CHAPTER 16

Developing Sustainable Open Heritage Datasets

Henriette Roued-Cunliffe

This chapter will explore issues and methods involved in the development of sustainable open heritage datasets and how these datasets may be used. To do so, it includes discussions of concepts such as openness, crowdsourcing and copyright. The chapter will include a case study based on the collection "Aagaards Photos" from Kolding City Archives (Denmark). The dataset consists of image-files, tables of identified people, and geographical locations for people and places. It is currently stored in three different locations that are openly accessible: Flickr, Google Maps, and Google Sheets. This chapter will present a step-by-step guide to how data is extracted from each of these datasets through readily-available and well-described Application Programming Interfaces (APIs) and how these datasets can be combined for the purpose of analyses or visualization. Open heritage data combines the idea of heritage openness on one hand and working with heritage data on the other. Openness can be understood in broad terms as involving participatory

H. Roued-Cunliffe (✉)
University of Copenhagen, Copenhagen, Denmark
e-mail: roued@hum.ku.dk

l. levenberg et al. (eds.), *Research Methods for the Digital Humanities*,
https://doi.org/10.1007/978-3-319-96713-4_16

approaches, community collaborations, crowdsourcing, and audience engagement. Working with heritage data typically falls under the domain of Digital Humanities, archeological computing, heritage informatics, and collection building. The combination of these two ideas constitutes a growing area that is identifiable in a range of initiatives, such as opening-up museum collections to heritage hackathons. Yet, many heritage professionals and humanities students are still seeking concrete methods to deal with open heritage data.

First, I will briefly discuss the concept and practicalities of open heritage data. Then I will illustrate the approach through an example of accessing, transforming, and combining three different types of data. The example is taken from the "Aagaards Photos" digitization and crowdsourcing project, which is conducted through Kolding City Archives in Denmark and incorporates photos, geographical locations, and biographical metadata.[1]

Openness in Heritage Data

There are a lot of things to consider in relation to openness in heritage data. For instance, all Digital Humanities researchers working with heritage data face choices pertaining to formats, platforms, and other technical specifications. When researchers intend to open their data to other users they must also consider issues regarding users, reuse, misuse, commercial use, legal requirements, funding and copyright. Courtney Ruge, et al. identify several issues that make heritage institutions in Australia reluctant or unable to share heritage collections online:

- lack of technological knowledge, resources, and limited funds within the organization;
- questions of ownership (where material has been donated, there are various sources and copyright holders, or where many of the collected texts are **orphaned works**);
- the belief that physical access to the collections meets most needs;
- concerns regarding the unauthorized reuse of images;
- concerns over losing revenue streams (the ability to sell digital copies of images or access to material);

[1] See http://aagaardsbilleder.tumblr.com/english.

- concerns regarding infringement of copyright or **privacy** regulation;
- and, the ethics of sharing material about identifiable people.[2]

Although some major heritage institutions around the world have the resources, know-how, and political stamina to rise above these issues, most community groups, as well as smaller and medium-sized institutions share these concerns.

First, the issue of copyright stands out as likely to have the most impact on heritage institutions. It is currently a cause of fear and uncertainty, because many heritage professionals have so little experience with it. Graham Cornish suggests that:

> Copyright law aims to protect [the growth of writing, performing and creating] but, at the same time, tries to ensure that some access to copyright works is allowed as well. Without this access creators would be starved of ideas and information to create more copyright material.[3]

Thus, copyright must balance the protection of current creative work with the support of future creative work. In essence, copyright undermines its stated intention when it suppresses new creative work without a compelling reason. Furthermore, Peter Drahos and John Braithwaite observe that "Copying and imitation are central to our process of learning and the acquisition of skills. […] The creator of innovation is also always the borrower of ideas and information of others."[4] Fear of copyright is, thus, one of the biggest risks to future innovation because, contrary to the intentions of copyright itself, it can stifle creativity, research, and invention.

[2] Courtney Ruge, Tom Denison, Steve Wright, Graham Willett, and Joanne Evans, "Custodianship and Online Sharing in Australian Community Archives," in *Participatory Heritage*, ed. Henriette Roued-Cunliffe and Andrea J. Copeland (London, UK: Facet Publishing, 2017), 82–83.

