**DCMS Open Data Forum.**

**Monday 14th April**

Attendees:

Bill Thompson (BBC)

Neil Wilson (BL)

Daniel Evans (Science Museum)

Alex Pilcher (Tate)

Richard Keyte (DCMS)

Tom Knight (DCMS)

Ben Horan (DCMS)

Laura Clayton (EH)

Rob Scott (BFI)

Katelyn Rogers (OKF)

Steve Benton (Wikimedia)

Annette Ure (NHM)

Vince Smith (NHM)

John Jackson (NHM)

Jane Audas (IWM)

Andrew Turton (GC)

Max Beverton (OfCom)

**Introduction**

Tom provided a re-cap of the first event, which went well and was generally useful for people. It was agreed then to focus in April on economic and social business case for open data. DCMS has interest in and understanding of intrinsic value of culture and does a lot of work around it, however there is need for a better understanding and explanation of the benefits and business case for Open Data. DCMS are keen to encourage collaboration across the group and to know whether what we are doing is helping – and help in plotting the way forward

**Richard Keyte** provided an update from DCMS more generally:

Still very interested in helping to explore opening and sharing data, and to identify common difficulties, and see if the group or Department can help overcome them – identify common challenges and effective/tried solutions. If anyone has interest in DCMS data, or other contacts the department could provide (for example with other departments of their ALBS) please let us know and we will do our best to help.

In terms of increasing buy-in we are keen to understand how DCMS can leverage position, either in the area or with individual organisations. We appreciate ALBs all face different challenges – with some similarities and differences in opportunities and issues. Also, if people think smaller groups would be of more use DCMS would be happy to facilitate; for example could look at instigating a regulators group with contacts from other Departments if that would be of use.

**Bill Thompson – Head of participation development in archive group at BBC**

*BBC Memorandum of Understanding*

Bill gave an introduction to the current situation at the BBC and the background to the signing of the MoU. There is a general understanding that openness is a good thing, and orientation towards open data, but no policy to support this in dealing with colleagues or informing thinking/decision making, thus this approach to openness is not reflected in policy/practice or funding. The MoU was an attempt to clarify and lock down the BBC’s position, and act as a starting point from which to explore and share thinking, and also act as leverage in future discussions with colleagues. Can point to ‘agreement’ to show that we are moving that way, have support and partners and it has approval.

The MoU covers:

* Open Data Institute
* Open Knowledge (formerly OKF)
* Mozilla Foundation
* Europeana

Open Data is important to ‘discoverability’ – making it indexed, visible and available to people who might be interested. Releasing under an Open Data licence – don’t have rights, but catalogued entry is important as a means to start conversation about availability or making it available – it is very difficult at present to find what they have.

The MoU was signed last year, as documented on the wiki page: <https://wikimedia.org.uk/wiki/Potential_memorandum_of_understanding_and_collaboration_with_BBC>

This meant the public/community could have say and comment – this input was valued. Crucially it meant that colleagues had seen press coverage – couldn’t then ignore emails or approaches from other parties, and also had searchable external reference via the Wikimedia page.

The MoU sets out a public statement covering their commitment to open data, web and internet, and also won the support of James Purnell (BBC's Director of Strategy and Digital) which gave added impact in internal discussions and highlighted top level support. Example of work starting - Speakerthon – voice samples to R4 archive (sample = data). 400 samples added - <http://www.bbc.co.uk/rd/blog/2013/11/speakerthon-uploading-voice-samples-from-the-radio-4-archive-to-wikipedia>

Bill is actively looking for projects to work with at the moment - if anyone is interested please contact him (bill.thompson@bbc.co.uk ).

**Natural History Museum** – looking at European common position, with declaration around big six in Natural History signing up – this has been productive so far. There is no binding commitment, but outlined aspirations, and encouraged benefits helping to work towards a ‘critical mass’ of enthusiasm. *Could DCMS do similar to set out a common direction/basic best practice?* It would help to give departmental approval to general approach to openness in ALBS.

