

Challenges and perspectives on the use of open cultural heritage data across four different user types: Researchers, students, app developers and hackers

Laursen, Ditte¹[0000-0002-07386562], Roued-Cunliffe, Henriette²[1111-2222-3333-4444] and Svenningsen, Stig¹[0000-0001-7949-0740]

¹ Head of Department, Royal Danish Library

² University of Copenhagen

dila@kb.dk

Abstract. In this paper, we analyse and discuss from a user perspective and from an organisational perspective the challenges and perspectives of the use of open cultural heritage data. We base our study on empirical evidence gathered through four cases where we have interacted with four different user groups: 1) researchers, 2) students, 3) app developers and 4) hackers. Our own role in these cases was to engage with these users as teachers, organizers and/or data providers. The cultural heritage data we provided were accessible as curated data sets or through API's. Our findings show that successful use of open heritage data is highly dependent on organisations' ability to calibrate and curate the data differently according to contexts and settings. More specifically, we show what different needs and motivations different user types have for using open cultural heritage data, and we discuss how this can be met by teachers, organizers and data providers.

1 Introduction

Open data in the heritage sector is a combination of the idea of openness in heritage (such as public engagement and access) and the practicality of heritage data (in other words the increasing amount of digitised heritage material as well as the dissemination of this). GLAM institutions (Galleries, Libraries, Archives, and Museums) throughout the Nordic countries (as well as in many parts of the world) have made a considerable investment in digitising the documents and photographs held in their physical collections and then subsequently providing access to these digital materials to a global audience [1]. This online access to digital heritage material has meant a massive step towards more openness in heritage across the world. However, the extent to which these datasets are open for reuse on platforms and circumstances other than those provided by the institution is limited by organisational policies as well as technological constraints. In addition, little is known about how the data are utilised by different users.

One of the biggest challenges when venturing into providing open heritage data is to strike a good balance between data and interface. Not only in terms of how heritage material is presented to the public but also in terms of how open data is presented to potential end-users through the API. API or Application Programming Interface is an interface that enables data access for the use in application programming. As with any

interface design an API comes in many shapes and sizes and each of the cases show different needs in terms of interface access to heritage data. Another challenge is balancing the needs of end-users against the needs of institutions and data providers. Each has a different perspective and the cases will illustrate some of the needs and potential uses of end-users and in each case how this corresponds to the organisational perspective.

In this paper, we analyse and discuss from a user perspective and from an organisational perspective four cases where we as teachers, organizers and/or data providers have interacted with four different user groups: 1) researchers, 2) students, 3) app developers and 4) hackers. We deal specifically with the challenges and perspectives of the use of open cultural heritage data available as structured data as curated data sets or through API's. The cases we will present illustrate the range of challenges and perspectives that are currently a large part of heritage institutions work with open data. Each case provides a different view of the challenge of balancing datasets vs. interfaces as well as end-user perspective vs. organisational perspective. The cases range across four different potential user groups of heritage data and will show how the institution interacts with these different end-users as well as the challenges involved with each.

2 User cases

2.1 Researchers

In spring 2016, the Royal Danish Library invited researchers to join a series of data sprints in the exploration of digitized material related to the former Danish colonies, eg. photographs, maps, census records, newspapers, and toll registrations. The library's university lab helped organize the sprints and used their university networks for recruiting approx. 45 participants. The participants had different skills, for instance some were subject experts in history, others had programming skills, and some just had an interest in combining and learning about using digital data in new ways. Most participants came from humanities, fewer had a technical background and even fewer were from social science. In turn, we as data providers and curators brought a variety of competences (eg. in the archival material and in GIS and programming).

Some researchers came with specific research questions in mind, and others with more explorative approaches. In both cases, the data was critically examined and questions related to the data's reliability and trustworthiness sparked a continuous dialogue with curators and data providers. In several cases, research questions were reframed in an iterative process according to newly gained insights about the data. In the same way, tools were tested and discussed among researchers and curators, eg. methodological implications were discussed when analysing data with the tool Open-Refine to reduce the complexity of data by clustering and merging almost similar data. Researchers also had specific needs for working with the data, for instance some wanted to work with high resolution files for visual exploration, so this was facilitated by the library. The researchers emphasised the need to be able to search for material across collections through the API. Furthermore, the data sprint showed the need to improve

the API documentation and interface. This motivated the library to further develop the API with a how-to guide. In summary, the case illustrates researcher's continuous need for knowledge about data and tools during a research process and the challenge for the organization to meet unique and specific demands for original research questions while at the same time providing a generic service.

