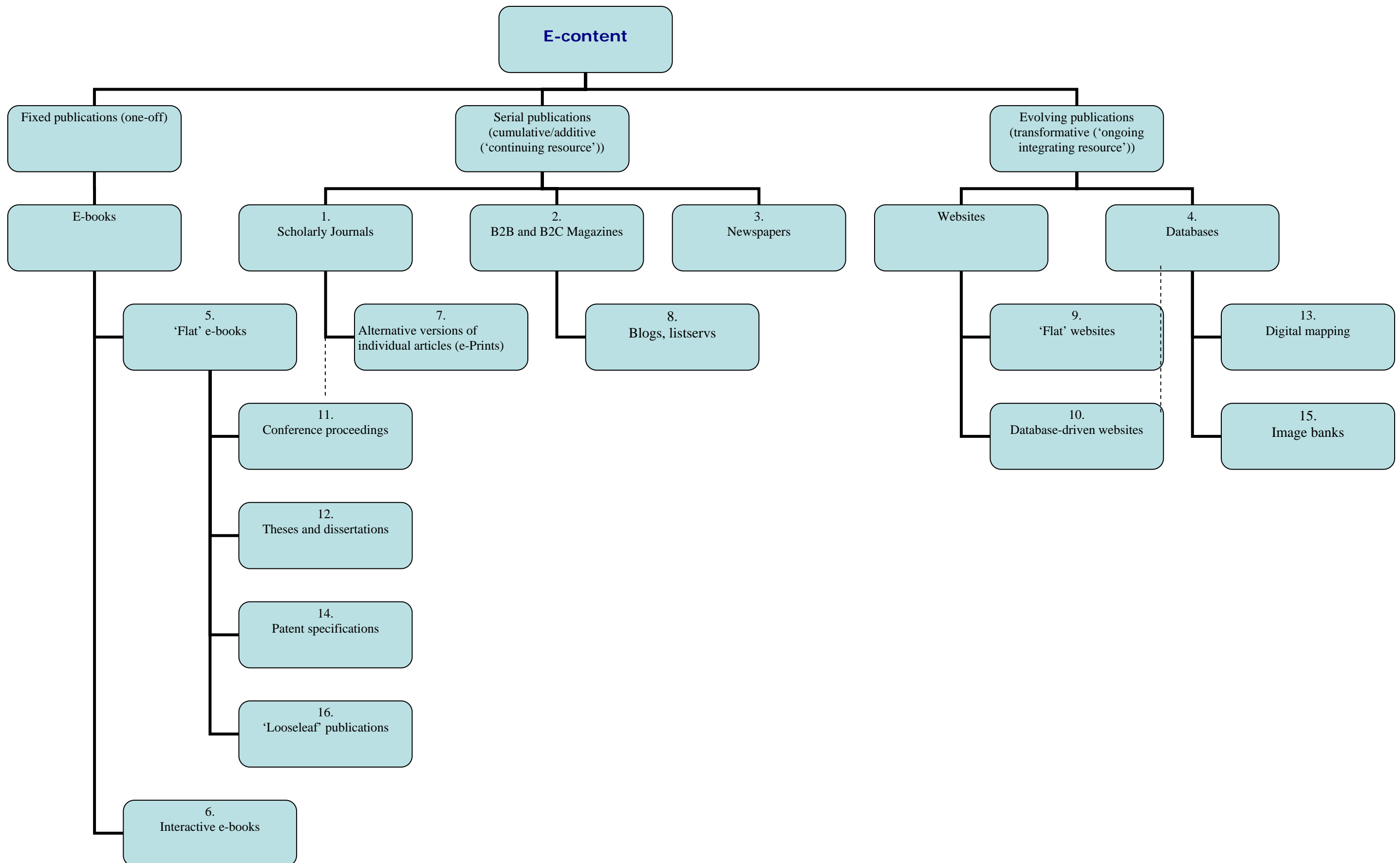


Chart 2: Updated version of the original map

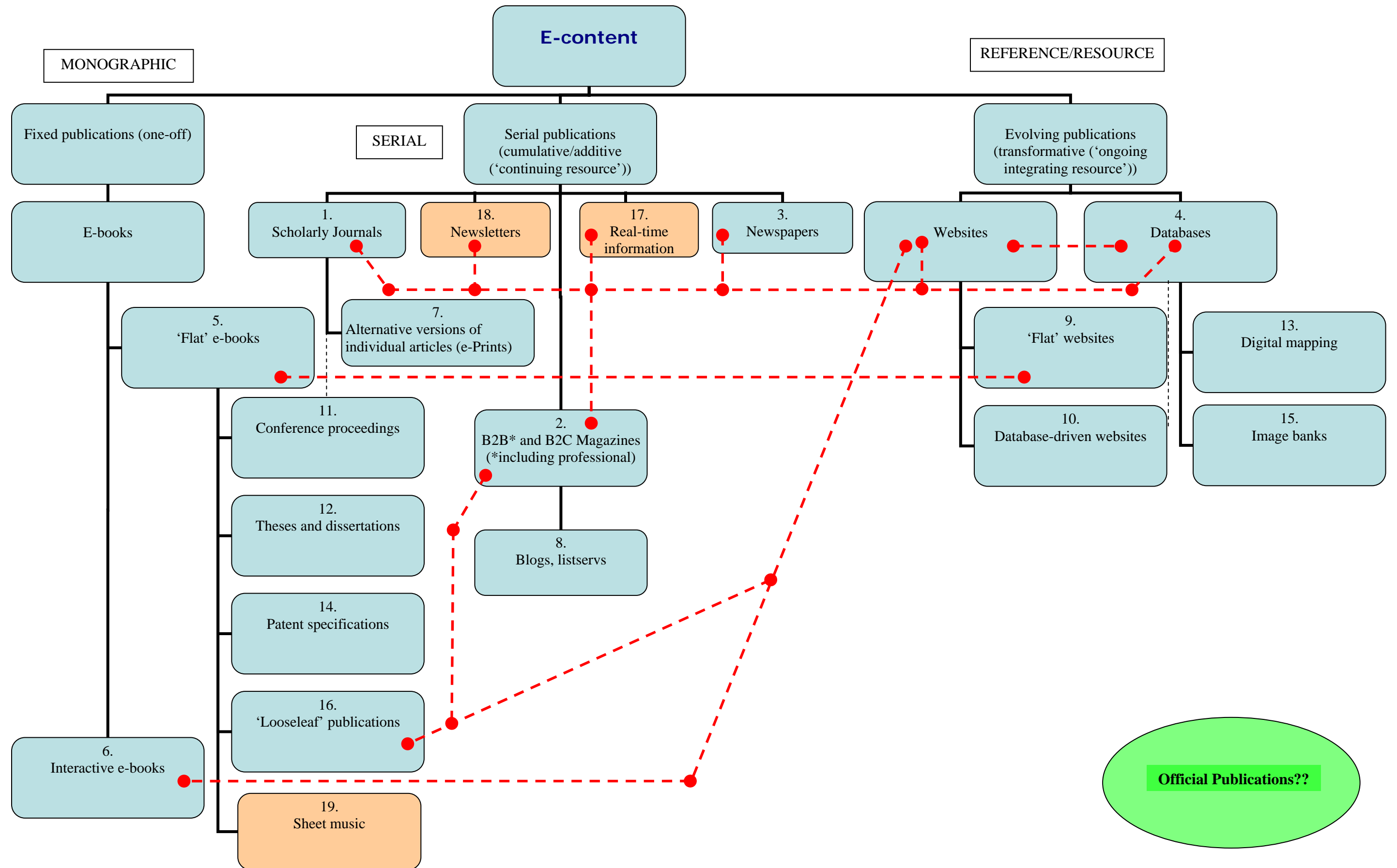


Chart 3: Complementary map of the universe of e-publications

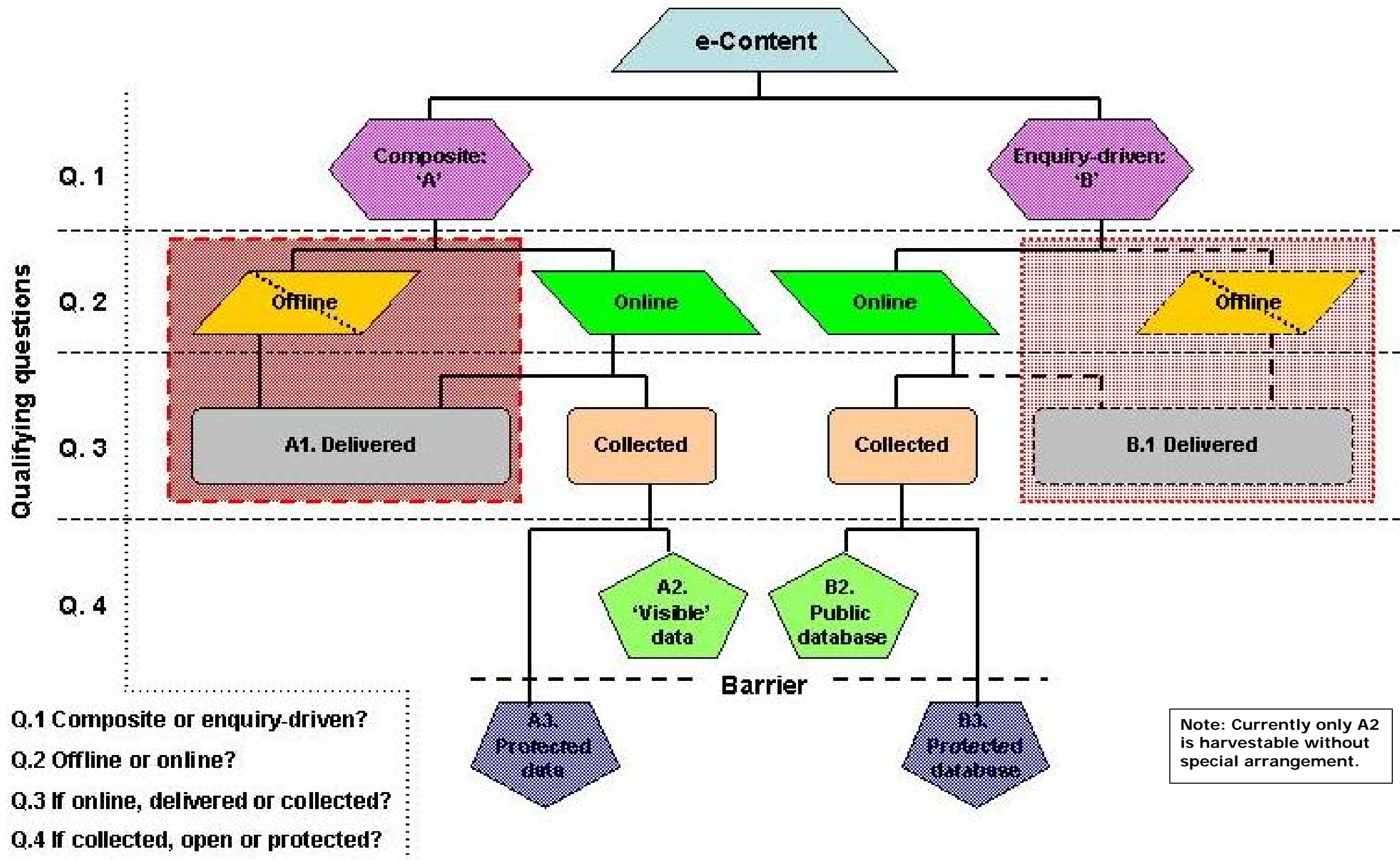


Chart 4: Alternative view of the complementary map

22 October 2006

1. Is it composite or enquiry-driven?	Composite: A		Enquiry-driven: B			
2. Is it offline or online?	Offline	Online	Online	Online		
3. Is it delivered or collected?	Delivered: A1		Collected		Delivered: B1	
4. Is it open?			Visible: A2	Public: B2		
... or protected?			Protected: A3	Protected: B3		

Chart 5: Matrix showing relationship of original and complementary maps

22 October 2006

		1 Scholarly journals	2 B2B/B2C magazines	3 Newspapers	4 Databases	5 Flat e-books	6 Interactive e-books	7 e-Prints	8 Biographies	9 Flat web sites	10 Database-driven web sites	11 conference proceedings	12 Theses/dissertations	13 Digital mapping	14 Patent specifications	15 Image banks	16 Loose-leaf publications	17 Real-time information	18 Newsletters	19 Sheet music
Granular	A1. Delivered	?	?	?																
	A2. Visible data																			
	A3. Protected data																			
Enquiry-driven	E2. Public database																			
	E3. Protected database																			

Key:

Normal supply method

Occasional supply method

Rare supply method



Appendixes

Appendix A: Terms of reference

SPECIFICATION

Legal Deposit Advisory Panel

Specification for a study to refine the map of the universe of online digital publications potentially eligible for legal deposit

Summary

This is an invitation to tender for a commission to refine the map of the universe of online digital publications potentially eligible for legal deposit.

The total cost of the commission, including VAT and expenses, will not exceed £19,500.

To be considered for this commission, intending bidders must return their proposals by 13:00 on Monday 15 May 2006.

1 Background

Following the enactment of the Legal Deposit Libraries Act in 2003 the Legal Deposit Advisory Panel ('the Panel') was established in 2005 as an Advisory Non-Departmental Public Body with the task of advising the Secretary of State for Culture, Media and Sport on the operation of the law of legal deposit in the United Kingdom. The Panel is engaged in making recommendations to the Secretary of State on the content and timing of Regulations for implementing the Act with reference to the deposit of particular classes of non-print material. These were not within the scope of legal deposit before the 2003 Act.

The Panel is currently working on recommendations on three formats: offline publications, electronic journals and websites.

As a preliminary, however, to preparing its recommendations the Panel wishes to develop a 'map' of the entire range ('universe') of online digital publications currently issued in the United Kingdom that are potentially eligible for legal deposit. This map will help to identify discrete categories of such publications, and the relationships between them. It will then be possible to produce a programme of recommendations that will, over time, cover the whole range of publications, minimising overlap and omission. (It is recognised that future publication types will emerge which cannot yet be anticipated today, let alone described; the mapping exercise is not, therefore, a once and for all process.)

