

Summary Record of the
"Public hearing on guidelines on recommended standard licenses, datasets and charging
for the reuse of public sector information"

25th November 2013, 10:00 – 16:30

Euroforum (EUFO) building
10, rue Robert Stumper, L-2557
LUXEMBOURG

AGENDA

10.00 – 10.15 Registration and coffee

10.15 – 10.30 Context and objectives of the hearing – European Commission, CNECT G3 Unit

10.30 – 11.30 **SESSION I: Recommended datasets.** Interventions by:

European Commission - Yvo Volman

The National Archives (United Kingdom) - Jim Wretham

Austrian Economic Chambers - Georg Hittmair

LAPSI - Prodromos Tsiavos

11.30 – 12.30 **SESSION II: Charging for re-use. (1)** Interventions by:

European Commission - Szymon Lewandowski

Alliance Big Data - Denis Berthault

Wikimedia Deutschland - Mathias Schindler

PSI Alliance - Nikolaus Futter

Dutch Chamber of Commerce - Jan Leering

12.30 – 13.30 LUNCH

13.30 – 15.00 **SESSION II: Charging for re-use. (2)** Interventions by:

W3C – Phil Archer

Federation of European Publishers – Anne Bergman

15.00 – 15.30 **SESSION III: Standard licenses.** Interventions by:

National Library of Luxembourg - Patrick Peiffer

EDPS - Zsuzsanna Belenyessy

LAPSI – Prodromos Tsiavos

Discussion

15.30 – 15.45 Closing remarks

Ms. Marta Nagy-Rothengass (Head of Unit CNECT.G3) welcomed the participants and informed them about the organization of the public hearing. She underlined that opening up public data resources for re-use is one of the key actions of the Digital Agenda for Europe whose overall aim is to deliver sustainable economic and social benefits from a digital single market. Mr. Yvo Volman (Deputy Head of Unit.G3) drew attention of the participants to the fact that policies around open data are receiving more attention internationally, underlining the importance of the Open Data Charter released by the G8 last June. In addition, he remarked that **the hearing was not about changing but clarifying concepts in order to make the PSI Directive as useful as possible.**

The hearing is part of a broader consultation process on recommended standard licensing, datasets and charging for the re-use of public sector information (PSI) launched by the European Commission on 30.08.2013. This consultation included an online survey, accessible via 'Your Voice in Europe', where all interested parties were invited to share their views. The final objective of both the hearing and the online survey is to contribute to the adoption of a common EU-wide approach to licensing terms, charging practices and the publishing of high-value datasets.

The meeting was divided into three thematic blocks: recommended datasets, charging for re-use and standard licenses; and included the participation of 42 stakeholders. Although some individual presentations were announced in the agenda, the organization was flexible in order to allow new interventions from the participants.

Morning session

Block I: Datasets

The European Commission presented the overall figures and preliminary results of the online consultation, regarding datasets. The consultation, that included almost 300 submissions, had been open since 30.08.2013 and was closed three days before the meeting (22.11.2013). The key outcomes of the online consultation are as follows:

- There is a general perception that defining core datasets could lead to a commitment for ensuring updates.
- Slight majority in favor of machine-readability and quality-guarantees as the main desired features of core datasets.
- Value for both commercial and non-commercial re-use as equal criteria to identify core datasets.

Interventions:

Jim Wretham, The National Archives (United Kingdom)

Mr. Wretham summarized the latest open data developments in the UK, underlining the following:

- The commitment of the UK Government with open data and transparency, and its actions in order to adopt the G8 Open Data Charter during the last G8 Summit (17-18 June 2013).
- The current importance of activities adopted as part of the data strategy: a portal (data.gov.uk) that includes more than 10.000 datasets, the Data Strategy Board, the National Information Structure and the Shakespeare review, which has had a huge emphasis on PSI through listing data and prioritizing datasets.
- The UK Action Plan 2013 that includes 4 principal uses of data: location, performance and delivery, fiscal and operational. He also pointed out the benefits of open data on three levels:
 - Open data generates innovation & savings, that will improve public services.
 - It enables greater scrutiny, a positive outcome for civil society.
 - It enhances services in the private sector.

