

Towards Generous Interfaces for Archival Collections

Mitchell Whitelaw

Faculty of Arts and Design, University of Canberra

mitchell.whitelaw@canberra.edu.au

Abstract

This paper considers the challenges and opportunities of access to digital archival collections. After the era of digitisation, collecting institutions hold vast digital treasure-troves. But the techniques we use to access these collections have been fundamentally unchanged for decades. Search remains the dominant access tool. This paper makes the case for the limitations of search, and proposes instead the concept of “generous interfaces” for digital collections. Generous interfaces offer rich, browsable views; provide evocative samples of primary content; and support an understanding of context and relationships. Based on experimental work to date, the paper outlines some design principles and patterns for generous archival interfaces, as well as considering challenges to their implementation.

Introduction

After more than a decade of effort in digitisation, archival collections - and cultural heritage collections more broadly - are increasingly digital collections. As the name of the National Library of Australia’s popular *Trove* service implies, digital collections are our cultural treasure-houses. We now take for granted the ability to access these collections from any web browser - and with increasing network speeds, we can only expect faster access to higher-quality resources. With digital cultural collections and network infrastructure in place, the time is right to reconsider the interfaces to our collections. Techniques for accessing and presenting digital collections have changed little in two decades; at the same time, mainstream and commercial information experiences have become richer and more diverse. Ten years ago Clifford Lynch (2002) recognised that cultural collections could play a crucial role in a broadband economy. The same potential exists today, but I would argue that if it is to be realised, we must rethink our collection interfaces.

Accessing Digital Collections

For digital collections - in archives and other institutions - interfaces attain new significance. In a digital collection, the interface is now often a user’s only point of contact with the collection - it is the manifestation of that collection, for the user. This is a significant change from the traditions of the computerised catalogue in libraries and archives - where the purpose of the digital interface is merely to connect us with the physical object we seek. Interfaces matter now more than ever, because many users, in many collections, will never encounter the physical object (if there is one); increasingly, our only experience of a collection will be digital.

In that light it is important to consider our current interfaces critically; and for current digital collections, that interfaces is almost entirely search. Search is such a familiar, powerful and ubiquitous interface method that we barely notice it. It has a deep heritage in the history of computing, and especially the field of library and information science; it is deeply ingrained in the thinking and practices of digital collections, but equally its limitations have long been documented in the scholarship around information retrieval. As Belkin, Oddy and Brooks wrote in 1982, “one can identify two assumptions basic to the best-match [search] principle: that it is possible for the user to specify precisely the information that he/she requires; and, that information needs ... are functionally equivalent to document texts.” As Taylor’s foundational work on information need explains, a search query is a compromised and imperfect expression of a feeling that arises as a

“vague dissatisfaction” (Taylor, 1962). Rather than an information need, Belkin et al (1982) propose an “anomalous state of knowledge”, and observes that “in general, the user is unable to specify precisely what is needed to resolve that anomaly.” Some twenty-five years later, little seems to have changed: writing on “grand challenges for information retrieval”, Belkin writes that “only considering specified search as the basis for IR models and techniques, is clearly inadequate, and inappropriate” (Belkin 2008).

Despite its flaws, search has been so effective that other techniques have been largely neglected - at least in practice. “Browse” is often offered as an alternative to search - yet browse-based collection interfaces are typically impractical and uninspired. By contrast, researchers in information retrieval suggest that browsing is “a rich and fundamental human information behaviour” (Chang and Rice, 1993). More recently, researchers have sought to extend or augment standard paradigms of search. Marchionini et al propose “exploratory search”, arguing that while conventional “lookup” search is effective for returning precise results to a well specified query, modern searchers are engaged in more complex, open-ended tasks: learning, investigating, analysing and synthesising information. Marchionini argues that search interfaces should therefore be more dynamic and interactive, enabling users to explore by selecting facets (or browsing), rather than requiring a specified query. Marchionini (2006) writes that “better tools to support exploratory searching are needed” - and the same is largely true today. Faceted search is one exploratory technique that has been widely adopted in digital collection interfaces; however while this is a significant improvement on “pure” search, facets are generally provided only after a traditional query. Search remains the only way in to these collections.

One limitation in theoretical models of search and information retrieval is the highly constrained, functionalist philosophy that they reflect. The task-based paradigm embedded in information retrieval is increasingly out of step with our ever-more-ubiquitous, casual and everyday experiences of information systems. Happily, alternative models are emerging. Marian Dörk’s “Information Flaneur” models a contemporary information-seeker on the urban flaneur of 1840s Paris:

Following the flaneur’s attitude toward the city, the information flaneur sees beauty and meaning in growing information spaces. By envisioning the information flaneur as a curious, creative, and critical persona, we promote a shift from negative concepts such as needs and problems towards positive information experiences (Dörk, Carpendale and Williamson, 2011).