[3] Graham Cornish, *Copyright: Interpreting the Law for Libraries, Archives and Information Services* (London, UK: Facet Publishing, 2015), 1, http://public.eblib.com/choice/publicfullrecord.aspx?p=2073251.

[4] Peter Drahos and John Braithwaite, *Information Feudalism: Who Owns the Knowledge Economy?* (Abingdon, UK: Earthscan, 2002), 2.

Open heritage approaches offer an alternative to this fear. The Open Definition project states that: "Open means anyone can freely access, use, modify, and share for any purpose (subject, at most, to requirements that preserve provenance and openness)."[5] In a heritage context the OpenGLAM[6] initiative has developed principles regarding how institutions should engage with the public about the reuse of their open data.

The first OpenGLAM principle is to "release digital information about the artifacts (metadata) into the public domain using an appropriate legal tool such as the Creative Commons Zero Waiver."[7] Releasing metadata can improve the discoverability of your data, whether your data consists of finds, findspots, maps, images, videos, interpretations, official documents, letters, artwork, etc. Institutions have different traditions of metadata. Libraries have always been at the forefront of metadata collection and sharing, mainly due to their fairly homogenous material as well as the obvious value in libraries sharing the metadata for items in their collections. However, when it comes to material that is more unique and scarcer metadata sharing becomes less common. This is an area where initiatives such as the Europeana digital platform for cultural heritage have made a great difference.[8] Europeana has demonstrated the value of sharing metadata, so that cultural heritage material can be discovered and shared across Europe.

The second principle prompts researchers and archivists to "keep digital representations of works for which copyright has expired (so works that are in the public domain) in the public domain by not adding new rights to them" ('OpenGLAM Principles').[9] This principle contains a practical criticism of copyright law and the institutions which use it to maintain control over the material that they digitize. In some countries,

[5] Open Definition, "The Open Definition - Open Definition - Defining Open in Open Data, Open Content and Open Knowledge." 2017. Accessed May 10, 2017, http://opendefinition.org/.

[6] According to their website, "**OpenGLAM** is an initiative coordinated by Open Knowledge that promotes free and open access to digital cultural heritage held by Galleries, Libraries, Archives and Museums."

[7] "OpenGLAM Principles." OpenGLAM. 2013. Accessed April 8, 2017, https://openglam.org/principles/.

[8] The Europeana.eu platform links thousands of organizations from across the European Union to aggregate digitized cultural heritage works: https://www.europeana.eu/portal/en.

[9] "OpenGLAM Principles." OpenGLAM. 2013. Accessed April 8, 2017, https://openglam.org/principles/.

interpretations of copyright law stipulate that when the process of digitizing analog material constitutes a creative process by a human being (for example, adjusting light, etc.) and is not just an automated machine process (for example, a scanning), then the individual or institution acquires copyright over the digitized version.[10] One could argue that this practice is ethically irresponsible, especially if it is done using public funds or by a public institution. A counter argument is that heritage institutions sometimes rely on the revenue that this practice produces. However, Simon Tanner found that license and service revenue from digital images does not cover the actual costs of digitization practices and services in museums, and concluded that external funding is essential.[11]

The third OpenGLAM principle states: "when publishing data make an explicit and robust statement of your wishes and expectations with respect to reuse and repurposing of the descriptions, the whole data collection, and subsets of the collection."[12] This means that the institution making data available online needs to make sure that they are knowledgeable about the copyright status of the material with which they are working and they are clear about how it can and cannot be reused. Reuse forms a significant part of creative and innovative processes, which are facilitated through the web. Publishing online has, therefore, come to mean that everyone with the technical ability can reuse this material whether they are permitted to or not. Nonetheless, with the movement for openness growing, as it currently is, and the growing availability of heritage material for reuse, the bigger risk is that the material you put online without clear reuse statements is more likely to be passed over and, thus, forgotten. There is also the question of commercial use. It is now generally accepted that commercial use can be anything from posting on a blog with advertising revenue to use by large multibillion-dollar industries. This has the consequence that saying no to commercial reuse not only means saving your material from "big bad corporations", but also stops new creative reuse from artists, bloggers, and others who could potentially use it to make small earnings.

