Documentation
An Experimental Hackaton Workshop for Visual Material Testing Recogito among humanities reserachers'
January 2020 in collaboration with Uppsala University Library, Sweden

Coordinators:
Karolina Andersdotter, Digital Librarian, Uppsala University Library
Karin Hansson, Associate Professor, Computer and System Sciences, Stockholm University
Anna Dahlgren, Professor of Art History, Stockholm University

Keywords: crowdsourcing, researcher-generated metadata, image archives, annotation, data curation, hackathons, recogito

BACKGROUND

Hackathons and expert-based crowdsourcing

The general description of a *hackathon* – a portmanteau of *hack* and *marathon* – is a "highly engaging, continuous event in which people in small groups produce working software prototypes in a limited amount of time." (Komssi et al., 2015) While "[h]ackathons vary wildly in their purpose and execution but generally have a common structure and characteristics" (Komssi et al., 2015), the wilderness is not so broad as to not prompt Thornham and Gómez Cruz to define their broader use of the term in a footnote, saying that "[a]lthough we have participated in a number of variously named and labelled events, and although we are aware of the particular history the term 'hackathon' evokes – sometimes uncomfortably for participants – we will refer to these events collectively as 'hackathons' for the duration of the paper." (Thornham & Gómez Cruz, 2016)

Uppsala University Library has organised several *cultural heritage hackathons* since the first one in May 2018 (see e.g. (Svensson, 2018)).The purpose of these hackathons has not been to work on software projects, but rather to enrich the cultural heritage collections so that they may offer better opportunities for research. Another purpose has been to teach data literacy to the library users attending the event, as part of the information literacy courses the library offers in various formats. The latter purpose is also the reason the term *hackathon* was used; because of its connotations, the intention was to de-dramatise data for students and researchers, especially those who are in the field of arts and humanities.

Crowdsourcing participants motivations and work modes differs depending on the set-up and the aim of the crowdsourcing activity. Hansson et al. (2017) mentions power differences and (lack of) transparency as some factors that have an impact on the crowdsourcers' motivation, and they divide crowdsourcing into three distinct types depending on the technologies used: human computation, peer competitions, and open collaboration. Of these three types, the crowdsourcing hackathon is most similar to *open collaboration*, which is defined by that "participation is voluntary and there are no prizes or money involved. Participants are often known to each other or at least have public profiles within online social network." (Hansson et al., 2017). Because facilitators and participants meet in the same setting, power differences and transparency can be approached (and possibly negotiated) by facilitators and participants immediately.

Expert-based crowdsourcing has previously been used in e.g. crowdsourced policymaking. Aitamurto and Chen (2017) writes that "expert sourcing (only experts invited) and public sourcing (the broader public invited) could be useful in cases in which the crowdsourcer wants to receive knowledge from a more narrow set of participants." Depending on the aim with the crowdsourcing activity this can be very useful; the drawback is that the data becomes less diverse if only experts are employed and for some projects this would be detrimental.

Using expert-based crowdsourcing can also be a way to harmonise crowdsourced data; knowing your crowd beforehand means that you can anticipate irregularities in the data input. In public crowdsourcing, there's examples of how the data input is irregular depending on the participant. Rawson (2016) writes about such an experience in the *Curating Menus* project, where volunteers worked to transcribe historical restaurant menus digitised by New York Public Library:

"Because untrained volunteers typed the menu item data in a free-text field, it contains an array of orthographic variations. The menu data, much of which was created by an earlier team of volunteer transcribers working from handwritten catalog cards, is also highly irregular" (Rawson, 2016, p. 59)

Recogito – a digital humanities platform for annotation
Recogito is described on its webpage as a tool for "Semantic Annotation without the pointy brackets" (*Welcome to Recogito*, n.d.), which implies a simplified way of using the markup/annotation methods frequently used in digital humanities (e.g. corpus linguistics). Martin Weisser, using the words *markup* and *annotation* synonymously, notes that "[a]lthough there are actually many different ways of marking up a document, one fairly standard method is the use of the kind of *tags* that we use for writing HTML documents /.../ generally represented in markup languages by pairs of opening and closing angle brackets, i.e. < & >, with the name of the [tag] appearing in between the two" (Weisser, 2016, p. 244).

