succeed★   DELIVERABLE SUBMISSION SHEET

To: Cristina Maier _____ (Project Officer)

DG CONNECT G.2
Creativity Unit
EUFO 01/162 - European Commission
L-2557 Luxembourg

From: ___Support Action Centre of Competence in Digitisation___

Project acronym: ___Succeed_____ Project Number: ___600555___

Project Manager: ___Rafael C. Carrasco Jiménez_____

Project Coordinator: ___Universidad de Alicante_____

The following deliverable:

Deliverable title: _Recommendations on common licensing scheme for tools and
resources_____

Deliverable number: _D4.2_____

Deliverable date: _M22_____

Partners responsible: _PSNC_____

Status: [X] Public   [ ] Restricted   [ ] Confidential

is now complete

[X]   It is available for your inspection

[X]   Relevant descriptive documents are attached

The deliverable is:

[X]   A document

[ ]   A website   Url:_____

[ ]   Software

[ ]   An event

[ ]   Other   _____

Sent to Project Officer: **Sent to functional mailbox:**   **On date:**

cristina.maier@ec.europa.eu   CNECT-ICT-600555@ec.europa.eu   29/10/2014

D4.2 Recommendations on common licensing scheme for tools and resources

Succeed

28/10/2014

Abstract

This deliverable was prepared as part of the WP4 of the Succeed project. The objective of WP4 is to support the EC in the scope of activities identified in the Digital Agenda for Europe, by recommending a set of guidelines, formats, standards and licenses for digitization activities, both in terms of data and tools. The aim is to facilitate the implementation of digitization activities in the European institutions, by making the necessary tools and resources more interoperable, easily accessible and usable. This report provides a set of recommendations on common licensing scheme for tools and resources in the context of digitization and related activities.

Document information

Deliverable number	D4.2	Start: M13	Due: M22	Actual: M22
Deliverable name	Recommendations on common licensing scheme for tools and resources			
Internal/External	External			
Activity type	SUPP			
Participant	UA, INL, IAIS, PSNC, BVC, BnF, BL			
Estimated person months for this deliverable	16.00			
Dissemination level¹	PU			

Document history

Revisions

Version	Status	Author	Date	Changes
0.1	Draft	Tomasz Parkoła	6.06.2014	Document structure and initial content
0.2	Draft	Bob Boelhouwer	24.06.2014	Subject outline of Chapter 3
0.3	Draft	Tomasz Parkoła	24.07.2014	Chapter 4 added
0.4	Draft	Bob Boelhouwer	30.07.2014	Ch 3 adapted, small addition to Ch 1
0.5	Draft	Quique Mollá	31.07.2014	Chapter 6 added
0.6	Draft	Tomasz Parkoła	12.08.2014	Chapter 2 added
0.7	Draft	Ourdia DJAOUT	29.08.2014	Chapter 5 added
0.8	Draft	Quique Mollá	02.09.2014	Chapter 6 moved to chapter 5
0.9	Draft	Bob Boelhouwer	08.09.2014	Enhanced Chapter 1, 4, 6
0.10	Draft	Tomasz Parkoła	09.09.2014	Comments and improvements
0.11	Draft	Bob Boelhouwer	18.09.2014	New division of chapters. Removed summaries of license types. Plenty other changes.
0.12	Draft	Tomasz Parkoła	24.09.2014	Small corrections and comments/suggestions.

---

¹ PU Public; RP Restricted to other programme participants (including Commission Services); RE Restricted to a group specified by the consortium (including Commission Services); CO Confidential, only for members of the consortium (including the Commission Services)

0.13	Draft	Bob Boelhouwer	08.10.2014	Conclusions and summary.
0.14	Draft	Tomasz Parkoła	09.10.2014	Review. Changes in recommendations and summary chapters. Appendix A added.
0.15	Draft	Alicia Blaya	19.10.2014	Review. Changes, suggestions modifications.
0.16	Ready for internal review	Tomasz Parkoła	20.10.2014	Reflected Alicia's comments and suggestions.
0.17	Reviewed	Stefan Eickeler	21.10.2014	Comments, suggestions, improvements.
1.0	Final	Tomasz Parkoła	22.10.2014	Improvements according to review comments. Formatting.
1.1	Final	Tomasz Parkoła	28.10.2014	Improvements based on comments from coordinator.

Approvals

Version	Date of approval	Name	Role in project	Signature
0.17	21.10.2014	Stefan Eickeler	Supervisor D4.2	

Distribution

This document was sent to:

Version	Date of sending	Name	Role in project
0.1	6.06.2014	WP4 participants	Project partners
0.3	24.07.2014	WP4 participants	Project partners
0.9	08.09.2014	WP4 participants	Project partners
0.11	18.09.2014	WP4 participants	Project partners
0.12	24.09.2014	WP4 participants	Project partners
0.13	08.10.2014	Alicia Blaya, Tomasz Parkoła	Project partners
0.14	09.10.2014	Alicia Blaya, Bob Boelhouwer	Project partners
0.15	19.10.2014	Bob Boelhouwer, Tomasz Parkoła	Project partners
0.16	20.10.2014	Stefan Eickeler, WP4 participants	Project partners
0.17	21.10.2014	Tomasz Parkoła	Project partners

1.0	22.10.2014	Isabel Martinez, Rafael Carrasco, WP4 participants	Project partners
1.1	28.10.2014	Isabel Martinez, Rafael Carrasco	Coordinator
1.1	29.10.2014	Cristina Maier (EC)	Project Officer

Table of Contents

1. INTRODUCTION

Galleries, Libraries, Archives, Museums, Research institutes and Universities see it as their task to provide access to information[2], share knowledge and provide expertise. In the Internet age, that involves providing access to digital content and software components via various channels (e.g. website, dedicated API). In contrast to conventional publication methods, digital content and software tools are very easy to copy and to redistribute. Mentioned institutions have to consider the consequences of this. There are two main reasons why they have to consider the protection of digital content. First, if the rights on the content lie with other parties, e.g. the author, the institution offering the digital content should inform the user about these rights and prohibit their violation. Second, if the right on the content lies with the institution offering it (e.g. original texts, databases or tools) then the institution probably will like to protect its credits to the material. Commercial companies may have a slightly different approach. Since revenue streams drive them, they are generally more intent on monetizing their assets. It does not mean that all of those assets need to be closed. By opening parts of their work (e.g. software tools) they can build a greater trust and therefore bring more customers.

This document provides general information on licensing methods and forwards recommendations for certain licensing schemes. Even though some legal information is included, it is important to underline that this work is not intended to provide legal advice[3].

This document investigates licenses that particular institutions might want to attach to their products. This document therefore does not investigate licenses for products received/obtained from other parties (e.g. content from publishers, software from developers). For the latter subject we would like to direct to a publication of the `European Bureau of Library Information and Documentation Association` (EDLIBA) called "Licensing Digital Resources – How to avoid the Legal Pitfalls"[4]

There are basically two ways to regulate the use of content:

---

[2] http://www.ifla.org/publications/the-glasgow-declaration-on-libraries-information-services-and-intellectual-freedom

[3] In no event shall the Succeed Project, any of its consortium partners, or sponsors be liable for any damages whatsoever (including, without limitation, direct, indirect, special, incidental, consequential, punitive or similar damages) arising from any use of the information provided in this document. In cases where licensing of materials occurs, and in particular where the licensor considers it might incur damage to him/her or others, we strongly suggest to seek professional legal advice.

[4] http://www.eblida.org/activities/advocacy-and-lobbying-for-libraries-in-europe/publication.html

- For the end user. In this case a document stating the terms of use should be provided.
- For redistribution or repurposing. The user may distribute the content (or parts of it) to others or use it for some purpose. In this case, a license stating the restrictions should be provided.

Strictly speaking, both methods can be seen as licensing. However, this document will be restricted to the licensing for distribution or repurposing.

2. ASPECTS OF INTELLECTUAL PROPERTY LEGISLATION

Intellectual property (IP) rights, very broadly, means the legal rights which result from intellectual activity in the industrial, scientific, literary and artistic fields. Countries have laws to protect IP for two main reasons. One is to give statutory expression to the moral and economic rights of creators in their creations and the rights of the public in access to those creations. The second is to promote, as a deliberate act of Government policy, creativity and the dissemination and application of its results and to encourage fair-trading, which would contribute to economic and social development[5].

There is an extensive body of information on IP legislation. We will not attempt to cover all of it. Some notions, however, are important.

2.1. Copyright on original works

The basic legal framework for copyright at the international level is the 'Berne Convention for the Protection of Literary and Artistic Works', 1886, usually known as the 'Berne Convention'[6]. Signatory countries are required to recognise copyright of works originating in other signatory countries in the same way as they recognise the copyright of their own nationals. Copyright protection is automatic, not subject to any formality. In addition, the agreement establishes certain minimum standards of protection concerning the rights acknowledged (moral rights and economic rights), limitations to the exclusive rights and the duration of the copyright.

Whenever an author produces a creative work, under the Berne Convention he/she automatically becomes owner of the copyright of this work[7]. Owners of a creative work have the exclusive right to use the work and to authorise others to use it on agreed terms.

---

[5] Chapter 1, WIPO Intellectual Property Handbook: Policy, Law and Use (http://www.wipo.int/about-ip/en/iprm).

[6] http://www.wipo.int/treaties/en/ip/berne.

[7] This may be different when the author is hired by a person or organisation to produce the work.

Under the Berne Convention, the general minimum duration of the copyright is life of the author plus 50 years after his/her death[8], but contracting parties are free to provide longer terms. The European Union extended that term with the 1993 Directive on harmonising the term of copyright protection[9]. In the European Union a work is protected for 70 years after the death of the author[10]. If the author is not known, it is protected for 70 years after its first lawful publication[11]. The Berne Convention authorizes countries to allow certain free uses of copyrighted works. This includes, for instance, the reproduction of limited parts of copyrighted material for certain purposes (e.g., review, news reporting, teaching or scholarly research) without obtaining permission from the author and without paying a fee or royalty.

2.2. Copyright on derivative works

Also works derived from original works can bear copyright[12]. In general, institutions do not publish much original work. However, institutions put in a lot of effort to digitise works in their catalogues for online publication. We can consider some digitised versions as derivative works. To be eligible for copyright, a derivative work must be different enough from the original to be considered a "new work" or must contain a substantial amount of new material. So, simply reproducing a public domain text in digital form would not create a derivative work. But, by enhancing the work with links, annotations, sound or images it likely will. It is important to note that when the original work is not in the public domain, its copyright holder has to authorise the creation and exploitation of the derivative work.