[10] Cornish, 7.

[11] Simon Tanner, *Reproduction Charging Models & Rights Policy for Digital Images in American Art Museums* (KDCS Digital Consultancy, 2004), https://kclpure.kcl.ac.uk/portal/files/48081293/USMuseum_SimonTanner.pdf.

[12] "OpenGLAM Principles."

The fourth principle is "when publishing data use open file formats which are machine-readable."[13] This means avoiding proprietary formats, which can only be accessed by certain computer programs or require special licenses. There are open, machine-readable formats available to suit a myriad of projects and types of material. In essence, a **CSV** file is open and machine-readable, but a webservice using **JSON** or **XML** data format would be even more so. Data documentation is also an important consideration. Without good documentation and a user-friendly interface for accessing the dataset, your well-intended open data is as unlikely to be used as the many boxes in your remote storage.

The last principle states that "opportunities to engage audiences in novel ways on the web should be pursued."[14] This requires funding and an overall strategy for open data. It is a significant aspect of open data and a challenge at which projects most often fail. We put data on the web, we say people can use it, but we do not facilitate and actively engage with the people who would potentially want to use it. This is not only about going to hackathons and encouraging programmers to develop stuff with your datasets. This is also about thinking outside the box and considering who could benefit from your data. School children and teachers, perhaps? In this case, you need to contact some to see how they could use it. University students? Then you will need to reach out to relevant departments or lecturers teaching relevant subjects. Amateur historians? To reach them you can find groups that already exist on social networking platforms that are organized around subjects such as local history or amateur archeology. The likelihood of these groups coming to you is very small if you are just starting out with open data. Further, this is not a task for a temporary student placement. It requires ongoing funding and dedicated staff that have time to reach out and engage people about using your open data. As is exemplified in the Kolding City Archive's Aagaard Photos project, the benefits and outcomes from this approach are many and varied. They include local and national media exposure, interest from scholars and students, app development, new artwork, and a solid network of people interested in heritage (many of whom have contacted the archive and enriched the project with their own funds and time).

[13] Ibid.

[14] Ibid.

Aagaards Photos

The Aagaard Photo project began suddenly after New Year, 2015 when project member Maria Wehde discovered the photographs of Dines Christian Jochum Pontoppidan Aagaard. Aagaard used the early technique of collodion negatives to take portraits of local Kolding citizens, between 1857 and 1880. The early photography technique made the collection historically significant and the name labels on the glass negatives made it particularly useful for family historians in the area.

The project quickly evolved into an ad hoc crowdsourcing project through which family historians helped identify the people in the photos by transcribing the labels, and looking up individuals in the census and parish records. From the beginning, it was a goal of the project to keep all materials, old and new, openly accessible simultaneously with their digitization or collection. This has resulted in a dataset that includes 2103 portraits of which 733 have been identified. All the images are hosted on Flickr and the information about each individual featured in the photographs is hosted in a Google Sheet. Furthermore, all places noted in the customer registry are stored in a custom Google Map. As part of the crowdsourcing effort, volunteers use a simple form to send information they have discovered about the individuals in the photos with references to the sources in which they are found. The archive then checks and approves the information and adds it to the constantly growing list. Like in many other crowdsourcing projects there is also a long-tail distribution of contributors' workloads; a few volunteers contribute most of the information and many contribute only a little.[15] Connection to the genealogy and local, amateur history communities through active Facebook groups has led to various new efforts at the archive, such as ongoing history workshops, a mash-up past and present workshop, and a repair and remake festival with a photographic atelier inspired by Aagaard's photos.

[15]Tim Causer and Melissa Terras, '"Many Hands Make Light Work. Many Hands Together Make Merry Work': *Transcribe Bentham* and Crowdsourcing Manuscript Collections," in *Crowdsourcing Our Cultural Heritage*, ed. Mia Ridge (Surrey, UK: Ashgate, 2014), 73.

Why APIs?

