

## APIs and Researchers: The Emperor's New Clothes?

Jennifer Edmond  
Trinity College Dublin

Vicky Garnett  
Trinity College Dublin

**Abstract**

As part of the Europeana Cloud (eCloud) project, Trinity College Dublin investigated best practice in the use of web services, such as APIs, for accessing large data sets from cultural heritage collections. This research looked into the provision and use of APIs, and moreover, whether or not more customised programmatic access to datasets is what researchers want or need. In order to understand whether current patterns of API usage reflect a skills gap on the part of researchers or a mismatch of tool to purpose, we looked not only at the creators and developer/users of APIs, but also at humanists already re-using big data; approaches in cultural heritage institutions and other research infrastructures to bring API use to non-technical audiences; and the kinds of training and other support services available or emerging within the data-intensive humanities research lifecycle. We conducted both desk research and a series of 11 interviews with figures working as researchers, developers or data providers, including figures from both the API development and the data usage communities. This research, conducted under the eCloud project and supported by the European Commission's ICT Policy and Support Programme (Grant number 325091), was begun in March 2014 and is now in its concluding validation stage. The results of the research are not yet finalised, but the contribution is already emerging of this work to the debate about APIs being either the way forward for digital cultural heritage collections, or the Emperor's New Clothes (or maybe a bit of both).

*Received* 16 January 2015 | *Accepted* 10 February 2015

Correspondence should be addressed to Jennifer Edmond or Vicky Garnett, Trinity Long Room Hub, Trinity College Dublin, College Green, Dublin 2, Ireland. Email: [edmondj@tcd.ie](mailto:edmondj@tcd.ie) or [garnetv@tcd.ie](mailto:garnetv@tcd.ie)

An earlier version of this paper was presented at the 10<sup>th</sup> International Digital Curation Conference.

The *International Journal of Digital Curation* is an international journal committed to scholarly excellence and dedicated to the advancement of digital curation across a wide range of sectors. The IJDC is published by the University of Edinburgh on behalf of the Digital Curation Centre. ISSN: 1746-8256. URL: <http://www.ijdc.net/>

Copyright rests with the authors. This work is released under a Creative Commons Attribution (UK) Licence, version 2.0. For details please see <http://creativecommons.org/licenses/by/2.0/uk/>



## Introduction

Digital curation is an activity beholden to multiple, sometime competing, forces. For example, the need to preserve and migrate digital data requires different tools and strategies from the need to make data accessible to users in appropriate formats over the long term. One of the most popular tools currently in the curator's arsenal is the API, or Application Programme Interface, which holds out the promise of offering bespoke access to data while displacing the design and development costs for that access on to the users. Whether this promise is regularly realised or not, from the institutional perspective, the pressure to develop APIs seems high.

Over the past several years, the Europeana Digital Library has also been investing in the development of its API. Europeana is a European-wide resource that provides open access to cultural heritage data from museums, archives and cultural organisations across the continent. Its primary public access point has been via the Europeana search portal, but now an equal development focus is forming around its APIs – a RESTful API, basically to retrieve items from the collection, and a 'more experimental' API allowing for entire metadata sets to be retrieved in SPARQL-query language. Although Europeana does not collect information on how the API keys they issue are used or by whom, the raw usage statistics clearly indicate the value of these developments: so far, around 1,863 API-keys have been issued for access to the Europeana API, with steadily increasing numbers being requested, and nearly 5.2 million requests made to the API in August 2014 alone. Of the API keys issued, around 100 are used regularly on at least a monthly basis, according to Europeana Labs (personal communication, 30 October 2014).

Much of this development took place in the context of the Europeana Labs project, Europeana's space for developers and entrepreneurs to build and share interesting and innovative tools or apps, largely for the creative industries, for whom the high quality, easily reusable images of cultural objects aggregated by Europeana are of great value.

However, more recently Europeana has been evaluating how another key user group for Europeana, namely scholarly researchers in the Humanities and Social Sciences (HSS), might use the Europeana APIs to access Europeana's data as a 'platform, rather than a portal.' As part of the Europeana Cloud (eCloud) project, a team based at Trinity College Dublin has therefore been investigating this potential, looking into the provision and use of APIs, and whether or not more customised programmatic access to datasets is what researchers want or need.