2.3. Database rights

The so-called "sui generis" database right is also a property right, which is to certain extent comparable to but distinct from copyright. According to the Directive 96/9/EC

---

[8] There are a few exceptions to this general term for certain categories of works, like cinematographic works (minimum protection of 50 years after the work has been made available to the public, or, if not made available, 50 years after the making of such a work) or photographic works and works of applied art (min. 25 years from the making of the work).

[9] Directive 93/98/EEC was repealed and replaced by Directive 2006/116/EC, amended by Directive 2011/77/EU. See http://ec.europa.eu/internal_market/copyright/term-protection/index_en.htm

[10] Same general protection term applies also in the United States since the Copyright Term Extension Act, 1998.

[11] In the digitisation field, the issue of "orphan works" is important. Orphan works are those works still protected by copyright but whose authors or other right holders are not known or cannot be located or contacted to obtain copyright permission. Directive 2012/28/EU sets out common rules on the digitisation and online display of this category of works. See http://ec.europa.eu/internal_market/copyright/orphan_works/index_en.htm

[12] Derivative work: http://www.wipo.int/tk/en/resources/glossary.html#19; http://en.wikipedia.org/wiki/Derivative_work; http://www.publicdomainsherpa.com/derivative-work.html

on the Legal Protection of Databases[13], a database is a collection of independent works, data or other materials arranged in a systematic or methodical way and individually accessible by electronic or other means. As such, it can be protected by copyright if there is originality in the selection or arrangement of the contents, and/or by the sui generis right. The latter right is granted to the maker of the database, and for it to apply, there must have been a substantial investment in obtaining, verifying or presenting its contents. It is possible that a database will satisfy both these requirements so that both copyright and the sui generis right apply.

There is no registration required for database rights - it is an automatic right like copyright. However, the term of protection under database right is in principle much shorter than under copyright, as it lasts for 15 years from the making of the database[14]; if made available to the public before expiry of that period, then the term is 15 years from the making available to the public.

Many databases are a collection of copyright works, such as a database of poetry from the last fifty years where each poem will also be protected by copyright. So people compiling databases need to make sure that they have permission from the copyright owners for use of their material and people using databases need to be aware of the rights of the owners of underlying works as well as database rights owners.

3. COMMERCIAL LICENSING

There are no general frameworks for commercial licenses available. The reason for this might be that the conditions of commercial licenses are dependent on many complex factors like the business models of both licensor and licensee, the nature of the data to license and the market dynamics in which both parties operate.

A licensor might develop a general license for some products with no room for negotiation for the licensees. But also in that case the licensor should carefully consider possible business models for licensees in order to make licensing the product enticing.

Conditions that could be part of a license are:
- Type of usage of the product. In case of textual data, can it only be used in printed books and not in eBooks?
- Period of usage. A license may expire after a certain date. Will the licensee still be able to sell existing stock after that date?

---

[13] http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31996L0009:EN:HTML

[14] It should be noted that any substantial change to the contents of a database, which would result in the database being considered to be a substantial new investment, shall qualify the database resulting from that investment for its own term of protection (art. 10.3 Database Directive).

- Territory. In which countries is the licensee allowed to make the derived products available?
- Exclusiveness. Can other parties not license the same product? Does that provision hold for a certain territory?
- Fee, royalties. Is an initial payment involved? How much is that? Will the licensee also have to pay a fee to the licensor for every sold item based on the product?
- Sublicensing. Is the product used for a semi-manufacture? In what way is the licensee allowed to make that available?
- Attribution. Must the name of the licensor be provided on the derived product and/or on the packaging? What should the reference look like?
- What to do in case of conflict. Under which governing law must disputes be resolved? Will arbitration take place?

The safest way to develop commercial licenses is to involve marketing experts and legal experts.
For more detailed information we like to point to the following resources:
- Licensing Agreements: The basics http://www.inc.com/encyclopedia/licensing-agreements.html
- Key aspects of IP License Agreements *www.jurisdiction.com/lic101.pdf*
- How to license Intellectual Property http://www.inc.com/guides/201106/licensing-in-intellectual-property.html
- Licences for Europe http://ec.europa.eu/internal_market/copyright/licensing-europe/index_en.htm

For those who want to experiment with licenses, there are several so-called license generators available on Internet:
- http://www.binpress.com/license/generator
- http://www.useplus.com/pluslicensegenerator/Steps/Start.aspx?AspxAutoDetectCookieSupport=1

3.1 Public Private Partnership

Of special concern is the re-use of public sector information. The European Union provides a common legal framework for a European market for government-held data (public sector information) as a Directive[15]. Such partnerships often involve commercial partners.

There are also no special licensing frameworks for Public Private Partnership. In general, such partnerships will involve tailored contracts and licensing that will very much depend on the particular roles and business models of all parties involved. For such licenses the considerations of the previous section will also hold. In Appendix B

---

[15] http://ec.europa.eu/digital-agenda/en/european-legislation-reuse-public-sector-information

we provide a description of the policy that the British Library uses in partnering with commercial entities.

4. NON-COMMERCIAL DATA LICENSING

In this chapter we will present options for institutions to license materials that they want to publish online. We will assume that there are no other parties that have copyright on those materials. Publishing material that is copyrighted might need a license that is agreed on by the rights owner and it is not possible to give general recommendations for such licenses. In the following sections we make a distinction between content (original creative material), data (collection of information) and metadata (descriptions of data).

In this section we will present a number of licensing frameworks with a non-commercial nature. The organisations that provide these frameworks all have the ideal that sharing content, data and metadata is advantageous for the common good.

4.1 Licensing content

As was mentioned in Chapter 2, the owner of a work can authorize others to use that work. Usually such authorization is formalized in the form of a license. A license is a contract between the owner (or it's representative) of copyrighted material and its user. The license specifies what the user can do with the material and also what the user is not allowed to do. If the user has to pay a fee for the use of the material, we consider it a commercial license. If no fee is involved, we consider it a non-commercial license. Licenses typically differ in the types of usage they allow to the user and the types of usage that are prohibited.

4.1.2. Creative Commons (CC)
Creative Commons[16] is a non-profit organisation that has released several copyright licenses known as Creative Common Licenses free of charge to the public. With these licenses the owner of a work can express which rights they want to reserve and which rights they waive for the benefit of users of the work. All licenses grant the user the right to redistribute the work under certain conditions.

The six licenses are combinations of four additional conditions. Below we will explain these.

Attribution (BY). Users of the work need to credit the author.

NoDerivs (ND). The work should be passed along unchanged and in whole.

---

[16] http://creativecommons.org

NonCommercial (NC): The work may not be sold or used in commercial products.

ShareAlike (SA): The work may be changed and build upon, as long as the users license their new creations under identical terms.

All six licenses have at least the 'Attribution' condition (which is an expression of one of the basic moral rights of the author, i.e., the paternity right). The combination of 'NoDerivs' and 'ShareAlike' does not occur because these conditions exclude each other. So, the six licenses are:

CC BY: *Attribution.* This license lets others distribute, remix, tweak, and build upon the work, even commercially, as long as they credit the author for the original creation. This is the most accommodating of licenses offered. Recommended for maximum dissemination and use of licensed materials.

CC BY-ND: *Attribution-NoDerivs.* This license allows for redistribution, commercial and non-commercial, as long as it is passed along unchanged and in whole, with credit to the author.

CC BY-NC-SA: *Attribution-NonCommercial-ShareAlike.* This license lets others remix, tweak, and build upon the work non-commercially, as long as they credit the author and license their new creations under the identical terms.

CC BY-SA: *Attribution-ShareAlike.* This license lets others remix, tweak, and build upon the work even for commercial purposes, as long as they credit the author and license their new creations under the identical terms. This license is often compared to "copyleft" free and open source software licenses. All new works based on this one will carry the same license, so any derivatives will also allow commercial use.

CC BY-NC: *Attribution-NonCommercial.* This license lets others remix, tweak, and build upon the work non-commercially, and although their new works must also acknowledge the author and be non-commercial, they don't have to license their derivative works on the same terms.

CC BY-NC-ND: *Attribution-NonCommercial-NoDerivs.* This license is the most restrictive of the six main licenses, only allowing others to download the work and share it with others as long as they credit the author, but they can't change them in any way or use them commercially.

In addition to the licenses above, Creative Commons provides a tool (CC0) to waive all rights on a work. It is difficult to place a work in the public domain as long as the automatic copyrights and database rights have not yet expired. But CC0 enables scientists, educators, artists and other creators and owners of copyright- or database-protected content to waive those interests in their works and thereby place them as completely as possible in the public domain, so that others may freely build upon, enhance and reuse the works for any purposes without restriction under copyright or

database law.

Creative Commons also provides a tool[17] that helps in choosing the right license for a certain application.

4.2 Licensing data

4.2.1. Creative Commons version 4.0
Version 4.0 of the CC license suite addresses database rights in addition to copyright and the other copyright-like rights covered in earlier versions. Because database rights can impede a user's ability to share, reuse, and modify a work in the same way that copyright can, 4.0 makes it clear that these permissions apply to works that would otherwise be restricted by database rights as well. CC 4.0 provides the same set of license types as described in section 4.1.2.

4.2.2. Open Data Commons (ODC)
Open Data Commons[18] provides a set of legal tools to provide open data. The licenses make an explicit distinction between 'data' and 'content'. Data would pertain to collections of content or information typically organised in a database (see section 2.3 on database rights). The ODC database license (ODC-ODBL) grants the user the rights to copy or redistribute the database, to produce works from it, and to modify, transform or expand it. There are, however, provisions to these rights. If a derived work is published, the user must attribute the original database. Derived (published) work should also be shared under the ODC database license and it is possible to redistribute the database (or a derived version) in a closed form as long as an open version is made available as well.

Connected to the license for databases, ODC provides a separate license for database content (ODC-DBCL). That license is to be used in combination with the database license.

Next, ODC provides an attribution license (ODC-BY). This license gives the user the same options for application of the database as the ODC database license, but with less provisions. Only attribution is required.

Finally, ODC provides the Public Domain Dedication and License (ODC-PDDL). Using this license will place the database in the public domain. So, users can use and redistribute the database without provisions.

---

[17] http://creativecommons.org/choose/

[18] http://opendatacommons.org/

4.3 Licensing metadata

Metadata is "data on data". There are several applications for this term[19], but we will consider the descriptive application, so metadata as far as it is used for describing the content and the context of data files. Metadata in this sense is especially useful for archival and search purposes.

The data that a metadata set comprises of will in general not be eligible for copy protection since it will mainly contain factual information. The collection of metadata, however, might resort to database rights (see section 2.3). Before making metadata publicly accessible one should carefully consider who is the owner of the metadata.

Commercial licensing of metadata is not very common, especially among institutions, but it is possible. As is the case with data licensing, there are no standard frameworks available for commercial licensing. Basically, the same considerations are in place as for commercial licensing of data (see Chapter 3).