Understanding and using APIs is essential to optimize the openness and usefulness of the online publishing methods that researchers and archivists use. I have formulated a list of questions that can be asked about material that is shared online:

1. Is the material published with metadata so that it can be searched and found?
2. Is the material published with an open license or in the public domain, and is this clearly communicated in conjunction with the material?
3. Does the creator actively encourage reuse of the material and provide support for anyone who wishes to reuse it, free of charge?
4. Is a static version of the material available in an open, machine-readable format, that anyone can export/download?
5. Is the material available as a live data stream through a well-described API, that anyone can access?

There are many ways to make heritage material available online and some are more open than others but, currently, the most open method is through an API. An API is, as the name suggests, a set of protocols and tools for building applications. There are many ways to access and use heritage material through APIs. You can download datasets to visualize and analyze in existing software or view it with command line prompts. Another method is to access the live stream of datasets remotely. This method will be described in the following guide.

Many online services (e.g. Twitter, Instagram, Facebook, etc.) provide APIs with the intention that developers around the world can use them to build new applications that complement and support the main service. In this guide we will access two Google APIs and Flickr's API to reuse heritage data stored on these two services. As mentioned previously, one of the most extensive heritage APIs has been built by Europeana, which provides access to 53,455,335[16] artworks, artifacts, books, videos, and sounds from across Europe through different methods and protocols. Europeana labs, which is the hub for apps and app development, includes a strong program for encouraging the use and reuse of this

[16] www.europeana.eu, as of October 1, 2017.

enormous heritage dataset. Among other things it showcases 74 API implementations, such as a bulk downloader, new ways of browsing the dataset, host challenges, and in other ways encourage reuse.

Guide

This guide will demonstrate how to access and combine heritage data from three linked datasets in order to present and share it online. Additional information and guides on the coding presented in this guide can be found at W3Schools.com (this site is continuously updated to reflect the latest coding uses).[17] The steps will use server-side programming language, **PHP**, web-markup language, **HTML**, **XML**, and **JSON** data formats. This requires limited technical specifications:

- a browser to view and execute the code;
- a server or localhost that runs **PHP** and has **SimpleXML** and **JSON** installed;
- and, a text-editor and a way of transferring the php files to the server.

Extracting Photos from Flickr

This guide shows how to extract photos from Flickr, which is currently the most comprehensive image sharing site. The biggest advantage of using Flickr for heritage photo sharing as an individual, group, or heritage institution is that it has a built-in and very well-documented API, which is relatively easy to use without signing up. In order to make full use of the API on a different server you will need to apply for an **API-key** from Flickr. The following example will extract the photo URL, title, and id from the identified Aagaard photo album on Flickr.

Step 1: Identify the methods and parameters needed to call the API and retrieve the dataset. In this case, we are using the Flickr method flickr.photosets.getphotos, which will allow us to retrieve data for each photo in the album. When we call this method we include the following parameters (Table 16.1):

[17] "W3Schools Online Web Tutorials," W3 Schools. 2017. Accessed May 29, 2017, https://www.w3schools.com.

Table 16.1 API Parameters

Parameter name	*Value*	*Description*
photoset_id	72157650969693930	This parameter is required and identifies the photoset/album we want to call. The id can be found at the end of the photoset album url
format	Rest	The format of the data returned (in this case RESTful **XML**)
api_key	[YOUR API-KEY]	The API-key you received from Flickr
extras	url_s	Extra data we need to extract (in this case the url for the small-sized image)

Flickr's API explorer helps us put the method and the parameters together into a URL that can return the dataset as XML:

Flickr.com API

URL: https://api.flickr.com/services/rest/?method=flickr.photosets.getPhotos&api_key=[YOURAPI-KEY]&photoset_id=72157650969693930&extras=url_s&format=rest

Box 16.1: Returned XML

```
<?xml version="1.0" encoding="utf-8"?>
<rsp stat="ok">
 <photoset id="72157650969693930" [...]
title="Identificerede personer">
  <photo id="22297400674" [...] title="0001 Ane
Marie Pedersen, Eltang 73" [...]
url_m="https://farm6.staticflickr.
com/5741/22297400674_97d034a93b.jpg" [...] />
  <photo id="24005467813" [...] title="0002
Andreas Lauritsen, Vejstruprød" [...]
url_m="https://farm2.staticflickr.com/1656/
24005467813_12df393f6d.jpg" [...] />
[...]
</photoset>
</rsp>
```

