



OPEN DATA SUPPORT

Training Module 2.5

Data & metadata licensing



PwC firms help organisations and individuals create the value they're looking for. We're a network of firms in 158 countries with close to 180,000 people who are committed to delivering quality in assurance, tax and advisory services. Tell us what matters to you and find out more by visiting us at www.pwc.com. PwC refers to the PwC network and/or one or more of its member firms, each of which is a separate legal entity. Please see www.pwc.com/structure for further details.

Presentation metadata

Open Data Support is funded by the European Commission under SMART 2012/0107 'Lot 2: Provision of services for the Publication, Access and Reuse of Open Public Data across the European Union, through existing open data portals'(Contract No. 30-CE-0530965/00-17).

© 2013 European Commission

This presentation has been created by PwC

Authors:

**Makx Dekkers, Nikolaos Loutas, Michiel De Keyzer
and Stijn Goedertier**

Disclaimers

1. The views expressed in this presentation are purely those of the authors and may not, in any circumstances, be interpreted as stating an official position of the European Commission. The European Commission does not guarantee the accuracy of the information included in this presentation, nor does it accept any responsibility for any use thereof. Reference herein to any specific products, specifications, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favouring by the European Commission. All care has been taken by the author to ensure that s/he has obtained, where necessary, permission to use any parts of manuscripts including illustrations, maps, and graphs, on which intellectual property rights already exist from the titular holder(s) of such rights or from her/his or their legal representative.
2. This presentation has been carefully compiled by PwC, but no representation is made or warranty given (either express or implied) as to the completeness or accuracy of the information it contains. PwC is not liable for the information in this presentation or any decision or consequence based on the use of it.. PwC will not be liable for any damages arising from the use of the information contained in this presentation. The information contained in this presentation is of a general nature and is solely for guidance on matters of general interest. This presentation is not a substitute for professional advice on any particular matter. No reader should act on the basis of any matter contained in this publication without considering appropriate professional advice.

Learning objectives

By the end of this training module you should have an understanding of:

- The importance of licensing.
- The meaning of “open” in Open Data.
- The licensing requirements in the revised PSI Directive.
- Creative Commons and the Open Data Commons.
- The licensing options for data and metadata and the consequences for sharing and reuse.
- The Europeana Licensing Framework as a practical example.

Content

This module contains...

- The importance of licensing
- Licensing in the Open Data Principles
- Licensing in the revised PSI Directive
- Licensing options and good practice for reuse of data
- Licensing options and good practice for reuse of metadata
- A scenario for reuse of metadata
- Case study: Europeana

The importance of Licensing

Clear licence information is important because...

- It **tells users** and **reusers** exactly **what they can do** with your data and metadata.
- It **encourages the use and reuse** of your data and metadata the way you want them to be used and reused.
- It **creates visibility** of your efforts downstream (if you ask for attribution).

If no explicit licence is provided, a user does not know what can be done with the data/metadata – the default legal position is that nothing can be done without contacting the owner on a case-by-case basis.

Licensing in the Open Data Principles

*How licences appear in the basic principles of open data
and why licensing of open (meta)data is important.*

The Open Data Definition

It also covers metadata

*“A piece of data or content is open if **anyone is free to use, reuse, and redistribute** it — subject only, at most, to the requirement to **attribute and/or share-alike**”*

-- opendefinition.org

This means, according to the Open Knowledge Foundation:

- **Availability and Access:** the data must be available as a whole and at no more than a reasonable reproduction cost, preferably by downloading over the internet. The data must also be available in a convenient and modifiable form.
- **Reuse and Redistribution:** the data must be provided under terms that permit reuse and redistribution including the intermixing with other datasets.
- **Universal Participation:** everyone must be able to use, reuse and redistribute - there should be no discrimination against fields of endeavour or against persons or groups. For example, ‘non-commercial’ restrictions that would prevent ‘commercial’ use, or restrictions of use for certain purposes (e.g. only in education), are not allowed

Licensing is the first star...