This turned out to be a more difficult question to answer than the team expected for the simple reason that 'researchers' don't seem to use 'APIs'. This is not to say that their work was never data intensive, only that the manner in which the data was accessed was either individually negotiated or not perceived as a key part of the research process, compared to what they did with the data once accessed (e.g. analyse it on a local machine). Many research projects and digital infrastructures used by HSS researchers do, of course, have APIs at their heart: indeed the CENDARI project, which is coordinated by one of the authors of this paper, is one such example. But for Europeana, the question was very much about how researchers themselves, rather than developers building for research purposes, could be empowered through the API.

For this reason, we were required to redefine the initial terms of our investigation, and approach the problem via more easily identifiable factors that could predict or

enable usage of an API by a research community, including: common and best practices in the content and context offered by cultural data APIs; technical developer perspectives on working with cultural data APIs generally, and Europeana specifically; emergent patterns in the information seeking behaviours of humanities and social science research, in particular as perceived by library-based support staff working with digital collections; the practices of data intensive humanists and social scientists, including what skills they had/needed, and what they would want to do with data delivered by an API.

In this way, the outcomes of the research would triangulate a response to the questions of whether current patterns of API usage reflected a skills- or information-gap on the part of researchers or a mismatch of tool to purpose. We also decided to supplement our desk research with a series of 11 interviews with figures working as researchers, developers, library support staff or data providers, to ensure we were obtaining a rich account of current and potential future practices. Who is developing APIs? What do they offer, how are they used and how are they promoted?

The following examples illustrate how varied the practice of developing these tools can be, and point toward possible definitions of a ‘successful’ API implementation.

### Exemplar Providers

Of the API providers that we looked at, Trove (National Library of Australia<sup>1</sup>), the HathiTrust Digital Library (Downie and Bainbridge, 2013), the Digital Public Library of America (DPLA), and the British Library stood out as API providers who have been successful in promoting their APIs and making them available for app development. However, each has flourished in a different way. Trove provides an API<sup>2</sup> for users to enable searches within their metadata and is supported by a dedicated website with detailed information about the API, including its purpose, and examples of blogs with information on building on the Trove API<sup>3</sup>. It promotes its API through examples and experiments that have been carried out by the Trove staff in order to showcase what can be done<sup>4</sup>.

The HathiTrust Digital Library has two APIs available to potential developers. The ‘Bibliography API’<sup>5</sup> offers metadata about bibliographic, copyright and volume information. The ‘Data API’<sup>6</sup> offers images of webpages, OCR text and associated metadata. The HathiTrust Research Center allows researchers access to all public domain works within its corpus. However, it divides this public domain work into two main categories – those that have been digitised by Google, which can only be used for non-commercial scholarly activity, and those that have not, which are freely available. Customised datasets can be created in agreement with the HathiTrust.

The Digital Public Library of America (DPLA) works on a model very similar to that of Europeana. It draws in digital data and content from libraries, museums and archives from across the United States of America, and makes them available for public use via their portal and platform infrastructure. The DPLA’s main draw to many researchers and curious members of the public is the vast array of apps associated with the resource. The Apps Library on the website showcases the various apps and widgets

---

1 Trove: <http://trove.nla.gov.au/>

2 Trove API Overview: <http://help.nla.gov.au/trove/building-with-trove/api>

3 Building with Trove – Examples: <http://help.nla.gov.au/trove/building-with-trove/examples>

4 Trove Experiments: <http://help.nla.gov.au/trove/building-with-trove/experiments>

5 HathiTrust Bibliographic API: [http://www.hathitrust.org/bib\\_api](http://www.hathitrust.org/bib_api)

6 HathiTrust Data API: [http://www.hathitrust.org/data\\_api](http://www.hathitrust.org/data_api)

that have been built using the DPLA API, and that can be used to access content within the portal. These apps can be used for academic purposes (for example the Library Observatory<sup>7</sup> app from Harvard) or simply to amuse (such as the ‘Historical Cats’ Twitterbot<sup>8</sup>).