Step 2: We now have the URL and can use it to call the live dataset for this album. At this point there is the option to save the dataset as a static file on our server or computer and analyze it there. Instead we will manipulate and analyze the data as entities on the server, so that our application is always using the latest dataset updated by the city archive. To do this we need to call the dataset into our PHP file using the simplexml_load_file() function:

Box 16.2: Calling the dataset

```
// URL built with Flickr API
$flickr_api_url =
"https://api.flickr.com/services/
rest/?method=flickr.photosets.
getPhotos&api_key=[YOUR API-KEY]&photoset_
id=72157650969693930&extras=url_s&format=rest";
// Load XML from Flickr URL
$flickr_xml = simplexml_load_file($flickr_api_url);
```

Step 3: We can now treat our dataset like a PHP array, which we can loop through in order to transform id, title, and URL into HTML that can be presented in a browser. The title of the photo in Flickr is made up of the original photo number and the photo title, so we need to use the explode() function to split these two entities. In order to present the image files as actual images in the browser we add the URL to the HTML <*img*> tag.

Box 16.3: Present the image files in a browser

```
// Loop through the photos and add each one to a variable
foreach ($flickr_xml->photoset->photo as $photo){
        // Get title from Flickr
        $flickr_title = $photo->attributes()->title;
        // Split id from title
        $explode_title = explode(" ", $flickr_title, 2);
        // Grab the integer value of the aagaard photo id
        $aagaard_photoid = intval($explode_title[0]);
        // Output photo id:
```

```
        echo "<b>Aagaard photo number:</b> " . $aagaard_
        photoid . "</br>";
        // Grab the title-text of the photo:
        $aagaard_title = $explode_title[1];
        // Output the title text with a linebreak after:
        echo "<b>Title:</b> " . $aagaard_title . "<br/>";
        // GRAB URL for photo
        $img_url = $photo->attributes()->url_s;
        // Output photo-url in a HTML image tag:
        echo "<img src='" . $img_url . "'/>";
        // Output a horizontal line after each image
        echo "<hr/>";
}
```

This enables us to output the image (see Fig. 16.3) below the text: Aagaard photo number: 1, Title: Ane Marie Pedersen, Eltang 73.

Extracting Locations from Google Maps

This guide shows how to extract geographical locations from a custom Google map. The archive chose to use a custom Google map to pinpoint the location of individuals in the photos based on the photo registry, because it has a simple interface for doing so. Flickr also allows for geolocating uploaded photos, however, the level of detail for this geotagging is too low.

Step 1: Build the URL to extract the custom Google map data as a KML XML format. First, we need to find the map id in the map URL. It is located with the parameter mid:

https://www.google.com/maps/d/viewer?mid=1axcGJpqIjtHUfo-MUhKhc3Z-2L74&ll=55.38810642013983%2C9.992526799999955&z=8

In this case we are simply requesting the map dataset as KML instead of viewing it through the Google map viewer. An API-key is not necessary for this yet:

https://www.google.com/maps/d/kml?mid=1axcGJpqIjtHUfoMUhKhc3Z-2L74&forcekml=1

Step 2: Again, we use the simplexml_load_file() function to access this dataset as KML, which is a variant of XML.

Box 16.4: Accessing the dataset as KML

```
// Get id for the custom map
$maps_id = "1axcGJpqIjtHUfoMUhKhc3Z-2L74";
// Build the URL requesting KML (a form of XML)
$maps_api = "https://www.google.com/maps/d/kml?mid=" .
$maps_id . "&forcekml=1";
// Call the KML as an XML file
$maps_kml = simplexml_load_file($maps_api);
```

Step 3: As with the photos we need to loop through the dataset, but this time we need two loops. Custom Google maps are made up of layers and in this case the archive has made two layers (Fig. 16.1). The first layer contains all the place names found in the photo registry and the second contains all the identified people, which are organized by photo. It is the second layer we want to use here, so first we loop through the layers (stored