**DCMS** will consider issuing a statement, perhaps to ‘expect that where possible that ALBs will move toward greater openness’ or similar, based on the examples above. Will be sure to get steer from ALBs but at least have stuff on table, and key principles.

**Daniel Evans - Science Museum**

Daniel gave an overview of the Science Museum’s background in the area. Helping to get other people get visibility and awareness of content and engagement – openness helps. Approach to deriving value from content –*value = quality vs reach*

Letting other people do things with data – can be a mental step but is very important. At present some of their data does make significant money – proceeds from the picture library roughly equate to the budget of the web team. However overcoming internal reluctance can be supported by highlighting successes – helps to foster a general attitude towards openness. Gave the example of the *Climate Change game* (2011) – users could rip it back to their own sites. It was initially seeded on 12 websites – before long 1000 sites had it, with 6 million hits. Multiplied value and impact of what they were doing and massively broadened range.

Second example - picture blogged by the Science Museum– 2500 hits. Ultimately was used by the Huffington post – meant that they hit their five year target for web traffic in two and a half years – and this all started with one post. This success, and the way it has been shared to colleagues has helped build a very good reputation within the organisation.

Formed presumption of openness – very light and non-binding but effective. Releasing with non-commercial CC license - putting it out there has enabled them to show it is making not losing money. Discussed which organisations made money from their picture libraries – there was a split– often they cover costs but little else. Those that do make money tend to make most of the profit from a small number of (corporate) customers.

**DCMS** currently looking at image tracking project – would be interested in discussing further with anyone interested. Want to look at demonstrating reach as value for organisations.

**British Library** – trying to play this situation both ways. Combination of images online sales, and 1 mil images on Flickr freely available. Has broad reach – images have been used on a huge range of things, even on custom skateboards! This approach is nuanced – different releases (variety of sizes and formats) – some have been monetised and some not, depending on uses and values, achieving ‘best of both worlds’. Paper written some years ago Jonathan Drury – encouraging digital access to culture, would be worth looking at again.

**Richard Keyte** – knowing info. is there in first place is important as mentioned by Bill and David – the National Information Infrastructure provides a useful way of making data more visible, both published and unpublished.

**Comments:**

**Annette Ure**: Value internally – being joined up can improve ways of working within organisations

Linked data way for internal users - this can also be really useful get own house in order in terms of data cleansing etc. before wider release.

Then persuade people – two ways

* More integrated online experience for collections….
* Synchronised content intending to different apps

Want to look at ingesting other people’s data, and making most of what is available to enrich own data. Then encourages moving forward with own data. Archive hub example (<http://archiveshub.ac.uk/> ): correspondence between two individuals shared between two archives, leading to difficulties in accessing and using – presenting clear benefit to users.

**British Library**

Commercial side of metadata - currently sell and very keen to open. Keen to encourage use of CC licence but also not to disrupt commercial relationships. BL tactics – embargo data for a period and then release openly after period. They have proprietary format – if released in bulk – big issue. Other stuff is released as a staged release, in a nuanced way – works well and provides distinction between two - bulk data in one format, selectively available for those that want it.

2 year embargo on data works well for them. Doesn’t have to be immediate (clearly has more value, hence cost), but still useful to have legacy available - has inherent value, and certainly better than nothing.

**Katelyn Rogers - OPEN GLAM Working Group**

Katelyn presented on the successes of the Open Glam working group (full presentation here: )

Open Glam is starting to build a real community around the working group. The general perception within the field that had previously been promoted: ‘build it and they will come’. In reality this is not the case and never has been, and while there has been lots of discussion about benefits, and releasing social and commercial value, ultimately to get results we need people doing and working with it. Need people building apps, using data and exploring new directions.

Many people within organisations are getting tired – where are the benefits that we have been led to expect? Oversold – good stuff happening but won’t necessarily save work or create huge visible benefits.

OpenGLAM is bringing together advocates and people who want to work with organisations and reuse data themselves. This public mission - makes huge sense: public accessible hugely important.