There are several guidelines for dealing with right issues of metadata. Below we will mention some of the most important ones.

Europeana is a partnership between more than 2000 European institutions. It realized an internet portal that provides access to millions of books, paintings, films, museum objects and archival records that have been digitised throughout Europe. Europeana provides a Licensing Framework[20] for all contributors. Contributors need to agree to the Europeana Data Exchange Agreement[21] which states that all metadata will be made available according to CC0.

The Open Archives Initiative (OAI) is an organization to develop and apply technical interoperability standards for archives to share catalogue information (metadata). OAI recommends using Open Data Commons (see Section 4.2.2) for metadata made available through OAI repositories[22].

The UK-based organisation Discovery also propagates the use of a standard open licensing framework for metadata that is placed online[23]. This organisation promotes a.o. ODC-PDDL and CC0.

---

[19] http://en.wikipedia.org/wiki/Metadata

[20] http://pro.europeana.eu/documents/900548/380f8794-6db3-45de-acf4-3d5721138d26

[21] http://pro.europeana.eu/documents/858566/7f14c82a-f76c-4f4f-b8a7-600d2168a73d

[22] http://www.oaforum.org/otherfiles/oaf_d48_cser3_foullonneau.pdf

[23] http://discovery.ac.uk/profiles/principlesprofile/

5. TOOL LICENSES

5.1 Introduction

In general software license provides information on the possibilities and restrictions related to usage, modification or distribution of the software. There is a large variety among available licenses. For example, a license can restrict the number of computers the software can be used on or can even restrict the number of processors used to run the software. On the other hand there are licenses with almost no restrictions, e.g. allowing a user to even modify the source code. Depending on the restriction level of the license under which the software tool is released, we can distinguish several types of software tools (as depicted on Figure 1):

- Free/Libre and Open Source Software (FLOSS) – this is the most liberal approach to software distribution and it means that the software is open and free. Open means that the source code is available for any user. Free means freedom in terms of usage, modification and distribution of the software.
- Multi-license software – it can be used in various scenarios to provide software to different groups of users with different licenses.
- Proprietary software – it usually has a lot of restrictions put on the usage and there is usually no access to source code. For example, users cannot modify the software and can use it for personal purposes only.
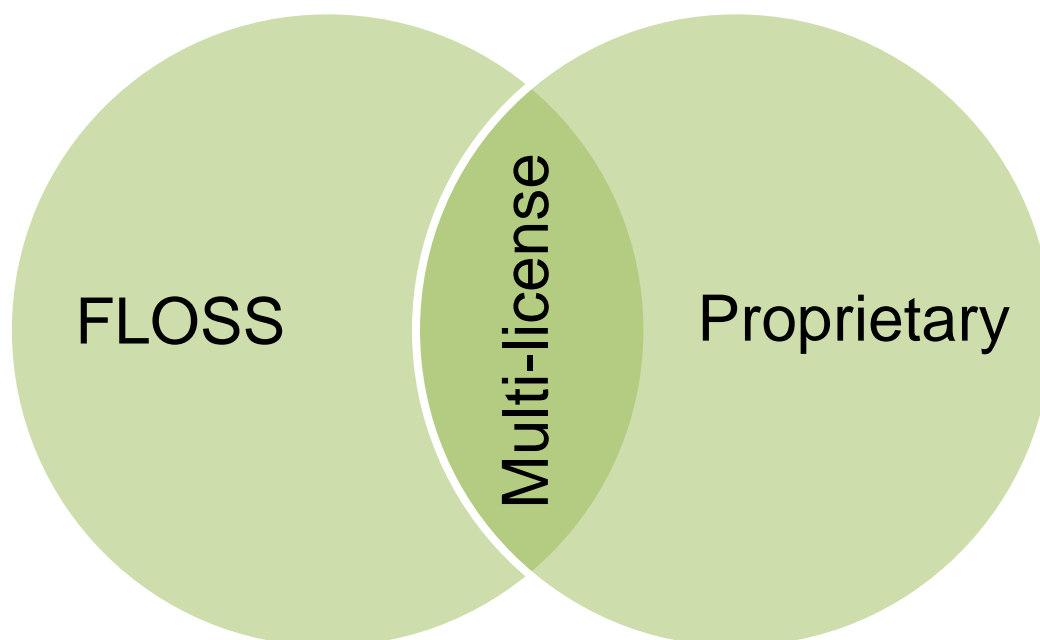
Figure 1. Types of software depending on the licenses used for release

It is important to note that the type of the license under which the tool is released does not determine whether the software is commercial or not. It is a common mistake to interpret proprietary software as commercial software and open source software as

a free one (free meaning at no cost). In fact proprietary software, multi-license software and FLOSS can all have a commercial purpose. For example proprietary software can be provided at no cost – it is then called freeware. And also open source software can be provided with the main goal to get revenues, e.g. the license itself is provided at no cost, but new features or additional services like support are paid.

There is also a separate type of tools that are in the public domain. If a tool is in the public domain it means that it is not under copyright (e.g. copyright period elapsed). Then anyone can use it in any way possible.

5.2 Free/Libre and Open Source Software

There are two major initiatives which support and identify/approve Free/Libre and Open Source Software (FLOSS) licenses. The two initiatives are: Free Software Foundation[24] (FSF) and Open Source Initiative[25] (OSI). Both initiatives support and advocate making software available to the users in a way that they can use, modify and share it. The differences are in the strategy for doing that and terminology. The FSF is more focused on emphasizing the freedom aspect of the software by promoting free software term, while OSI has taken more practical approach thus promoting open source term. The on-going discussion between the free software and open source software resulted in a neutral form, which is FLOSS, and this term is mostly used in this report. To get detailed information on how particular initiatives interpret and understand similarities and differences in their approaches, please refer to their arguments[26] [27].

The two initiatives mentioned in the previous paragraph divide currently available licenses into several categories. Based on these categories it is possible to identify several characteristics that are important for copyright holders (tool providers) which intend to release software publicly. The most important are:

- The level of restrictions put on the licensee – that is especially emphasized by the FSF. In general there are two types of FLOSS licenses: permissive and copyleft. Copyleft licenses assure that any software that is based on a tool released with copyleft license will be distributed using the same license. This kind of restriction ensures that the licensed software (modified or not) will be always distributed using a FLOSS license and also derivative works will be distributed in the same manner. On the other hand so called permissive license does not put such a restriction on derivative works, which means that, e.g. the software can be used in a commercial product as a proprietary and closed source.
- Compatibility with the FLOSS idea. There are many licenses that by mistake can be understood as a FLOSS license. Therefore both FSF and OSI provide a

---

[24] http://www.fsf.org/

[25] http://opensource.org/

[26] http://opensource.org/faq#free-software

[27] https://www.gnu.org/philosophy/open-source-misses-the-point.en.html

list of licenses that are approved as FLOSS licenses. Additionally, non-FLOSS (non-free) licenses have been listed[28].

- Adaptation of the license – there is a multitude of FLOSS licenses available. This differentiation causes compatibility issues and therefore can limit the (re)use of software. In this context OSI provides a comprehensive list of licenses that are commonly used and popular within the FLOSS community[29].

When releasing software publicly, it seems to be necessary to comply with the FLOSS idea, therefore the basic criteria for selecting the license is the FLOSS compatibility. Practically it means that selected licenses should be already approved by the OSI and FSF, or at least they should be approved either by OSI or FSF. In order to reach wider community of users, it is advisable to consider licenses that are already in use. This is especially important for software provided in the form of programming libraries. These libraries are usually integrated into other software, or first modified and then integrated. From this perspective, a wide adaptation of the license under which the software is to be distributed is an important criterion.

5.3 Multi-license software

Multi-licensing is usually used in scenarios that aim to provide software to different groups of users to which different licenses apply. There are three main scenarios for using multi-licensing to distribute software. The first one is related to open source projects and aims at removing barriers that arise from license compatibility. Imagine that software A is licensed under license L1. Imagine also that there is an open-source project B which is licensed under license L2 and wants to leverage the A software by putting it as a component in the B system. Let's assume that licenses L1 and L2 are not compatible (discussed in Section 5.6). In such situation project B cannot use software A despite the fact that both are open source. The reason is legal – the licenses are incompatible. In order to overcome this barrier software A can be distributed not only with license L1, but also with license L3 which is compatible with license L2. Thanks to that it is then possible to use software A inside the B system, using the package distributed under L3 license. An example of this is the Perl[30] language, which is distributed under all versions of GNU General Public License (GPL)[31] plus the artistic license[32].

The second scenario for using multi-licensing is related to software tools developed as commercial products, but still with the open source model in mind. In such case the usual scenario is that a commercial company holds the copyright of the software and distributes it in several ways, e.g. distributes the software at no cost with a GPL

---

[28] http://www.gnu.org/licenses/license-list.en.html#NonFreeSoftwareLicenses
[29] http://opensource.org/licenses/category
[30] http://www.perl.org/
[31] http://www.gnu.org/licenses/licenses.en.html
[32] http://dev.perl.org/licenses/artistic.html

license *and* distributes it with a proprietary license that allows inclusion of the software in other commercial software (e.g. a closed-source product). An example of this is the well-known MySQL database, which is licensed under GPL (version 2) and which also can be used under a commercial license agreement with copyright holder (which is Oracle)[33].

The third scenario for multi-licensing pertains to proprietary software that is distributed with different licenses for different users and different usages. The main idea is to provide different levels of restrictions and also different functions corresponding with the height of the fee. An example is Microsoft® Office 365, which has several licensing models, including Home, Personal, Student, Business and Professional. In this example only the Business and Professional versions can be used for commercial purposes, the others are only allowed for home use[34].

5.4 Proprietary software

Proprietary software usually puts a lot of restrictions on the way it can be used. Moreover, it is usually only available as an executable tool without access to the source code. This means for example that users are prevented to modify or distribute the software. Additionally, users can be also restricted to use the software only for personal purposes and not commercial (for commercial usage another license is required). The most critical issue with proprietary software is the so-called vendor lock-in, which means that only the copyright holder can decide to further develop it or not and when to stop supporting the software (e.g. stop fixing errors in the software). It is one of the biggest risks for the licensee, as shifting from one software solution (which is not further developed/supported) to another one (e.g. from different vendor) can be very costly or sometimes practically not possible within a given time-frame.