What Recogito does is provide an interface which will let the user interact with the research material in a *graphical user interface* (GUI) rather than adding metadata in a *text-based user interface*. The GUI makes the tool more accessible to users that have less experience in digital research tools and methods, as it makes the tool similar to other, more commonly used digital tools such as web browsers or word processors.

Recogito also allows for export of the created metadata (i.e. the annotations) to linked data. Linked data "holds the promise to create meaningful links between objects of disparate collections, but the actual implementation tends to be quite complex" (Hooland & Verborgh, 2014) and one of the aims with adding researcher-generated metadata to cultural heritage collections is that it can improve links between different collections, through a research lens. In theory, the exported files from Recogito could be the content that is uploaded to cultural heritage collections (as references or attached files).

METHODOLOGY

To explore what type of metadata researchers want to attribute to images, we organised a metadata hackathon to which we invited an interdisciplinary group of researchers. Our crowd of experts consisted of twelve researchers from nine different disciplines who we asked to supply metadata to five images representing the materials typically found in digitised cultural heritage collections. We invited them to perform this task during a three-hour long metadata hackathon where they were asked to add metadata to the images in the annotation platform Recogito.

The methodology was chosen through an iterative process where we, within the research group, discussed and tried different platforms, images and processes to discern what would be the best way to answer our research question: what type of metadata does researchers want to attach to images in cultural heritage collections?

Methodological exploration phase: images, platform, process

To explore the potential of harvesting researcher-generated metadata of images, we did a test run at a higher seminar at the art history division of the Department of Culture and Aesthetics at Stockholm University. We asked the attendees, mainly art historians, to take a few minutes to write down on post-its whatever they thought was interesting in an image displayed on the screen.[1] The image was choose randomly, by clicking through public domain images at Wikimedia Commons. The information written on the post-its are for all intents and purposes to be viewed as researcher-generated metadata to this image. The results showed that the attendees had very different answers to what they found interesting in the image[2], despite them belonging to the same academic discipline.

Based on this result, our research project work group discussed which images would be interesting to use in a similar exercise, if the purpose was to explore the type of metadata researchers want to add to images in special collections of research libraries. The delimitation from *cultural heritage collections* to *special collections of research libraries* were made since the case study was a collaboration between Stockholm University and Uppsala University Library. Initially, we wanted to focus on images in printed materials (such as newspapers and book illustrations) as this is an under-researched part of imagery in special collections. In addition to this, printed materials often make out the lion's share of special collections in research libraries.

Test hackathon

At this stage our working idea was to arrange a metadata hackathon with researchers where we could collect the aforementioned post-its digitally, in a more structured manner. The next step in the process was therefore to try out a digital tool or platform for harvesting the metadata. Based on the open science gap of tool use mentioned by Mosconi (2019) in the introduction, as well as practical experience from teaching digital tools to researchers in a university library setting, we wanted to use a simple-to-use platform for our metadata harvesting so that the focus would be on collecting metadata rather than navigating through new software.

The platform we first tried out was Recogito, which is an open source online platform for collaborative document annotation (*Recogito in 10 Minutes*, n.d.) that allows for two types of document uploads: texts and images. The platform is still under development and its purpose is to provide "[a platform] for digital annotation that any researcher can use." (Simon et al., 2017)

The Recogito project derives from the Pelagios initiative, whose main focus is linking online resources documenting the past with places (place names and coordinates). Therefore, geographic annotation is well developed in the platform, with automatic linking to several gazetteers[3], and

---

[1]The image was Caspar David Friedrich's painting *Klosterruine Eldena bei Greifswald*, 1825.