A draft of the map has already been prepared, together with a provisional schedule or list of the characteristics of each category. This has been done, in collaboration with the Panel, by the Joint Committee on Legal Deposit (JCLD), a voluntary body that brings together publisher representatives and the legal deposit libraries. The map shows the different categories of *online* digital publication, and their inter-relationships. It is supplemented by a draft schedule of the characteristics of each category. These characteristics include: a definition or description of the category; publishing dynamic; mode of availability; proposed method of capture by the deposit libraries; links to external data; library collection priorities; issues for rights holders; special issues for libraries and users; estimation of priority for making recommendations. The schedule does not yet contain quantitative information about the size of the published output in each category.

The most recent version of the existing map and schedules is attached as an appendix to this invitation.

2 Aims of the commission

The organisation selected for this commission ('the researcher') will be expected to carry out the following tasks:

- to refine the map itself, testing the categories and their inter-relationships with members of JCLD, the Panel and any other relevant stakeholders
- to review the information already supplied in the supplementary tables, with the help of JCLD members, LDAP members and other stakeholders
- to add to the information in the tables, with the help of JCLD members, members of the Panel, and other stakeholders. In particular, to add quantitative information such as the number of publications in each category, and qualitative information about publishing dynamic, modes of capture, collection priority, priority for legislation and special factors
- to take account of relevant work already done by the Panel, eg on 'territoriality'

- to draw out the attributes of the different categories in the map affinities and the affinities they share, as a preliminary to identifying coherent packages of recommendations for regulation (this is especially important in the case of e-journals, so that early progress can be achieved on that category)

- incorporate the results in a written report for the Panel, to incorporate revised versions of the map and of the schedules attached to it.

It is expected that about a half of the available research time will be spent on analysis and data collection (reviewing the map, populating the schedule, etc.), and a half on synthesis (considering the attributes of each categories and the links between them).

It is recognised that it will not be possible to gain full and authoritative quantitative data for all publication categories: the researcher should be prepared to supply reasonable estimates where appropriate, and to point to gaps in the evidence and how they should be filled in future.

The resulting report will aid the Panel in drawing up a coherent programme of recommendations, and will supply information useful in framing the content of recommendations on individual formats.

3 Outputs and deliverables

The researcher is expected to produce by 31 August 2006 a final report, in electronic form with three copies on paper, to include the following three elements:

- a final version of the map (diagram) showing the categories of online publication potentially eligible for legal deposit, and the relationships between them
- a final version of the schedule, detailing for each category of publication a series of qualitative and quantitative information of assistance in preparing for recommendations. This information will include the following (the second set does not feature in the current version of the map):
 - description/definition
 - publishing dynamic (static, accumulative, dynamic)
 - mode of capture (publisher deposit, harvesting, other)
 - links to other sites
 - collection priority
 - special issues for rights holders
 - special issues for libraries
 - special issues for users
 - priority for legislation
 - other issues (eg territoriality)

 - number of items currently published in each category
 - number of items collected by the legal deposit libraries in each category at present

- a commentary drawing out the common attributes linking different categories of publication included in the map, and the affinities between them, to aid the Panel in compiling a coherent and, for the moment, complete list of packages of recommendations.

The researcher will also provide an interim report, with provisional versions of these three elements, by 31 July 2006.

The researcher should be prepared to give a presentation, if requested, at a meeting of the Panel scheduled for 12 September 2006 in London.

4 Management of the commission

DCMS is responsible for all administrative and financial aspects of this commission.

The Panel will designate one of its members as a Contract Manager, whose role will be to maintain contact with the researcher throughout the period of the commission and monitor its progress and submission.

The researcher will be able to ask the Contract Manager for advice on the conduct of the research as it progresses.

5 Timetable

28 April 2006	Tender invitation issued
15 May 2006	Deadline for receipt of tenders
w/b 22 May xx	Tenders assessed and invitation to interviews sent Interviews (in London)
31 May 2006	Commission awarded by the Panel
w/b 5 June	Start-up meeting to clarify operation of commission
31 July 2006	Interim report delivered to the Panel
31 August 2006	Final report delivered to the Panel

6 Skills and experience required

Knowledge of the principles and practice of existing legal deposit in the UK

Familiarity with the full range of existing digital publications and their attributes

The ability to co-operate effectively with librarians and publishers in order to elicit views and information from them

The ability to collect, analyse and synthesise substantial amounts of qualitative and quantitative data within a short period of time

The ability to present data, results and conclusions systematically and comprehensibly, in written and oral form.

7 Costs

The total cost of the commission, including VAT and expenses, will not exceed £19,500.

8 Evaluation of the tenders

Tenders will be evaluated according to the following criteria:

- the feasibility of the work plan
- the strength of the methodology
- the experience and skills of the researcher(s) and the organisation
- the proposed costs and their distribution
- value for money

9 Delivery of the tender / Contact details

Tenders, which should not exceed ten A4 pages in length, should include the following information:

- CV(s) of the researcher(s) responsible for carrying out the commission
- a work plan for the operation of the commission, with milestones
- the research methodology proposed
- an analysis of risks associated with carrying out the commission
- the number of days proposed for each researcher
- a statement about the qualifications for selection for the commission
- costings for carrying out the commission including a detailed breakdown including staff costs, travel costs, VAT etc.
- the names and contact details of recent clients for reference purposes

For more information on this commission, contact Susan Fairweather 020 7211 6482

Appendix B: List of interviewees

Table of candidate interviewees, target quotas, and interviews completed

The full list of candidate and actual interviewees is set out below. The candidate list was compiled from suggestions put forward by members of LDAP, to which EPS added a couple of names. Between formal meetings of the project steering group, telephone conversations were also held with Andrew Green and Andrew Yeates.

	Quota	Done	Comments
	18:	19	
Journals: Commercial	2:		
Cliff Morgan, Wiley			e-mail comments
David Hoole, Nature		Y	
Steve Hall, Journal Sales & Marketing Director, Blackwell		Y	
Ian Bannerman, T&F			
Journals: Learned Societies and University Presses	2:		
Terry Hulbert, IoPP		Y	
Desmond Reaney, IoPP (re data structure)			
David Smith, CABI			
Robert Parker, RSC			
Martin Richardson, OUP		Y	
Michael Holdsworth, CUP			
B2B and Professional	2:		
Trevor Fenwick, Euromonitor		YF*	
Keith Jones, RBI			
Dick Greener, Thomson Legal			
Jo Howard, Marketing Director Business Info., Experian		YF	
Educational and Reference	1:		
Peter McKay, Hodder Arnold		Y	
Simon Lake, Pearson Education			
Libraries	2:		
Clive Field/John Tuck, BL		YF	John Tuck twice
Michael Jubb, RIN		YF	
Robin Adams, TCD			
Peter Fox, Cambridge UL			
Simon Tanner, KCL		YF	
News organizations	2:		
Cathy Smith, BBC			
Mike Seery, Economist			
Stephen Dunn, Guardian		YF	
Guy Black, Telegraph			
Henry Manisty, Reuters		YF	
Magazines	1:		
Aimee Nesbitt, National Magazines			
Trade Associations	3:		
Sally Morris, ALPSP		YF	
Michel Mabe, STN			
Alex White, AOP		Y	
David Newell, NS			
Judith Dunn, NPA		Y	
Ian Locks, PPA			

Brian Green, BIC/Editeur		YF	Stipulated by LDAP
Government	1:		
Michel Woodman, DTI		YF	
Other	2:		
Christopher Roper		YF	Mapping specialist
David Worlock		YF	General overview
F Face-to-face interview			
* Informal conversation before the discussion guide had been finalized			

Appendix C: Updated schedules attached to original map

The Universe of e-Content for potential Legal Deposit

(Updated from version produced by Sally Morris – v10, 27 March 2006)

Introduction

The following schedules provide detailed descriptions of each of the categories of e-publication which were identified in the 'map' originally produced by Sally Morris in March 2006. Descriptions have been reviewed and updated by EPS between July and October 2006.