5.5 Licenses used in the FLOSS community

There are many licenses that can be used in the context of multi-licensing or FLOSS and it is useful to identify the most common ones. Table 1 presents a list of popular[35] licenses that are both approved by OSI and FSF. The column named "Level of Restrictions" provides information on the restrictions that are included in a particular license. There are three levels of restrictions identified in this summary:
- Permissive – indicates that the license poses minimal requirements about how the software can be distributed/used or modified. All permissive licenses in the summary were approved by the Copyfree initiative[36].
- Copyleft – the most restrictive license type assigned in the summary to the GPL license. The reason for that is that only GPL puts the obligation on the licensee that any derivative work need to be released under GPL. Derivative

---

[33] http://www.mysql.com/about/legal/licensing/oem/
[34] http://office.microsoft.com/en-001/buy/compare-microsoft-office-products-FX102898564.aspx
[35] Popularity level indicated by the OSI portal
[36] http://copyfree.org/

work can be both modified version of the software as well as the software that uses the licensed tool. Because by definition a certain copyleft license requires derivative work to be licensed under the very same license, it is incompatible with other copyleft licenses. This is why, in practice, GPL licensed code cannot be used in a software tool that is not licensed with GPL.

• Partial copyleft – it is a type of copyleft license, which is less restrictive and allows licensed software to be used in software licensed differently. In other words software tools that are released under partial copyleft license can be used in other software tools which are licensed using a different licenses (but still some restrictions may apply, e.g. in the case of the Mozilla Public License (MPL), the source code licensed with partial copyleft license needs to be released under the same license, but other components can be licensed with different licenses).

The last column provides additional notes on the license. One of the especially interesting remarks is GPL compatibility, meaning whether the tool licensed with a certain license can be used in a software tool released using GPL. Not all licenses are compatible with GPL, especially those posing more restrictions on usage than GPL. Obviously, GPL is not compatible with any other license, except GPL itself.

Table 1. Summary of most common OSI-approved licenses for software tools

License	Level of restrictions	Additional notes
Apache License v.2.0	Partial copyleft	Compatible with GPL v.3.0
3-Clause BSD license[37]	Permissive	Compatible with GPL v.3.0
2-Clause BSD license[38]	Permissive	Compatible with GPL v.3.0
GNU General Public License v3.0 (GPL)	Copyleft	The most popular copyleft license for software tools
GNU Library or "Lesser" General Public License v.3.0 (LGPL)	Partial copyleft	Compatible with GPL v.3.0; modified source-code need to be released under the same license
MIT license	Permissive	Compatible with GPL v.3.0
Mozilla Public License 2.0 (MPL)	Partial copyleft	Compatbile with GPL v.3.0; source code files which are under MPL need to remain under MPL
Common Development and Distribution License	Partial copyleft	Incompatible with GPL v.3.0; due to additional restrictions on notes in the source code it is incompatible with GPL
Eclipse Public License v.1.0	Partial copyleft	Incompatible with GPL v.3.0; it is more restrictive than GPL in the context of patent retaliation thus not compatible with GPL

---

[37] Also known as New/Revised/Modified BSD license
[38] Also known as FreeBSD or Simplified BSD license

Within the context of Succeed an inventory of tools for digitization has been created, containing more than 200 items[39]. The most common FLOSS licenses used by these tools are GPL (23% of tools) and Apache License (10% of tools). The other tools did not have a clear statement about the license (36%) or had a commercial license (22%). A few tools were released using MIT, MPL or Creative Commons licenses (altogether 9%). Figure 2 presents an overview of licenses used in the tools from the Succeed inventory.
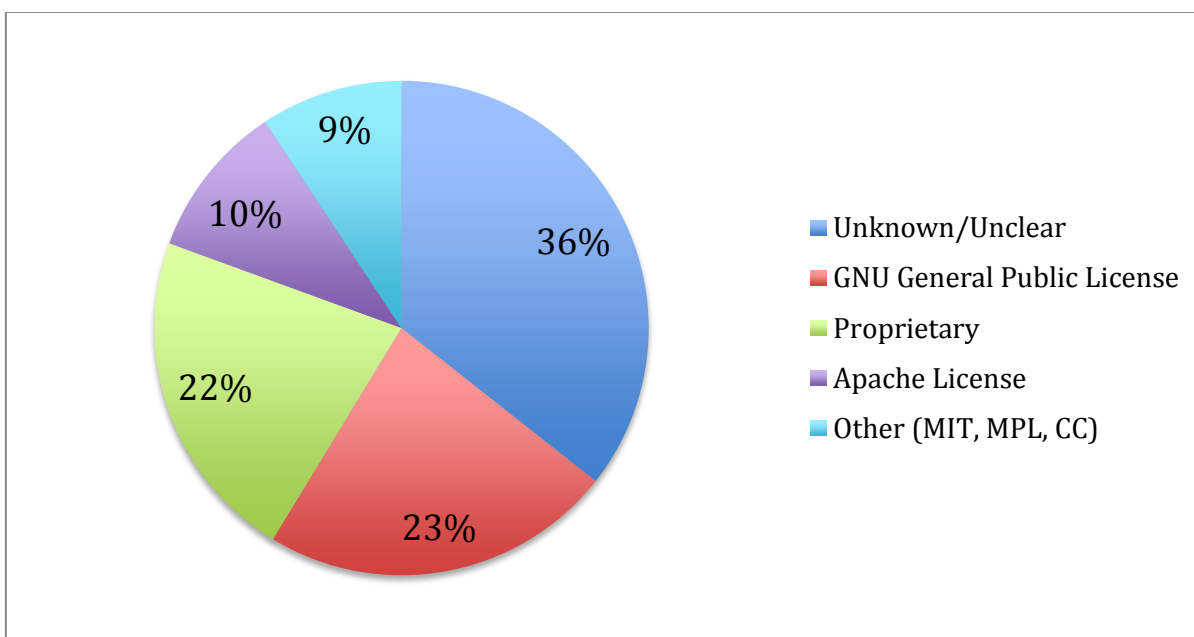
Figure 2. Licenses used for the Succeed inventory of tools for digitization

Additionally, various hosting platforms provide information on the licenses used for the hosted projects. For example, at one of the most popular platforms, SourceForge, 94% of the projects use licenses that are OSI-approved. Within those projects, the most common licenses are from GPL family (approx. 77% of the projects use the GPL or LGPL license). The other licenses commonly used are the BSD License, Apache License v.2.0 and MIT License. Figure 3 shows the usage of OSI-approved licenses on SourceForge.
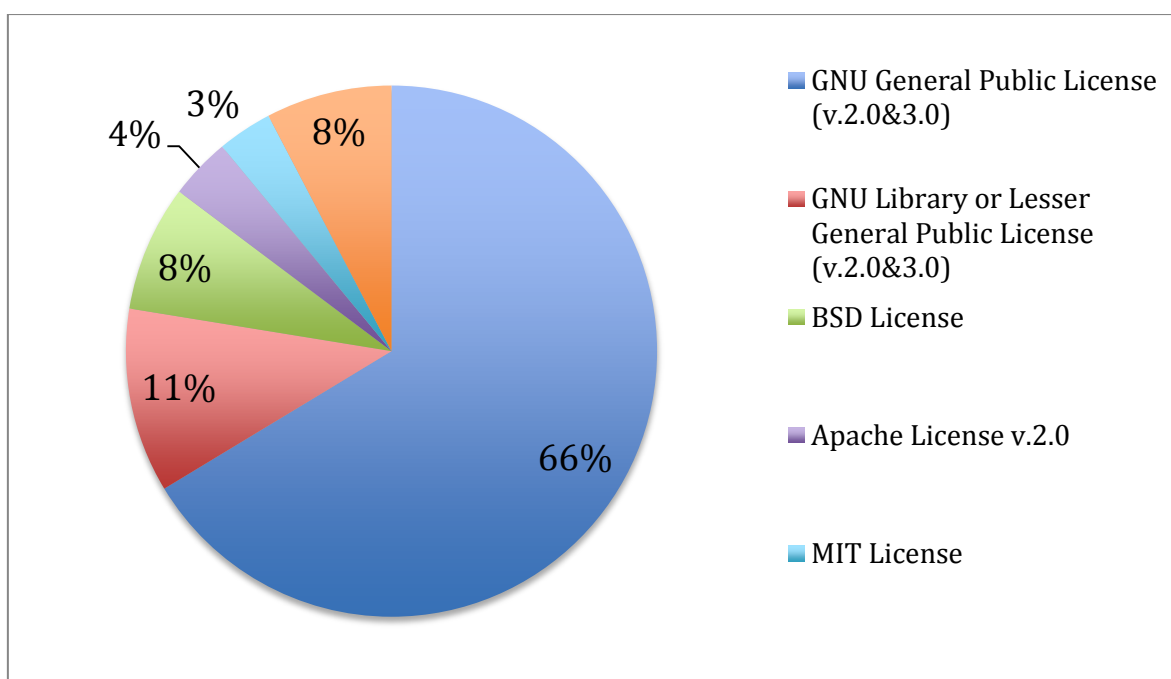
---

[39] http://www.digitisation.eu/tools-survey/index-coc

Figure 3. Usage of OSI-approved licenses on SourceForge

5.6 Licenses compatibility/proliferation

Compatibility of licenses is one of the biggest troubles that tool developers face when releasing software. It is important to remember that when one develops a new tool it is relatively common to use various software libraries to support particular functions, e.g. logging, mathematical computations, linguistic analysis, etc. It is convenient to use software libraries that are open source. Unfortunately, it is not enough that the tool is open source, because one can only choose from those software libraries that are compatible with each other in the sense of licenses. For example if a tool A is licensed using Apache License v.1.1 and one wants to use it in the software tool B which will be licensed using GPL then it is legally impossible as the licenses are incompatible. This is why various licenses have been enhanced and modified to make them more compatible with each other. Take for example the Apache License. Its version 2.0 is compatible with GPL, but only one-way. That means that software released with Apache License v.2.0 can be used in GPL licensed tools, but not vice versa. This is because GPL is a copyleft license and allows derivative works to be licensed only with the use of GPL. The common name for this problem is license proliferation. Although the problem still exists, especially in large tools consisting of many components, it is presently reduced by various recommendations and suggestions from the developers community. An example is Google Developers platform, which restricts the types of licenses that can be used in projects and also recommends the GPL or Apache License. To have deeper understanding of the differences, similarities and compatibility issues between the most common licenses a short summary of each is provided in the following subsections. The intention is to characterise the newest versions of the licenses, therefore only those versions are considered in the descriptions.