The British Library (McGregor, 2013) has taken a different route to opening up its data for reuse. Initially, access to digital collections was available via a public access API. However, in the past year they have removed this API, and offered access via different means. First, they provided access to more than a million digital images from their digitised collections via Flickr Commons<sup>9</sup>, and have also begun reviewing the way in which they provide access to metadata. The British Library Labs project (BL Labs) has been looking into how to make their collections available online in a format that will suit most researchers. In doing so, they have come to the conclusion that a ‘one-size-fits-all’ API is not fit to do the trick. Instead, they encourage researchers to work with them to build a purpose-built data retrieval method. By their own admission, these purpose built tools may be short lived, but they don’t build them to do anything other than the job for which they are intended. The analogy BL Labs uses is building a footbridge, rather than a suspension bridge to cross a stream.

It should also be noted that these are not the only examples of creative promotion or implementation of APIs – indeed, the variety of contextualising and support approaches we discovered was quite astounding. These approaches range from the practical, such as the Victoria and Albert Museum’s one-page layperson’s guide to their API, to the exemplary use of social, media, blogging and videos by EDINA’s AddressingHistory project. But the most striking method would have to be the role we referred to as the ‘data evangelist,’ a person whose own projects speak volumes for the potential of the API. Australia’s Trove has made good use of their ‘poacher turned game-keeper,’ Tim Sherratt, but HathiTrust has also been very well served by the work of Ted Underwood. In many cases, however, these relationships involve working together on data sets in a way that could not be facilitated only through an API, such as the work of John Bradley (Bradley, 2008) at King’s College London and the Pliny project<sup>10</sup>, or indeed the current approach taken by the BL Labs team.

## Researcher Behaviours with Regards to APIs and Data Generally

### Research Methodology and Process

As discussed in the introduction to this report, it was initially not clear where we should draw the boundaries of a research project investigating API usage by researchers in the humanities and social sciences, as API usage is simply not a characteristic that researchers use when describing their projects, even in the (already relatively rare) occasions when it does occur. There are plenty of researchers using cultural data (Cohen, 2006; Terras, 2009) that they could have obtained through an API. However, most of the researchers we were able to identify really only cared about the data, and

7 Library Observatory: <http://dp.la/apps/1>

8 Historical Cats Twitterbot: <http://dp.la/apps/20>

9 Flickr Collections – British Library: <https://www.flickr.com/people/britishlibrary/>

10 Pliny: <http://pliny.ech.kcl.ac.uk/>

had no specific opinions about how that data was accessed and no particular need for some of the special functionality an API could offer, like continuous updates or read the information. The term API seems therefore to be *a priori* restricted to the use of and by developers (Gibbs, 2011). We therefore decided to pursue a series of research questions that we considered to be slightly ‘upstream’ from the central concern of API use itself, in order to capture not just API usage happening below the visible level, but also behaviours and requirements that would predict the likelihood of API use, should an ideal tool and environment be available. We chose to complement this direct contact with researchers and the results of research projects with perspectives from developers and library/infrastructure professionals. Through this approach, we hoped to be able to gain a full picture of what an ideal environment for API use by researchers might look like.

Our initial desk research indicated that very little published literature would be available to contribute directly. We used what we did find as a baseline, incorporating it with the results of the eCloud Project Expert Fora<sup>11</sup>. To elicit specific information, we conducted a series of 11 full interviews supplemented by one additional email correspondence with a key individual not available for interview (indicated in Table 1 below with ‘\*’). The list of interviewees was always intended to be representative, rather than comprehensive, across the profiles we had identified. The interviewees were first selected, incorporating people that we had previously worked with, or had particular knowledge of APIs. We were then able to put these people into three main categories:

**Table 1.** Profile and coding of interviewees.

Profile of Interviewee	Number of People Interviewed	Interviewee References
Data intensive Humanist and Social Science researchers	4	R1, R2, R3 (Humanities) and R4* (Social Science)
Developers using Humanities data, including ‘data evangelists’ and participants in API-developed hackathons	4	D1, D2, D3 and D4
Librarians, infrastructure specialists and curators involved in supporting or promoting APIs	4	L1, L2, L3 and L4