Georg Hittmair, Austrian Economic Chambers

Mr. Hittmair identified two features as the "crown jewels" of data [for small Austrian private re-users]. Data should be:

- Easy to implement. Data that allows proven business models with short lead times and simple structures.
- Easy to obtain. Long-lasting procedures are not affordable.
- Main obstacle: member states often defend what they regard as "their" data and revenues.

Prodromos Tsiavos, LAPSI

Key ideas:

- To exclude personal data of the datasets is not viable, as some of the most important datasets (registers) include personal data.
- Privacy becomes very difficult when you have data licensing.
- Users need more information about how we actually give the info: APIs (what users want).

Discussion

The EC was asked to clarify the difference between use and re-use of data, referred to situations where a particular use can be made of a document but the document itself is not openly available for re-use. The EC pointed to the fact that under the rules of the revised Directive, this dilemma will mostly not apply any more, as all material that is accessible shall automatically be reusable.

The EC asked Mr. Wretham about the consequences of defining information as "core dataset" in the UK. UK perspective: core data is all data that should be open, in line with the government's policies.

Block II: Charging issues(1)

Szymon Lewandowski (EC, CNECT G3 Unit) presented the results of the online survey regarding charging. He underlined that in the new PSI Directive to cover the **marginal cost charging is the rule and not the exception.**

The survey shows:

- Low support for Telecoms, Archiving and Maintenance cost to include in marginal costs.
- Broad support for calculation of costs on database or organization level.
- Acceptable level of reasonable return on investment in the lowest ranges.
- Support towards the publishing of re-use revenues.

Interventions: *(The order of the planned agenda was altered during this block in order to attend the participant's needs)*

Denise Berthault, Alliance Big Data

Key ideas:

- Focus should be not in the already produced data sets (of which approx. 10% is available for re-use) but in the ones that have yet to come out.
- Quality is the key. The maintenance of quality has costs so charging is necessary. To establish **annual fees is the best option.**

Discussion

The EC asked all the participants about the definition of marginal costs. Key ideas from the participants:

- For defining marginal costs it is necessary to define, first, unified cost account systems and rules. (In some countries, apparently, accounting practices leave a lot to be desired).
- Software cost should be included when calculating marginal cost (general perception).
- If Federal Ministries are recognized as exceptions, the whole marginal system is inoperative. Therefore, it is important to interpret exceptions narrowly.

- If Data Collection Cost is calculated, all associated revenues should be taken into account.
- No sense in establishing a "common reasonable proportion of costs" as marginal cost.

Mathias Schindler, Wikimedia Deutschland

Mr. Schindler presented the findings of a survey organized by Wikimedia Deutschland (based on written reply from the German Government on specific questions). Main findings:

- Income produced by PSI re-use is virtually nonexistent (0.0018%)
- Licensing terms are not enforced – no public institution reinforces them.
- (Strong hypothesis) Administrative effort will possibly be larger than the expected income when enforcing licensing terms.
- Conclusion: **Do not charge, it is not worth the trouble. / Do not add conditions to re-use if you do not intend to reinforce them.**

Nikolaus Futter, PSI Alliance

- Mr. Futter stressed the four main needs of the market:
 1. To have access to re-usable raw data.
 2. To avoid competition with PSI holders.
 3. Less intellectual Property Rights connected to public data.
 4. Fair Pricing: marginal cost as key element. However:
 - What markets need is reliability, long-term partnerships and service level agreements. And companies are willing to pay a fair price in order to have them. Therefore, **not only marginal cost but fair charging is an option.**
 - Charging should be accountable and transparent.
 - To establish an effective redress mechanism is vital.

Discussion

Noel Van Herreweghe pointed out to (in his opinion) unclear meaning of recitals 23 and 25. His suggestion for the EC would be to draft clear guidelines on marginal cost and a definition of what costs can be included in the term.

Jan Leerling, Dutch Chamber of Commerce

- Key idea: marginal costs are accountable; we should specify in guidelines what marginal costs are - But the **guidelines cannot set in detail how different member states will comply with their own accounting rules** .
- It is necessary to differentiate between Open Data / PSI and GEO PSI.