One star: publish data under an **open licence**



Two stars: publish in **machine-readable** format



Three stars: publish in **open format**



Four stars: assign **URIs** to data



Five stars: create **links** to other data

See also:

<http://www.slideshare.net/OpenDataSupport/introduction-to-linked-data-23402165>

Licensing in the revised PSI Directive

Obligations of Member States according to the PSI Directive

Public sector bodies are obliged, for all information they hold related to their public task and that is not explicitly mentioned in one of the exceptions, to:

- Make information reusable for commercial or non-commercial purposes under non-discriminatory conditions.
- Process requests and provide access within 20 days (or 40 if request is complex); justify negative decision and inform about how to appeal.
- Charge no more than the marginal cost incurred for reproduction, provision and dissemination; the charges (if any) and the calculation basis for those charges are to be pre-established and published, through electronic means where possible and appropriate.
- Publish licences in digital format.
- Make information available in pre-existing format and language, and where possible and appropriate, in open and machine-readable format together with their metadata. Both the format and the metadata should, in so far as possible, comply with formal open standards.
- Put in place practical arrangements that facilitate the search for documents available for reuse, such as assets lists, accessible preferably online, of main documents, and portal sites that are linked to decentralised assets lists.

Directive 2013/37/EU says...

- Any licences [...] should [...] place **as few restrictions on reuse as possible [...]. Open licences** available online, which grant wider reuse rights without technological, financial or geographical limitations and relying on open data formats, **should play an important role** in this respect. Therefore, **Member States should encourage the use of open government licences [...].**
 - *Recital (26)*
- Public sector bodies may **allow reuse without conditions or may impose conditions**, such as indication of source, **where appropriate** through a licence. These conditions shall not unnecessarily restrict possibilities for reuse and shall not be used to restrict competition.
 - *Article 8, paragraph 1*

See also:

<http://www.slideshare.net/OpenDataSupport/the-psi-directive-and-open-government-data>

Consequences of the PSI Directive with regards to licensing

Make information re-usable for commercial or non-commercial purposes under non-discriminatory conditions

Conditions need to be pre-established, transparent and the same for everyone. The Directive encourages the use of an **open licence**.

Publish licenses in digital format

Explicit licences should be associated with the available data.

Facilitate search for information, preferably online (e.g. portal)

This implies **public availability** of descriptions of available data.

Licensing options and good practices

The case of data – different options exist for licensing your data depending on its nature.

Licensing datasets

- If you want to attach no restrictions to your data: Say it!
- Every dataset should have a licence associated to it.
 - Without an explicit licence, reuse is restricted.
- The **objective** should be to make data(sets) **as openly available as possible**, within the boundaries of the law.

But how can I know which licence is fit for purpose?

Different data have different licensing needs

- Some data(sets) may be required to be **openly available**.
 - e.g. subject to a Freedom of Information Act
- Some data(sets) may be **subject to restrictions**.
 - e.g. privacy, national security, third party rights
- Some data(sets) may be **available for reuse but not for modification**.
 - e.g. legal texts, public budgets (if modifications are made, it must be made clear that the data is not the actual authentic version)
- Some data(sets) may be published **allowing derivations** with attribution of authoritative source.
 - e.g. legal commentary, translations

Licensing approaches: Creative Commons



Public Domain Mark/No rights reserved – Others may freely build upon, enhance and reuse the works for any purposes without restriction under copyright or database law.



Attribution – Others can distribute, remix, tweak, and build upon your work, even commercially, as long as they credit you for the original creation.



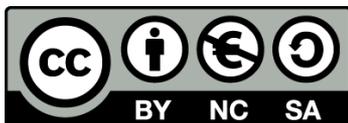
Attribution-ShareAlike – Others can remix, tweak, and build upon your work even for commercial purposes, as long as they credit you and license their new creations under the identical terms.



Attribution-NonDerivs – Allows for redistribution, commercial and non-commercial, as long as it is passed along unchanged and in whole, with credit to you.