[2]E.g. *ruin, oil painting, time layers, secular – religious, crackling, building, gothic, how did the house end up in the ruin?, where in the world is it?, odd depth.*

[3]A gazetteer is a geographical dictionary (Aurousseau, 1945, p. 61).

Recogito provides two different modes: text annotation or image annotation. The creators of Recogito describes the image annotation interface as such:

It provides a zoom- and pan-able view for navigating high resolution images, along with drawing tools for marking points and regions and attaching annotations. Since the interface has been designed specifically with digitized maps in mind, the view allows the user to rotate the image freely. There is also a unique drawing tool for selecting a tilted box (which retains orientation information, i.e., which side is "up" and which "down"), specifically for the purpose of annotating and transcribing place names. (Simon et al., 2017)

Because the image annotation interface was specifically designed for maps, we expected this might cause some issues we would have to work around when it came to regular images. These issues would be best explored by testing the platform.

Since we were working with the scope *printed materials in special collections of research libraries* we decided to use two issues of 19th century fashion magazines for our experiment; printed media with a a lot of images. We uploaded two full issues to Recogito as two separate documents.

When we tested the platform we simultaneously meta-analysed our work from a perspective of usability, usefulness and compatibility: Could Recogito be used for tagging images despite it not being its intended function? How intuitive was it to use? Were there any obstacles to using the platform?

Our conclusions were that we *could* use the platform for the metadata hackathon, though we would have to have a *smaller set of images* as it was not ideal to work with more than approximately ten images in the interface. Despite some limitations and ambiguity in the tools and functions (for example, we discussed whether we needed to define beforehand how the participants should use the *tags* and *comments* in Recogito, or if we should leave it up to them to negotiate this amongst or by themselves) we thought the simplicity of the tool outweighed these aspects. We therefore decided on using Recogito in our metadata hackathon.

The Hackathons

The benefit of using the crowdsourcing method at a hackathon is that the event format allows for a continuous semi-structured discussion with the participants about the metadata collection process. The power relations within the hackathon is close to what is defined by Hansson et al. (2017) as crowd deliberation, which is "a more consultative mode of participation, [where] workers are viewed as experts and production is a way to get in tune with public views and values, garner good ideas, and develop consensus through deliberative dialogues. The worker has a communication channel to the other workers, they share a public; be it a newspaper, a mailing list or similar forum that makes communication with the other workers possible."

Previous experience from organising cultural heritage hackathons indicates that the setting helps de-dramatise the unfamiliar, whether it is using a new platform, identifying concepts like *data* and *metadata* (which take on new definitions when removed from one's own knowledge domain) or just doing something you haven't done before. This is because the setting allows for a continuous conversation where peer-to-peer learning and sharing of one's own knowledge and experiences are encouraged, appreciated and happen simultaneously as the participants work.

Rockwell writes that a successful crowdsourcing activitity has "lots of little tasks that carry little risk, are relatively quick and easy to complete, and do not call for lots of collaboration between participants" (2012, p. 147). To us, this meant that we needed to formulate a clearly defined task for the participants in preparation for the hackathon, as well as clear goals and intended outcomes. The goals do not need to be measurable; the cultural heritage hackathons previously organised by Andersdotter had the intended goals of *teaching data literacy*, *improving research material*, and *communicating the wealth and opportunities of the library's cultural heritage collections*.

In our invitations we formulated the task/*goal as: "discover how researchers' competence and results can be used/reused in cultural heritage collections"*, "explore how digital platforms can be used to harvest qualified metadata from researchers", and "as a researcher you will benefit from

getting an introduction to a digital platform which might be useful to your own research in the future".

We also needed to select dates, participants, images, and to think through all technical things properly so that the limited time of the hackathon would not be taken up by tech support.