We structured each interview according to the interviewee’s background and experience of APIs, data-centric humanities research or other web services to date. In the case of API Providers, we would discuss their API initially, and how they viewed users. We would then discuss if and how they participated in the research being carried out using the API, and if they knew of any projects that had used it recently. For Humanists using APIs, we would ask them about their topic, and how the API or web service had helped them to investigate questions that wouldn’t be possible otherwise. If

<sup>11</sup> eCloud Expert Fora are a series of four events within the eCloud project that bring together researchers from the Humanities and Social Sciences to discuss tools and content that might be useful for Europeana. Three of the four fora were conducted in 2013, and we were able to draw on the resulting discussion papers. The final forum is due to take place in mid-2015.

they had used a particular API or web service, or had a relationship with a particular project or provider, we asked them the strengths and weaknesses of that service.

Over the course of the interviews, three main topics emerged as key drivers of data reuse across APIs or indeed across any services or sources. These topics were as follows:

- **Data:** What researchers want it to be and what they want to do with it;
- **Technical Expertise:** How to develop it or get access to it;
- **Environments:** Social and technical preconditions to data reuse.

Each of these issues will be discussed in turn below.

### Data

Researchers want data, and several researchers commented that what they really want from a digital data service is that it should ‘just give them everything.’ Completeness of data would therefore be a key requirement, which has implications both for what an API should encompass, but also places a high bar for the quality of the metadata underpinning any filtering that occurs. The consistency and quality of metadata can therefore make or break a project’s utility.

This is a persistent challenge for data providers. D2, who is based in a data-intensive research infrastructure, also describes his institution’s frustrations at setting up reliable metadata:

‘We have a metadata task force which is creating tools to benchmark quality. It’s a continuous process that requires some effort. But on the other hand if you have a unified easy-to-use tool, people get confronted with the poor quality of the metadata.’

This might occur as inaccurate or incomplete metadata, or even as a broken link. But a project can also attempt to achieve too high a standard with metadata as well: L1 at a major university library reports on an occasion when they investigated crowdsourcing as a means to enhance their data. This was less than successful because:

‘We were too librarian about it, and tried to say that for things to be fully described, participants had to fill in 35 fields, which isn’t really what Zooniverse users go in for!’

The project ended up having to re-evaluate its metadata fields and reduce them in number in order to make it more user-friendly, but according to L1 it was still too many.

It must be remembered as well that reuse of data via an API is a subsidiary question to the larger issue in the research community of the reuse of data. This is a known problem in the digital humanities ecosystem, one which funding agencies have approached via reuse programmes such as DeDeFi<sup>12</sup> in the UK and the international consortium behind the ‘Digging into Data’ challenges. But in spite of the investment, the problem of data silos and under-utilised resources remains. Building an API therefore doesn’t guarantee use of the API or re-use of the data, nor does it fix a bad platform or bad metadata, or implement its own solutions (McGraw, 2014). R2, for example, admits his

---

<sup>12</sup> Digital Equipment and Database Enhancement for Impact (DeDeFi): <http://www.ahrc.ac.uk/Funding-Opportunities/Pages/Digital-Equipment-and-Database-Enhancement-for-Impact.aspx>

pessimism around this, having seen many projects invest in a great resource, only to see it used almost exclusively by the person who built it. He suggests that the research question is the thing that fuels the investigation, and it is rarely the other way around.

### **Technical Expertise**

The question of whether social science and humanities researchers need to code seems to be a polarising one. Many of the humanities and social science researchers we spoke to were very adept at using tools like Python (R1), R (R2) or Perl (R4) to write software code. But despite this, these researchers took their datasets, usually acquired through a download or data transfer from a trusted human partner, and structured and manipulated them offline. Some of those working on the software development side often consider a researcher with a little knowledge to potentially be a dangerous thing. Researchers with coding skills generally take the opposite perspective: for example R2 tells us:

‘I’ve come to the belief over the last 20 years that we really have to get over the idea that programming is a foreign occupation for Humanists. It’s ridiculous that we discuss if someone needs to be a programmer or not.’