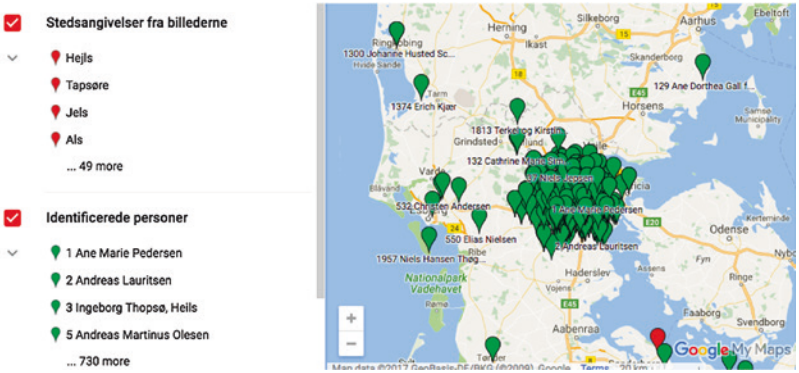

Fig. 16.1 The dataset in Google maps with locations

in <*folder*> tags in the KML) and if the layer has the right name we move on. Then we loop through the placemarks. For each placemark we want the name, which is made up of the original photo id and the title of the photo. This has to be split using the explode() function. The coordinates for each photo are found together in the <*coordinates*> tag and need to be split too.

Box 16.5: Looping the dataset

```
// Loop through map folders
foreach ($maps_kml->Document->Folder as $folder){
      // Retreive the right map folder
      if ($folder->name == 'Identificerede personer'){
            // Loop through placemarkers
      foreach ($folder->Placemark as $placemark){
            // Get placemarker name
                  $map_title = $placemark->name;
                  // Split photo id from photo title
                  $explode_title = explode(" ", $map_title, 2);
                  $map_photoid = intval($explode_title[0]);
                  $map_title = $explode_title[1];
                  // Output id
                  echo "<b>Aagaard photo number:</b> " .
                  $map_photoid . "</br>";
                  // Output title
                  echo "<b>Title:</b> " . $map_title . "<br/>";
                  // Get lat/lon
                  $coordinates_string =
                  $placemark->Point->coordinates;
                  $coordinates_explode = explode(",",
                  $coordinates_string);
                  $longitude =
                  ltrim($coordinates_explode[0]);
                  $latitude = $coordinates_explode[1];
                  // Output coordinates
                  echo "<b>Coordinates:</b> " . $latitude .
                  " / " . $longitude . "<br/>";
                  echo "<hr/>";
            }
      }
}
```

The output for this is:

Aagaard photo number: 1
Title: Ane Marie Pedersen
Coordinates: 55.5277318 /9.5367765

Extracting Biographic Metadata from a Table

The archive decided to use Google sheets to present the crowdsourced biographic data about the individuals that were identified (Fig. 16.2).

Step 1: Google sheets has a well-documented API and can be used to access the data in the tables in the JSON format. In order to do so we apply for an API-key from the Google API console and enable the API for Google Sheets.

Step 2: Spreadsheets are normally private in terms of access and permission like other Google documents, so to share this sheet through the API we need to publish it to the web. In the sheet go to File > Publish to the web > Publish. Now we need the sheet id, which can be found in the URL:

https://docs.google.com/spreadsheets/d/1krJlfljyXUY6orNIkgOo8LIHlLnYf5NfII81_5GDvEE/edit#gid=681292462.

Then identify the range we want to extract. In this case, the range includes column A to column V (i.e. A:V). Adding the API-key, sheet id, and range as parameters to the URL we can now retrieve this dataset as JSON:

https://sheets.googleapis.com/v4/spreadsheets/1krJlfljyXUY6orNIkgOo8LIHlLnYf5NfII81_5GDvEE/values/A:V?key=[GOOGLEAPI-KEY].