Participants
To use the expert-based crowdsourcing method we needed experts as participants. Within the context of the research project, we defined an expert as a researcher who works, or could work, with images from cultural heritage collections. We wanted the group of participants to represent different disciplines, so as to capture as many aspects as possible. We also made a geographical delimitation, and reasoned that since the hackathon was to take place at Uppsala University Library, participants should ideally not have more than a few hours travel time to get there. Based on these criteria we made a list of 25 researchers known from our personal networks.

We initially planned for one hackathon but due to availability of participants we decided on running two hackathons. In the planning phase, participants were (more or less) equally spread on the two events but due to unforeseen circumstances we ended up having four participants at the first hackathon and eight participants at the second hackathon. Below is a list of all participants, their discipline, and which of the hackathons they attended.

List of participants

Alias	Event
Cesar	1
Doris	1
Emir	1
Fanny	1
Ivar	2
Jenny	2
Khaled	2
Laila	2
Matti	2
Nora	2
Olof	2
Pippi	2

Images
Five images were used as working material in our experiment hackathon. They are all printed images, typically found in library collection. They represent different periods (1698-1985) and different printing methods such as etching, xylography, photography and half tone prints in colour and black and white reproducing drawings, photographs and paintings.

Moreover, these images were chosen because they display a variety of genres and display a diversity in iconography. The material includes an etching from 1698 depicting L'Académie des Sciences et des Beaux-Arts by Sébastien Leclerc which include a plentitude of tools and machines, dress, architectural elements and personage. In addition, there are pages from three different journals. First the 19th century illustrated *FamiljeJournal* (no 20, 1882, p. 20) with a visual collage depicting the area around the Gabon River in Gabon. Also here there are a great variety of elements in the picture, landscape, objects, humans, and events are depicted. The second printed page is taken from the daily journal *Svenska Dagbladet (*10/3 1957, p. 16.) and is a typical editorial page which include a great variety of images in the same page. Here is for example drawings in adverts, reproductions of art works, photographs, satires and visual vignettes. The third journal page is from the magazine *Vecko Revyn* (nr 1 1985, s. 32-22) targeting a young female audience. It includes colour fashion photographs which include human, dress and different outdoor milieus. Finally, a photograph is included which is a colour photograph taken 1939-40 depicting two men having a

meal in a city landscape. The existent metadata on this image is scarce yet. It's title is 'Luccheon in harbour' (Lunch i hamn) and it is probably depicting two scientist as the image stems from Centre for Science (Centrum för vetenskapshistoria, Kungl. Vetenskapsakademien) and the Ahlmanns collection. As the above images this photograph include a multitude of objects depicted, humans, food, furniture, and a cityscape with buildings, a harbour, streets etc.

Taken together these five images have been chosen because they are rich in details and kinds of objects depicted. In sum they are chosen both because they display the complexity of visual resources (it does not suffice to use 1-2 words to describe them) and to meet the great variety in special knowledge in the invited group of scholars.
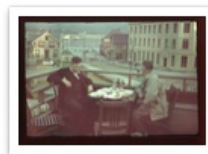
All of images were collected from digital cultural heritage platforms (images 1-3, 5) or ordered from the special collections stacks and scanned by us (image 4).

Image Selection: 1. Sida ur FamiljeJournal nr 20, 1882, s. 20.; 2. Sida ur Svenska Dagbladet 10/3 1957, s. 16.; 3. Lunch i hamn. Troligen från Ahlmanns samling. Fotograf och ägare: Centrum för vetenskapshistoria, Kungl. Vetenskapsakademien. Ca 1939-1940.; 4. Uppslag ur VeckoRevyn, nr 1 1985, s. 32-22.5; . Sébastien Leclerc: L'Académie des Sciences et des Beaux-Arts, 1698

Preparations

In preparation of the hackathon we upload the images to Recogito. The images were numbered 1-5 and they had no metadata attached to them other than the image itself and the filename (e.g. *1.tif*). Our reasoning behind this was to allow for the images to speak for themselves when analysed and tagged by the researchers; the purpose for our experiment was not to identify the origins of the pictures but to find out what the researchers found interesting in the images *as images*.

