

How good is good enough?

A common language for quality?

Focus

This session will look at the requirements and possible solutions for defining, measuring, expressing and communicating quality of published Public Sector Information. The outcome of the session will be submitted to [W3C's Data on the Web Working Group](#) to inform the development of the [Data Quality vocabulary](#).

Rationale

There is a lot of talk about the need to publish "high-quality" PSI. While it is certainly important that data has sufficient quality to make it useful and usable for users and re-users, we currently lack a common or standard way to express what the quality of data is. The question is whether it is necessary to have such a common way, and, if so, what a "quality vocabulary" could look like.

Issues for discussion

- **Determining perspectives on quality**

Participants are asked to express views from various perspectives, e.g. what a publisher considers important versus what a user/re-user would expect or is willing to accept.

- **Prioritising quality aspects**

Using a catalogue of quality dimensions (e.g. the one [here](#)), participants will discuss whether a ranking of quality aspects is possible, in general or related to specific types of data.

- **Measuring quality**

For a limited number of priority aspects, participants will discuss how the level of quality can be measured and how those levels could be conceptualised in a way that makes sense to (re)users.

- **Standardising terminology**

Participants will discuss how the conceptual levels of quality can be expressed human-readably in disclaimers/proclaimers (e.g. as in [ODI's Open Data Certificates](#)) and which metadata properties and controlled vocabularies could provide a way to make those levels machine-readable.

The Questions

- 1. What X is the thing that should be done to publish or reuse PSI?**

Unambiguously express and communicate quality levels of published data.

- 2. Why does X facilitate the publication or reuse of PSI?**

This will allow re-users to take informed decisions on whether and how to re-use data.

- 3. How can one achieve X and how can you measure or test it?**

Provide a standard set of terms that can be used by publishers. The number of sites that use those terms is an initial test; an increase in re-use of data covered by such standard terms can be measured over the medium term.