



Share-PSI 2.0

Deliverable 7.1

**Initial draft of the evolving
best practices document
demonstrating impact of
preceding workshops**

**Share-PSI 2.0
Standards for Open Data and Public
Sector Information**

Table of Contents

- Table of Contents 2
- Executive Summary 4
- 1 Introduction..... 5
- 2 Progress in DWBP..... 6
 - 2.1 Group Dynamics and Membership..... 7
 - 2.2 Resolutions at TPAC..... 7
- 3 The Samos Use Case..... 9
- 4 The Best Practices Document12
- 5 Conclusion.....13
- Appendix – Data on the Web Best Practices Working Group Charter.....14

Contractual Date of Delivery to the EC:		November 2014	
Actual Date of Delivery to the EC:		14/11/2014	
Editor:		Phil Archer, W3C/ERCIM	
DOCUMENT HISTORY			
Version	Version date	Responsible	Description
0.01	12 November 2014	W3C	Draft
0.02 – 0.03	13 November	W3C	Minor tweaks based on feedback from partners

The information and views set out in this publication are those of the author(s) and do not necessarily reflect the official opinion of the European Communities. Neither the European Union institutions and bodies nor any person acting on their behalf may be held responsible for the use which may be made of the information contained therein.

Executive Summary

At the time of writing, one Share-PSI 2.0 workshop has been held and its report published. The W3C Data on the Web Best Practices Working Group, whose output forms the multiplier for Share-PSI, has not yet published any draft of its best practices document. However, it has published two drafts of an unforeseen and extensive set of Use Cases and Requirements. The next iteration of this document is due to be published imminently and does include a significant contribution from Share-PSI. This is tied in with significant progress made by the working group at its recent face to face meeting, during which its direction finally became clear.

1 Introduction

The primary output of the Share-PSI 2.0 network will be a set of best practices developed by the key players across Europe and beyond. To achieve this, rather than develop its own best practice document, the network is feeding the output of its workshops and discussions to the W3C Data on the Web Best Practices Working Group. That group is constituted separately and runs under the usual W3C process. This document amounts to a report on progress within that group as much as within Share-PSI itself and on the relationship between the two.

In brief, the W3C WG is still working on its Use Cases and Requirements document and work on the resultant Best Practices document is only just getting under way properly. Nevertheless, Share-PSI is positively influencing the working group through six common people involved and twelve requirements derived from the Samos workshop conclusions taken into consideration by DWBP.

2 Progress in DWBP

Like any standards-setting working group, Data on the Web Best Practices is governed by its charter – a statement of what the community has asked it to do¹. The charter explicitly calls for the group to take account of Share-PSI 2.0 (in its Dependencies & Liaisons section) and the full text of the charter is included as an appendix but from a Share-PSI 2.0 point of view the key wording is as follows:

The mission of the Data on the Web Best Practices Working Group is:

- to develop the **open data ecosystem**, facilitating better communication between developers and publishers;
- to provide **guidance to publishers** that will improve consistency in the way data is managed, thus promoting the re-use of data;
- to **foster trust in the data** among developers, whatever technology they choose to use, increasing the potential for genuine innovation.

The guidance will take two forms: a set of best practices that apply to multiple technologies, and vocabularies currently missing but that are needed to support the data ecosystem on the Web.

As well as the Best Practices document, the group is chartered to produce some vocabularies: one to cover quality and granularity of a dataset and another to cover usage of data.

As the working group soon realised, the charter could be interpreted in many ways and the resultant output *could* be enormous. In vernacular terms, the charter asks the WG to boil the ocean. As a consequence, the decision was taken early on to first develop a Use Cases and Requirements Document that would clarify the scope of the work. This was not explicitly foreseen in the charter.

The first version of that document was published in June² with an update in October³. A new version is under preparation⁴ and it is this latest version that includes specific features informed by the Share-PSI 2.0 event in Samos. For example, references to locations should be consistent across multiple data providers at all levels of government. Full details are given in section 3. It is anticipated that the updated version of the use case document will be formally published before the end of the year. The latest version is always available at <http://www.w3.org/TR/dwbp-ucr/>.