The open (Geo)data principles are: web license (geolizenz.org), registration, reliable standards & liability. For this data it is better to have some cost than no reliable data.

It is necessary to establish optimistic legal framework conditions for economic use of GEO PSI (could be up to 80% of all PSI), that will contain:

- *Practical and easy rights for using and re-using Geo data.
- *Pricing taking into account the market.
- *Charging mostly for delivering to user.
- *Fair data protection / privacy (as around 50% can be accounted as personal data).
- *Easy and reliable availability.

Discussion

Debate ensued regarding licensing and personal data. For some attendants licensing is the only way for dealing with personal data, for others excluding such data from the scope of licenses is a better option. Some participants noted that open data and personal data are not incompatible and there could be cases where personal data might be re-used in conformity with data protection regulations.

LUNCH 12:30 – 13:45

Afternoon Session

Block II: Charging for re-use (2)

Phil Archer, W3C

Mr. Archer recalled the "open standards philosophy": based in the promotion of open data as a principle, not only for transparency and economic reasons. He distinguished between giving access to information in general and giving access to a certain data (commercial use).

Key ideas about charging:

- The public task always includes collection of data. This is a sunk cost that is not related to re-use. Thus, **marginal cost shall always be close to 0 in a digital environment.**
- When charging, it is necessary to take into account the countries' culture, political views... so it is necessary to **advance step by step in every individual country.** [Side remark: at present, there is not even a standard for web payments, hopefully to be agreed within 2 years, which shows the difficulties in reaching consensus on such key topics for digital economy]
- Data should be simple, but in a level that allow people to use it further.
- If you have to use private data: anonymize it.

Anne Bergman, Federation of European Publishers

Key ideas:

- **More transparency and accuracy** - Important for publishers.
- **Guidance on what culture institutions would be allowed to charge for PSI** is necessary. When culture institutions are charging, it should be the cost used in digitizing and providing access to data. [In this regard, Mrs. Bergman asks for clarification in the EC guidelines].

Discussion

- **Marginal cost can be defined as the difference between what you produce anyway and what you need to implement in order to give access to re-users, it does not include the initial investment.**
- Debate around the necessity of defining "investment" in order to clarify concepts as "reasonable return of investment".
- Clarification by the EC: As far as pricing is concerned, according to the logic of the new directive, cultural institutions will be subject to the existing regime (2003 PSI Directive).
- The public data is initially intended for public sector (PS) use and increasing its quality is something the own PS bodies should have to pay for. **Common understanding: There should be no marginal cost charges for improving quality.**

BLOCK III: LICENSING

Mr Daniele Rizzi (EC) presented the results of the online survey regarding licensing:

- Even distribution of preferences with regard to the three choices: license, disclaimer or no conditions at all.
- Acknowledgement of attribution and of modification most in line with the Directive.

- Confused feedback with regard to black-listed conditions, although the majority of respondents is against defining black-listed conditions.
- Stakeholders are in favor of existing standard models. (Not only Creative Commons but also others).
- Regarding interoperability, majority choose machine-readability as the key element.

Interventions:

Patrick Peiffer, National Library of Luxembourg

Key ideas:

- **Avoid creating additional national/sectorial licenses.**
- If your data is not covered by copyright, you should treat it as public domain.
- **Interoperability is the maximum goal.**
- Experience from the discussion on licensing of Europeana content (EU's cultural heritage online) could be an example: It led to quick results and even non-EU countries (US, South Korea) and companies have adopted its licensing approach. Europeana played a crucial role by asking the right questions in the beginning of the process. In the case of cultural sector metadata, adopting open licensing under CCO was the best solution.

Zsuzsanna Belenyessy, EDPS

- Regarding privacy - as PS bodies can take three different decisions when dealing with personal data, there are three different scenarios:
- Personal data cannot be published: Sensitive data.
- Personal data can be published. Example: land registration.
- Personal data can be published but as aggregated info in statistics. [If we aggregate data: danger of re-identifying data].
- **Licenses cannot solve this situation of different scenarios, but they can establish a set of conditions and legal obligations – limitations of use.** They can also reinforce the applicable regulation in force.