2.2 Students

In spring 2017 the Department of Information Studies, University of Copenhagen ran an elective course on open data (not only in heritage) for second year BA students. The course included an introduction to political and ethical issues through for example copyright and data protection legislation as well as practical exercises in using HTML, CSS, PHP and JavaScript to transform open datasets into new applications. The students met with open data providers, such as the Royal Danish Library, in order to gain a deeper understanding of institutional issues in providing open heritage data. Furthermore, the students were invited to join a boot camp organised in order to kickstart the idea generation and development process.

The course resulted in a handful of projects and prototypes developed in groups which presented different types of open data in various applications. In terms of heritage material one project, for example, used data from the Danish register of listed buildings provided by the Danish Agency for Culture and Palaces in order to build an app prototype for that uses principles of serendipity to lead users through the city of Copenhagen. The brief for the students was to use existing datasets to build new interfaces and visualisations of these data. They were instructed to clearly define a target audience for their app and could also include knowledge on system evaluation from other courses. One of the biggest challenges faced by the students was not so much using the data but rather understanding the content of the datasets through the API. The data interfaces they wanted to use were generally speaking not very well documented, with the exception of the Flickr API.

The students' overall aim, in this case, was to pass the course exam. In order to do this they had to produce a low or high fidelity prototype and discuss their use of open data. However, they were not reliant on being able to access a well-documented API as the discussion element of the course allowed them to also reflect over the challenges of using a particular dataset.

This case shows in particular the value of presenting open data in an interface that provides well documented content. Flickr is a good example of this with very specific methods for accessing data (fx. `Flickr.photos.getinfo` which returns information about specific photos if it receives the photo id). At this point in time with open heritage data at this early stage one of the best outcomes of this case was to get the students in the same room with data providers. Furthermore, getting students to test and discuss open heritage datasets is a prime opportunity to get feedback to data providers on the usefulness and usability of their APIs and data documentation.

2.3 App developers

This case reports on the use of digitised historical aerial imagery from the Royal Danish Library by app developers, retrieved through an API. The collection holds more than 900.000 historical aerial photographs, which have been digitised and geo-located through a crowdsourcing platform using volunteered geographical information (VGI) approach, where users can geo-locate the aerial images. The analogue collection has historically been intensively used by consultancy firms and government authorities to screen for potential environmental hazards, such as soil pollution and landfills. However, such use has been expensive due to the high cost of manual retrieval of the information. The first phase of the project concentrated on the island of Fyn. By 2015 more than 95% of the aerial photographs from Fyn were geo-located, thus constituting a comprehensive dataset of historical aerial imagery for this area [2].

The collection is accessible through an API in addition to the crowdsourcing platform. In January 2017 the Royal Danish Library was contacted by Geo Fyn A/S, a company established by the municipalities on Fyn, to develop and maintain a jointly operated Spatial Data Infrastructure (SDI) for administrative purposes among the local authorities on Fyn (<http://www.geofyn.dk/>). Key personnel from Geo Fyn A/S used the aerial photo API to harvest the coordinates and metadata, thus creating a GIS dataset of points with metadata and links to the specific imagery. This layer was then embedded on the Geo Fyn A/S data portal as well as a data layer in different systems of the various municipalities, making the historical information instant available on screen. Although only 5000 records could be retrieved pr. server request, this was easily navigated by Geofyn by developing a small script. Geo Fyn A/S plans to make yearly reviews of the data due to the potential adding of new metadata and adjustment of the geo-location by the library's volunteers. The Geofyn programmers did not have any issues accessing and using the API and did not need further documentation than was already provided.

This case illustrate that open heritage data, with sufficient metadata and accessible through an API, can provide app developers from industry and authorities with access to datasets, which can be used at low cost for developing new systems or adding new layers to existing decision support systems. App developers from these sectors, generally have the technical skills to extract data through an API and thus probably do not need a high degree of documentation in order to do so. However, their interested in heritage datasets is probably limited to those that hold an instant and easily accessible value for their operations, such as easily extractable location points and limited metadata.