1. Scholarly Journals

Description	Publications produced at regular or irregular intervals, intended for the academic and/or professional market, containing one or more items from one or more of the following categories: (i) original articles reporting research; (ii) articles reviewing previous articles in a given research field; (iii) book (or other) reviews; (iv) educational articles; (v) editorials and opinion pieces. Original research articles (and sometimes review articles) are generally peer reviewed. Some (the minority) contain advertising, which may be regionally specific.
Fixed/additive ('continuing resource')/transformative ('ongoing integrating resource')	Additive ('continuing resource'). There are strict conventions about not changing previously published articles – errata may be published and linked to the original article.
Free or Restricted access	Hitherto usually restricted, though a small but growing number of titles is published under an 'Open Access' model which enables some or all articles to be available free of charge from the date of publication. Also, publishers may make back issues available free of charge after a (variable) period.
Pull/harvest or push/deposit	Push/deposit – possibility of authorised pull
Linkages	Errata (see above). Cited and citing articles (see below). Letters to the editor/post-publication comment may be linked to individual articles. Raw or organised data on public or publisher's site. Journal website. Publisher/owner (e.g. Society) website. Book reviews may be linked to book publisher's and/or online retailer's site. May also contain links to: Instructions to contributors; Copyright assignment/licence forms (and other publishing-related administration). Comments forum (raw or moderated) – pre- or post-publication 'open peer review'.
Collection priority: [1 = essential (national), 2 = essential (regional), 3 = discretionary, 4 = sampling only, 5 = do not collect]	1
Special issues for rightsholders (e.g. commercial, legal)	May make use of third-party software under licence. Individual articles are the normal unit of consumption, so widespread easy availability of single articles could undermine the business model. Cost of modifying systems and/or converting outputs to LDL-friendly formats.
Special issues for libraries (e.g. technical, ephemerality)	Publishers use a variety of formats (PDF – a proprietary format – is widespread either as well as or instead of HTML or, increasingly, XML) and may or may not be susceptible to standardisation (see e-journals pilot). Smaller publishers who host their own journals may lack technical and financial resources to provide content in optimum formats for preservation. (Very small publishers tend to use third-party hosts, so not a problem.) See also issues arising from e-journals pilot. Associated files such as DTDs, schemas, macros, style sheets etc may be required in order to make sense of the publisher's tagging. There will be cost implications of formatting non-standard content such as supplementary materials.
Special issues for users	
Notes	May replicate print journal – in case of intentional or unintentional variation, electronic version is increasingly often deemed definitive. May also include non-printable material (e.g. manipulable images/data, moving images). Frequently include links to cited articles in other journals (usually using the DOI and CrossRef system, which means that the item linked to is more likely to be found when web addresses change). Some journals may make pre-publication versions of articles freely available until publication (see also 7, e-prints). Although the e-version may replicate the print version, individual articles may be published electronically ahead of print, with the print issues essentially acting as archives.
Feasibility of moving to Regulations	
Quantitative indicators	UlrichsWeb lists a total of 22-24,000 journals globally, and indicates that

	<p>a total 6,740 academic or scholarly journals are published in the UK. Research by the ALPSP some years ago suggested that (depending upon what is understood by published in the UK) UK journals constituted c. 25 per cent of global journals. If applied to the Ulrichs total, this percentage results in a figure of between 5,500 and 6,000. This figure corresponds well with the Ulrich's figure of 6,740. According to ALPSP, around 90 per cent of journals are available electronically.</p>
Composite or enquiry-driven	Composite

2. Magazines

Description	Publications produced at regular or irregular intervals, intended for corporate or individual customers. Contain commissioned articles, illustrations; may contain letters from the general public, with an implicit licence to publish and archive. Frequently contain advertising, which may be region-specific.
Fixed/additive ('continuing resource')/transformative ('ongoing integrating resource')	Mostly additive ('continuing resource'); however, potential to become transformative ('ongoing integrating resource'), through wiki-style reader responses
Free or Restricted access	Often restricted; some may be provided free of charge to a specified list of recipients ('controlled circulation'); others may be free to all, funded by advertising or sponsorship
Pull/harvest or push/deposit	Probably push – free titles could be pull
Linkages	Related website (see 9 & 10, websites). Advertiser websites.
Collection priority (1 = essential (national), 2 = essential (regional), 3 = discretionary, 4 = sampling only, 5 = do not collect)	1, 2
Special issues for rightsholders (e.g. commercial, legal)	Visual presentation/layout may be important – not just a question of content structure, but also presentation.
Special issues for libraries (e.g. technical, ephemerality)	Likely to be a wide variety of non-standard formats used in comparison with scholarly journals. Titles may come and go quickly. Publishers may be small and have few resources to meet the requirements of e-deposit especially if these involve standardisation. Need to determine whether reader-contributed elements (such as wikis) should form part of the archive.
Special issues for users	
Notes	Print and electronic content may differ widely. In addition, magazines may have associated websites which do not necessarily replicate the printed content. Magazines may have different regional or customer-driven format (such as handbag-sized) print versions.
Feasibility of moving to Regulations	
Quantitative indicators	A survey by PPA Marketing (http://www.ppamarketing.net/cgi-bin/wms.pl/60#business) indicates a total of 5,108 professional and business magazines published in the UK (November 2005). A survey by PPA Professional (<i>Professional Media: Connecting Business 2005</i>) indicates that print remains the bedrock of B2B magazine publishing - 47% of publishers reported that none of their titles are published online, whilst 37% of publishers produce all titles online. However, with online investment and activity strongest amongst larger publishers, who publish a large percentage of B2B titles, the total number of magazines and leading titles online is likely to be higher. The number of B2C magazines in the UK is estimated at 3,366 according to the PPA Marketing survey (November 2005).
Composite or enquiry-driven	Composite