Attribution-NonCommercial – Others can remix, tweak, and build upon your work non-commercially, and although their new works must also acknowledge you and be non-commercial, they don't have to license their derivative works on the same terms.



Attribution-NonCommercial-ShareAlike – Others can remix, tweak, and build upon your work non-commercially, as long as they credit you and license their new creations under the identical terms.



Attribution-NonCommercial-NoDerivs – Only allows others to download your works and share them with others as long as they credit you, but they can't change them in any way or use them commercially.



Good practices for licensing your data

Good practices:

- ✓ If the original data is in the public domain (e.g. by law), keep it there – use for example the Creative Commons Zero Public Domain Dedication or the Open Data Commons Public Domain Dedication and License (PDDL)
- ✓ For some documentation integrity needs to be protected – use a No-Derivatives licence, for example Creative Commons Attribution-NoDerivs, but only if really necessary
- ✓ Avoid Non-Commercial licences if at all possible, as these seriously restrict reuse.

Licences for data should provide appropriate security and control (but not more than that).

UK Government licence for PSI

You are free to:

- ✓ copy, publish, distribute and transmit the Information;
- ✓ adapt the Information;
- ✓ exploit the Information commercially for example, by combining it with other Information, or by including it in your own product or application.

You must, where you do any of the above:

- ⚠ acknowledge the source of the Information by including any attribution statement specified by the Information Provider(s) and, where possible, provide a link to this licence;

If the Information Provider does not provide a specific attribution statement, or if you are using Information from several Information Providers and multiple attributions are not practical in your product or application, you may consider using the following:

Contains public sector information licensed under the Open Government Licence v1.0.

- ⚠ ensure that you do not use the Information in a way that suggests any official status or that the Information Provider endorses you or your use of the Information;
- ⚠ ensure that you do not mislead others or misrepresent the Information or its source;
- ⚠ ensure that your use of the Information does not breach the Data Protection Act 1998 or the Privacy and Electronic Communications (EC Directive) Regulations 2003.

These are important conditions of this licence and if you fail to comply with them the rights granted to you under this licence, or any similar licence granted by the Licensor, will end automatically.

⚠ Exemptions

This licence does not cover the use of:

- personal data in the Information;
- Information that has neither been published nor disclosed under information access legislation (including the Freedom of Information Acts for the UK and Scotland) by or with the consent of the Information Provider;
- departmental or public sector organisation logos, crests and the Royal Arms except where they form an integral part of a document or dataset;
- military insignia;
- third party rights the Information Provider is not authorised to license;
- Information subject to other intellectual property rights, including patents, trademarks, and design rights; and
- identity documents such as the British Passport.

<http://www.nationalarchives.gov.uk/doc/open-government-license/version/1/open-government-license.htm>

Using an open and unrestricting licence for your data

Whenever data is licensed for open and unrestricted access, reusers can create new knowledge from combining it.

For example,

- **Cross-referencing** public spending with geographic data to visualise which regions are better funded.
- **Matching** public transport timetables with GPS data to be able to give real time information on delays.
- **Measuring performance** of public services based on transaction counters and waiting times.
- **Deriving recommendations** for prevention policies relating accident statistics with weather data and road maps.

Protecting against liability

Liability risks are related to:

- Infringement on third-party rights (personal data, copyright, database rights etc.)
 - Rights must be cleared and data must be anonymised
- (In)correctness of data
 - A disclaimer makes clear in how far the publisher guarantees correctness of the data
- Unfair competition to market parties already selling the information
 - In such cases, market parties need to be consulted, e.g. providing a phasing-in period

Source: Marc de Vries. Open Data and Liability. EPSIplatform Topic Report No. 2012/13.
<http://epsiplatform.eu/sites/default/files/Final%20TR%20Open%20Data%20and%20Liability.pdf>

Licensing options and good practices

The case of metadata – publish your metadata under a public domain licence to ensure wide distribution & reuse.

Which licences are suited for metadata?