Dörk shows that the Information Flaneur is already with us, in the diversity of our current information experiences, including exploratory browsing, leisurely curiosity, aesthetic pleasure and serendipity. He sets out useful principles for the design of information-seeking interfaces that can fluidly accommodate both high-level exploration of large information spaces, and immersion in their detail.

Generous Interfaces

In the context of this work on the limitations of search, and the need for richer, more exploratory modes of information seeking, this paper proposes a single principle or ethos that can provide a critical perspective on current and future collection interfaces. That ethos is generosity.

Generosity is normally thought of as a human attribute, a quality of interpersonal interactions. The argument here rests on the assertion that the same qualities can apply to our digital interactions with cultural collections. The qualities of generosity I am interested in here are “to be liberal in giving or sharing”; also to be “large, abundant, ample” . Both of these qualities seem well aligned with the aims and missions of cultural collections. Our digital collections are certainly large, abundant and ample; and the charters of our cultural institutions place a high value on sharing these riches liberally with the public. Generosity seems to be very much in line with the aims of our cultural collections.

Our digital collections already embody the second sense of generosity - “large, abundant, ample”. However collection interfaces often fail to be generous in the second sense - to be “liberal in giving or sharing”. In considering what generosity means in the context of collection interfaces, and through analysing features of existing interfaces and some experimental prototypes, the aim of this paper is to provide a critical framework for analysing collection interfaces, and set out a vision for the next generation of digital collection interfaces.

Current collection interfaces are dominated by search. A cursory survey of the collection interfaces of the national Archives of [Australia](#), [New Zealand](#), the [United Kingdom](#) and the [United States](#) shows that the search box is universally prominent. Generosity here provides an additional perspective to the critiques of search already discussed; it allows us to consider the interface as the site of a (crucial) relationship between visitor and collection. Viewed in this way search alone is, I would argue, an ungenerous interface. Search is ungenerous or inhospitable, in that it demands the user make the first move, and enter a query. Rather than offer information about the collection, the search box conceals or withholds it. Thus search favours expert users, for those who understand a collection’s contents and can query it effectively. It is most ungenerous, ironically, to those most in need of generosity: visitors unfamiliar with a collection.

Imagine the analogous situation at a museum or gallery. The visitor enters the building, whose collection she hasn’t encountered before. Instead of expansive exhibition halls however, she finds a small, drab lobby with an attendant at a desk. The attendant asks the visitor to write her query on a small slip of paper. The visitor invents a query, and the attendant disappears for a moment before returning with a line of artworks on trolleys, which are paraded - ten at a time - through the lobby.

While search alone may be ungenerous, the principle of generosity can be applied as a continuum, rather than a hard criterion. Contemporary collection interfaces begin to show how we can be more generous. Faceted search results - a hallmark of Marchionini’s exploratory search - provide an excellent example. Facets make search more generous by offering rich information and developing a sense of context. The UK National Archives provides results faceted by subject, date, collection and catalogue level; so after entering a search, we immediately learn more about the collection’s structure and contents. Unfortunately, again, these facets become available only after searching; so we can explore within the returned subset of records, but not beyond them.

Current interfaces also make some use of selection-based or browse techniques. Many archival collections offer something like the US Archives’ [Topics](#) page - a categorised list of subject headings, leading to relevant resources and collection subsets. The UK Archives provides a similar [Categories](#) page with a selection of online resources. Such approaches are potentially more generous again than faceted search: they convey the scope and diversity of a collection, and often present evocative previews or samples of collection contents. Yet as currently implemented they also have significant limitations. In the case of the UK National archives collection the trail of exploration often ends abruptly at a search box - specific collection subsets (such as the [Seamens’ Wills](#)) demand a search in order to examine individual records. Another major limitation is that they are often partial representations of a collection - perhaps selected or curated highlights or exhibitions, or the “most popular” holdings. In their current form these devices play a valuable and generous role in offering engaging content to a wide audience, but they are not comprehensive interfaces to digital collections. This need not be the case: for example the Art Gallery of New South Wales provides a hierarchy of [collection categories](#) that covers the entire collection. Like facets, these categories show their relative size, such that we can immediately see, for example, the relative distribution of Australian paintings, drawing and photographs. As well as quantity these categories show samples of the collection content; the combination of evocative sample and structural and quantitative information effectively orients the curious browser within the collection; we can continue exploring all the way down to individual item level.