3. Newspapers

Description	Generally daily, weekly, or fortnightly publications containing news and comment and produced in combinations of timed and regionalised or local editions. May include syndicated and other content licensed from third parties. Almost always contain advertising, which may be region- or edition-specific. Electronic version may be updated continuously and may publish content that does not appear in printed editions; web sites or sub-sites may be database-driven.
Fixed/additive ('continuing resource')/transformative ('ongoing integrating resource')	Additive ('continuing resource'). In some cases, search may enable recreation of the digital equivalent of much of the paper edition. Generally, however, continuous updating may require 'snapshot' approach akin to that of dynamic databases – see 10.
Free or Restricted access	Both models are in operation, sometimes within the same site. PDF versions of print pages not available via the web.
Pull/harvest or push/deposit	
Linkages	Related website may not (or not only) parallel print content. It could contain more or less than the print edition. Advertiser links
Collection priority (1 = essential (national), 2 = essential (regional), 3 = discretionary, 4 = sampling only, 5 = do not collect)	(1) and (2) Need to replicate comprehensive coverage of national and local titles maintained by BL in print environment.
Special issues for rightsholders (e.g. commercial, legal)	Not all content in printed newspaper may be licensed for e-publication. Special issues of regionality – adverts, even content may be very locally specific. Potentially major issues of legal liability (e.g. defamation, contempt of court, third-party intellectual property - 2003 Act already provides for defamation); content may need to be withdrawn completely at very short notice, either temporarily or, if undertakings have been given that it will not be re-published, permanently – need reassurance about how LDLs would handle this. Differently sized print issues (such as tabloid or broadsheet) may also be published, which may or may not contain all the same content.
Special issues for libraries (e.g. technical, ephemerality)	E-edition may vary considerably from paper edition. Should PDFs or online content be collected? Content may vary through the day; at which point is the edition to be considered 'definitive'. Local variations exist of both regional and national titles.
Special issues for users	
Notes	'Newspapers' are now produced in a variety of formats; printed, E-Newspaper (based on PDF format as surrogate for printed page) and online. Newspaper websites increasingly demonstrate the characteristics of evolving/transformative ('ongoing integrating resource') publications. See records of discussions prior to and during process of finalising 2003 Act.
Feasibility of moving to Regulations	
Quantitative indicators	There are 1,301 regional and local newspapers in the UK, including 110 dailies, 22 Sundays, 522 paid weeklies, and 647 free weekly newspapers. (www.newspapersoc.org.uk/Default.aspx?page=8) Newspapers are increasingly deepening their penetration of local markets and growing audiences by developing their presence online. This can be in the form of a website providing basic information, or a more comprehensive stand-alone news-site which is subscription or advertising-based. Stand-alone news-sites are likely to publish news on a more frequent basis. The content on these sites is typically different from that published in the print version. The Newspaper Society estimates at least 800 websites of both these types (www.newspapersoc.org.uk/Default.aspx?page=1227) There is also a trend towards newspaper replications, or e-editions, which preserve the look and feel of the print edition whilst offering features that capture the benefits of reading news online.

Composite or enquiry-driven	Composite
-----------------------------	-----------

4. Databases

Description	Collections of content organised for retrieval and viewing. Content may be 'pure' data (e.g. numerical data points), metadata (information about other objects or natural/legal persons), text, images, or other object types. Entries in 'secondary' databases may consist of summaries of and/or links to content elsewhere. Some information services are actually driven by multiple databases which generate custom output 'on the fly' according to users' requests. Note that compendious collections of discrete items (such as the collected works of authors) are not considered to be databases in this sense.
Fixed/additive ('continuing resource')/transformative ('ongoing integrating resource')	Transformative ('ongoing integrating resource') – updated data points tend to replace previous versions
Free or Restricted access	Both models in operation.
Pull/harvest or push/deposit	Concept of regular snapshots (i.e. record of entire content – or samples thereof - as it stands at a particular date and time) may have to be considered as an acceptable way of capturing change, if unable to capture by streaming. Only free databases may lend themselves to harvesting.
Linkages	Widespread. Secondary databases frequently link to summarised content; others may link to website of person or organisation described; etc.
Collection priority (1 = essential (national), 2 = essential (regional), 3 = discretionary, 4 = sampling only, 5 = do not collect)	Many will be 1, but 3 would possibly be the general rule, as priority will depend upon the nature of the content, the relationships/linkages to other forms of publication, and existing (alternative to LDLs) 'trusted repository' provision for long-term archiving and access.
Special issues for rightsholders (e.g. commercial, legal)	Identification of rightsholders is an issue.
Special issues for libraries (e.g. technical, ephemerality)	File formats and data structure very important – changes will have significant effect on content management; absence of standards presents a problem. Database template will need to be delivered in addition to database itself. May not be able to access or be authorised to access proprietary formats; preservation may thus not be possible
Special issues for users	
Notes	Snapshot may be only viable method of collection? Frequency of updating may vary widely, and be hard to ascertain automatically – thus frequency of snapshot may also need to vary.
Feasibility of moving to Regulations	
Quantitative indicators	Probably not higher than tens of thousands, possibly fewer.
Composite or enquiry-driven	Enquiry-driven

5. 'Flat' e-books (including e-reports)

Description	Self-contained online publications which resemble printed books in that their content is fixed permanently (until the next edition), and which have no additional functionality. Also, published separately, rather than in a serial publication. Could be born digital or retrospectively digitised from existing print copy. Often, but not always, presented in PDF format, anticipating that users will print out rather than read online. Alternatively, may be presented in dedicated e-book format such as Mobipocket; may become hard to distinguish between delivery platform and content. Often mirror print version exactly, but some (particularly research or publicity reports for niche audiences) may no longer be published in print at all (or only printed on demand). Some e-books are essentially repackaged compilations of a basket of printed materials, eg for an e-learning market. Additional functionality (e.g. workgroups) could be added.
Fixed/additive ('continuing resource')/transformative ('ongoing integrating resource')	Fixed
Free or Restricted access	E-books are mostly commercially produced and usually restricted to subscribers or individual purchasers (who pay for downloads or print on demand); some may be provided free by publisher in order to drive print sales. E-reports are increasingly available free for downloading or printing from public, private or charitable sector websites, even if a paper edition is still available for purchase.
Pull/harvest or Push/deposit	E-reports might be harvested individually (if of special importance) or as part of a more general harvesting of the parent website. As discrete items, delivery could be required.
Linkages	Publisher's site, bookseller's site. Educational books may link to supporting website for teachers/parents/pupils; popular books may link to 'fan' site. Non-fiction books may include 'further information' links. Some e-reports may link to original research data.
Collection priority (1 = essential (national), 2 = essential (regional), 3 = discretionary, 4 = sampling only, 5 = do not collect)	E-reports will be (1), since the BL maintains a National Reports Collection. E-books will vary between (1) and (4) depending upon the degree of originality of content. (5) if e-book is simply PDF of printed book.
Special issues for rightsholders (e.g. commercial, legal)	Management of new editions and updated chapters an issue.
Special issues for libraries (e.g. technical, ephemerality)	Some e-books may be dependent on use of specific hardware. Software may be required to navigate around the book. Some e-books can be stand-alone hardware devices or require dedicated hardware to read.
Special issues for users	
Notes	Links may be only differentiation from print version.
Feasibility of moving to Regulations	
Quantitative indicators	Information obtained from consultant Linda Bennet, of Gold Leaf, is as follows [edited, not <i>verbatim</i>]. As regards institutional supply, Nielsen Book Data is not a good source, as it has only just started categorizing e-books. Similarly, Bowker's records are incomplete (though probably better than Nielsen's). There is no one single reliable source of information, but Gold Leaf research indicates that about 100,000 titles are now available in the UK to institutions. As far as retail goes, Lightning Source Inc, which supplies 'e-tailers' throughout the world, currently has about 110,000 titles in its digital warehouse, probably more than any other digital warehouse. Of these, about two thirds are available for sale in the UK. Library suppliers categorize e-books and print books together for their customers.
Composite or enquiry-driven	Composite