We employ the type of crowdsourcing that can be referred to as "expert-sourcing" - Bekker and Felus define this as "Expert-sourcing uses methods similar to crowdsourcing except that it collects information from experts and professionals." (Bekker & Felus, 2019)

Bekker and Felus also reasoned about the interface used for the crowdsourcing: "The user interface should be simple and intuitive but still meet all the functional requirements. It should be one that requires no training or a user manual." (Bekker & Felus, 2019) We considered similar things and therefore decided on using Recogito after a test hackathon with our project group.

Technical considerations

The software solution being used had to be compatible with all attending users, which takes into account considerations about usability, compatibility and ethics. Usability is how easily users can navigate with the software (this part is closely related to the task you want to perform), compatibility is that it can be run on the users' computers (or the computers you provide for them) without too much effort, and ethics concerns choices regarding how the data is stored, whether it is an open source tool, and which agreement you – in relation the two previously mentioned points – want to make with the participants.

After deciding on the platform, we decided to use a private instance of Recogito. The main reason for doing so was that the public instance of Recogito had an upload limit of 200 MB whereas the local instance allowed for 20 GB of data to be uploaded. Because we were using high resolution images, the total size of our project (five images) was 472.66 MB.

Event timeline and format

The total running time of the hackathon was three hours, which we divided in five different sections. We started the event with a 60 minutes introduction to the research project as a whole and an introduction to the basic functions in Recogito. The participants also had the opportunity to introduce themselves to one another (many of them had not met before). After the introduction the participants worked for 45 minutes in Recogito. Then followed a 30 minutes coffee break where we had a semin-structured discussion with the participants (questions: *What are your impressions so far? Which issues have you run into? Do you have any general questions or reflections you wish to discuss with us or the other participants?*) Thereafter followed 15 minutes to continue working in Recogito (perhaps with new insights/considerations from the discussion) and we concluded the hackathon with another 30 minutes of discussion and reflections.

Exporting the collected metadata

We downloaded the data from each participant in Recogito in CSV format and then manually merged together the CSV content from all participants in one spreadsheet. From the CSV export we extracted comments and tags to two different sheets. In the comment sheet we put the columns *alias*, *image filename, comment text* and a *y/n indicator of if the annotation also had tags atatched to it*. In the tags sheet we put the columns *alias, user, tags, image file name* from the CSV files and added the columns *total frequency, user frequency* and *image frequency* and manually counted the amount of times each tag was used in those contexts. In both the comment and the tag sheet there was one row of data for each tag added (i.e. if *participant 1* used *tag 1* three times this makes three rows in the sheet).

We also made separate sheets for the images where each participant's tags and tag frequency were listed in two separate columns. We chose to merge *Gabon river* and *Gabon river* as they refer to the same entity/name and were assigned to two different annotations (same with *barn/Barn*, btw). We chose not to merge *instrument* and *instrument*, despite the latter clearly being a mistyping of the first word, since they are two different words (how something is typed can have a big impact in search and retrieve systems).

RESULTS

In total, the participants made 282 annotations, consisting of 704 tags (503 unique tags) and 127 comments. All participants used the comment function at least once; eight participants made 1-10 comments, four participants made 11 or more comments. One user only used the comment function, opting out of the tag function completely.