¹ <http://www.w3.org/2013/05/odbp-charter>

² <http://www.w3.org/TR/2014/WD-dwbp-ucr-20140605/>

³ <http://www.w3.org/TR/2014/WD-dwbp-ucr-20141014/>

⁴ <http://w3c.github.io/dwbp/usecasesv1.html>

2.1 Group Dynamics and Membership

Aside from W3C/ERCIM, the official list of DWBP WG members⁵ includes a number of individuals or organisations in Share-PSI 2.0:

- Martín Álvarez-Espinar (CTIC)
- Makx Dekkers (in the group as an individual invited expert)
- Deirdre Lee (NUIG)
- Raj Sing (Open Geospatial Consortium)
- Erik Mannens (iMINDS, unfunded Share-PSI partner)

OGC's representation in the WG is in flux at the moment as Raj Singh has left that organisation but that is expected to be resolved soon when he'll be replaced by Ingo Simonis who has recently joined the OGC staff. Other names that are familiar in the European open data world include Antoine Isaac, Christophe Gueret, Riccardo Albertoni, Carlos Iglesias, Raphaël Troncy and Bart van Leeuwen (who presented a paper in Samos).

Unusually for W3C, the Working Group includes a substantial number of members from Brazil and has 4 co-chairs. Two of these, Deirdre Lee and Hadley Beeman, take primary responsibility for driving the work forward, following the W3C process etc. Hadley Beeman does this in a private capacity distinct from her role within the UK Cabinet Office/Government Digital Service. The other chairs, IBM's Steve Adler and Yaso Córdova from Nic.br, take responsibility for engagement and technical leadership respectively. Finally, it's worth noting that Manuel Tomas Carrasco Benitez from the European Commission is also an active participant.

The group includes many individuals new to the standards process and unfamiliar with standards such as DCAT which added to the inertia and lack of clarity of direction the group had in its early stages.

2.2 Resolutions at TPAC

The DWBP WG holds a weekly teleconference and met face to face for the first time at the end of March 2014. A second face to face meeting was held very recently (30 – 31 October) at W3C's big annual meeting known as TPAC (Technical Plenary and Advisory Committee). This two day meeting proved to be extremely useful. The personal dynamic of the members was greatly improved and, most importantly, a number of resolutions were taken that has set the group on a more clearly defined course.

An issue the group struggled with is the dividing line between policy and technology. Where does one end and the other begin? A simple resolution adopted by the group was that the Best Practices should be technical in nature but more clarity was needed on what that means. In the end, three criteria were adopted by which the WG can assess whether a specific requirement is in or out of scope:

1. **Is it unique to publishing data on the Web?** This comes from the fact that this is a World Wide Web Consortium Working Group. The Web itself defines the outer scope such that if a requirement applies in other spheres then it is not in the competence of W3C to make statements about it.
2. **Does it encourage reuse or publication of data on the Web?** This helps the WG to decide whether a Best Practice is specifically within its own area of interest as

⁵ <http://www.w3.org/2000/09/dbwg/details?group=68239&public=1>

opposed to other working groups. It further ensures that the requirements are relevant.

3. **Is it testable?** If it's not possible to test whether a Best Practice has been followed or not then it's requirement is either irrelevant or redundant.

Only requirements meeting these three criteria will be addressed in the Best Practices.

It was in this spirit that a number of requirements that were included in the existing drafts of the use cases document will be removed from future versions.

R-SelectHighValue

Datasets selected for publication should be of high-value, which should be indicated in a quantifiable manner/property.

R-SelectDemand

Datasets selected for publication should be in demand by potential users, which should be indicated in a quantifiable manner/property.

R-CoreRegister

Core registers should be accessible

Are all to be removed as they are too policy-oriented and not technical requirements.