Somewhere in the middle would be Fred Gibbs, Assistant Professor of History at the University of New Mexico, who commented in the context of a workshop on APIs: ‘In terms of bridging the humanist/technology gap, isn’t it easier to slide the APIs a bit closer to the humanists than the other way around?’ (Gibbs, 2011).

But these opposing positions do not assist us with the very practical question of what skills and level of ability a humanist or social scientist might need to allow them to use coding as a part of their methodology, to experiment enough to understand the implications of a resource like an API for their work? L2 describes an example of a Bioinformatics researcher who was already reasonably familiar with coding, and taught himself how to use an API to create his own workflow. Equally, D1 relays how, three years ago, two very traditional humanists of his acquaintance took it upon themselves to learn how to use digital tools for network analysis and transcription purposes. They are now highly adept at using these digital tools. These examples show what can be achieved with a little support and a compelling research application, but as D1 says, a positive attitude towards learning a new skill is required.

### **Environments**

How data is created, stored and made accessible constitutes an environment which, viewed at a macro scale, includes not just technical elements, but human and infrastructural/institutional elements as well. What are the optimal characteristics of the environment that surrounds an API to be supportive to research or scholarly activity? L2 has studied workflow, and identified the point at which software and web services (in this case, the archival tool, Dataverse<sup>13</sup>) becomes most useful. He describes an example where he and his team at the Library documented the workflow of PhD students within the humanities and social sciences while they conducted work on an informal archival project. This provided not only a dataset of the collection created by the students, but also of workflows and practices they undertook to reach completion of the project.

As a Humanities librarian, L1 believes that:

---

<sup>13</sup> Dataverse: <http://thedata.org/>

‘Libraries shouldn’t be the gatekeeper. We shouldn’t be making judgement calls, and we shouldn’t be doing the work for people. These are the methods and workflows for the research – we can’t be making decisions as to how the data should be used.’

But as the nature of the data the researcher is using and creating changes, one would expect that the role of the library would change as well. L2 feels that this could involve a greater support by libraries in the future for methodologies, whatever they are. But in terms of having the skills or human resources to provide that, we are still very far from this potential future.

The rise of new model digital infrastructures alongside the old ones offers different options for addressing the environmental issues of data reuse and API uptake. According to D2, the work at a major research infrastructure suggests that more remote means might be the answer. This research infrastructure provides access to data via different web servers from centres across its network and online web services available via the infrastructure’s site that allow users to build and tailor their own tool chains for data collection and reuse. These tools make use of tools in a customisable ‘chain’, which can be deployed from the infrastructure’s graphical interface, providing a visual way of creating tools to access data. R2, however, is not such a fan of ‘Humanist-friendly interfaces’, as he believes that ultimately they don’t work. He feels instead that a Humanist needs to understand the modules in systems in order to properly get to grips with the data. A Humanist needs coding expertise, and knowledge of some tools to be able to understand this. Graphical interfaces are, he says, good for smaller jobs, but will not be able to handle big data as the builders of such interfaces can’t predict all the different ways in which to manipulate data; ‘the tweaking is where the innovation comes in.’ D1 also expresses a hesitation about tools that do too much, or operate at the wrong level. Certain tools, able to batch process otherwise time-intensive or tedious processes, could certainly bring value, but he, like R2, does not necessarily feel that most tools gave more than they took away.

## Key Learnings for Developing APIs in CHIs

### Key Learnings Regarding Data, Metadata and Content

For a repository, platform or portal to be successful, the quality of the content must be paramount. As we’ve seen, the promise of content can lead to frustration when actual content isn’t available. This has become an issue when a researcher is trying to use a repository to access anything other than metadata, but also when results don’t match user expectations.

We have also seen how other organisations have responded to this need for concise and clear metadata. Crowdsourcing has been used as an approach to enriching metadata, but in order to ensure the data is trustworthy, clearly this must be deployed carefully, both in terms of facilitating crowd contribution and of making its provenance clear. One approach could be to recruit participants to enhance/correct metadata and then ensure that crowdsourced and official data does not become mingled and confused. The first of these can be managed via productive user interaction and the second by clever UI design.