Annotation Data Input Type
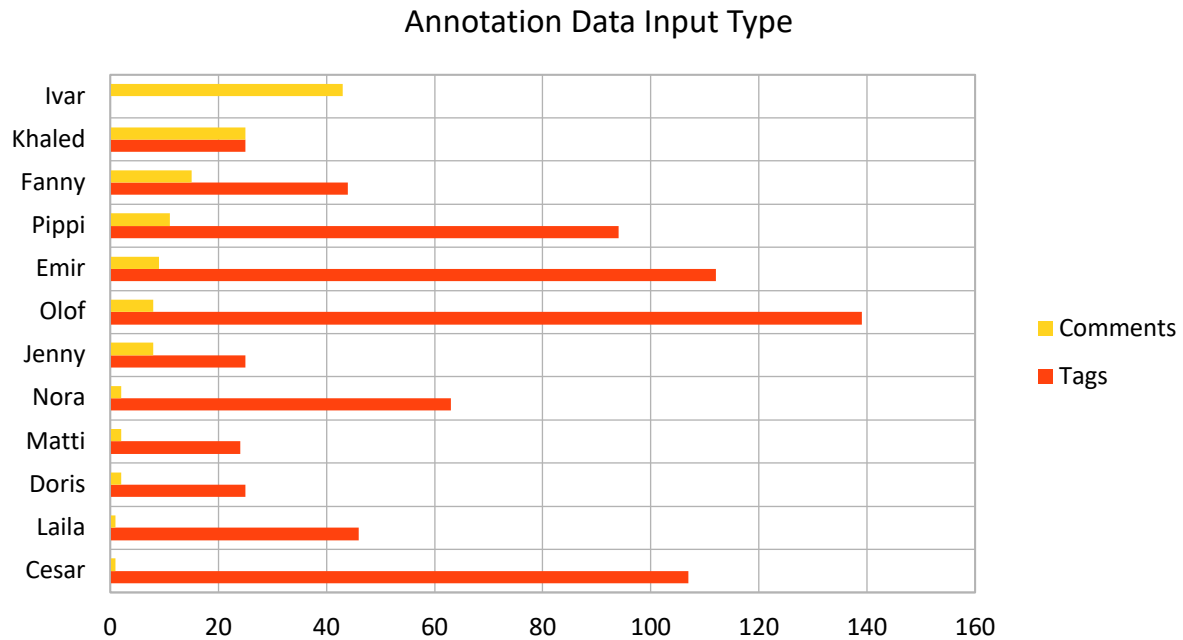
The participants seems to want to use *comment* for different purposes, e.g. to transcribe textual elements from the images, to subcategorise tags, to write explanatory comments (from which tags potentially could be extracted), to comment the image as "a whole" (e.g. "page from newspaper", "photograph of docklands"), to note questions or uncertainties about the annotated detail. Excluding the comments of the participant who only used the comment function, 6 of 82 annotations are comments without any tag; 3 of the 6 comments are descriptions of the page as "a whole".

The collected metadata display a great diversity in how different concepts are worded. The standardised vocabularies used by information professionals are not used by occasional metadata providers. Standard ways of writing dates or time periods, the use of singular or plural forms, or choice of language seems to be subordinated to the participants. Contrary to our hypothesis that the participants would discuss and negotiate how to write the annotations, the only one of these issues that were discussed by the participants during the hackathon was which language to use, and that discussion was raised because of the languages on their screens (three of the images had text in Swedish, the platform has text in English) and not because they wanted conform their annotation methods.

The annotation principles participants *did* discuss were whether they should only use terms of the time period when the image originated, modern terms, or both? An example is the use of the tag *manspread* on image 3:

Illustration 1: Was a manspread a manspread before the concept existed? And is it relevant?

This question does not have a right or wrong answer, it's rather a question of how to approach historical images and how to contextualise them for an unknown user. How do we decide terminology, and how do we signal to the users what we decided on? Concerning geographical entities, this is an issue that has been partially solved by gazetteers; geographical dictionaries created for a specific purpose.

In Recogito, the platform we used, several gazetteers are available, e.g. *Digital Atlas of the Roman Empire*, *Map of Early Modern London*, *GeoNames.* The first lists placenames in the forms used in sources from the Roman Empire (e.g. *Lutetia* for *Paris*), while the London gazetteer lists name forms used in early modern history (Jenstad & McLean-Fiander, 2016), and GeoNames lists contemporary name forms. It's easy for a user to filter and switch between the different gazetteers, changing "world view" as we go, and a question is whether similar time (or discipline) layers could be employed for navigating through cultural heritage collections? Or is going beyond proper names – which by nature are very precise – an impossible task to accomplish?