R-SynchronizedData

Dynamic generation of Data on the Web from non-Web data resources and automatic update when original data source is updated

Is to be removed partly because it is too close to one that is being retained (R-Access Up To Date which is about update cycles) but also because it refers to aspects other than the Web.

3 The Samos Use Case

It was the recent face to face meeting that provided the best chance to provide input based on the Samos workshop: *Uses of Open Data Within Government for Innovation and Efficiency*. W3C Working Group meeting time is severely limited and so the group has made it clear that they do not wish to receive presentations and the original plan to present each workshop report to the WG cannot be pursued therefore. However, it is possible to add a use case and to discuss the issues and requirements arising from it and this has been done.

One of the key messages from the Samos workshop was that in order to be effective and sustainable, public sector bodies must have a plan and that plan must have high level support (this reflects earlier work such as the G8 Open data Charter). Whilst true and recorded as an important conclusion in the report, as can be seen above, this aspect is out of scope for the DWBP WG. However Share-PSI's primary objective is unaffected. As stated in the first line of the description of work:

"The objectives of the Share-PSI 2.0 Thematic Network exactly match those of the call, namely to bring together a very broad range of stakeholders in the re-use of public sector information and to help them to reach consensus on *technical standards*, complementing existing and ongoing initiatives in the domain."

(emphasis added). Furthermore, many of the requirements arising from Samos are definitely in scope for DWBP.

The general pattern for use cases in the document is that there is a short block of descriptive text followed by some combination of lists of the challenges posed, positive and negative aspects of the use case and then, crucially, a list of requirements arising from the use case. For the Samos 'use case' the document includes a short description of the event and the Share-PSI 2.0 network and then links to the many requirements that can be derived from it.

The requirements from all the use cases are collected towards the end of the document with hyperlinks back to the use cases that motivated them. In the case of the requirements from the Samos workshop, the links point to specific sections of the report, each of which is itself derived from a paper and presentation delivered at the event. The table below shows the requirements arising from the Samos use case and the sections of the report that each one links to.

Requirement	Requirement expansion	Report section and fragment	Relevant Paper(s)
R-VocabReference	Existing reference vocabularies should be reused where possible	Statistics (#stats)	Valentina Janev, IMP; George Papastefanatos, IMIS; Sarven Capadisli, 270a.info
R-MetadataStandardized	Metadata should be standardized. Through standardization, interoperability is also expected.	Metadata (#minhap)	Dolores Hernandez, MINHAP

Requirement	Requirement expansion	Report section and fragment	Relevant Paper(s)
R-MetadataDocum	Metadata vocabulary, or values if vocabulary is not standardized, should be well-documented	Metadata (#at)	Peter Parycek and Johann Höchtl, DUK
R-ProvAvailable	Data provenance information should be available.	Statistics (#sarven)	Sarven Capadisli, 270a.info
R-AccessRealTime	Where data is produced in real-time, it should be available on the Web in real-time	Transport (#gijon)	Martin Alvarez-Espinar, CTIC
R-Access Up to date	Data should be available in an up-to-date manner and the update cycle made explicit	Transparency and Anti Corruption (#fire)	Bart van Leeuwen, Netage
R-SensitivePrivacy	Data should not infringe a person's right to privacy	Plans and Implementations (#difi)	Heather Broomfield and Steinar Skagemo, Difi
		Transparency and Anti Corruption (#snap)	Amanda Smith, ODI
R-SensitiveSecurity	Data should not infringe an organization's security (local government, national government, business)	Transparency and Anti Corruption (#fire)	Bart van Leeuwen, Netage
R-Citable	It should be possible to cite data on the Web	Transparency and Anti Corruption (#albania)	Julia Hoxha, AIS
R-QualityComparable	Data should be comparable with other datasets	Transparency and Anti Corruption (#fire)	Bart van Leeuwen, Netage

