

Open Standards Board Meeting Minutes

Date: 9 December 2019 14:00 - 16:00

Location: Room Room 611, The White Chapel Building, 10 Whitechapel High Street

Attendees

Board Members	Non Members
John Strudwick, GDS Adrian Hepworth, Atos Chris Francis (SAP) Daniel Appelquist (Samsung) Dave Cridland (Forward Health) Matthew Dovey (Jisc)	Beata Lisowska (ONS) Rosalie Marshall (GDS) Mark Barrington (GDS) Rhiannon Lawson (GDS) James O'Neill (GDS) Lawrence Greenwood (GDS) David Reed (MOJ)

Apologies.

Simon Kerridge (Uni of Kent)
John Sheridan (TNA)
Kathy Farndon (Carbonel Consulting)
Zaheda Borat (Amazon)

Main Outcomes

The Board decided that

- CSV on the Web be accepted as a recommended standard, with condition that the metadata vocabulary mentioned in proposal is refined for the syntax of tables only.
- Schema.org accepted as a recommended standard, with conditions that profile is limited to recommending the dataset schema for this challenge and should be decoupled from tabular datasets to include data.
- Dublin Core accepted as a recommended standard for tabular data with the condition the Board see the final profile and recommendations for comment before publication.

1. Introductions

John Strudwick welcomed everyone to the meeting

John asked everyone to introduce themselves and gave apologies for those not attending.

John explained that we don't have any new members and that Edafe Onerhime has decided to leave the Board.

The board was asked of any conflicts of interests in the standards proposed in the Schemas for Tabular Data Challenge. None were given.

2. Update on the last meeting

John updated on the property and street reference numbers (UPRN/USRN). They are still subject to the new Public Sector Geospatial Agreement (PSGA) being signed.

There are no other outstanding issues.

3. Schemas for Tabular Data Challenge

[link to challenge](#)

John Introduced the proposal Presented by challenge owners Rosalie Marshall Beata Lisowska

Rosalie gave a history of the challenge and how the three standards came to be proposed. There are three proposals for three use cases.

The timeline of these proposals -

This started when JS wanted GDS to look at how to make data sharing in government easier. Having hired data architects to research data sharing in government. They first focussed on private data sharing between individuals internally through spreadsheets. They looked at how to make sure that the data is of high quality and can be found, catalogued and analysed more easily. They built a prototype to describe this tabular data.

The next phase was to research tabular data that was published.

The issues found around open data on data.gov.uk and gov.uk was that the users couldn't find the data and didn't know that it had been published.

They came up with a data description language to make all of these easier. These description languages were taken from:

- Dublin Core tags for private sharing
- Schema.org tag for describing published data
- CSV on the Web for describing the contents of tabular data

A series of workshops were held on this descriptive language and meta language. We spoke to a number of departments with a varying level of maturity in terms of data sharing to understand their needs.

There were 3 main issues: found

Private data and cataloguing it

Private data and integrating it

Publish data and how it is catalogued and could be found

If the three proposals are approved guidance relating to each of these standards will be written and published.

After the original three proposals were published on GitHub there was a lot of useful input

from the community, including some raising the issue of GEMINI and INSPIRE. In response we refined our proposal to reflect that if organisations are already aligning data with GEMINI to comply with INSPIRE they should continue to do so rather than follow our recommendation.

The revised drafts were shared with the Board members for their consideration at this meeting.

Beatta explained how the Office for national statistics (ONS) have been involved initially through cross-government conferences looking for solutions to common problems in the community of practice. The biggest problem was there is a large difference between departments in the level of information being published with the data. The departments that were not advanced did not use any metadata and need help to publishing data.

A minimum set of metadata was the chosen solution that can be used as a starting point for departments that need help. This is where the idea to use Dublin Core (DC) came about. This is a well established metadata standard that is the foundation of other standards including DCAT.

The idea is that you start with DC if you have nothing else then move on through the maturity framework. For example, in ONS we have our own models created with statistical needs in mind but based on DCAT and DC. The idea is to have DC as the base metadata standard.

Schema.org was selected for metadata when publishing data into the public domain. Many departments are already using this schema when they publish on data.gov.uk. schema.org included the 15 core elements from DC.

CSV on the web (CSVW) is a structural metadata standard for publishing data. There are other competing standards but in ONS we decided to use CSVW and have been developing best practices around using the standard. The beauty of CSVW is that allows a structural metadata file to be attached to a data file which among other things eases the scraping of data. MOJ are also adopting CSVW having first started on their own schema and arrived at something very close to CSVW. Building a community around a standard will help adoption and it is very important for helping departments struggling with data.

The challenge owners handed over to the Board for questions

Matthew Wanted clarification on Inspire and Gemini issue brought up on GitHub. They apply to spatial data. We are talking about tabular data across different domains not just spatial

Beatta confirmed that this is the case and that this is more about “what good looks like” for data in general and building on that. It is not just geospatial which is well developed.

Matthew queried that the discussion was limited to a standard for any tabular data.

Rosalie confirmed that if you are following Gemini you don't need to apply DC to your data set too. This is in our revised proposal that has been shared with you.

Matthew - there are two aspects here - the ability to describe a structure (CSVW) and the other is the vocabulary for the tag library. He wanted to distinguish between the syntax versus semantics.