5.6.1   GNU General Public License (GPL)

The GPL license is one of the most popular licenses used for releasing FLOSS software. The main characteristic of this license is that it is a strong copyleft license. It means that if a tool is GPL licensed then any derivative work need to be licensed with GPL as well. One of the key issues with the GPL license is the interpretation of the term derivative work. There is still debate whether linking a GPL-licensed program to another one (using as a static or dynamic library) yields a derivative work or not. For the time being the safe approach (same as interpretation of FSF) is to assume it is a derivative work, although courts under a certain jurisdiction might decide differently. Because GPL is strong copyleft, it is not compatible with licenses that pose similar restrictions (e.g. Mozilla Public License). On the other hand all permissive licenses are compatible with GPL (e.g. the MIT License). Because GPL is a popular license, some licenses (which are still in use, attract new developers and gain synergy) were modified to become compatible with GPL (e.g. Apache License). The current version of GPL is 3.0 which was released in 2007. GPL v.3.0 is recommended by the FSF.

5.6.2   GNU Library or Lesser General Public License (LGPL)

LGPL was created with a special focus on software libraries and it was created as a compromise between strong copyleft GPL license and more permissive ones like the MIT License. The main idea is that the software licensed with LGPL can be used in (linked with) another software tool. The new software can be licensed with a different license, including a proprietary one. Nevertheless, LGPL is not fully permissive since all derivative work from the LGPL-licensed software need to be released with the same license. LGPL is recommended by the FSF for special cases only, like when the functionality of a library is already available in other software libraries licensed with a more permissive approach. In such situation there is no reason to apply GPL, as it will limit the number of users of the specific software library (because there are other libraries released with permissive approach, meaning that proprietary tools can use them).

5.6.3   Apache License

The Apache License is a free software license developed by the Apache Software Foundation and initially based on BSD License. The license is commonly known as a permissive one because a modified version of the Apache licensed software can be released using a different license and the licensed software can be used in proprietary tools. Nevertheless, the Apache License requires that every file that has not been modified is licensed using the original Apache License and also that special notes need to be present in the modified files. Apache License v2.0 is the current one. It is compatible with GPL v.3.0, meaning that software licensed with Apache License can be used in tools released with GPL v.3.0 (but not vice versa). FSF recommends Apache License v.2.0 when there is a necessity to use a non-copyleft license (permissive).

5.6.4   BSD License

The BSD License is one of the best-known permissive licenses. It has three main versions, but only two are approved by the OSI and FSF. The two approved versions

are 3-Clause BSD License and 2-Clause BSD License. Initially the BSD License had 4 main clauses posing restrictions. Due to one of those clauses (so called advertising clause) OSI rejected it[40]. In 1999 the advertising clause was removed and a so-called New/Modified/Revised BSD License was created. There is also a 2-Clause BSD License (also known as FreeBSD license), which omits the so-called non-endorsement clause. Both the 2-Clause and 3-Clause are GPL compatible. For the BSD License it is important to compare the different variants and decide carefully which to choose.

5.6.5   MIT License

The MIT License is a permissive license created by the Massachusetts Institute of Technology. It is sometimes called the X11 License as it was designed for the X Window System. The MIT License is similar to the 2-Clause BSD License. It is compatible with the GPL license.

5.6.6   Mozilla Public License (MPL)

The Mozilla Public License is maintained by the Mozilla Foundation. It is a weak copyleft license, somewhere in between the Apache License v.2.0 and the GNU General Public License. MPL licensed software can be used by a differently licensed tool, but a modified version of the software needs to be released with MPL license. The current version is MPL 2.0 and this version is compatible with GPL (in contrast to prior versions).

5.6.7   Eclipse Public License (EPL)

The Eclipse Public License is maintained by the Eclipse Foundation. It is a weak copyleft license that is weaker than GNU General Public License. In case of EPL, additions and modifications to the software can be licensed with a different license only if these cannot be considered to be a derivative work. In case of derivative work, the software needs to be licensed with EPL and it should be made available to everyone. The EPL is currently in version 1.0. EPL is incompatible with GPL because it is too restrictive (e.g. due to patent retaliation).

5.6.8   Common Development and Distribution License (CDDL)

Common Development and Distribution License was created by Sun Microsystems based on the MPL 1.1. It is a weak copyleft license since it allows CDDL licensed software to be used in differently licensed tools. But it still requires derivative works to be released under CDDL conditions.

6. SURVEY ON LICENSING OF CONTENT AND TOOLS

In order to gather information on current practices of licensing digitation tools and content it was decided to conduct a survey. For greater impact, synergy, and better response level, the survey was co-organized with work package 7 of the Succeed

---

[40] https://www.gnu.org/philosophy/bsd.en.html

project dealing with the roadmap for future research and development activities in the context of European centers of competence related to digitization. Therefore, the survey consisted of two questionnaires – one pertaining to licensing (work package 4 questionnaire) and one pertaining to emerging usages of digital content (work package 7 questionnaire). In this chapter we present the results of the questionnaire on licensing.

6.1 Purpose and scope

The purpose of this survey was to get a general overview on the current practices on tools and resources licensing at various institutions interested in digitization. Therefore the questionnaire was composed of three main parts:

- General information – contact details of the participant, in case further discussion would be needed.
- Content licensing – questions about practices at the respondent's institution related to content publishing, including main barriers for publishing data as well as requirements that are connected with the process of determining appropriate license for content, data or metadata to be published.
- Tools licensing – questions about practices at the respondent's institution related to tools licensing, including main barriers for releasing tools as well as requirements that are connected with this process.
-

The survey was disseminated via various channels and addressed to different organisations, institutions and companies that are focused on digitization in terms of tools and content. All the questions can be found in Attachment A, while the online survey is available at:
https://docs.google.com/forms/d/1LXEjvbgd6hzpY8blv1PWofGgWTm5HscN12oLhRTHPUA/viewform

6.2 Methodology

The survey had the form of an online questionnaire. The questionnaire has been prepared in a series of consultations with Succeed project partners, based on their experience with content and tools licensing. There were two types of questions:

- Option questions – a question consisting of several options to mark (one or many).
- Open questions – question consisting of an input field where respondent can answer with free text.

The option questions were used when there was a set of most probable options to choose from by respondents. An example is a question about important features of licenses for tools. Open questions were used in cases where there is no clear answer to the question, e.g. barriers for opening content or releasing tool.

It was decided to create an online survey to reach a wider community and simplify the procedure of answering to the survey. For efficient and successful dissemination of the

survey a list of dissemination channels has been created. It was composed of two main parts:

- List of institutions to directly ask to fill in the survey – it includes 31 institutions to which Succeed partners have direct contacts and can with high probability obtain answers to the survey.
- Other dissemination channels, such as mailing lists, blogs, etc. – a list of 15 channels (hundreds of institutions) to which information about the survey should be sent.

A special focus was given to the institutions such as companies and commercial archives (also for sound or vision) that had responded to the previous survey on formats and standards in digitisation. Dissemination activities were done in two rounds, each of them lasted for approximately one week. In each round all partners from Succeed project were asked to disseminate information about the survey to the channels they were assigned. The survey gathered in total 37 responses. Having in mind that the survey was a demanding one (a lot of questions that required free text input) the result is satisfactory.

6.3 Analysis of results

The survey was filled out by 37 institutions, including commercial companies, research centres, data centres, libraries, archives, museums as well as institutions dealing with sound and vision. Figure 4 provides an overview of the institution types that participated in the survey. It is visible that cultural heritage institutions (20 respondents) were most willing to share their opinions tools licensing and content publishing. Also research institutions (including research centres, data centres and universities – 14 respondents) are eager to discuss their approach for tools and content licensing. The poorest response rate comes from the commercial sector, where only several companies were willing to share their approach and experience with tools and content licensing. The reasons for that is probably proprietary nature of commercial companies, which are usually focused on in-house knowledge sharing rather than public dissemination of critical information (such as content or software). This also transpired from comments provided by respondents from companies. For example one of the companies stated that the developed content (or potentially a tool) is very valuable, therefore it cannot afford to publish this content without specific requirements, not necessarily aligned with the commonly used open licenses.
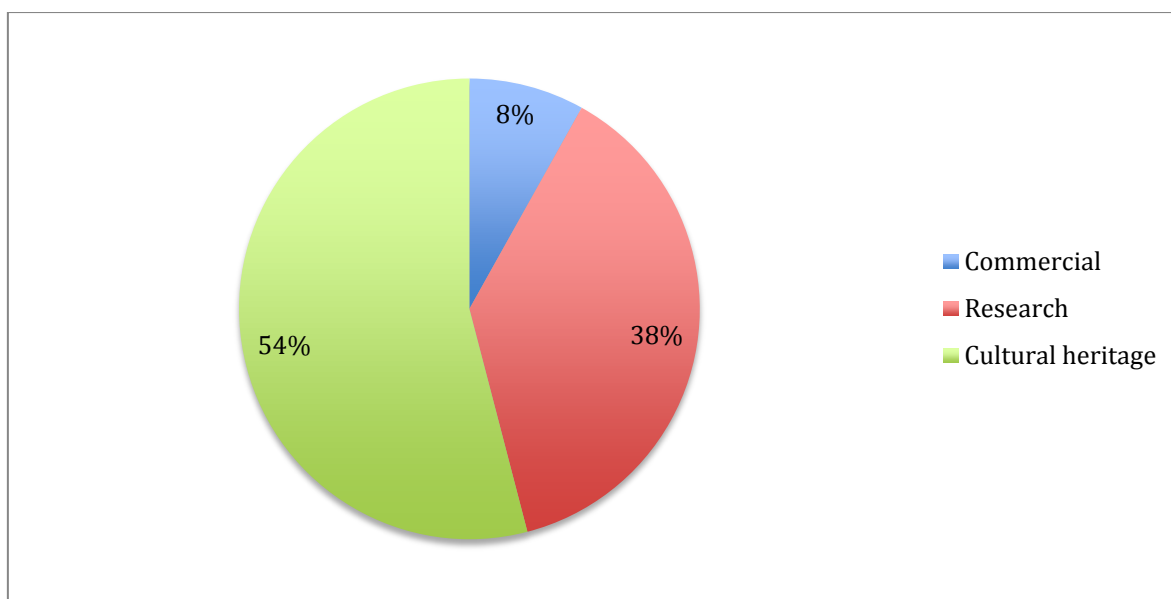
Figure 4. Percentage of respondents by institution type