**Institute of Sound and Vision in Netherlands** - Open images project (<http://openbeelden.nl/.en> )

Opened 0.015% of their collection, resulting in reuse in over 1,600 Wikipedia Articles, generating 40,000,000 pages views. API received 169,000 requests for reuse in a number of applications.

**Wikimedian in residence scheme**

Wanted to make group aware of availability of Wikimedians in residence, and possible support they could provide.

What can community do?

* Curate data
* Enrich and improve
* Provide content

Katelyn ran through several examples of open data in action and the resulting benefits:

**HistoryPin**

Allows users to submit content to a timeline and map - <http://www.historypin.com/> . Follow timeline – enables organisations enrich own collection with crowd sourced data. Collaborating with over 200 cultural institutions worldwide. Lets users tell stories about their history

**Brooklyn museum**

Tag, you’re it - <http://www.brooklynmuseum.org/opencollection/tag_game/start.php> . Generated 70,000 tags in 10 months. 7,385 of these tags were challenged by ‘freeze tags’, demonstrating the capacity for peer review in projects of this type.

**Rijksmuseum**

Make your own masterpiece competition –inspiring creations based on museum holdings’. Winner – designs for make-up boxes to be sold in the gift shop <https://www.rijksmuseum.nl/en/rijksstudio-award>

**RLUK (Research Libraries UK)**

Spoke to OKF, who were able to give advice and content for their first hackathon which worked very well and got people excited internally and externally - <http://openglam.org/2014/04/16/rluk-hack-day-14-may-2014/>

Also worthy of note: Coding da Vinci Hackathon - <http://blog.okfn.org/2014/04/03/coding-da-vinci-open-glam-challenge-in-germany/>

**The Public Domain Review**

Curates stuff from public domain. Generates 20-30 x more hits than if elsewhere – opportunity to get people involved with public domain through established channels – get in touch if you have an unusual collection you would like to feature.

Engaging with public is key, to then share more widely. Before releasing - find out what the audience want; go to events, engage with groups – generate interest. Effective community building, with the right people is hard, but crucial to engagement. As a rule we don’t communicate that well outside of the community but it is an area we can easily improve on and share experience in.

Understanding the magnitude of content and significance is also an important step. Content governs who will be using it, and who it will interest. We need engagement, as well as proactive disclosure / and work to take this forward.

**Comments:**

The Science Museum’s **Zooniverse project** is working on data, and asking for key specific tasks to be undertaken. This is a good example of a practical/clever interface, and manages interaction very well. From there some people will look at further involvement or invent projects.

Help to build a ‘pyramid of engagement’

* Help with small tasks– 20 minutes
* Help with larger task - 2 hours
* Help with wider projects on specific areas
* Then move on to generating own ideas/projects

Ensure that pilot projects work and have deliverables. They might not accurately replicate bigger release, but will still help build community, internal experience and acceptance/familiarity.

**NHM** – raised the transcription example as a potentially tricky but also very useful direction.

Each project different – have to build with specific requirements in mind.

**Jane Audas** – provided an IWM example. Moderation proved difficult and heavy going, especially concerning handwriting in documents. Looks like fun until…

Engagement key, plus getting scope accurately understood.

**UCL** have also been running an open transcribing project - <http://blogs.ucl.ac.uk/transcribe-bentham/>. It is clearly a very interesting concept but has suffered issues with engagement. (Incidentally it is useful to define ‘engagement’ at different steps, and understand how much a user can reasonably and usefully provide. Initial cost/benefit then makes more sense.)

Risk of overselling benefits though – will impact later engagement internally if people do not see steady benefits.

Experience has shown it is very important to have ‘safe’ community to begin working with, and get some early success stories to tell. Rather than having huge community – not going to be practical at present and without considerable work to manage and support.

Spectrum of data represented by DCMS ALBs – Gambling Commission and Ofcom – ultimately want to get out. Need to be careful around release and scope of data though as some premises/operating data may be sensitive. Merging with Lottery Commission – sales of different games can be very interesting balancing sensitivities.

**Wikimedian in residence (comments)**

Other ALBs already have wiki assistance.