6. Interactive e-books

Description	Self-contained publications with which the user can interact. This may take the form of clicking on page to display further information. May also result in generation of personalised (parts of) content, e.g. inputting data (e.g. answers to questions, personal information which elicits context-specific content). These products will be driven by an underlying database/content management system (see Databases), which may run under licensed proprietary software. May include non-printable media (sound, moving image etc) which may be under narrow licence from a 3 rd party located either within or outside the UK.
Fixed/additive ('continuing resource')/transformative ('ongoing integrating resource')	
Free or Restricted access	Both models in operation
Pull/harvest or Push/deposit	Regular periodic snapshots (as with databases) may be the only feasible method.
Linkages	Publisher/owner site. Support site (educational products); 'fan' site (popular products). External information resources, book or other retailer sites. Very hard to validate that all links are valid, especially if they point to the open internet where they may well be ephemeral
Collection priority (1 = essential (national), 2 = essential (regional), 3 = discretionary, 4 = sampling only, 5 = do not collect)	3
Special issues for rightsholders (e.g. commercial, legal)	Software and third-party content licences may be an issue.
Special issues for libraries (e.g. technical, ephemerality)	Technical challenges of multimedia content. What exactly is preserved? Is the content permanently changed by a user interaction? Could copies be made to allow users to interact with the content but without affecting the original content?
Special issues for users	
Notes	
Feasibility of moving to Regulations	
Quantitative indicators	See: 5. 'Flat' e-books, above.
Composite or enquiry-driven	Composite

7. E-prints

Description	Versions of individual scholarly journal primary research articles (see e-journals), at various stages of their life-cycle (early pre-publication draft, version as submitted to journal, as revised after peer review, as edited by publisher, or final PDF as prepared by publisher), deposited elsewhere than in the journal of publication – e.g. institutional repositories, subject repositories, departmental or individual web pages.
Fixed/additive ('continuing resource')/transformative ('ongoing integrating resource')	Should be fixed, but repository rules vary about ability to alter or remove
Free or Restricted access	Usually free
Pull/harvest or Push/deposit	Usually pull
Linkages	Link to definitive published version, with any associated functionality, on publisher's site highly desirable; some publishers make this a requirement of deposit, but even then may not always be complied with. Links to data; citations less likely to have live links than in the published version (if they do, URLs more likely than DOIs). Links to author's own page (and thence to his/her other work) and/or department/institution also a possibility
Collection priority (1 = essential (national), 2 = essential (regional), 3 = discretionary, 4 = sampling only, 5 = do not collect)	Generally 3; but probably 5 for non-definitive versions. E-prints on subject repositories are likely to be of higher priority than on institutional repositories. Will depend in part upon extent to which a repository can be regarded as 'trusted' in terms of its long-term commitment to archiving and access, and the robustness of its arrangements for ensuring such.
Special issues for rightsholders (e.g. commercial, legal)	Publishers extremely concerned about risk posed to journal subscriptions if significant proportion of a journal's content is freely and easily available elsewhere (even if in a less-than-final version). Some articles are illegally deposited in contravention of agreements signed with publishers, although other publishers allow it (sometimes after an embargo period).
Special issues for libraries (e.g. technical, ephemerality)	Duplication of published content highly likely. Organised repositories probably reasonably secure, but individual/departmental pages highly ephemeral. Repositories will contain mixture of UK-published and overseas (latter not subject to Legal Deposit); hence selection challenges. E-prints will not necessarily reflect version of record. Likely to be substantial complexity in defining and building appropriate links between different versions in metadata terms. Generally resource-hungry. Need clarity over the roles of institutional repositories and the role of legal deposit; they do not necessarily have the same aims, and some institutional repositories are likely to evolve more as 'portals' pointing to resources rather than trying to preserve material for all time.
Special issues for users	Note problem of uncertainty about whether or not they are viewing the 'definitive' version.
Notes	Almost certainly (unless not submitted or not accepted for publication) will represent alternative versions of published articles; however, the journal of publication may or may not be subject to UK deposit, depending on the location of the publishing entity. NISO has started a project to devise standard naming conventions and, ultimately, technical standards to identify variant versions of the same article, but this is probably some years away from completion. NB some institutional repositories also contain a wide range of other material (see Databases).
Feasibility of moving to Regulations	
Quantitative indicators	EPS obtained data from ePrints-UK, which harvests metadata from approximately 30 institutional repositories on a daily basis. It is developing a series of national, discipline-focused services through which the HE and FE communities can access the collective output of e-print papers available from compliant Open Archive repositories, particularly those provided by UK universities and colleges. The ePrints-UK Service

	Demo (http://eprints-uk.rdn.ac.uk/stats/) has harvested a total of 1,715,539 e-prints since January 2004 (last count August 2006). Not counting the first seven months of harvesting (when the number of e-prints deposited was rather lower) an average of 397,165 e-prints has been deposited annually.
Composite or enquiry-driven	Composite

8. Blogs, discussion lists/listservs

Description	'Amateur' publications, generally informal in style and structure. Blogs usually (but not always) written by one person, commenting on a more or less frequent basis on issues, other publications, etc of current interest; vary widely in their levels of moderation/independence. Listservs usually (but not always) a multi-person discussion under various topics ('threads'); all postings may be visible ('bulletin board') or archived in some way.
Fixed/additive ('continuing resource')/transformative ('ongoing integrating resource')	Additive ('continuing resource')
Free or Restricted access	Almost always free; some may be restricted to a given group to which access may be controlled by registration, or some other means.
Pull/harvest or Push/deposit	
Linkages	Blogs in particular frequently link to other content items to which they refer.
Collection priority (1 = essential (national), 2 = essential (regional), 3 = discretionary, 4 = sampling only, 5 = do not collect)	3; potentially important witness testimony (amateur or expert)
Special issues for rightsholders (e.g. commercial, legal)	Copyright transfers/publication licences rarely in evidence, so rightsholders may be disparate and un-contactable. Most use proprietary (though often free) software.
Special issues for libraries (e.g. technical, ephemerality)	Some arguably constitute an important record of social history. Libel and other liability could be an issue. Selection is labour-intensive.
Special issues for users	
Notes	
Feasibility of moving to Regulations	
Quantitative indicators	<p>Calculating the number of blogs in the UK is challenging, as no organisation dedicated to this task exists. Technorati claims to be the 'recognised authority on what is going on in the world of weblogs' (www.technorati.com/about). An initial review of Technorati reveals that in July 2006, it recorded 20 million blogs written in English, but this figure relates to all Anglo-speaking countries, and is not broken down by country. Furthermore, the accuracy of this count is compromised by the fact that this only reflects the number of blogs that Technorati has indexed and not private or restricted blogs. It also does not specify how many of these blogs are active, or indeed what would classify as an active blog. It is highly likely that a large percentage of these 20m blogs will be abandoned or dead blogs.</p> <p>EPS obtained user numbers for the most popular UK blogging services: MSN Spaces and Live Journals. In June 2006, there were 1.5 million UK residents using MSN Spaces, and 200,000 UK users on Live Journal. This gives us an initial figure of 1.7million UK blogs.</p> <p>Nielsen/Netratings reports a total of 35.8m internet users in the UK (June 2005). Pew (www.clickz.com/showPage.html?page=3453431), indicates that 7% of online Americans (8m) had created a blog. There are no similar figures available for the UK, so EPS has used this 7% (on the basis that US and UK internet usage is likely to be comparable) to calculate that 2.5m British Internet users may have created a blog.</p> <p>Taking the initial figure of 1.7 million UK blogs, and acknowledging that a number of blogs will be created using American services, the lesser known UK blogging services, and non-hosted blogs (i.e. blogs that individuals maintain on their own servers using their own tools) it does not seem unreasonable to estimate that there are in the region of 2.5m UK blogs.</p>
Composite or enquiry-driven	Composite