There were two similar sets of questions related to content and tools licensing, which are described in the following part of this sub-section. Figure 5 presents information on the current content licensing practices in the surveyed institutions. The results show that the most common approach is to use Creative Commons licenses (approx. 49%). Other approaches include individual licenses, public domain, non-commercial use or Public Domain Mark. Except for individual licenses, all other approaches can be covered with Creative Commons licenses (at least to some extent). Therefore, it is quite clear that Creative Commons could be an important licensing method for content distribution, at least from the point of view of the survey respondents.
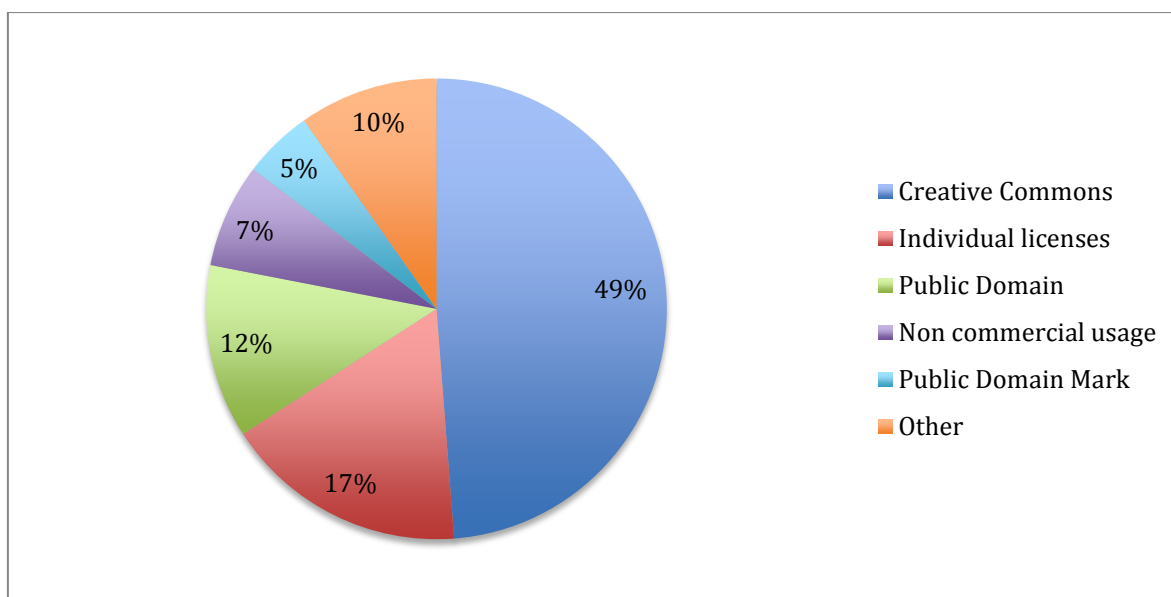
Figure 5. Licenses used for publishing digital content online

The main obstacles, which, according to our respondents, prevent institutions from publishing content online, are depicted in Figure 6. One of the current biggest challenges is the copyright issue. Because of copyright most of the content cannot be published online. Even if it can be published, then additional restrictions apply; e.g. content is available only from the computers which are on the institution's premises. The second obstacle is rights clearance, which is a time and money consuming process. It requires a lot of effort to respect legal requirements and often these efforts lead to nothing (e.g. not all parties holding rights are identified therefore permission is not given). Finally, several respondents complained that they have data available to be published, but the technology required to make that content online available does not exist (or is insufficient – lacks the required functionality). For example, there are databases that could be published, but there are no tools that can handle them and make them available online. Additionally, state of the art tools do not always have the required functionality (e.g. to restrict some actions on the content).
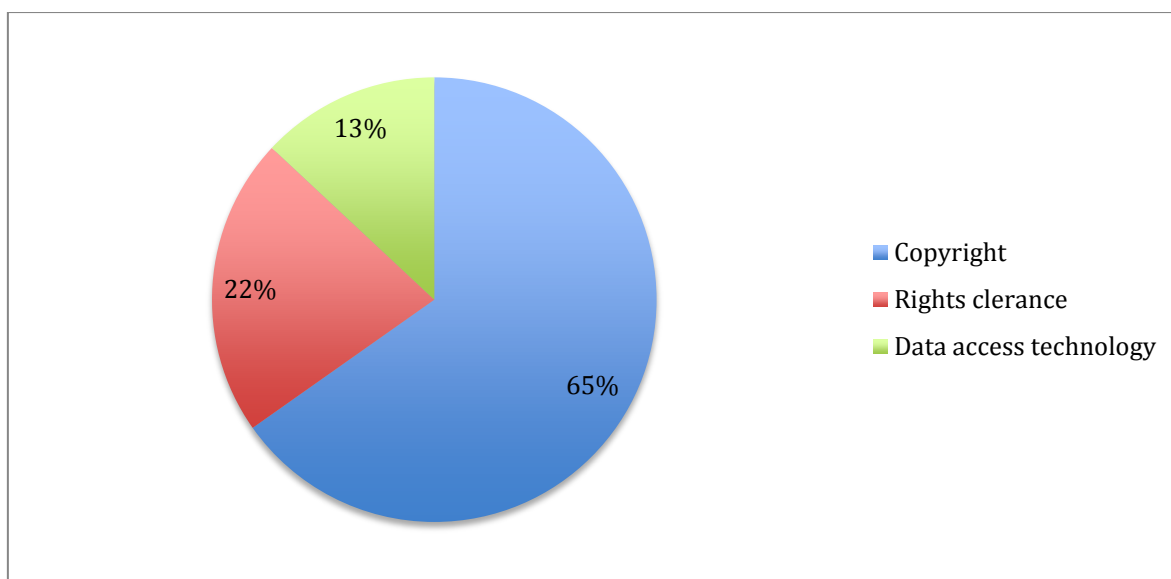
Figure 6. The most important issues preventing from publishing digital data online

Finally, Figure 7 provides detailed insight into the requirements of the respondents related to features of licensing schemas they would most likely use. The most important aspect for respondents is to have an option to assure attribution for reused content (24% of respondents). Distinction between commercial and non-commercial use is also important (23% of respondents). Other characteristics include: requirement to have feedback on how the content is used, preventing content modifications, distinguishing between research non-commercial and public non-commercial use as well as preventing further distribution of the content.
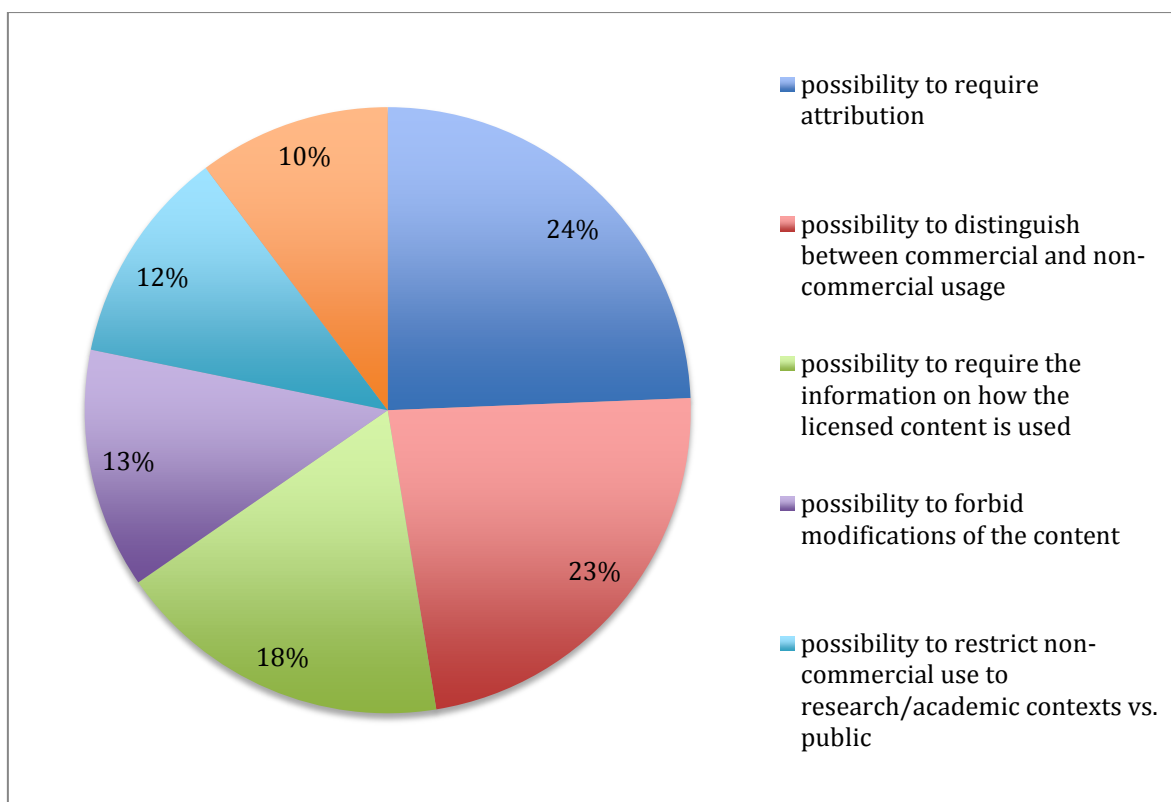
Figure 7. Most important features/characteristics of content licensing

Practices for tools licensing are summarized in Figure 8. These results show that GNU licenses are most popular (GPL itself reaches 27% plus LGPL which reached 9%). Other licenses used by the respondents include Creative Commons, BSD Licenses and Apache Licenses. An interesting observation is that respondents do use Creative Commons licenses for tools although it is discouraged by the Creative Commons itself[41].

41

http://wiki.creativecommons.org/Frequently_Asked_Questions#Can_I_apply_a_Creative_Commons_license_to_software.3F
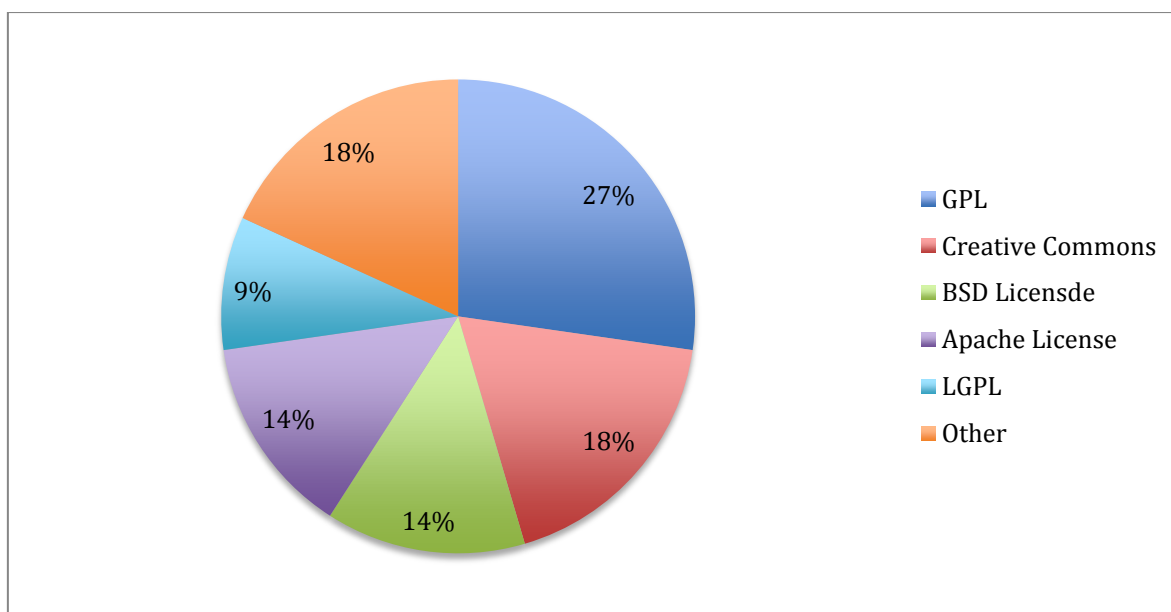
Figure 8. Licenses used to release tools

Issues preventing institutions from publishing tools were mostly related to the additional technical requirements to the source code. In order to release a tool, good documentation is a necessity, together with a good quality of the code itself. These additional efforts hold back the release of tools.