**BBC** experience of them adding real value – having a data literate individual on the team can be a big asset and help unlock capacity internally. Making sense of data models – having a data architect in residence can not only add beautification to what you have but make maximum use of data, and again supports the internal case by showcasing what can be done, and that there is external interest.

Also helps to raise political agenda internally. Helped to shake things up, get museum to think more generally to look at openness - raising profile/political advocacy – really useful. Can work with open/meta data, and typically stay with an organisation for 3/6 months. Depending on their specific skillset they can:

* Help explain benefits and encourage releases of data sets
* Build internal and external communities
* Look at work organisation does and linking it to external community

However there is no specific pattern or skillset; it will vary by organisation and individual. Wikimedia can circulate case studies to anyone who is interested- <http://outreach.wikimedia.org/wiki/Wikipedian_in_Residence>

**Wikimania**

Wikimania – global conference annual. This year lucky to have it at the London Barbican, and also that it has a broader focus than purely wiki related topics, looking at wider open data/content/licensing. It runs from 9th-11th August, should be really useful, and will include specific focus on open data, providing opportunities to see and meet users and advocates. Will be good to have broad involvement - DCMS will try to attend to demonstrate support as well.

Main site: <http://wikimania2014.wikimedia.org/wiki/Main_Page>

The main event is also preceded by an Open Data hack – July 5th/6th which again people are welcome at - <http://wikimania2014.wikimedia.org/wiki/Fringe/Open_Data_Hack>

**Suggestions for any exercises around finding consumers of data?**

Different levels depending on nature of datasets. Best to look at use cases for data linked to specific research output e.g. digitalisation activity – led by demands, and on case by case basis. Best to research use case for specific areas, and then look at level of granularity etc. (not expecting extra effort).

Open GLAM – really interested in any support or interaction they could offer – come along and ask for help. OpenGLAM – keen to get UK involvement from a range of areas. If you have any questions regarding OpenGLAM please contact Katelyn (katelyn.rogers@okfn.org )

**What can DCMS do better?**

The group was asked for comments on what DCMS could do to better support members. The following were suggested:

* Be vocal about events and work – do help ALBs to work under the banner of DCMS
* Start to help identify the demand

DCMS is currently setting up a public page on data.gov.uk which should be live before the next session. It will include content and minutes etc. from Forum meetings, any contributions from participants, share news, point from which to answer queries or requests. Also be useful to help showcase work under the DCMS umbrella – useful internally to show support and engagement.

DCMS will also look at coming up with a draft MoU/Memo regarding our suggest approach to Open Data for ALBS

* Progressing cases – get case studies written up and distributed
* Investigate exploring policy data
* Building communities- next steps

Also happy to encourage collaborations with European colleagues – GLAM

How DCMS are supporting – what we could do more of - please let us know if you think of anything.

**DCMS request** – if releasing data or tools – please link to www.data gov.uk – we can provide support/training on the system but it is fairly simple/intuitive, and will help flag that it is available, what organisations are releasing, and make it more publically visible. It is a useful tool for showing what data is held, attracting interest and also demonstrating DCMS support

**Ofcom – advice for releases**

Ofcom – interested in releasing but had question regarding long term or short term releases - International broadband tool – currently used once /add onto press releases.

**Steve Benton** – be surprised who’s interested: for example Broadband speeds – data around this would generate interest from campaigners/local and national businesses/housing developers etc. as well as political or technical organisations or groups looking to monitor progress. People may be interested in very different ways/reasons

Would advocate prioritising releasing data over building tools – can look to find people once data is released but more difficult the other way around. Also helps if you have ideas to propose to communities – let them know how to engage as a starting point. That will then generate things they can do.

**Ideas for July/August meeting:**

Ofcom, BBC, British Library and Science Museum all volunteered facilities for future events

* MoU/Statement for DCMS ALBS
* Further specific case studies for discussion
* Building communities – specific guidance and examples
* Make Open Data business as usual
* Useable metrics that people can agree too (including KPIs etc)
* Usability and spread of open data
* Wikimania – lead-in about contributions and planned content, or feedback following main event
* OfCom – possible content on upcoming releases