The following licenses allow **fully open reuse**:

- Public Domain Mark
- Creative Commons Zero Public Domain Dedication
- Open Data Commons Public Domain Dedication and license (PDDL)

The following licenses are also used but lead to “**attribution stacking**”, the need to keep track of a chain of attributions:

- CC-BY
- ODC Attribution
- ISA Open Metadata licence 1.1
- Open Government licence

Other licences (non-commercial, non-derivatives, share-alike) are less suited; they make reuse of metadata for Linked Data applications difficult because they place restrictions on how metadata can be shared, used and enhanced.

Example: Discovery Open Metadata Principles

Discovery Open Metadata Principles

Discovery, the metadata ecology for UK education and research, invites stakeholders to join us in adopting a set of principles to enhance the impact of our knowledge resources for the furtherance of scholarship and innovation.

1. **Open metadata** creates the opportunity for enhancing impact through the release of descriptive data about library, archival and museum resources. It allows such data to be made freely available and innovatively reused to serve researchers, teachers, students, service providers and the wider community in the UK and internationally.
2. **Libraries, archives and museums** exist in an environment where they, their stakeholders and partners share an interest in exposing and repurposing metadata.
3. **Government and the wider community** are increasingly embracing a commitment to openness, sharing and reuse of metadata, especially where resources have been collected using public funds.
4. **We therefore recognise** the importance of promoting a clear endorsement of open metadata with practical guidance about licensing.
5. **We recommend** that institutions and agencies should proceed on the presumption that their metadata is by default made freely available for use and reuse, unless explicitly precluded by third party rights or licences.
6. **We strongly advocate** that all metadata releases require licensing, for which institutions and agencies should adopt a standard open licensing framework that is suited to their purposes.
7. **Reference to permissible usage** under the terms of a standard open licence will promote confident and appropriate use. When licensing open metadata in the majority of circumstances, the standard Open Data Commons Public Domain Dedication & Licence ([ODC-PDDL](#)), the broadly similar [Creative Commons CC0 licence](#) or the [UK Open Government Licence](#) (OGL) will be appropriate.
8. **Avoidance of variations** to such standard licences will make it easier to combine data from different resources and will reduce repeated requirement for legal advice.

Good practices for licensing your metadata

What you need to think about:

- ✓ Metadata helps people to discover your data.
- ✓ The wider your metadata is distributed, the higher your visibility is.
- ✓ Others may want to add to it, enhance it, link to other resources.

Good practices:

- ✓ Licences for metadata should be as open as possible.
- ✓ A public domain licence allows the widest reuse.
- ✓ An attribution licence ensures you get credit downstream, but may cause problems if data is shared multiple times (attribution stacking).

A scenario for reuse of metadata

A reuse scenario for metadata published under a public domain licence.