Experiments in Generous Interfaces

Current practice in archival collections shows that while search remains dominant, glimmers of generosity are evident. What would a truly (and deliberately) generous collection interface look like? Is such a thing even possible? There are a number of examples within the cultural collections sector that strongly suggest a turn towards more generous interfaces. For example [Discovering Mildenhall's Canberra](#), a collaboration between the National Archives of Australia and the Museum of Australian Democracy at Old Parliament House, presents an immersive, exploratory view of some 7700 historic photographs. The National Library of Australia's [Australian Womens' Weekly](#) project is a similarly rich, browsable interface; and the Irish in Australia Wall, from the National Museum of Australia, shows how multiple collections can be combined into a generous mosaic of intriguing connections and contexts (Figure 1).



Figure 1. The Irish In Australia History Wall - National Museum of Australia, developed by Tim Sherratt.

Drawing on this work, as well as our own prior work in visualising digital collections (see for example Whitelaw 2009; Hinton and Whitelaw 2010), my colleagues and I have been testing the limits of generous collection interfaces in the web browser. This work is informed by the principles and approaches of data visualisation - in particular Schneiderman's influential Visual Information Seeking approach (1996). Schneiderman's "mantra": "overview first, zoom and filter, details on demand" is a powerful recipe for generous interfaces. These prototypes demonstrate the feasibility of creating rich, explorable interfaces to large collections, and also illustrate some characteristic problems, strategies and challenges in the creation of such interfaces.

The Prints and Printmaking Collection of the National Gallery of Australia contains records of some 40,000 works by more than 4,000 artists. We have developed several different, parallel interfaces to this collection, with the aim of revealing different aspects of the collection structure through multiple views.

The All Artists interface sets out to provide a rich overview of the collection, focusing on the creators, printers, publishers and artists involved. The interface shows each artist as a small

rectangular tile; a compact display that includes name, birth and activity dates, gender, and number of works in the collection. Visualisation techniques are used to represent gender and number of works: gender is represented in a colour-coded left-hand border, while number of works is reflected in the width of the tile. Thus on scanning a page of these tiles, we can immediately see the distribution of males and females (as well as companies, classified in this dataset as a “gender”). Similarly the distribution of works within the collection is clear; we can see that most artists contribute only a small number of works, while a few artists contribute many works - this “long tail” pattern is evident in the texture of the display.

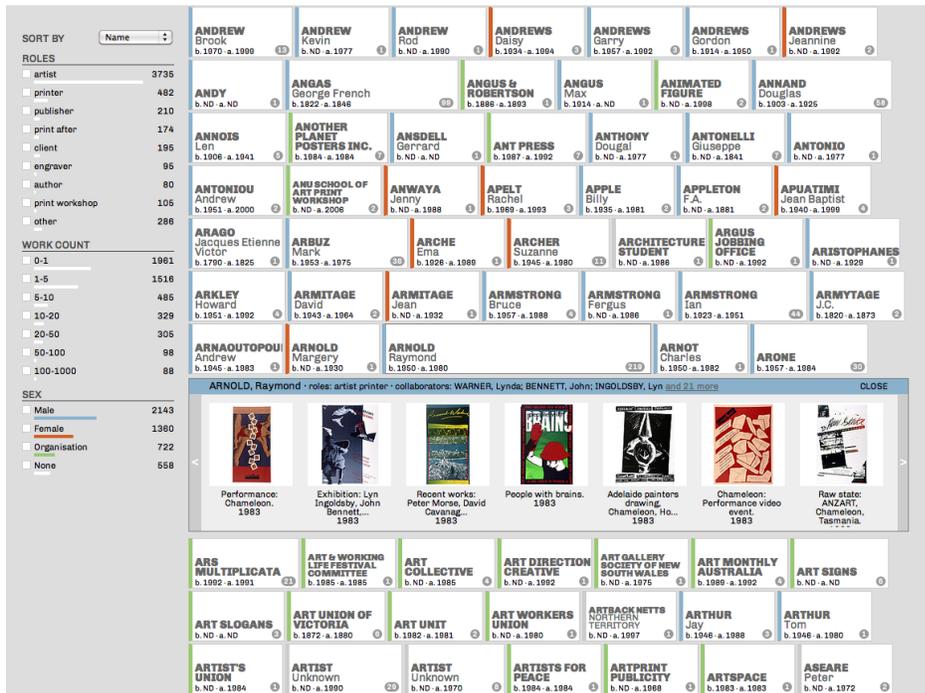


Figure 2 - All Artists interface prototype, for the Prints and Printmaking Collection of the National Gallery of Australia

While it’s possible to browse all 4,000 artists using this interface, a panel of facets or filters enables us to limit the display according to role, work count, and gender; so for example we may focus on female artists, or female printers, or only female printmakers with more than ten works in the collection. In this way the scope can be quickly narrowed from several thousand to ten or a hundred records. These facets also show distribution and relationships within the collection: selecting one facet updates the others dynamically, showing, for example, the different distributions of male and female artists for different roles.