	A	B	C	D	E	F	G	H	I
1		Photoid	Firstname	Lastname	Title/job	Born	Birth place	Death	Death place
2	02/08/2016 10.01.04	1	Ane Marie	Pedersen		12/08/1852	Gudsø Mark, Eltang		
3	02/08/2016 10.54.27	2	Andreas	Lauridsen		25/09/1863	Vejstruprød, Vejstrup		
4	10/04/2015 11.01.48	3	Ingeborg Margrethe	Topsøe		21/05/1869	Store Heddinge Sogn	18/02/1953	Vallekilde
5	02/08/2016 10.57.56	5	Andreas	Olesen		24/06/1843	Fredericia		
6	10/04/2015 11.10.01	9	Poul Jepsen	Bekker	Kludesamler	20/09/1828	Øster Starup		
7	10/04/2015 11.12.21	9	Else Marie	Jepsen		05/09/1829	Bøgvad, Egtved		

Fig. 16.2 The dataset in Google sheets

Step 3: As with the photos from Flickr and the coordinates from the custom Google map, we request the data from the API. However, this time we use the get_file_contents() and json_decode() functions to pull the dataset into a PHP array.

Box 16.6: Requesting data from the API

```
// API key - apply to Google, remember to enable
Google Sheets API:
https://console.developers.google.com/apis
$sheets_api_key =
"AIzaSyAZXM3NNUF_zVBEncYzVNU4byo3cOVrqvw";
// Sheet id - remember to publish sheet: File >
Publish to the web >
Publish
$sheet_id = "1krJlfljyXUY6orNIkgOo8LIHlLnYf5NflI81_5GDvEE";
// Range to extract from sheets - use spreadsheet
syntax, fx. A:D or A2:B32
$range = "A:V";
// Combine URL to call sheet
$sheets_api = "https://sheets.googleapis.com/v4/
spreadsheets/" . $sheet_id . "/values/" . $range .
"?key=" . $sheets_api_key;
// Call the JSON file
$sheets_json = file_get_contents($sheets_api);
// Decode JSON file into a PHP array
$sheets_array = json_decode($sheets_json);
```

Step 4: Because this data originated as a table where the first row consists of headers for each column, we begin by setting a counter before the loop and start the output after the first row in the dataset. Unlike the XML data, this data is not contained in meaningful tags and therefore we need to consult the original table to find the right column. For example, the original photo number is in column B and the equivalent of this in the $rows array is index number 1 (array indices always begin at 0). Note how this dataset is built up around individuals in the photos, rather than the photo themselves. This is why this dataset contains more than one row of information with the same photo id number (Fig. 16.2).

Box 16.7: Setting a counter and starting the loop

```
$counter = 0;
foreach ($sheets_array->values as $rows){
      // Add a counter to skip first row of labels
      if ($counter++ == 0) continue;
      // Find photo number in column B (array value 1)
      $sheets_photoid = $rows[1];
      // Output photo number
      echo "<b>Aagaard photo number:</b> " . $sheets_photoid
      . "</br>";
      // Find other information in the columns C=2, D=3,
      F=5, G=6
      $firstname = $rows[2];
      $lastname = $rows[3];
      $born = $rows[5];
      $born_place = $rows[6];
      // Output information
      echo "Full name: " . $firstname . " " . $lastname. "";
      echo "Born: " . $born . ", " . $born_place. "";
      // Add line after each individual
      echo "<hr/>";
}
```

The output from this is a list of photos with the id, full name, and birth date and place:

> Aagaard photo number: 1
> Full name: Ane Marie Pedersen
> Born: 12/08/1852, Gudsø Mark, Eltang

Combining Datasets

The datasets have two components: identified individuals and photographs. The two components have a many-to-many relationship; identified individuals can be in more than one photo and each photo can have more than one identified individual. Because only photographs have an id number, we will use them as the basis to combine the three datasets.

Step 1: The first step is to extract data from each dataset and loop through the data. Instead of outputting the data, we add it to one array called *$photographs*. To do this we use the photo id from the flickr dataset (*aagaard_Photoid*), *themapdataset* (map_photoid), and the sheets table ($sheets_photoid) as the key to combine the information from the three datasets into one array.

Box 16.8: Extracting Data

```
$photographs[$aagaard_photoid]['img'] = "<img src='" .
$img_url . "'/>";
              $photographs[$map_photoid]['latitude'] = $latitude;
              $photographs[$map_photoid]['longitude'] = $longitude;
$photographs[$sheets_photoid]['people'][$name] = $birth;
```

Step 2: Loop through the array and output the information about the two components as HTML in the form of a single listing for each photo (Fig. 16.3).