What can reusers do with metadata in the public domain

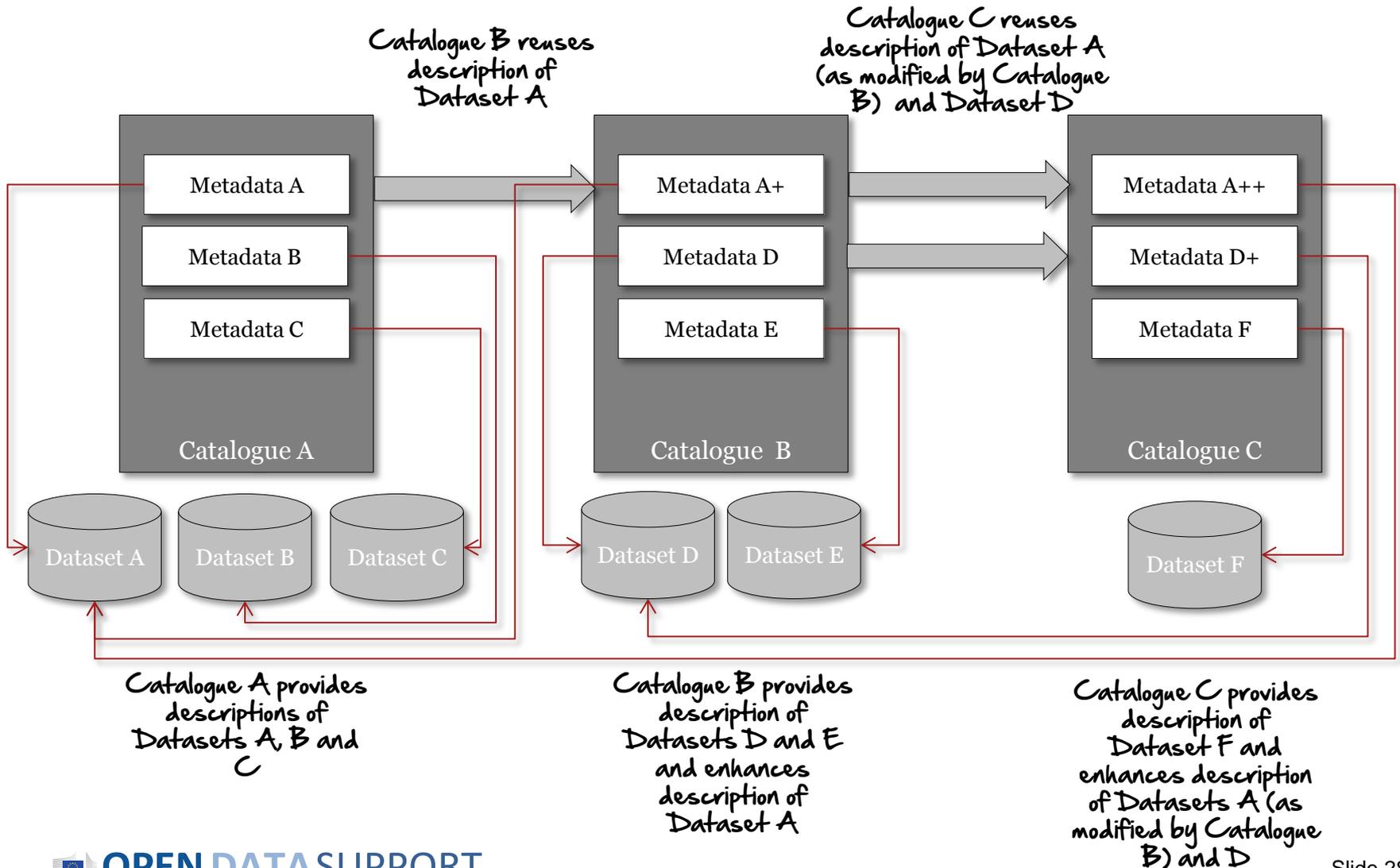
- **Copy & include** your metadata in **search engines and brokers** that refer back to the location of your data.
- **Correct** them if the original metadata contains errors.
- **Enhance** your metadata, for example by converting text strings to links.
 - This can happen if the reuser is aware of URI sets for organisations, subjects and other things to which your metadata refers
- **Augment** the metadata by making additional assertions about your data.

By harvesting metadata back from the reusers, you may be able to increase the quality of your metadata.

See also:

<http://www.slideshare.net/OpenDataSupport/promoting-the-reuse-of-open-data-through-odip>

Reuse scenario of metadata for datasets (1/2)



Reuse scenario of metadata for datasets (2/2)

Original metadata in Catalogue A

```
:dataset-A
  a dcat:Dataset ;
  dct:title "Dataset A" ;
  dct:description "This is example dataset A" ;
  dct:issued "2013-06-17"^^xsd:date ;
  dct:modified "2011-12-05"^^xsd:date ;
  dct:publisher :exampleOrg ;
  dct:language <http://publications.europa.eu/resource/authority/language/ENG> ;
  dcat:contactPoint :exOrg-vCard ;
  dcat:theme :theme-A ;
  dcat:distribution :dataset-A-csv ;
.
```