Requirement	Requirement expansion	Report section and fragment	Relevant Paper(s)
R-IncorporateFeedback	It should be possible to incorporate feedback on the data	Impact Studies (#feedback)	Jan Kučera, UEP
		Transparency and Anti Corruption (#feedback2)	Mateja Prešern, and Gašper Žejn, MNZ
		Transport (#france)	Philippe Mussi, Open Data France
R-Location	Locations (countries, regions, cities etc.) must be referred to consistently	Plans and Implementations (#helinkiLocation)	Ville Meloni, FVH

Table 1: The 12 Data on the Web Best Practices requirements derived from the Share-PSI 2.0 workshop in Samos with links to relevant sections in the report. Fragment identifiers can be appended to <http://www.w3.org/2013/share-psi/workshop/samos/report> for reference.

At the time of writing, very soon after the resolutions referred to above have been taken, the requirements in the DWBP document are under close scrutiny. In addition to those derived from Samos, use cases have also been added covering scientific research data which is in scope for DWBP but largely out of scope for Share-PSI. The new use cases and requirements, and the refinement of the existing ones, mean that this work is expected to continue for the foreseeable future. It is notable that the Location requirement was motivated by the most recently added use cases, including the Share-PSI one.

All the remaining Share-PSI workshops are of course expected to motivate new requirements as well as adding to existing ones beyond these 12.

However, the working group is not waiting until the use cases and requirements document is finished before working on the best practices document.

4 The Best Practices Document

Since the Data on the Web Best Practices Working Group was formed there have been numerous attempts to 'make a start on the BP Doc.' Whilst some work has been done, this has largely been piecemeal, is largely scattered across the working group's wiki and is generally not very coordinated. The recent face to face meeting was as helpful in resolving this as it was for the use cases and requirements.

The group spent a good deal of time considering what an individual Best Practice should look like and how the eventual document should be structured. The discussion around what is and isn't in scope was very influential of course and, to elide much of the unnecessary detail, a skeleton document – no more than that at this stage – is now in the WG's GitHub repository⁶.

As discussed, the Best Practices will not include policy issues, such as how to prioritise datasets for publication, but it will include a non-normative description of a data lifecycle that can be used as a structure. Individual Best Practices will have the structure shown below.

Best Practice N

Best Practice Title

Why

Why this is unique to publishing data on the Web

How this encourages publication or reuse of data on the Web

What

Full text description, including diagrams if necessary.

Intended outcome

What the implementer should be aiming for.

Possible Approach to Implementation

A description of one or more means of achieving the intended outcome.

How to Test

Information on how to test the BP has been met. This might or might not be machine testable.

Evidence

Relevant requirements from the use cases document.

It is anticipated that content will have been added to the document and that at First Public Working Draft will be formally published before the end of 2014.

⁶ <http://w3c.github.io/dwbp/bp.html>

5 Conclusion

The description of work indicates that by the end of year 1 the Best Practices document should be 40% complete and include at least 5 examples of standards that have been agreed by all the Share-PSI 2.0 partners. Given the skeletal state of the document this clearly has not been achieved. However, after just one workshop, Share-PSI 2.0 is providing direct, useful and citable input to the working group. The latter is now well positioned to make rapid progress and make progress in catching up with its own timeline, thus multiplying the impact of Share-PSI as planned.

Appendix – Data on the Web Best Practices Working Group Charter

The term 'open data' has evolved rapidly since it came into common usage less than a decade ago. From presidential executive orders to pan-national directives; from cultural heritage to scientific data, it's clear that more and more of the data that historically has been hidden away is now coming into the open. There is widespread consensus that this is a good thing, but there is also a lot of data that is not freely open to all. How can open data be interoperable not just with other open data but with data made available at marginal cost, under less permissive license terms, or with enterprise data?

As experience of publishing and using data on the Web grows there are further issues that need addressing. Developers would like easy access to data that is 100% accurate, regularly updated and guaranteed to be available at all times. Data publishers are likely to take a different view. There are disparities between different developers too: for many, data means CSV files and APIs, for others it means linked data and the two sides are often disparaging of each other.