Beatta confirmed there was a distinction between the two in the proposal.

John confirmed that we have three standards which we have to work together. He also asked if these standards would be "glued" together or are they going to be published individually. How will we publish them and package them on gov.uk?

Chris said he was concerned about the difference between data that was produced by an automated system and ad hoc data. There is also nothing about the licensing or legal structure of the data.

Rosalie - process and ad hoc. What we are doing is to appeal to the owners of the data. This is about applying a minimum set of metadata. At the moment we don't have anything, we don't catalogue data in any way and therefore it is not analysed and cannot be integrated. This won't solve everything but will solve the three issues we have identified.

How we will publish it on gov.uk is as three separate asks. The right terminology needs to be thought up.

Chris - it only works with curated tabular data. This is the limit of the scope. It is a recommendation for a minimum set of metadata so it is neither complete or sufficient necessarily

Beatta - No, depending on the different use cases you may need a different set of tag but this is about the first step. DCMS had the same issue with a data inventory and asked what is the minimum that we need to describe data. From my point of view if I have that minimum core metadata with the data I'm given I can build on it.

Matthew are we stepping on the toes of the data board

John - there isn't a data standards board at the moment.

Rosalie confirmed that DCMS is aware of this work.

Dan - citing Postels law, questioned if proposal this is about minimum data set or minimum metadata set to be useful for parties to interoperate.

Rosalie confirmed

Dan - stated he was familiar with and supported CSVW

John returning to Mathew's point; confirmed that there will be data standards authority at some point. At the moment data standards are only set through the tech code of practice and the Open Standards Board to set a minimum set of data standards.

Matthew acknowledged there were two conversations regarding syntax with CSVW and metadata applied to a dataset.

Chris questioned if the Board could confirm a minimum sufficient standard for metadata that would meet all use cases. There may be good reasons why sometimes this minimum would not be reached.

John agreed and asked how this would be dealt with.

Rosalie explained that this would be made clear in the guidance. Guidance will be needed to be linked to these standards and will be important for clarifying some of the issues which the board have raised.

Chris queried how we can explain better which bit are requirements and recommendations. He said it sounds like they are all recommendations.

Rosalie suggested that it is getting confused as we have Three standards for Three different types of data being exchanged. The standards say use tags and that the guidance will help them narrow down the tags people use. There are hours of discussions from the workshops to be incorporated into the guidance.

Chris suggested that the terminology needs to be carefully considered. for example there is always a publisher even for internally shared data. He also wanted clarification on which parts of the standard are a requirement and which were a recommendation.

Adam - need a clearer understanding and an explicit statement how you would pick the different schemas. The proposals have answered that they are open standards but they don't explain how you would pick which standards you use under which conditions.

Rosalie confirmed that these are three different standards for separate circumstances and this is a recommendation to use the DC tags and not another standard. Which tags used will be explained in the guidance.

Adrian requested confirmation that it will be stated where and when to use each standard and what is the order of precedence.

Beatta commented much of this will be in the guidance

John asked that as a board we must look at them separately to accept, accept with conditions or reject these standards to be published on GOV.UK as profiles.
I can almost be certain that these would be recommended not mandated

Matthew stated that he thinks we need more information to be sure that we are only talking about tabular data. I think we can accept these standards we a little more explanation.

Dan - questioned if it would be best to accept the standards given all the work put into them with some conditions, that they were completed.

Dave - commented these are easily separate decisions - how do we transmit the scheme and two is what schema vocabulary do we use? CSVW seems uncontroversial and the Board is happy with it. Decisions on recommending the entirety of the extensive scheme.org is rather a different decision. If we are saying scheme.org is an acceptable source for schemas, which particular schema should we mandate.

Chris - thought that all these standards meet our definition of open. CSVW is well defined the others are more complex due to the use cases.

John brought the discussion to a final decision.

CSVW accepted as a recommended standard, with condition that the metadata vocabulary mentioned in the proposal is refined for the syntax of tables only.

Schema.org accepted as a recommended standard, with conditions that profile is limited to recommending the dataset schema for this challenge and should be decoupled from tabular datasets to include data.

DC accepted as a recommended standard for tabular data with the condition the Board see the final profile and recommendations for comment before publication.

5. DDAT functional leaders and governance

Presentation from Rhiannon Lawson about FLG.

Rhiannon explained that part of the governance of the Open Standards Board falls under the DDAT functional leaders group.

Chris asked if we report in and what is the Boards relationship to the FLG.

John confirmed we are not being stirred by them but we report into the group. Open standards are independent.

6. Any other business

Communication to the board

John informed the board that following the permanent role that Terence Eden has secured in NHSX, James O'Neil (Senior Technology Policy Advisor - GDS Tech Policy Team) will be taking on the role of GDS Head of Open Standards and thanked Ravinder Singh for his support over the last 12 months).

John asked how best keep in touch with the board between meetings and shared a few ideas about how we could improve communications.

The board agreed that a monthly update including a summary of GitHub activity similar to a

GDS weeknote would be helpful.

John advised that the open standards team would invite members to subscribe to the GDS technology blog.

John informed the Board that there would be a review the Terms of reference at the two yearly anniversary (due early 2020) and members would be asked if they wish to continue with their membership, and advertise for new members.