Modified metadata in Catalogue B; added local identifier “CatB-IdX”

```
:dataset-A
  a dcat:Dataset ;
  dct:title "Dataset A" ;
  dct:description "This is example dataset A" ;
  dct:issued "2013-06-17"^^xsd:date ;
  dct:modified "2011-12-05"^^xsd:date ;
  dct:publisher :exampleOrg ;
  dct:language <http://publications.europa.eu/resource/authority/language/ENG> ;
  dcat:contactPoint :exOrg-vCard ;
  dcat:theme :theme-A ;
  dct:identifier "CatB-IdX" ;
  dcat:distribution :dataset-A-csv ;
.
```

Modified metadata in Catalogue C; added keyword “example”

```
:dataset-A
  a dcat:Dataset ;
  dct:title "Dataset A" ;
  dct:description "This is example dataset A" ;
  dct:issued "2013-06-17"^^xsd:date ;
  dct:modified "2011-12-05"^^xsd:date ;
  dct:publisher :exampleOrg ;
  dct:language <http://publications.europa.eu/resource/authority/language/ENG> ;
  dcat:contactPoint :exOrg-vCard ;
  dcat:theme :theme-A ;
  dcat:keyword "example" ;
  dct:identifier "CatB-IdX" ;
  dcat:distribution :dataset-A-csv ;
.
```

Pros & cons of public domain licence

Release of ownership:

- ➖ No-one will know that you created the initial metadata.
- ➕ Enables community-driven enhancement.

Loss of control:

- ➖ You will not know what statements are made about your data.
- ➕ Quality control will emerge from the community (cf. Wikipedia).

Reliability:

- ➖ A user will not know whether the metadata is accurate and up-to-date.
- ➕ Network partners (chains of aggregators) will be able to track quality.

Misrepresentation:

- ➖ Additions and modifications may be wrong or not to your liking.
- ➕ Anyone can say anything about anything in any case; if based on your original metadata there is a higher chance the metadata is correct.

Case study: Europeana

How Europeana overcame data and metadata licensing challenges.

Europeana – original approach

In 2009, Provider and Aggregator Agreements were signed. These included the restriction that metadata could only be used for non-commercial purposes.

However, this **made it impossible for metadata** to be:

- Published as Linked Open Data.
- Used on sites that include advertisements.
- Shared with Wikipedia (which does not allow such restriction).
- Used by commercial companies, e.g. for listing in search engines.
- Used by commercial apps.

Perceived risks of providing open metadata? (1/2)

- 1. Loss of quality:** the high-quality metadata provided will be divorced from the original trusted source and corrupted by third parties.
- 2. Loss of control:** institutions will no longer be able to control the metadata if anyone can reuse or distribute it.
- 3. Loss of unity:** metadata will get scattered across the digital universe while it should be (contextually) kept together.
- 4. Loss of brand value:** by releasing data openly the institution risks being associated with reusers that they do not want to be associated with.
- 5. Loss of attribution:** by releasing data under an open licence institutions will not be credited as the source/owner of the metadata.

Perceived risks of providing open metadata? (2/2)

6. **Loss of income:** institutions are afraid that they cannot replace current revenues from metadata with other sources of income.
7. **Loss of potential income:** in the future, institutions may think of a way to make money from metadata, but if they release it openly now someone else may do this.
8. **Unwanted spill-over effects:** institutions find it unfair that others make money with the metadata that they provide.
9. **Losing customers:** if data is openly available customers will go elsewhere to get the information they are looking for.
10. **Privacy:** there are privacy restrictions on the use of certain data.

Identified benefits of open metadata (1/2)

- 1. Increasing relevance:** open metadata can be used in places where online users congregate (including social networks). This helps providers to maintain their relevance in today's digital society.
- 2. Increasing channels to end users:** providers releasing data as open metadata increase the opportunities that users see their data and their content.
- 3. Data enrichment:** open metadata can be enriched by Europeana and other parties and can then be returned to the data provider. Opening the metadata will increase the possibility of linking that data and the heritage content it represents with other related sources/collections.
- 4. Brand value** (prestige, authenticity, innovation): releasing data openly demonstrates that the provider is working in the innovation vanguard and is actively stimulating digital research.
- 5. Specific funding opportunities:** releasing metadata openly will potentially grant providers access to national and/or European funding (European and most national governments are actively promoting open metadata).