Finally, we can also explore and investigate these artists in greater depth, within the same display. Selecting an artist opens an inset display showing their works, as well as additional data such as collaborators. We can “open” multiple artists, and compare their works; when the display is sorted chronologically, this enables us to explore the work of contemporaries or observe stylistic changes and distinctive period aesthetics. Works are displayed “in situ” - that is, maintaining the global context of the artists display. Like the infinite scroll technique described earlier, this removes a potential obstacle to exploration, allowing the user to browse the collection without opening a new page and adjusting to a new context. By maintaining the user’s exploratory context, this “in-situ” display also enriches it: the works enrich our understanding of the collection, even as the collection provides a context that frames our encounter with the works.

A second interface to the Prints and Printmaking collection focuses on works, rather than artists. The Decade Browser presents an overview of the collection organised by decade, with a horizontal bar graph showing the relative number of works in each. The graph is segmented in turn into different work types, such as intaglio, monoprint, and relief. Undated works are also represented here: experience has shown that collection metadata is often imperfect and incomplete, so generous interfaces must accommodate this imperfection, rather than ignore or exclude it. The resulting graph is informative in itself; it shows both the overall distribution of the collection and the relative distribution of different print types through time - we can see the boom in stencil printing in the 1970s and 80s, for example. It also acts as an interface; on selecting one the segments, the corresponding works are loaded into the right hand pane and displayed as small cropped thumbnails. These thumbnails provide a rich and immediate impression of the visual character of each group of works, as well as a further exploratory interface. Hovering on the thumbnails reveals full work images and details; selecting a work loads a larger image in the central pane. Thus the Decade Browser provides a single-screen interface to the entire 40,000-work collection. By displaying three levels of scale concurrently - the entire collection, the selected segment (by decade and print type), and a selected work - the interface again works to establish and maintain a contextual understanding of the collection.

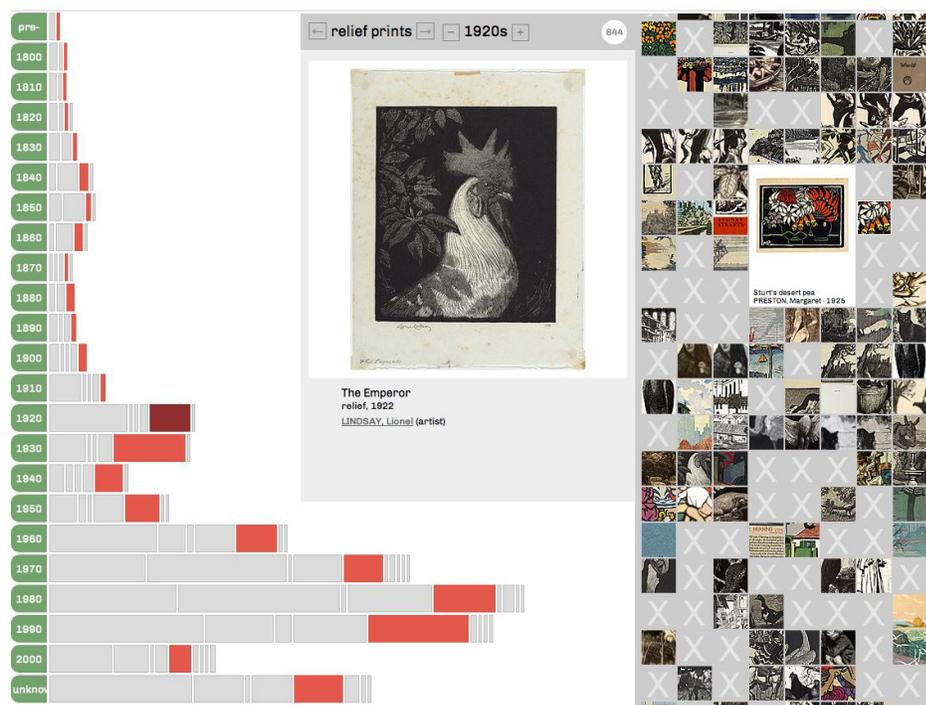


Figure 3 - Decade Browser prototype, for the Prints and Printmaking Collection of the National Gallery of Australia