The **mission** of the Data on the Web Best Practices Working Group, part of the Data Activity, is:

- to develop the **open data ecosystem**, facilitating better communication between developers and publishers;
- to provide **guidance to publishers** that will improve consistency in the way data is managed, thus promoting the re-use of data;
- to foster **trust in the data** among developers, whatever technology they choose to use, increasing the potential for genuine innovation.

The guidance will take two forms: a set of best practices that apply to multiple technologies, and vocabularies currently missing but that are needed to support the data ecosystem on the Web.

Join the Data on the Web Best Practices Working Group.

End date	30 July 2016
Confidentiality	Proceedings are public
Initial chairs	Hadley Beeman (Invited Expert), Yasodara Córdova (NIC.br/W3C Brasil), Steven Adler (IBM)
Initial Team Contacts (FTE%: 20)	Phil Archer
Usual Meeting Schedule	Teleconferences: weekly, Face-to-face: twice annually

Scope

The Data on the Web Best Practices Working Group will be agnostic about the technologies it considers important. Specifically, whilst it will promote linked data using the 5 Stars of Linked Data paradigm, it will also promote best practices for data in other formats such as CSV and JSON.

Government data, scientific research data and cultural heritage data are all explicitly in scope.

Out of Scope

The Data on the Web Best Practices Working Group will not develop any new technologies. Neither will it develop licenses or license terms.

Deliverables

The titles of the deliverables are not final; the Working Group will have to decide on the final titles as well as the structures of the documents. The Working Group may also decide to merge some deliverables into one document or produce several documents that together constitute one of the deliverables.

Data on the Web Best Practices (Recommendation)

This will build on and extend the work done in the Government Linked Data Working Group, taking a domain and technology-agnostic approach to cover aspects such as:

- URI design and management for persistence;
- use of core vocabularies to improve interoperability;
- guidance on the provision of metadata;
- publishing and accessing versions of datasets;
- making controlled vocabularies accessible as URI sets;
- technical factors for consideration when choosing data sets for publication;
- technical factors affecting potential use of open data for innovation, efficiency and commercial exploitation;
- data preservation.

Evidence of implementation will be gathered from national or sector-specific guidelines that reference the Best Practices.

Vocabularies (Working Group Notes)

The working group will develop 2 new vocabularies to support the data ecosystem:

- **Quality and Granularity Description Vocabulary.** This is foreseen as an extension to DCAT to cover the quality of the data, how frequently it is updated, whether it accepts user corrections, persistence commitments etc. When used by publishers, this vocabulary will foster trust in the data amongst developers.
- **Data Usage Description Vocabulary.** This will describe the use made of one or more data sets. Where data is used in an application, it will facilitate a description of what the application does and what problem it helps to solve. This will improve discoverability of the application. Where data is used in other contexts, such as in research, it will facilitate provision of information about what data was used and how it was used during the research. This information can link to and be cited within

published papers. In these scenarios, and others, use of the Data Usage Description Vocabulary will encourage the continued publication of the data on which the usage depends.

Subject to its capacity, the working group *may* choose to develop additional relevant vocabularies in response to community demand.

Inputs

The Working Group will pay attention to independent efforts to develop standards and methods relevant to open data and may choose to refer to such efforts or incorporate them directly in its work. An indicative list of inputs is:

- Best Practices for Publishing Linked Data, B. Hyland, B. Villazón-Terrazas, G. Ateamezing
- DCAT, F. Maali
- Cool URIs for the Semantic Web
- Designing URI Sets for the UK Public Sector (PDF) and its forthcoming update.
- Study on Persistent URIs with identification of best practices and recommendations on the topic for the Member States and the European Commission.
- Kerstin Forsberg's notes on URI design.
- Open Data Institute training materials
- ODI Open Data Certificate
- Open Quality Standards' 72 Open Data Best Practices
- Open Knowledge Foundation standards work
- Hydra Community Group
- The WG should of course take note of relevant projects and initiatives that represent bodies of expertise and experience. For example, the European Commission's ISA Programme, and projects such as ENGAGE, PRELIDA, LAPSI, HOMER, EUCLID, CitySDK, RECODE etc. The forthcoming Share-PSI 2.0 Thematic Network is of particular relevance.
- The Declaration of Data Citation Principles.