## Key Learnings Regarding Technical Skills and Support for Researchers

Many of the sites we reviewed carried information about their API, with helpful notes as to how to use them. The ease with which the API can be used depends on the clarity of the documentation. To a complete novice embarking on their first forays into API use, this documentation can still be slightly baffling as they try to navigate the technical terms and envisage just what it is the API is going to do for them, and what the output will look like. Many API providers complement their documentation with tutorials with clear technical prerequisites and pointers toward other sources of technical training, and through hosting periodic training ‘workshops’ (either online or in person) for those keen to learn new digital techniques alongside their content. CHIs will need to inspire people to work with their data and give them the confidence that they will achieve results. They may achieve this by making sure that the platform features a few actual worked use cases, showing both how the research question was developed and the results delivered (and by whom). The Australians seem to have mastered this, with Tim Sherratt (Sherratt, 2011) acting as ‘data evangelist’ par excellence, and tools like StripySock’s Forte (MacMullin, 2013) demonstrating the use of the data and the API. Even the DPLA’s humorous Twitter feed contributes to the overall awareness of their content.

## Key Learnings Regarding the Research Environment

There have been discussions about an online community that could also support the Europeana API. Inspiration could be taken from the Software Carpentry model, whereby experienced users of software create lessons for other community users. Admittedly this is mainly within the sciences at present, but the model itself presents an opportunity for the Europeana audience, which from the content strongly suggests Humanists and Social Scientists working in areas such as Social History, Sociology, Anthropology and Art History (among others). While the idea of a forum has not proved popular, and it would of course create the need for increased staffing resources at Europeana to continually monitor. However, there is potential within Europeana Labs or Europeana Research to allow researchers to build their own communities around collections. These communities could be user-driven and monitored, similar to how Facebook operates groups and pages.

## Tools and Chains

Obviously, there are split opinions about the place of tools in the humanist’s data workflow. In most cases, it would be of benefit for API providers to offer tools and the potential to build tool chains, but without placing too many at the top level and being sure that each provides a unique and basic service. Others that the community developed could perhaps register for use via a Wordpress-like widget directory.

But it is beyond doubt that certain tools would certainly be of value, so long as they do not get lost in a morass of applications too specialised or under developed to be an asset to the collection. In thinking about tool chains, we also did a bit of research into the free service Yahoo Pipes<sup>14</sup>, as an exemplar for how data sources and manipulation tasks might be combined for non-specialist users. As a visual tool allowing users to ‘mash up’ or aggregate web feeds and other data from various online sources, the project shows the potential for user friendly, flexible environments within which to manage data streams.

<sup>14</sup> Yahoo! Pipes: <http://pipes.yahoo.com/pipes/>

But the experience of the Pipes community points once again to the question of whether an API might be too much engineering for a humanities and social science researcher cohort. Matthew Dinmore and C. Curtis Boylls (Dinmore and Boylls, 2010) conducted a study of more than 30,000 Pipes compositions to determine end-user behaviours. Their findings suggested that users of Yahoo! Pipes only made use of a small number of the features available to them, preferring linear straight designs to the multi-branch pipelines that were possible. Dinmore and Boylls suggest that rather than looking to make an elaborate data flow that could be re-applied in future work, users tended to go for as simple a model as possible that answered the question before them. As they put it ‘...users set design-time parameters – as few as possible – to achieve an objective, again asking the question “am I done?”’ This may be the Yahoo! Pipes equivalent of the researcher impulse to download everything and then filter and analyse data as a very separate step.

### Limitations in APIs and other Web Services

As we have seen, technological know-how can have an impact on a researcher’s willingness to use APIs and other data techniques. Many researchers still prefer a less-technical, more ‘analogue’ (non-digital) approach to their research, and don’t see the need for more digital means. And this is fine. Not all research *should* be digital. However, there is increasing use of digital items in research, and indeed digital approaches open up some very interesting questions for future research.