9. 'Flat' websites

Description	Websites containing a variety of textual content and possibly images, which is created and posted a page at a time. Pages are likely to be updated with greater or lesser frequency; links between pages provide navigation of the site. Range from 'online brochures' for products and organisations, to sophisticated information resources; however, organisations' internal 'Intranets' would not be subject to deposit. Websites as a technological form pervade most of the other publishing categories and cannot always be conceptually separated from them.
Fixed/additive ('continuing resource')/transformative ('ongoing integrating resource')	Transformative ('ongoing integrating resource') (+ additive ('continuing resource')) as new pages added)
Free or Restricted access	Both models in operation, but primarily free
Pull/harvest or Push/deposit	Primarily pull
Linkages	Very numerous and often integral to value of site – both internal (navigation) and external to other web pages
Collection priority (1 = essential (national), 2 = essential (regional), 3 = discretionary, 4 = sampling only, 5 = do not collect)	Likely to be a domain harvest at intervals, and hence 1. Otherwise, collecting will be selective, but subsuming 1, 2, 3 and 4. Will need to take account of archiving by other (non-LDL) 'trusted repositories', e.g. BBC, Internet Archive.
Special issues for rightsholders (e.g. commercial, legal)	May well be dependent on third-party software under licence.
Special issues for libraries (e.g. technical, ephemerality)	Highly ephemeral. Selection criteria very difficult. Difficulty in identifying all rightsholders, and in obtaining consent from all those identified. Some sophisticated sites may be difficult to crawl, dependent upon crawler capabilities and site configuration.
Special issues for users	
Notes	See Web Archiving Consortium pilot
Feasibility of moving to Regulations	
Quantitative indicators	Providing an estimate of the number of total UK web sites is extremely difficult – new web sites are constantly being developed and there is no one body or systematic method that calculates the figure. Another obstacle is that each UK web site could have a domain that ends in any number of ways, such as: .com, .biz, .org, and .info. There is no easy way to discern which of these domains might be from the UK. However, it has been possible to obtain a figure for the .co.uk domain. According to Webometrics (a research group that acts as an observatory of the academic and scientific research activities and publications on the web), the number of .co.uk domains stood at 32,200,000 in January 2006 (www.webometrics.info/Number_National_Domains_World.asp.htm). This figure in itself clearly very inflated indeed, as it is also likely to include multiple domain names to ward off competition, and to effect redirections, as well as un-built sites. The British Library has done some further analysis and is working on figures of 5.2 million for .uk sites and 1 million for .com sites used by UK organizations/individuals. There is no way of differentiating in the statistics between 'flat' web sites and 'database-driven' web sites.
Composite or enquiry-driven	Composite

10. Database-driven websites

Description	Websites whose pages present content from one or more fields of an underlying database (content management system), which may itself be updated with any degree of frequency. May also offer context-specific presentation in response to data provided (consciously or otherwise – e.g. cookies, domain names etc) by the user.
Fixed/additive ('continuing resource')/transformative ('ongoing integrating resource')	Transformative ('ongoing integrating resource') – snapshot only possible way to preserve
Free or Restricted access	Both models in use.
Pull/harvest or Push/deposit	
Linkages	Very numerous and often integral to value of site – both internal (navigation) and external to other web pages
Collection priority (1 = essential (national), 2 = essential (regional), 3 = discretionary, 4 = sampling only, 5 = do not collect)	Likely to be a domain harvest at periodic intervals, and hence 1. Otherwise, collecting will be selective, but subsuming 1, 2, 3 and 4. For-fee services, like stock market evaluations, might be valuable to preserve.
Special issues for rightsholders (e.g. commercial, legal)	Dependent on sophisticated software, often proprietary under licence.
Special issues for libraries (e.g. technical, ephemerality)	What is preserved?
Special issues for users	Original functionality very unlikely to be offered?
Notes	
Feasibility of moving to Regulations	
Quantitative indicators	Providing an estimate of the number of total UK web sites is extremely difficult – new web sites are constantly being developed and there is no one body or systematic method that calculates the figure. Another obstacle is that each UK web site could have a domain that ends in any number of ways, such as: .com, .biz, .org, and .info. There is no easy way to discern which of these domains might be from the UK. However, it has been possible to obtain a figure for the .co.uk domain. According to Webometrics (a research group that acts as an observatory of the academic and scientific research activities and publications on the web), the number of .co.uk domains stood at 32,200,000 in January 2006 (www.webometrics.info/Number_National_Domains_World.asp.htm). This figure in itself clearly very inflated indeed, as it is also likely to include multiple domain names to ward off competition, and to effect redirections, as well as un-built sites. The British Library has done some further analysis and is working on figures of 5.2 million for .uk sites and 1 million for .com sites used by UK organizations/individuals. There is no way of differentiating in the statistics between 'flat' web sites and 'database-driven' web sites. See also: 4. 'Databases', above.
Composite or enquiry-driven	Enquiry-driven

11. Conference proceedings

Description	Publication containing programme, abstracts and sometimes full papers from a conference. May be part of a series; may be published as a journal supplement. May be sponsored by commercial company. May replicate (or expand) printed proceedings. May be elusive ('grey literature').
Fixed/additive ('continuing resource')/transformative ('ongoing integrating resource')	Fixed
Free or Restricted access	Both exist
Pull/harvest or Push/deposit	
Linkages	May link to conference website; individual presentations may link to full version on author's (or other) site, and/or to data, images etc.
Collection priority (1 = essential (national), 2 = essential (regional), 3 = discretionary, 4 = sampling only, 5 = do not collect)	1 to maintain comprehensive BL collecting and indexing of conference proceedings
Special issues for rightsholders (e.g. commercial, legal)	Copyright ownership may be complex or unclear. May be originally distributed to conference attendees only. Papers may subsequently be published in journal form (see e-prints, e-journals). May be difficult to determine country of publication – may not be the same as country in which conference held or organizer resides.
Special issues for libraries (e.g. technical, ephemerality)	Moderately ephemeral
Special issues for users	
Notes	
Feasibility of moving to Regulations	
Quantitative indicators	On the basis of work done by EPS for the BL in early 2004, a working figure for conference proceedings would be in the region of 4-5,000, whilst for other 'grey literature' reports the figure could be in the region of 70,000.
Composite or enquiry-driven	Composite

