Best Practice: Encourage crowdsourcing around PSI

25 July 2016

This version
   http://www.w3.org/2013/share-psi/bp/ec-20160725/
Latest version
   http://www.w3.org/2013/share-psi/bp/ec/
Previous version
   http://www.w3.org/2013/share-psi/bp/ec-20160627/

This is one of a set of Best Practices for implementing the (Revised) PSI Directive developed by the Share-PSI 2.0 Thematic Network.

Share-PSI Best Practice: Encourage crowdsourcing around PSI by Share-PSI 2.0 is licensed under a Creative Commons Attribution 4.0 International License.

Outline

Preparing PSI for sharing can be time consuming, expensive and, sometimes difficult. Engaging the community in the task will increase the quality and quantity of available data as well as enthusing the potential users.

Links to the Revised PSI Directive

Policies and Legislation, Platforms

Challenge

To increase the quality and quantity of machine readable data within a constrained budget.

Solution

Crowd sourcing can be an efficient way to increase quality and availability of machine readable data, in particular for cultural heritage institutions. Innovative techniques, including gamification, can be used to harness the skill and enthusiasm of the community at large. On a practical level, datasets can be made avilable on platforms such as GitHub so that users can offer corrections (accepting such corrections remains under the control of the data owner). This is the approach undertaken by the City of Chicago. On a policy level, identifying community crowd sourcing projects outside government institutions can also be an indicator of valuable datasets that should be prioritised for open publication since the level of community involvement is generally proportional to the level of interest in that data.

Why is this a Best Practice?

Many institutions lack resources necessary to manually go through large collections of unstructured data that has been created over many years (e.g. in the cultural heritage sector). By engaging external communities to collaborate on this data it is possible to create more detailed machine readable data supporting a wider range of re-use cases.

More machine readable open data supports a wider range of use-cases in services and applications.

- Many institutions lack resources necessary to manually go through large collections of unstructured data.
- By engaging external communities to collaborate on this data it is possible to create more detailed machine readable data supporting a wider range of reuses.
- Crowdsourcing engages the community that the end product serves.

How do I implement this Best Practice?

- Identify the exact need first and then seek groups able to support solving that need via crowd sourcing.
- Think of crowdsourcing as another tool to create/improve data sets and think about the phases of your data collection project and where crowdsourcing could best fit in.
- Involve stakeholders who could benefit from a free source of certain data sets and have them provide funding in order to sustain crowdsourcing efforts.
- Tasks need to be small to be able to be completed by volunteers with limited time.
- Utilize a gamification approach if possible, that is, by playing a game, users perform a useful task.
- It is possible to use crowdsourcing without the user's knowledge. The best known example of this is the use of CAPTCHAs to solve the micro task of reading words that optical recognition software cannot and by that method digitising hard to read texts.

Where has this best practice been implemented?

Country	Implementation	Contact Point
Sweden	Guiding principles for digital cultural heritage (PDF) Digisam	
Czech Republic	Společně otevíráme data)	Michal Tošovský

References

- Dimitris Paraschakis, Crowdsourcing cultural heritage metadata through social media gaming, 2013, Malmö University
- Krems Workshop Session: Towards A Sustainable Austrian Data Market

Local Guidance

This Best Practice is cited by, or is consistent with, the advice given within the following guides:

- (Austria) Open-Government-Vorgehensmodell Open Government Process Model
- (Belgium) Open Data Handleiding Open Data Handbook
- (CzechRepublic) Standardy publikace a katalogizace otevřených dat veřejné správy ČR Open Data Standards
- (Finland) Avoimen Datan Opas Open Data Guide

- (Germany) [Open Government Data Deutschland](#)
- (Hungary) [Nyílt Adatok kézikönyv](#) Open Data Handbook
- (International) [Open Data Handbook, Solutions Bank](#)
- (Latvia) [Atvērto datu vadlīnijas](#) Open Data Guidelines
- (Lithuania) [Viešojo Sektoriaus Informacijos platinimo gerosios praktikos](#) Best Practices for Sharing Public Sector Information
- (Luxembourg) [Recommandations pour l'ouverture des données publiques](#) Recommendations for opening data
- (Serbia) [Open Data Handbook](#)
- (Slovenia) [Priročnik za odpiranje podatkov javnega sektorja](#) Manual for the opening of public sector information

Contact Info

[Peter Krantz](#) peter@peterkrantz.se

Related Best Practices

[Enable feedback channels for improving the quality of existing government data](#)

Issue Tracker

Any matters arising from this BP, including implementation experience, lessons learnt, places where it has been implemented or guides that cite this BP can be recorded and discussed on the project's [GitHub repository](#)