The lack of take up of web services among humanists perhaps shows a mismatch in what they want, and what developers think they want. D1 raised this point when he was describing how few opportunities there are for researchers to get to the data. He suggests that the leap to allow API access to data often reflects the developers’ perspective to answering a question. In other words, the limitations of APIs are often set by the lack of understanding developers might have of researchers’ needs, and appreciation for the challenges their research questions pose. On the other hand, resolutely non-technical humanists may not even have the capacity to imagine what such an approach might mean for their research, much less have a mastery over the vocabulary required to describe their desired research process to a developer. In deciding how and indeed whether to create a researcher platform for Europeana, these macro-level considerations must also play a role. Rolling out an API for researchers linked to Europeana will require fundamental change not only in Europeana itself, but also in its users. If viewed as a long-term investment, such a development could be instrumental for cultural research in the digital age. If viewed only in the short term as a technical development only, our research indicates it will almost certainly struggle to find its place, becoming a fashionable, but perhaps ultimately invisible set of ‘new clothes’ for cultural data.

## Acknowledgements

The authors, along with the Europeana Cloud project, acknowledge the generous funding from the EU Commission under the Framework 7 CIP- Best Practice Network grant no. 325091 eCloud.

We also wish to thank the very kind participation of our informants in this study.

## References

- Bradley, J. (2008). Pliny: A model for digital support of scholarship. *Journal of Digital Information*, 9(1). Retrieved from <http://journals.tdl.org/jodi/article/view/209/198>
- Cohen, D.J. (2006). From Babel to knowledge: Data mining large digital collections. *D-Lib Magazine*, 12(3). doi:10.1045/march2006-cohen
- Dinmore, M.D., & Boylls, C.C. (2010). *Empirically-observed end-user programming behaviors in Yahoo! Pipes*. Paper presented at the 22nd Annual Workshop of the Psychology of Programming Interest Group. Retrieved from <http://www.ppig.org/papers/22nd-EUP-3.pdf>
- Downie, J.S., & Bainbridge, D. (2013). *Integrating independent discover and analysis tools for the HathiTrust corpus: Enhancing fair use digital scholarship*. Paper presented at the 3rd Conference of the Japanese Association for Digital Humanities (JADH2013), Kyoto, Japan.
- Gibbs, F. (2011, February 28). The UI/API divide [Web log post]. Retrieved from <http://fredgibbs.net/posts/post/data/ui-api-divide/>
- Lehane, R. (2012). *Access to online archival catalogues via web APIs*. Paper presented at the 2012 International Council of Archives Congress. Retrieved from <http://ica2012.ica.org/files/pdf/Full%20papers%20upload/ica12Final00233.pdf>
- Macdonald, S. (2011). AddressingHistory: A Web 2.0 community engagement tool and API. *IASSIST Quarterly*, 35(3), 17–20. Retrieved from <http://www.iassistdata.org/iq/addressinghistory-web-20-community-engagement-tool-and-api>
- MacMullin, J. (2013, March 25). National Library of Australia's Forte [Web log post]. Retrieved from <http://stripysock.com.au/blog/2013/3/25/national-library-of-australias-forte>
- McGraw, D. (2014, August 11). What an API can't do [Web log post]. Retrieved from <http://www.oyova.com/what-an-api-cant-do/>
- McGregor, N. (2013, February 11). Mash-ups, APIs and linked data, oh my! [Web log post]. Retrieved from British Library Digital Scholarship Blog: <http://britishlibrary.typepad.co.uk/digital-scholarship/2013/02/information-integration-mashups-apis-and-linked-data.html>
- Sherratt, T. (2011). It's all about stuff: Collections interfaces, power and people. *Journal of Digital Humanities*, 1(1). Retrieved from <http://journalofdigitalhumanities.org/1-1/its-all-about-the-stuff-by-tim-sherratt/>
- Terras, M. (2009). The potential and problems in using high performance computing in the arts and humanities: The Researching e-Science Analysis of Census Holdings (ReACH) project. *Digital Humanities Quarterly*, 3(4). Retrieved from <http://www.digitalhumanities.org/dhq/vol/3/4/000070/000070.html>