2.4 Hackers

Since 2012, the Royal Danish Library has been co-organiser of the Danish national event Hack4dk. The event is an annual heritage hackathon with about 50 – 150 participants in average, organised by major GLAM institutions in Denmark. Participants are data providers, programmers, interaction designers, curators, etc. The outcomes of the hackathons are usually prototypes, products or concepts. The event is free and takes place during a weekend in autumn, from Friday afternoon till Sunday afternoon.

The goal of the event from the point of view of the organisers and data providers can broadly be defined as exploring the possibilities and usability of digital heritage collections, as well as testing the limitations and digital infrastructure of heritage institutions. At the same time, organisers break their monopoly of on creative use of their digital resources. Participants typically engage in the event for networking, learning and having fun. In their own words: "I'm here to have fun and do cool stuff", "I'm here to meet fellow geeks and talk about common interests", "I'm here to network and meet new people" [3].

While the process is facilitated by organisers, participants self-organise and work on their own projects, and in ways in which they prefer to work. Two participants explain the process in this way: "We are back this year because we have time again, and the past years have been really, really fun. It's great meeting new people and work on fun projects. We recently started our company with us two and another friend of ours and we came the three of us together and most likely we will end up working together, but we are open to ideas. At the moment we are just kind of exploring. First we are going get some inspiration also known as beer and then we are going to improvise something I guess." (video interview, hack4dk's facebook page).

The hackers in this case work on projects that do not necessarily have to end in a product and engage in activities that test the limits of skill, imagination and wits. The resulting products often are for fun. For instance, the winner project from 2017 was connecting Tin toys from KULMUS with AR and QR codes. By using the QR codes on the small blocks it is possible play with the toys without actually touching the objects and making stop-motion-movies (see presentation video at goo.gl/mwRmio).

This case shows that sometimes the process, the challenge and the collaboration is what counts. It shows in particular that from organizational point of view the data must be easy accessible, self-explanatory and inviting – and even so, whether the data will be used or not, is at the total mercy of the participants, dependent on their personal feelings, tastes, or opinions.

3 Conclusion

In this paper, we have analysed and discussed four cases where we as teachers, organisers and/or data providers have interacted with four different user groups: researchers, students, app developers and hackers. The four cases illustrate the range of challenges and perspectives that are part of heritage institutions work with open data. More specifically, we show that each of the four user groups have different motivations and skills for using open cultural heritage data. These findings can be summarized in this table, showing the user perspective, ie. users' motivation and users' technical skills:

Table 1. User motivation and technical skills among different user groups

| | Users' motivation | Users' technical skills |
|-------------|-----------------------|-------------------------|
| Researchers | Gaining new knowledge | Varies |

| | | |
|----------------|--|----------------|
| Students | Learning, pass the course | Varies |
| App developers | Make profit, development of own services | Highly skilled |
| Hackers | Have fun, be challenged | Varies |
| Researchers | Gaining new knowledge | Varies |

In turn, different user groups' motivations and technical skills require the cultural heritage organisation to curate its data according to specific user groups. These findings can be summarized in this table, showing the organisational perspective, ie. data content requirements and data accessibility:

Table 2. Data content requirements and data accessibility among different user groups

| | Data content requirements | Data accessibility |
|----------------|---|---|
| Researchers | Subject specific, unique data, trustworthy data | Varies (technical skills), Flexible (research questions) |
| Students | Data as training sets | Varies, depending on skills and learning goals |
| App developers | Subject specific, unique data | "Help to self-help" |
| Hackers | Subjective data/dependent on personal feelings, tastes, or opinions | Data should be easy accessible/data as mean to something else |
| | Data content requirements | Data accessibility |

Thus, successful use of open cultural heritage data is highly dependent on organisations' ability to calibrate and curate the data differently according to contexts and settings.

4 Reference

1. Tanner, S., & Deegan, M. (2013, October). Measuring the impact of digitized resources: The Balanced Value Model. In *Digital Heritage International Congress (DigitalHeritage), 2013* (Vol. 2, pp. 15-19). IEEE
2. Svenningsen, S. R., Brandt, J., Christensen, A. A., Dahl, M. C., & Dupont, H. (2015). Historical oblique aerial photographs as a powerful tool for communicating landscape changes. *Land Use Policy*, 43, 82-95. DOI: 10.1016/j.landusepol.2014.10.021
3. Wang, Jacob R. (2017): Hack4dk: Hacking our heritage 2012-2017, presentation at Hack4dk, Enigma, Copenhagen, Sept. 29.