DISCUSSION POINT
• Should we expert source *the metadata* or *the new folksonomical metadata standards*?
• Is there a point in having researchers supply the metadata and what do they then need beforehand (in terms of data literacy, critical thinking around data creation and data sustainability). What parts of the dataset can be used to make collections accessible for a random user and what is just useful for other researchers?
• Would it be enough to have a possibility to upload/link csv/rdf files etc to a cultural heritage object? Or does it need to be incorporated into search tools? I.e. do you search for an object first, and then its metadata, or are these pieces of information "equal"?
• Which tagging should be done by collections staff and which should be crowd/expert sourced? Again, we're moving in the area between crowdsourcing user and metadata expert – what room is there for a middle ground between the two extremes? The researcher-generated metadata represent a certain expertise, though it is not articulated in the type of metadata or cataloguing standard that professionals in museums, archives, and libraries are used to. For cultural heritage institutions that have a core mission to support research, it is relevant to think about how researcher-generated metadata can be incorporated in the (objects of the) collections.

REFERENCES

Aitamurto, T., & Chen, K. (2017). The value of crowdsourcing in public policymaking: Epistemic, democratic and economic value. *The Theory and Practice of Legislation*, *5*(1), 55–72. https://doi.org/10.1080/20508840.2017.1282665

Alcalá Ponce de León, M. (2015). Crowdsourcing in the memory institutions: Mass transcriptions*. *BiD: Textos Universitaris de Biblioteconomia i Documentació*. https://doi.org/10.1344/BiD2015.35.26

Aurousseau, M. (1945). On Lists of Words and Lists of Names. *The Geographical Journal*, *105*(1/2), 61–67. JSTOR. https://doi.org/10.2307/1789547

Bekker, I., & Felus, Y. (2019). QUALITY CONTROL FOR CROWDSOURCING LARGE SCALE TOPOGRAPHIC MAPS. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *XLII-2-W13*, 1201–1205. https://doi.org/10.5194/isprs-archives-XLII-2-W13-1201-2019

Briet, S. (2006). *What is documentation?* (R. E. Day, L. Martinet, & H. G. B. Anghelescu, Trans.). Scarecrow Press. http://ella.slis.indiana.edu/~roday/what%20is%20documentation.pdf

Causer, T., & Terras, M. (2014). "Many hands make light work. Many hands together make merry work": Transcribe Bentham and crowdsourcing manuscript collections. In *Crowdsourcing our Cultural Heritage* (pp. 57–88). Ashgate. http://www.ashgate.com/isbn/9781472410221

Dahlgren, A. (2009). Tankar om tillgänglighet och fotografier i arkiv. In A. Näslund Dahlgren, P. Snickars, & Kungliga Biblioteket (Sweden) (Eds.), *I bildarkivet: Om fotografi och digitaliseringens effekter*. Kungl. Biblioteket.

di Lenardo, I., Seguin, B., & Kaplan, F. (2016). Visual Patterns Discovery in Large Databases of Paintings. *Digital Humanities 2016: Conference Abstracts*, 169–172. https://dh2016.adho.org/abstracts/static/data/584.xml#osds

Dionísio, M., Fraternali, P., Martinenghi, D., Pasini, C., Tagliasacchi, M., Zagorac, S., Harloff, E., Micheel, I., & Novak, J. (2013). *Building social graphs from images through expert–based crowdsourcing*. https://doi.org/10.14236/ewic/SOHUMAN2013.5

European Commission. (2005, September 30). *i2010: Digital libraries {SEC(2005) 1194} {SEC(2005) 1195} /* COM/2005/0465 final */*. https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:52005DC0465

Hansson, K., Aitamurto, T., & Ludwig, T. (2017). *From alienation to relation: Examining the modes of production in crowdsourcing*. https://doi.org/10.18420/ecscw2017-13

Holterhoff, K. (2017). From Disclaimer to Critique: Race and the Digital Image Archivist. *Digital Humanities Quarterly*, *011*(3).