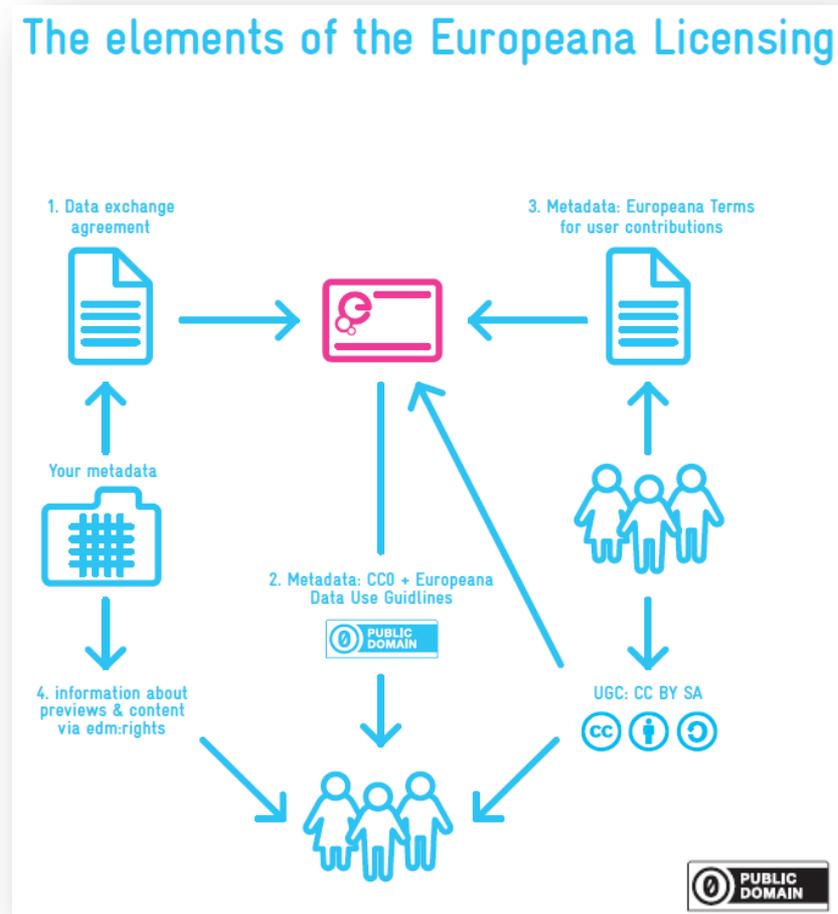
Identified benefits of open metadata (2/2)

- 6. Discoverability:** increased use and visibility of data drives traffic to the provider's website.
- 7. New customers:** releasing data openly offers new ways to interact with and relate to customers.
- 8. Public mission:** releasing metadata openly aligns the provider with the strategic public mission of allowing the widest possible access to cultural heritage.
- 9. Building expertise:** releasing metadata openly will strengthen the institution's expertise in this area, which will become a marketable commodity such as consulting services.
- 10. Desired spill-over effects:** institutions and creative industries will be able to create new businesses, which in turn will strengthen the knowledge economy.

Europeana Licensing Framework

Four layers:

1. **Physical objects:** *ownership or public domain* as appropriate.
2. **Digital objects** representing the physical objects: rights statement to be either *Public Domain*, or a *Creative Commons licence* or *Rights Reserved* (free, paid, or restricted access).
3. **Previews** (e.g. thumbnails): Europeana has *right to use but not to distribute unless licence allows this*.
4. **Descriptive metadata:** to be provided under *CC Zero Public Domain Dedication*, so reuse is unrestricted; in addition, provider should make best effort to correctly state intellectual property rights of digital objects.



Conclusions

- **Data and metadata** should be provided with **an explicit licence** so that reusers know what to do with the metadata and data and allow for maximum interoperability.
 - **Metadata** should be made **as open as possible**, ideally CC Zero or Public Domain Dedication to allow for network effects.
 - **Data** should be released under a **licence that enables appropriate protection** (but not more than necessary).

and don't forget...