Success Criteria

To advance to Proposed Recommendation, evidence will be adduced that each of the best practices have been recommended in at least two environments, such as data portals and formal policies.

The vocabularies that will eventually be developed by the working group will be published via the W3C Vocabulary Management Process that offers stability of terms but the flexibility for future evolution in response to community demand.

Milestones (Recommendation)

Milestones

Deliverable	FPWD	LC	CR	PR	Rec
Best Practices	May 2014	July 2015	November 2015	April 2016	June 2016

Milestones Vocabularies

Deliverable	FPWD	WG Note
Data Usage Description	April 2014	January 2015
Quality and Granularity	April 2014	January 2015

Note: The group will document significant changes from this initial schedule on the group home page.

Timeline View Summary

- December 2013: First teleconference
- February 2014: First face-to-face meeting
- March 2014: Vocabularies FPWD
- May 2014: Best Practices FPWD
- January 2015: WG Notes for the vocabularies
- November 2015: Best Practices document to Candidate Recommendation
- April 2016: Best Practices document to Proposed Recommendation
- June 2016: Best Practices document to Recommendation

Dependencies and Liaisons

- CSV on the Web Working Group
- Internationalization Activity
- Privacy Interest Group
- Data Activity Coordination Group

Furthermore, the Data on the Web Best Practices Working Group expects to follow these W3C Recommendations:

- QA Framework: Specification Guidelines.
- Character Model for the World Wide Web 1.0: Fundamentals
- Architecture of the World Wide Web, Volume I

Participation

To be successful, the Data on the Web Best Practices Working Group is expected to have 20 or more active participants for its duration. To get the most out of this work, participants should expect to devote several hours a week; for budgeting purposes, we recommend at least half a day a week. For chairs and document editors the commitment will be higher, say, 1-2 days a week. Participants who follow the work less closely should be aware that if they miss decisions through inattention further discussion of those issues may be ruled out of order. However, most participants follow some areas of discussion more closely than others, and the time needed to stay in good standing therefore varies from week to week. The Working Group will also allocate the necessary resources for building Test Suites for each specification.

Communication

This group primarily conducts its work on the public mailing list. Administrative tasks may be conducted in Member-only communications. Comments on the group's work will be welcome via the public comment list.

Information about the group (deliverables, participants, face-to-face meetings, teleconferences, etc.) is available from the Data on the Web Best Practices Working Group home page.

Decision Policy

As explained in the Process Document (section 3.3), this group will seek to make decisions when there is consensus. When the Chair puts a question and observes dissent, after due consideration of different opinions, the Chair should record a decision (possibly after a formal vote) and any objections, and move on.

A formal vote should allow for remote asynchronous participation—using, for example, email and/or web-based survey techniques. Any resolution taken in a face-to-face meeting or teleconference is to be considered provisional until 5 working days after the publication of the resolution in draft minutes sent to the group's mailing list.

This charter is written in accordance with Section 3.4, Votes of the W3C Process Document and includes no voting procedures beyond what the Process Document requires.

Patent Policy

This Working Group operates under the W3C Patent Policy (5 February 2004 Version). To promote the widest adoption of Web standards, W3C seeks to issue Recommendations that can be implemented, according to this policy, on a Royalty-Free basis.

For more information about disclosure obligations for this group, please see the W3C Patent Policy Implementation.

About this Charter

This charter for the Data on the Web Best Practices Working Group has been created according to section 6.2 of the Process Document. In the event of a conflict between this document or the provisions of any charter and the W3C Process, the W3C Process shall take precedence.

Phil Archer, Ivan Herman

Copyright © 2013 W3C ® (MIT, ERCIM, Keio, Beihang), All Rights Reserved.

\$Date: 2014-01-09 19:47:41 \$