Discussion

- Regarding public domain: it is possible to use the public domain mark or CCO, CCO is better because it proactively waives rights on data, whereas a public domain mark is more of an information notice.
- It is often necessary to define which licenses can be used for new datasets that are created by combining various 'old' datasets (as it is commonly made with GEO data).
- The EC asked the UK delegate for the reason behind not using Creative Commons but creating an Open Government License instead. UK view: as CC is an American creation it did not cover EU problems (database rights). In addition, UK felt their system was easier to use. They wanted something new and with impact (also in terms of PR). The interoperability was, nevertheless, a core part of the Open Government License in UK.
- General remark from the participants: Wherever you built open government license be extremely careful with interoperability.
- It is necessary to consider technological limitations.
- It is necessary to "make things attractive" and simple licenses are generally more usable. It may be necessary to spell out reasons for the need to mention the original source of material and to make sure the attribution is in the interest of the re-user (Example: cases where attribution makes your own work as re-user more visible).
- Germany's case is presented as "the potential disaster scenario": 9 licenses applicable for spatial data, with only one being open. The risk is that such a situation will defy standards and any interoperability will be impossible.
- Some participants consider that it is so vital to make the PSI data available that, although it is not possible to indicate how individual countries need to proceed, it should be possible for the guidelines to indicate a common objective with a set of clear principles.
- **Interoperability is the key issue for wider re-use within the European market.**
- Many participants are not in favor of licensing at all, and in case licenses have to be imposed, any restrictions should be imposed by the legislation rather than by the PSI holders themselves.

Prodromos Tsiavos, LAPSI

Mr. Tsiavos makes suggestions on PSI licensing interoperability.

- Clarification: Open licensing is a form of re-use licensing, that makes use of standard license and requires minimum transaction costs and maximum interoperability. It is not the same as re-use licensing (that is not by definition, or at least not always, open licensing).

- Most re-users want APIs, that have Terms of Service (ToS) that need to be standardized and this should be differentiated from content licensing. **Conditions applicable to APIs are therefore important too, not only data licensing.**
- The excess of license (**license pollution**) is perceived negatively. It negatively affects interoperability, which increases costs of re-use.
- If you create your Open Government License – INTEROPERABILITY is the key element.
- Alternative to licensing, a simpler regime that would generate less friction based on:
 - * Introduction of provisional "PSI as Public Domain" regime.
 - * Introduction of Smart Notices: that informs about the permanent position / Versioning / Attribution.
 - * Registration for Copyright: re-use / open.
 - * Use standard public licenses. Use CCO, alternatively use CC BY.
- Avoid:
 - * Noncommercial licenses only (definition issues)
 - * Copy left (Problems of compatibility)
 - * No derivatives. (Those are not really open)
 - * No redistribution. (Severely restrict re-use)
- Objective: to create EU Open Licensing Framework Model (Standard). How?
 - *Store licenses at a permanent location.
 - *Publicly announce revisions.
 - *Do not add a personal data license clause (use "Smart Notices" instead).
 - *Solve the Cultural Heritage Rights at the legislative level.
 - *Make sure you give the same thing – Do not contractually limit PD / Exemptions.
 - *Use standard restrictions. Attribution / No sub-licensing / No Endorsement / Indicate this is a modified version.
 - *Be explicit about interoperability.
- **Conclusion: PSI open by default - Avoid licenses - Use standards – make it interoperable.**

(Additional information on licensing provided in the LAPSI report, published on EC websites alongside this report)

Discussion

Key ideas mentioned during the brief discussion:

- Not having a standard license to use often makes it necessary to regulate.
- If it is not in the registry, it should go to Public Domain.
- Difficult to apply "one size fits-all" approach in detail.

Final Comments

This summary should be read in conjunction with the conclusions derived from the "online consultation on guidelines on recommended standard licenses, datasets and charging for the reuse of public sector information", and complemented by the slides presented by individual participants to the public hearing.

The summary record of the Public Sector Information Group meeting, which forms a part of the consultation process and took place on the 26th November 2013, is also available on the EC website.