Box 16.9: Loop through the array

```
// Loop photographs array
foreach ($photographs as $photoid=>$photoinfo){
       // Output photo id:
       echo "<b>Aagaard photo number:</b> " . $pho-
       toid . "</br>";
       // Output coordinates
       echo "<b>Coordinates:</b> " . $photoinfo['lat-
itude'] . " / " . $photoinfo['longitude'] . "<br/>";
       // Loop people in photo
       foreach        ($photoinfo['people']        as
       $name=>$birth){echo $name . ", " . $birth .
       "<br/>";
       }
       // Output image
       echo $photoinfo['img'] . "<br/>";
       // Add line after each photograph
       echo "<hr/>";
  }
```

Aagaard photo number: 5
Coordinates: 55.4890425 / 9.4724867
Andreas Olesen, 24/06/1843, Fredericia

Aagaard photo number: 9
Coordinates: 55.5856339 / 9.422493
Poul Jepsen Bekker , 20/09/1828, Øster Starup
Else Marie Jepsen, 05/09/1829, Bøgvad, Egtved

Fig. 16.3 The final output of the combined datasets

Conclusion

The Aagaards Photos archive and the step-by-step guide to use Flickr's API in a Digital Humanities project demonstrate opportunities for open heritage datasets. Researcher working with open heritage data still face a number of questions related to ownership, copyright, and data protection. However, other important and often neglected considerations involve the risks of losing data and failing to have social impact.

Even though we spend enormous funds digitizing and structuring heritage data, it is still as vulnerable to being lost or falling into disuse in our digital archives. In other words, digitization and online publishing do not ensure that heritage material will have a broader or long-lasting social impact. The motivation of this chapter is to reduce this risk through open data and APIs. It is still not as simple as building an API and sitting back to wait for the public to use it. As with any dissemination of a heritage project, it still requires facilitation. Good examples, which involve the public in open heritage data include heritage hacks and data sprints. The main goal is often to explain and motivate the use of heritage datasets to programmers and interface designers and encourage participation. Furthermore, teaching new generations of humanities students the basics of programming and data management will go a long way towards promoting deeper understandings of humanities materials through new digital platforms.

Acknowledgements I want to thank Maria Wehde for pushing the open agenda and including me in the Aagaard Project in the first place and City Archivist, Lene Wul, for continuing down this path with me. I also want to thank the students at University of Copenhagen for their willingness to learn to program with open data and for enabling me to develop my tutorials.

References

Causer, Tim, and Melissa Terras. "'Many Hands Make Light Work. Many Hands Together Make Merry Work': *Transcribe Bentham* and Crowdsourcing Manuscript Collections." In *Crowdsourcing Our Cultural Heritage*, edited by Mia Ridge, 57–88. Surrey, UK: Ashgate, 2014.

Cornish, Graham P. *Copyright: Interpreting the Law for Libraries, Archives and Information Services.* London, UK: Facet Publishing, 2015. http://public.eblib.com/choice/publicfullrecord.aspx?p=2073251.

Drahos, Peter, and John Braithwaite. *Information Feudalism: Who Owns the Knowledge Economy?* Abingdon, UK: Earthscan, 2002.

Open Definition. "The Open Definition—Open Definition—Defining Open in Open Data, Open Content and Open Knowledge." 2017. Accessed May 10, 2017. http://opendefinition.org/.

"OpenGLAM Principles." OpenGLAM. 2013. Accessed April 8, 2017. https://openglam.org/principles/.

Ruge, Courtney, Tom Denison, Steve Wright, Graham Willett, and Joanne Evans. "Custodianship and Online Sharing in Australian Community Archives." In *Participatory Heritage*, edited by Henriette Roued-Cunliffe and Andrea J. Copeland, 79–86. London, UK: Facet Publishing, 2017.

Tanner, Simon. *Reproduction Charging Models & Rights Policy for Digital Images in American Art Museums.* KDCS Digital Consultancy, 2004. https://kclpure.kcl.ac.uk/portal/files/48081293/USMuseum_SimonTanner.pdf.

"W3Schools Online Web Tutorials." W3 Schools. 2017. Accessed May 29, 2017. https://www.w3schools.com.