The main features of tools licensing that are interesting to respondents are summarized in Figure 9. Most important is to have the possibility to distinguish between commercial and non-commercial usage (28%). Attribution and information on how the software is used are also an important aspect (23% each). The remaining features include the possibility to prevent the distribution of the software (11%), to distinguish between non-commercial public use and research use (9%) and to prevent further modifications of the software (6%).
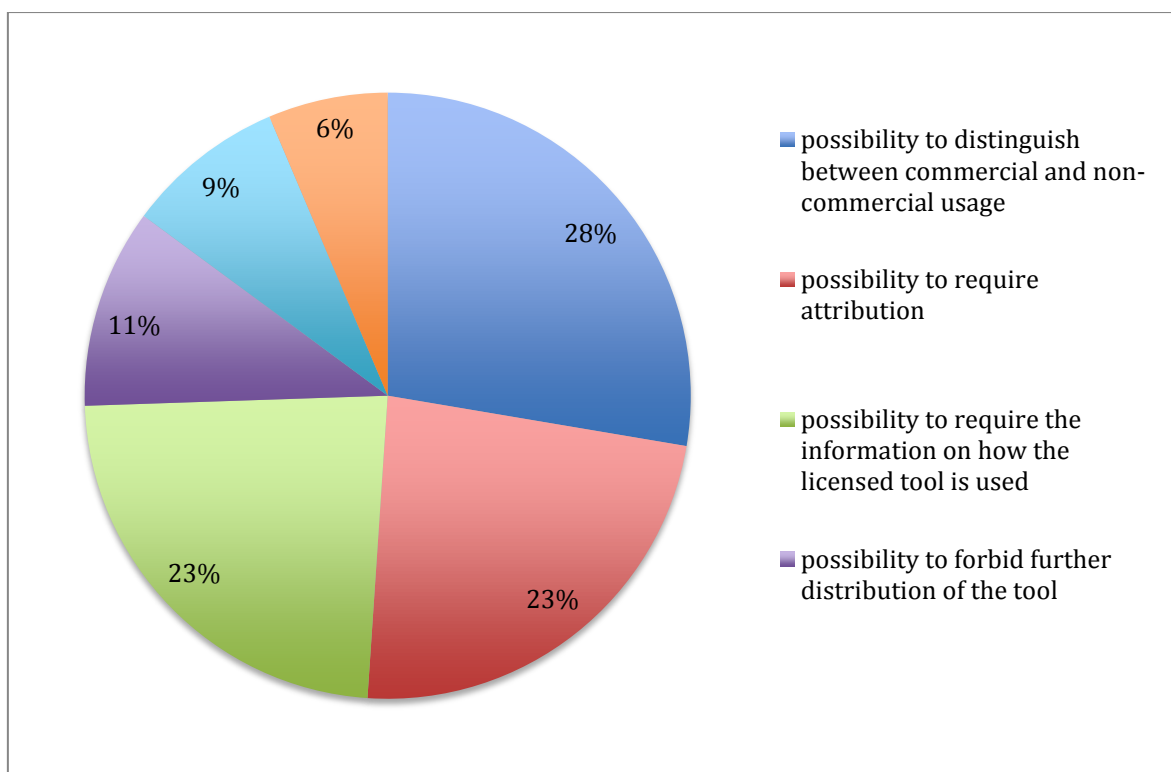
Figure 9. Most important features/characteristics of tools licensing

When comparing responses from content and tools licensing it is interesting to note when it comes to characteristics of the licenses for tools and content the requirements are quite similar. For example in both cases attribution and distinction between commercial and non-commercial use is a priority. And also three least needed characteristics are the same (possibility to forbid further distribution, possibility to restrict non-commercial user to research/academic vs. public ad well as possibility to forbid modifications).

7. SUCCEED RECOMMENDATIONS ON LICENSING SCHEMES

In this chapter we will summarize the results of our investigations. In the table below we first provide a summary of the licenses we discussed before. Like in the preceding chapters, we make a distinction between content (original creative material for human consumption), data/metadata (collection of information), and software (original creative material for computational use). To have a general overview of possible licenses and their applicability we provide summary table below. It provides information on the applicability of the investigated licenses with additional explanation comments if necessary.

Table 2. Licenses applicability matrix for digitisation tools and resources

License	Domain of aplication	Comments
Creative Commons CCZero (CC0)	Content, Data	Dedicate to the Public Domain (all rights waived)
Open Data Commons Public Domain Dedication and Licence (PDDL)	Data	Dedicate to the Public Domain (all rights waived)
Creative Commons Attribution 4.0 (CC-BY-4.0)	Content, Data	
Creative Commons Attribution (CC-BY)	Content	All versions 1.0-3.0, including jurisdiction "ports"
Open Data Commons Attribution License (ODC-BY)	Data	Attribution for data(bases)
Creative Commons Attribution Share-Alike 4.0 (CC-BY-SA-4.0)	Content, Data	
Creative Commons Attribution Share-Alike (CC-BY-SA)	Content	All versions 2.0-3.0, including jurisdiction "ports"; version 1.0 is little used and not recommended because it is incompatible with future versions
Open Data Commons Open Database License (ODbL)	Data	Attribution-ShareAlike for data(bases)
Apache License v.2.0	Software	See section 5.6.3.
BSD license	Software	See section 5.6.4.
GNU General Public License v3.0 (GPL)	Software	See section 5.6.1.
GNU Library or "Lesser" General Public License v.3.0 (LGPL)	Software	See section 5.6.2.
MIT license	Software	See section 5.6.5.
Mozilla Public License 2.0 (MPL)	Software	See section 5.6.6.
Common Development and Distribution License	Software	See section 5.6.8.
Eclipse Public License v.1.0	Software	See section 5.6.7.

7.1 Conclusions and recommendations

In the preceding chapters we have reported on existing frameworks for licensing content, data and tools. We have not been able to find a framework that covers all types of resources satisfactory. In general, license developers like Creative Commons advice not to apply their licenses to other resource types than those they were developed for[42].

In order to facilitate the adoption of tools and reuse of data, we recommend to adhere to the practice of using separate frameworks for data and software.

Also here, we will make a distinction between recommendations for licensing content, data/metadata and software. Some general considerations are useful for all types.

1. Check whether any rights are resting on the resources underlying the product. Sublicensing is only possible if the original license of the resources allows it. If sublicensing is allowed, the original licenses might impose restrictions on the type of license for (re-) distribution.
2. If the underlying resources are free of rights (see Chapter 2 for more information), it is not possible to license the product, unless the product qualifies as a derivative work. It that case, it is only possible to license the additional part of the product.

Content

Content should be provided with an explicit licence so that licensees know what they can do with it. Furthermore we recommend using popular licensing frameworks to allow maximum interoperability. The results of the survey presented in Chapter 6 show that Creative Commons was used most by the respondents. This is even more evident when considering results presented in Figure 7 (most important characteristics for the content licensing framework such as: attribution, non-commercial use and non-derivative works). This indicates that distinction that Creative Commons can provide is necessary (e.g. CC BY-NC-ND license).

Data/metadata

Data and metadata should be made as open as possible in order to optimise discovery of the content curated at libraries. We recommend using the Open Data Commons licenses or the Creative Commons 4.0 licenses. Since attribution is a strong requirement according to the results presented in Chapter 6, we specifically recommend the Open Data Commons Attribution License (ODC-By), or the CC (4.0) BY License to be used in this case. For metadata that is related to objects published online in digital libraries (that can be harvested by external parties) we recommend Europeana Licensing Framework. Such an approach will make it possible to add

---

42

https://wiki.creativecommons.org/Frequently_Asked_Questions#Can_I_apply_a_Creative_Commons_license_to_software.3F

metadata to the Europeana portal and brings the information about digital objects available to any other interested party. In practice using the Europeana Licensing Framework means that metadata is released under Creative Commons Zero Public Domain Dedication license (CC0).

Software

In general we recommend a strong copyleft license for non-commercial software. Since there are many FLOSS licenses available for software tools it is difficult to decide which particular license to use for a tool. Nevertheless, there is a current tendency towards a restricted set of licenses used by software developers. In Chapter 5 we have shown that among FLOSS licenses only several licenses are really popular (as listed in the Table 1 of Chapter 5) and just a few of them are widely used in the context of digitisation tools. These licenses are the GNU General Public License and the Apache License. In general, as is shown in the statistics of the SourceForge platform, there are a few other popular licenses in use such as the BSD License and the MIT License. Our investigation, however, points out the following three most applicable licenses for software tools in the context of digitisation:

- GNU General Public License v.3.0. This is the primary license to be considered when releasing a software tool. It assures that your work and any derivative work will be available in the same manner. It is especially recommended for tools that provide unique functionality - meaning that it does something no other already existing tool can do. In such a case the software tool can be of high interest to other parties and when they use or modify it they are obliged to release the derivative work under the same license, therefore contributing to the FLOSS community.

- GNU Library or Lesser General Public License v.3.0. This license should be considered when special circumstances appear, e.g. when our tool provides functionality that is already available in other tools licensed with a more permissive license than GPL. In such case making the software GPL licensed will limit the number of users (proprietary tools will not use our tool and will favour other solutions instead – this may happen even if the quality or readiness of our software is better).

- Apache License v.2.0. This license is most permissive and it should be used in cases where a software tool should be fully available to the user community, including for use in commercial products. It also means that the tool can be modified without the need to publish the modified source code. This license is sometimes used by commercial companies to make the FLOSS tool flexible for use in proprietary solutions. An example can be Android OS, which is licensed under Apache License v.2.0. Thanks to that it is possible for smartphone vendors to create special editions of Android that are not publicly available (the source code is not published).

8. SUMMARY

In this document we present the results of an investigation into licensing frameworks which can be used by institutions to publish content (original creative material), data (collection of information), metadata (description of data) or software tools online.