- If no explicit licence is provided, a user does not know what (if anything) can be done with the data.
- No reuse = no social and economic value.

Group questions



<http://www.visualpharm.com>

Do you have in your country a licence for open data and/or metadata? If not, which should, in your opinion, be the preferred approach?



<http://www.visualpharm.com>

What are/were the greatest barriers to publishing your data under an open licence?

Thank you!
...and now YOUR questions?

References

Slide 8:

- The Open Knowledge Foundation. Open Definition. <http://opendefinition.org/>
- The Open Knowledge Foundation. Open Data - An Introduction. <http://okfn.org/opendata/>

Slide 9:

- LOD Around The Clock (LATC). 5 ★ Open Data. <http://5stardata.info/>

Slide 11:

- Directive 2013/37/EU of the European Parliament and of the Council amending Directive 2003/98/EC on the reuse of public sector information. <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2013:175:0001:0008:EN:PDF>
- European Commission. Revision of the PSI Directive. http://ec.europa.eu/information_society/policy/psi/revision_directive/index_en.htm

Slide 17:

- Creative Commons. About the licenses. <http://creativecommons.org/licenses/>

Slide 19:

- (UK) National Archives. Government license for public sector information. <http://www.nationalarchives.gov.uk/doc/open-government-license/version/1/open-government-license.htm>

Slide 21:

- Marc de Vries. Open Data and Liability. EPSIplatform Topic Report No. 2012/13. <http://epsiplatform.eu/sites/default/files/Final%20TR%20Open%20Data%20and%20Liability.pdf>

Slide 24:

- Discovery. Discovery Open Metadata Principles. <http://discovery.ac.uk/businesscase/principles/>

Slide 31-36:

- Europeana. The Problem of the Yellow Milkmaid: A Business Model Perspective on Open Metadata. White Paper No.2. November 2011. <http://pro.europeana.eu/documents/858566/2cbf1f78-e036-4088-af25-94684ff90dc5>
- Jill Cousins. Europeana. Data Exchange Agreements. May 2011. <http://bit.ly/14Hwe5D>

Slide 37:

- The Europeana Licensing Framework. <http://pro.europeana.eu/documents/858566/7f14c82a-f76c-4f4f-b8a7-600d2168a73d>

Further reading



N. Korn and C. Oppenheim. Licensing Open Data: A Practical Guide.
<http://discovery.ac.uk/businesscase/principles/>



Europeana. The Problem of the Yellow Milkmaid: A Business Model Perspective on Open Metadata. White Paper No.2. November 2011.
<http://pro.europeana.eu/documents/858566/2cbf1f78-e036-4088-af25-94684ff90dc5>

Related projects and initiatives



Revision of the PSI Directive,

http://ec.europa.eu/information_society/policy/psi/revision_directive/index_en.htm



Europeana Licensing Framework,

<http://pro.europeana.eu/documents/858566/7f14c82a-f76c-4f4f-b8a7-600d2168a73d>



Creative Commons Licenses, <http://creativecommons.org/licenses/>



Open Data Commons – Licenses, <http://opendatacommons.org/licenses/>

LAPSI 2.0

The European Thematic Network on Legal Aspects of Public Sector Information, <http://www.lapsi-project.eu/>



EC ISA Programme, ISA Open Metadata licence v1.1.

<https://joinup.ec.europa.eu/category/license/isa-open-metadata-license-v11>



Be part of our team...

Find us on



[Open Data Support](http://www.slideshare.net/OpenDataSupport)

<http://www.slideshare.net/OpenDataSupport>



[Open Data Support](http://goo.gl/y9ZZI)

<http://goo.gl/y9ZZI>

Follow us



[@OpenDataSupport](https://twitter.com/OpenDataSupport)

Join us on



joinup

<http://www.opendatasupport.eu>

Contact us

contact@opendatasupport.eu