Hooland, S. van, & Verborgh, R. (2014). *Linked Data for Libraries, Archives and Museums: How to Clean, Link and Publish Your Metadata*. Facet Publishing. https://ezp.sub.su.se/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=nlebk&AN=969817&site=ehost-live&scope=site

IIIF Frequently Asked Questions (FAQs)—IIIF | International Image Interoperability Framework. (n.d.). Retrieved 19 February 2020, from https://iiif.io/community/faq/

Jenstad, J., & McLean-Fiander, K. (2016). *The MoEML Gazetteer of Early Modern London* [InteractiveResource]. https://mapoflondon.uvic.ca/gazetteer_about.htm

Komssi, M., Pichlis, D., Raatikainen, M., Kindström, K., & Järvinen, J. (2015). What are Hackathons for? *IEEE Software*, *32*(5), 60–67. https://doi.org/10.1109/MS.2014.78

Manzo, C., Kaufman, G., Punjasthitkul, S., & Flanagan, M. (2015). By the People, For the People: Assessing the Value of Crowdsourced, User-Generated Metadata. *Digital Humanities Quarterly*, *009*(1).

Mosconi, G., Li, Q., Randall, D., Karasti, H., Tolmie, P., Barutzky, J., Korn, M., & Pipek, V. (2019). Three Gaps in Opening Science. *Computer Supported Cooperative Work (CSCW)*, *28*(3), 749–789. https://doi.org/10.1007/s10606-019-09354-z

Rawson, K. (2016). Curating Menus: Digesting Data for Critical Humanistic Inquiry. In J. W. White & H. Gilbert (Eds.), *Laying the Foundation* (Vol. 7, pp. 59–72). Purdue University Press; JSTOR. https://doi.org/10.2307/j.ctt163t7kq.7

Recogito in 10 Minutes. (n.d.). Retrieved 18 November 2019, from https://recogito.pelagios.org/

Rockwell, G. (2012). Crowdsourcing the Humanities: Social Research and Collaboration. In W. Mccarty & M. Deegan (Eds.), *Collaborative Research in the Digital Humanities*. Routledge. http://ebookcentral.proquest.com/lib/uu/detail.action?docID=879680

Seguin, B., Costiner, L., di Lenardo, I., & Kaplan, F. (2018). New Techniques for the Digitization of Art Historical Photographic Archives—The Case of the Cini Foundation in Venice. *Archiving Conference*, *2018*(1), 1–5. https://doi.org/10.2352/issn.2168-3204.2018.1.0.2

Simon, R., Barker, E., Isaksen, L., & CaÑamares, P. D. S. (2017). Linked Data Annotation Without the Pointy Brackets: Introducing Recogito 2. *Journal of Map & Geography Libraries*, *13*(1), 111–132. https://doi.org/10.1080/15420353.2017.1307303

Svensson, J. (2018, December 11). Samarbete för bättre digitala resurser. *Universen : Uppsala Universitet Informerar*, *2018*(4), 9.

Thornham, H., & Gómez Cruz, E. (2016). Hackathons, data and discourse: Convolutions of the data (logical). *Big Data & Society*, *3*(2), 2053951716679675. https://doi.org/10.1177/2053951716679675

Trant, J. (2009). *Tagging, Folksonomy and Art Museums: Results of steve.museum's research*. https://repository.arizona.edu/handle/10150/105627

Weisser, M. (2016). *Practical Corpus Linguistics: An Introduction to Corpus-Based Language Analysis*. John Wiley & Sons, Incorporated. http://ebookcentral.proquest.com/lib/uu/detail.action?docID=4187356

Welcome to Recogito. (n.d.). Retrieved 5 February 2020, from https://recogito.pelagios.org/