

Best Practice: Dataset Criteria

25 July 2016

This version

<http://www.w3.org/2013/share-psi/bp/dc-20160725>

Latest version

<http://www.w3.org/2013/share-psi/bp/dc/>

Previous version

<http://www.w3.org/2013/share-psi/bp/dc-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: Dataset Criteria by [Share-PSI 2.0](#) is licensed under a [Creative Commons Attribution 4.0 International License](#).

Outline

This best practice sets out a number of criteria that can be used to prioritise the publication of some datasets ahead of others.

Links to the Revised PSI Directive

[Dataset criteria](#)

Challenge

To develop the criteria for ‘high-value datasets’ taking into consideration the likely re-use of open data and to help governments understand which datasets to prioritise for publication.

Solution

To follow this guidance on dataset criteria which has been developed through engaging with both users and re-users of the data. The characteristics of ‘high-value datasets’ are seen from three perspectives: re-usability, value for data owners, value for re-users.

Reusability

- High-value data should reach at least 3-stars on Tim Berners-Lee's [5 star schema](#) (making it available on the Web under an open licence in a non-proprietary, structured format).

Value for data owner

A dataset may be considered of high-value when one or more of the following criteria are met:

- sharing it contributes to transparency;
- the publication is subject to a legal obligation;

- the data directly or indirectly relates to their public task;
- sharing it helps with cost reduction.

Value for reusers

The value of a dataset primarily depends on its use and reuse potential, which can lead to the generation of business activity. The potential of the dataset is defined by:

- the size and dynamics of the target audience;
- the number of systems or services that could use the dataset.

Datasets contributing to transparency have a strong social impact and re-user's interest in these datasets is high.

Engaging with Reusers

It is important to engage directly with reusers to understand the value of your dataset.

Recommendations:

- establish a communication channel, for example, with a mailing list or a community on [Joinup](#) or on the Open Data Portal that could be used to make announcements to re-users and to gather feedback;
- use collaborative tools. This encourages collaboration between a community or re-users and the cross-fertilisation of ideas and business opportunities.

Why is this a Best Practice?

It's important to have a shared understanding of what can be considered to be high-value datasets so that publication of these datasets can be prioritised.

Why is there a need for this Best Practice?

Understanding which datasets should be published, under what criteria and priority, will help public authorities to see the benefits of publishing more high quality datasets.

How do I implement this Best Practice?

In order to implement this BP you need:

- an understanding of high-value data;
- communication channels with data users and data reusers.

References

- Timisoara Workshop Talk: [Good practices for identifying high value datasets and engaging with reusers: the case of public tendering data](#). Nicolas Loozen.
- Timisoara Session: [How good is good enough?](#) Makx Dekkers
- Krems Workshop Session: [Specific Critical Success Factors for Open Data Publication and Use](#). Anneke Zuiderwijk, Iryna Susha, Yannis Charalabidis, Peter Parycek, Marijn Janssen
- Open Data Institute white paper: [How to prioritise open data to drive global development](#)
- Open Data Institute guidance: [Engaging with reusers](#)

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
- (Finland) [Avoimen Datan Opas](#) Open Data Guide
- (Germany) [Open Government Data Deutschland](#)
- (Greece) [Εφαρμογή των διατάξεων του Κεφαλαίου Α' του ν. 4305/2014 \(ΦΕΚ 237/Α'\)](#)
Guidelines on the implementation of open data policy and l. 4305/2014
- (International) [Open Data Handbook, Solutions Bank](#)
- (Ireland) [Guide for publishers](#)
- (Italy) [Linee Guida Nazionali per la Valorizzazione del Patrimonio Informativo Pubblico](#)
National Development Guidelines for Public Sector Information
- (Latvia) [Atvērto datu vadlīnijas](#) Open Data Guidelines
- (Lithuania) [Viešojų Sektoriaus Informacijos platinimo gerosios praktikos](#) Best Practices for Sharing Public Sector Information
- (Luxembourg) [Recommandations pour l'ouverture des données publiques](#) Recommendations for opening data
- (Malta) [PSI Directive Implementation & Internal Data Sharing Platform \(draft\)](#)
- (Netherlands) [Handleiding bij openen van data](#) Guidance on Open Data
- (Portugal) [Guia Dados Abertos - AMA | Dados.gov](#) Open Data Guide
- (Romania) [Ghid Pentru Publicarea Datelor Deschise](#) Romanian Open Data Guide
- (Serbia) [Open Data Handbook](#)
- (Slovenia) [Priročnik za odpiranje podatkov javnega sektorja](#) Manual for the opening of public sector information
- (Spain) [Guía de aplicación de la Norma Técnica de Interoperabilidad de reutilización de recursos de información](#) Application Guide for Technical Interoperability Standard on PSI re-use
- (Spain) [Government Data Openness and Re-use](#)
- (UK) [Open Data Resource Pack](#)
- (UK) [Birmingham and West Midlands Localised Guide for Open Data](#)

Contact Info

[Nicolas Loozen](#), PwC

Related Best Practices

- [Develop an Open Data Publication Plan](#)
- [Open Up Public Transport Data](#)
- [Establish an Open Data Ecosystem](#)

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](#)