Before publishing a product, it is important to first carefully consider whether there are any right resting on (part of) it. If the product or part of it has been created by an agent outside of the institution it might be copyrighted. In that case it is necessary to investigate whether the rights on the product have been expired (depending on jurisdiction) and whether the product was originally published with a license.

In case the rights have been expired, the product in its original form cannot be licensed, but if it has been enriched with additional information, the additions can be licensed as a derivative work.
Databases and collections of metadata can also be licensed regardless of what they contain (e.g. copyrighted content or factual knowledge).

We have not been able to general frameworks for commercial licensing. This has probably to do with the fact that there are many complex factors involved like business models, market dynamics, customer segments. Therefore, commercial licenses in most cases need to be tailor-made. We present a number of conditions that could be part of a commercial license.

For non-commercial licensing there are more general frameworks available. We consider the cases of content, data/metadata and software in separation because we have not been able to find frameworks that cover them all.

For content we recommend the Creative Commons framework because Creative Commons is much in use and our survey among institutions point to a preference for the restrictions possible to impose by the Creative Commons license (e.g. BY-NC-ND). Using a license that is widely in use will contribute to the interoperability among institutions and other parties.

For data and metadata we recommend the Open Data Commons Attribution License (ODC-BY) or the Creative Commons (version 4.0) BY License. A special case is given to metadata that is published in digital libraries – we then recommend to use Europeana Licensing Framework to make the information available in the Europeana portal.

Finally, for software we recommend the GNU Public License (GPL) v3.0 in most cases. In cases where the tool provides functionality that is covered by other software that is more permissively licensed than with GPL, we recommend the GNU Library or Lesser General Public License (LGPL) v3.0. In cases where a wider uptake and the possibility of commercial applications is desired, we recommend the Apache License v2.0.

Table presented below summarizes recommendations for non-commercial licensing. It does not contain commercial licensing, because as already mentioned, commercial licensing is highly dependent on different factors (e.g. business models, customer segments) and did not create any publicly available and common approaches shared among different companies.

Table 3. Recommendations for non-commercial licensing

Type of asset	Recommendation	Comment
Content	• Creative Commons v4.0	Popular licensing framework, addresses most of the needs of content holding institutions.
Data and metadata	• Open Data Commons v1.0 • Creative Commons v4.0 • Europeana Licensing Framework	Open Data Commons and Creative Commons v4.0 are for general purpose use. Europeana Licensing Framework is recommended for metadata that can be harvested by external parties, it means in practice that metadata need to be published with Creative Commons Zero Public Domain Dedication license.
Software	• GNU General Public License v3.0 • GNU Library or Lesser General Public License v3.0 • Apache License v2.0	The license should be determined based on the requirements from the tools provider: GNU GPL v3.0 is a strong copyleft license, GNU LGPL v3.0 is a partial copyleft, while Apache License v2.0 is a permissive license.

GLOSSARY OF ABBREVIATIONS

API	Application Programming Interface
BSD	Berkeley Software Distribution
BY	Requirement of 'Attribution' in Creative Commons licenses
CC0	A tool provided by Creative Commons to waive rights on a work
CC	Creative Commons
CDDL	Common Development and Distribution License
EDLIBA	European Bureau of Library Information and Documentation Association
EPL	Eclips Public License
FLOSS	Free/Libre and Open-Source Software
FSF	Free Software Foundation
GPL	GNU General Public License
IP	Intellectual Property
IPO	Intellectual Property Office (UK)
LGPL	GNU Library or Lesser General Public License
MIT	Massachusetts Institute of Technology
MPL	Mozilla Public License
NC	Requirement of 'Non-commercial Use' in Creative Commons licenses
ND	Requirement of 'Non-derivative Work' in Creative Commons licenses
OAI	Open Archives Initiative
ODC-BY	Open Data Commons Attribution License
ODC-DBCL	Open Data Commons Database Content License
ODC-ODBL	Open Data Commons Database License
ODC	Open Data Commons
ODC-PDDL	Open Data Commons Public domain Dedication and License
OSI	Open Source Initiative
SA	Requirement of 'Share-Alike' in Creative Commons licenses

ATTACHMENT A. SUCCEED SURVEY QUESTIONNAIRE

Below you will find Succeed questionnaire. It is in form of screenshots for presenting
the same look and feel that respondents had.

If you were considering a new licensing scheme for your content, what features would most influence your choice?

☐ possibility to distinguish between commercial and non-commercial usage

☐ possibility to restrict non-commercial use to research/academic contexts vs. public

☐ possibility to require attribution

☐ possibility to require the information on how the licensed content is used

☐ possibility to forbid modifications of the content

☐ possibility to forbid further distribution of the content, so that only you can distribute it

Tools licensing

Please name the licenses you use when publishing your tools and describe the rationale for using them.

If you have tools, but do not provide them yet, please state why and what are the most important barriers.

If you were considering a new licensing scheme for your tool, what features would most influence your choice?

☐ possibility to distinguish between commercial and non-commercial usage

☐ possibility to restrict non-commercial use to research/academic contexts vs. public

☐ possibility to require attribution

☐ possibility to require the information on how the licensed tool is used

☐ possibility to forbid modifications of the tool

☐ possibility to forbid further distribution of the tool

New usages

Please indicate any new or emergent usages for digitised content which your
institution/company may be considering.

Please describe what are the issues, if any, that you need to overcome in order to put your
content to new use.

Please describe the quality requirements associated to these new ways of exploiting your
content.

Please indicate possible business opportunities connected to the new usage of your data.

Additional comments

Please indicate any additional comments or general remarks which you may have.

Submit

ATTACHMENT B. DIGITISATION AT THE BRITISH LIBRARY: PUBLIC PRIVATE PARTNERSHIPS

The British Library (BL) has developed mass digitisation partnerships with a range of publishers over the pasts 10 years. These include Cengage Learning, DC Thomson Family History, Google and others. The material digitised ranges from early English books (1500-1700), through electoral registers, to early 20th-Century newspapers.

Mass digitisation of significant parts of our vast collection lies at the heart of the British Library's strategy. Through digitisation we increase access to the collection, make the collections more usable and searchable, create a surrogate copy of content that allows us to minimise over handling of often fragile originals and makes content available in a form that is familiar to the 'born digital' generation.
The ambition is one thing but pretty soon you bump up against the issue of cost. Digitisation is incredibly expensive: for example, we estimate that producing a single page of newspaper content in searchable, digital form would cost us at least 1.2 Euros. To put this into context we have 750 million pages of newspaper content in total. Partnership, therefore, is one way to gain funding for this expensive process.

Ownership
It is vital that the licensor retains ownership of its digital assets. At the BL the primary driver for digitisation is not generation of income (though this is still important) but finding partners and working models that will fund the highly expensive business of digitisation whilst at the same time allowing you to retain control of the asset. This lies at the very heart of what you, as the licensor are trying to achieve – a fully owned digital asset. Under our model, the licensee, having paid for the digitisation, is allowed a period in which to commercialise the content. On expiry of that period (typically between 7 and 10 years), the BL is free to make choices as to how the database might be further exploited. One important consideration for a public institution is to make the content freely available to all.

Ownership (of both the scans and the OCR) then allows the licensor to operate within a virtuous circle in which the process is paid for and the ownership and control lies with the library.

Technical Standards
The speed, cost and quality of digital imaging gets better all the time.

It is vital to be realistic about what you are doing. If the project involves the scanning of millions of pages of mainly textual material, then the quality of the image that you will require is going to be very different from the short run, high resolution 'in facsimile' experience where the object itself is as high in interest as the words.

The operating principle here is 'good enough' for the purposes to which the digital asset is being put. In addition, in the light of the huge investments made by the licensee realism must rule and an insistence on boutique digitisation standards for mass digitisation is entirely unrealistic. In this area you really need to listen to what your commercial partner has to say.

Condition check
Time invested in a thorough audit of the material to be digitised always pays dividends.

Fragile materials will require special handling; awkward material (e.g. materials that have been tightly bound; materials that are of an odd shape) will also require extra time, care and attention. All of this means that extra time and budget will need to be set aside by the commercial partner. The partner should be encouraged to take a clear eyed view of the project and operate on the principle of digitisation often taking longer than was original anticipated.

Flexibility of business models
There is no single business model that works for all commercial partners.

Each partner will likely have rather different motivations for getting into a commercial relationship and each will have a preferred model (within the limitations prescribed by the level playing field under the PSI Directive). The driver (and the business model) for a partner such as Google will be different from those of a Higher Education publisher like Cengage Learning.

Transparency
Recent legislation has changed the rules. Under the Revised Public Sector Information (PSI) Directive[43] (2013) there is a requirement to operate transparently and on a 'level playing field'. This means that contractual terms and conditions will, on the whole, be available for one and all to review. This, coupled with the need for a level playing field means that going forward the licensor will need to be very certain in that what terms are offered to one partner is offered to another. Varying terms and conditions in critical areas such as the length of contract, royalty rates, ownership of the asset, indemnities and so on will be problematic and can be subject to challenge from other licensees if there is significant variance. It is important therefore to be up front about what your terms and conditions are, that these cannot be subject to wide deviation without justification. It seems therefore that we are moving towards greater standardisation of terms and conditions in digitisation partnership projects.

Cover yourself
You should not be in the business of providing indemnities to third parties. Your partner should assume all commercial risk and it is their responsibility to ensure that

---

[43] http://ec.europa.eu/digital-agenda/en/news/consolidated-version-psi-directive-now-available

they assess all other risks.

In areas such as data protection, re-publication of libel and breach of copyright, the contract should state that rigorous due diligence be exercised by the partner and responsibility for breach lies with the licensee. In addition, where there is the possibility that the interests of a third party may be breached, a vigorous notice and takedown policy should be written into the contract.

Maximising access and reuse

Modern communications encourage sharing, embedding, repurposing and so on of digitised content – indeed this is one of the great benefits of digital content. Many commercial partners will push back hard to restrict reuses to a tiny percentage of the material and to narrow the range of repurposing activities: the partner will see this 'free' use as a threat to their commercial model. The licensor should push back equally hard in the interests of the user and seek to maximise the uses to which the materials can be put.

Maximise market penetration

Different publishers may reach different parts of the market. It is quite possible to negotiate a 'share' of content where one publisher sells and markets to say the higher education market and another publisher sells to a more general consumer market.

You are not the commercial partner

Having gone to all the trouble of choosing a private sector partner you should allow them to do what they do best: act commercially. There is nothing more aggravating to a commercial partner than an overly controlling public sector partner. On matters relating to sales, marketing, pricing, distribution and so on, the partner must be free to make decisions. Other than three areas (brand, PR and content selection) the licensor is well advised to avoid meddling.