12. Theses and dissertations

Description	Extended papers (of widely varying length and originality) produced by students in partial fulfilment of the requirements of research or taught university degree courses. Normally created in electronic form but not necessarily submitted or preserved in electronic form. Research theses (doctoral or masters) deposited with the awarding university, sometimes in electronic form and for inclusion in a local repository. Dissertations for taught courses, at postgraduate or undergraduate levels, less consistently preserved by the awarding university.
Fixed/additive ('continuing resource')/transformative ('ongoing integrating resource')	Fixed
Free or Restricted access	Very few research theses are embargoed and, as originally submitted, none are offered on a commercial basis. Most are available for public consultation from the time of the award of the degree.
Pull/harvest or Push/deposit	Those on institutional repositories could be harvested.
Linkages	May include links to referenced web pages/sites; in some areas may also include links to data. Link to author's own page also a possibility.
Collection priority (1 = essential (national), 2 = essential (regional), 3 = discretionary, 4 = sampling only, 5 = do not collect)	(1) for theses awarded for research degrees; (5) for most dissertations for taught degrees. May depend upon extent to which institutional repositories (as alternative to LDLs) have 'trusted' status. Hitherto not subject to deposit, as not published. Presence on a university repository will change that status except where access to a thesis is embargoed.
Special issues for rightsholders (e.g. commercial, legal)	Author is rightsholder. Most universities have agreements with BL to lend a paper copy of doctoral theses which is subsequently microfilmed and made available by BL for onsite or remote use. Authors agree to this in writing at time of submission to their university. Arrangement could be extended to electronic deposit. In some instances, a thesis may subsequently have been published as a monograph or journal article, with greater or lesser degree of alteration (and with transfer of copyright or at least publication rights); in which case, the published form would not be regarded as a thesis.
Special issues for libraries (e.g. technical, ephemerality)	Range of formats; co-operation of university registrars key to success.
Special issues for users	Logically, as for public domain websites, there would be no particular reason why access to most theses should be restricted to legal deposit library premises.
Notes	JISC is currently funding BL and CURL to develop a pilot national e-thesis service, with a central metadata registry and full-text of theses held either centrally at the BL or on a university repository. Service may cover retrospective digitisation of microfilm/paper theses.
Feasibility of moving to Regulations	Good prospects. Increasing recognition by individual universities and funding bodies (Research Councils, HEFCE etc.) of importance of making research theses more widely available. UK lagging behind other countries in this regard. Increased visibility for authors, especially in humanities where no longer so easy to get theses accepted as monographs with academic or commercial publishers.
Quantitative indicators	EPS estimates that 15,520 theses are produced annually in the UK. This number represents the number of research doctorate qualifiers for 2004/5 recorded by HESA. We can assume that this is equal to the number of theses completed in that year. The submission of theses in electronic form is still in the early stages. A number of universities encourage the submission of e-theses, but this number is very low, and the universities also require hard copies. The Theses Alive! project (www.thesesalive.ac.uk and www.era.lib.ed.ac.uk), led by Edinburgh University Library with grant funding from JISC, was set up to promote the adoption of a management system for UK e-theses and dissertations.
Composite or enquiry-driven	Composite

13. Digital Maps

Description	Collections of geospatial datasets that in conjunction create a cartographic product. This can consist of vector and raster files as well as spatial and georeferenced databases.
Fixed/additive ('continuing resource')/transformative ('ongoing integrating resource')	Additive ('continuing resource') or transformative ('ongoing integrating resource')
Free or Restricted access	Restricted
Pull/harvest or Push/deposit	
Linkages	Generally no external linkages
Collection priority (1 = essential (national), 2 = essential (regional), 3 = discretionary, 4 = sampling only, 5 = do not collect)	1 for maps of national importance, e.g. Ordnance Survey.
Special issues for rightsholders (e.g. commercial, legal)	Software and third-party content licences may be an issue.
Special issues for libraries (e.g. technical, ephemerality)	Dependent on sophisticated software, sometimes proprietary.
Special issues for users	
Notes	Frequency of updating/snapshot will vary, depending on the production schedule of the publisher, which may be daily or intermittent.
Feasibility of moving to Regulations	
Quantitative indicators	There are three original mapping providers in the UK: the Ordnance Survey (OS), Bartholomews, and the AA. Each of these providers resells its maps to other companies who use them to create their own products, which in turn are sold to the end-user. Bartholomews original mapping data consists of London street level maps, customised mapping, and copyright maps. The OS resells its maps to 148 companies who are licensed to use these maps, create their own products and sell them on. Further information is not available.
Composite or enquiry-driven	Enquiry-driven

14. Patent Specifications

Description	The technical description, drawings and claims prepared to support a patent application. The specification will be published first prior to grant of the application. If the application is granted the specification will be republished. Amendments and translations may also be published during the life of the patent application/granted patent.
Fixed/additive ('continuing resource')/transformative ('ongoing integrating resource')	Additive: specification republished, amendments/translations published as patent application progresses through the system.
Free or Restricted access	Free
Pull/harvest or Push/deposit	
Linkages	Internal linkages (i.e. between different "versions" of specification) need to be made. External linkages might cover linking of "equivalents" (applications for patent protection for single invention in more than one country/state) into a patent family.
Collection priority (1 = essential (national), 2 = essential (regional), 3 = discretionary, 4 = sampling only, 5 = do not collect)	1
Special issues for rightsholders (e.g. commercial, legal)	UKPO technically copyright holder for UK specs but copyright is waived.
Special issues for libraries (e.g. technical, ephemerality)	To archive the specifications which cover patents which might be/are/were in force in the UK involves both UK specifications and European Patent specifications designating the UK.
Special issues for users	
Notes	Longstanding trend among major authorities to e-publication. European Patent Office moved to publication on dedicated public access server in Jan 05 (BL still receives on CD). UK Patent Office keen to cease production of CDs and pushing patent libraries to take specs via FTP which is delivery route for all other customers.
Feasibility of moving to Regulations	
Quantitative indicators	In 2005, according to the Thomson Scientific [Derwent] Database, the Patent Trade Office issued 48,152 British patent specifications compared with 52,071 in 2004. All specifications are now recorded in electronic form, although some specifications are recorded both in electronic and print.
Composite or enquiry-driven	Composite

15. Image Banks

Description	Collections of static or moving images (the latter outside the scope of the LDL Act)
Fixed/additive ('continuing resource')/transformative ('ongoing integrating resource')	Additive ('continuing resource')
Free or Restricted access	Either, depending whether access is controlled
Pull/harvest or Push/deposit	
Linkages	May be linkages to creator/owner
Collection priority (1 = essential (national), 2 = essential (regional), 3 = discretionary, 4 = sampling only, 5 = do not collect)	
Special issues for rightsholders (e.g. commercial, legal)	Original creators may be very concerned about protecting their images from manipulation
Special issues for libraries (e.g. technical, ephemerality)	Sophisticated search technology incorporated in some image databases. Issues of file formats, identification of images and of rightsholders
Special issues for users	
Notes	
Feasibility of moving to Regulations	
Quantitative indicators	EPS estimates that there are about 680 image banks in the UK. The British Association of Picture Libraries and Agencies (BAPLA) has about 450 members, and according to its own estimates this equates to about two-thirds of the picture libraries and agencies in the UK. The largest providers are not UK-owned and it is not known how many of their images might meet the criteria for deposit in the UK.
Composite or enquiry-driven	Composite

16. 'Looseleaf' publications

Description	Publications that in print form are provided in a looseleaf format so that individual leaves may be replaced by updated content. In electronic form, content is similarly replaced by updated material. Common in legal publishing and some STM areas, such as library protocols.
Fixed/additive ('continuing resource')/transformative ('ongoing integrating resource')	Transformative ('ongoing integrating resource')
Free or Restricted access	Restricted
Pull/harvest or Push/deposit	Push
Linkages	
Collection priority (1 = essential (national), 2 = essential (regional), 3 = discretionary, 4 = sampling only, 5 = do not collect)	
Special issues for rightsholders (e.g. commercial, legal)	Updating may be on an irregular cycle – should work be re-deposited at every update, or on a regular periodic basis?
Special issues for libraries (e.g. technical, ephemerality)	
Special issues for users	
Notes	
Feasibility of moving to Regulations	
Quantitative indicators	Probably no higher than 1,000
Composite or enquiry-driven	Composite