A key challenge in creating rich, exploratory interfaces is compression - the need to represent large, diverse collections in a compact, browsable form. The Decade Browser uses facets based on the collection metadata to provide a segmented overview. Without such well-defined segments, creating a rich overview is a more difficult challenge. In work currently underway, we are developing an exploratory interface to the Manly City Council Historic Image collection; a set of some 7000 images with more limited metadata. In the prototype shown here, we use terms from image titles to segment the collection. These segments are displayed as “tiles” whose size reflects the number of related records - again, we use basic data visualisation principles, conveying quantitative relationships in an immediately legible form. A key feature of this prototype is its use of

image “samples”: each tile shows an image from the group of records it represents. These images change over time, cycling through each record in the group; as a result, the interface as a whole provides a rich, gradually changing impression of the entire collection. Hovering over a tile displays the currently displayed record’s title, rewarding exploration with deeper information. Selecting a tile “opens” it, as in the All Artists interface, revealing a pane enables us to scroll through all records in the group. Again, this display appears in situ, with multiple levels of scale displayed concurrently.

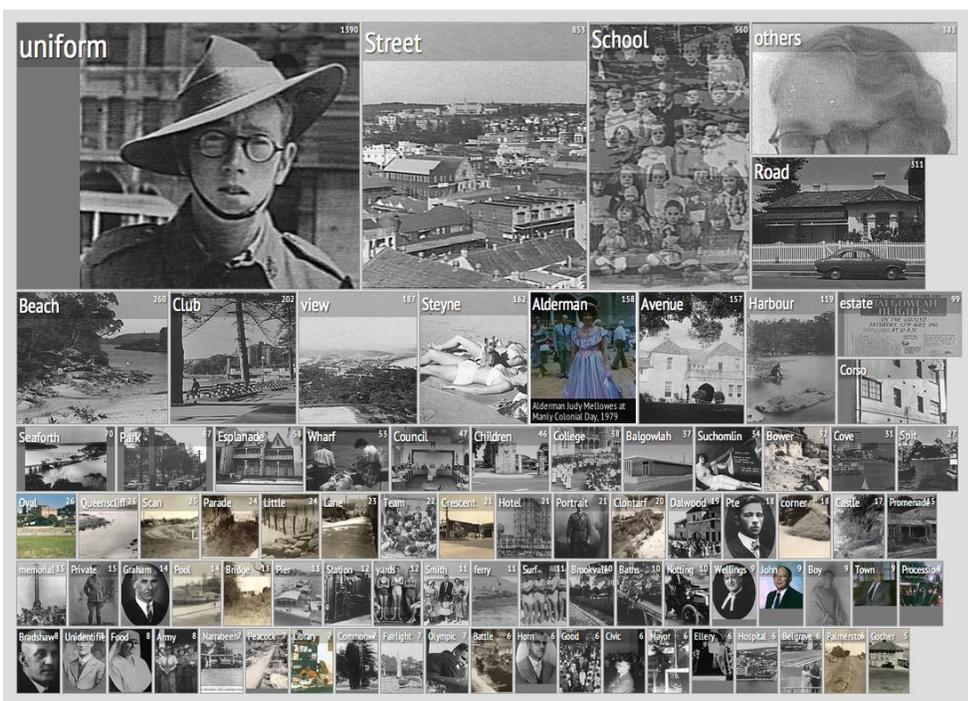


Figure 4 - Manly City Council Historic Image Collection prototype interface

Principles and Patterns

Based on the examples outlined here, we can begin to articulate some general principles for the creation of more generous interfaces for archival collections:

1. *Show first, don't ask.* Search-based interfaces demand the user enter a query, even in cases where they may not have a focused query, or may not be in a position to query the collection effectively. This is ungenerous and inhospitable. Generous interfaces should volunteer rich information that supports an audience's understanding of the collection.
2. *Provide rich overviews.* According to Schneiderman's information-seeking mantra, the "overview" is the foundation of exploratory data visualisation. Generous interfaces should provide rich, representative overviews of digital collections that characterise the collection as a whole and help orient the user's exploration.
3. *Provide samples.* Displaying an entire digital collection is rarely possible. Generous interfaces can use samples of primary content to represent and characterise a digital collection. Samples provide rich contextual cues, and invite exploration. To faithfully represent a collection samples should be exhaustive (as in the Manly images example above) or representative (eg a random sample) rather than hand-selected or curated.
4. *Provide context.* Generous interfaces should help us explore and understand digital collections by establishing and maintaining context - that is, displaying the structure and

relationships within the collection in a way that supports our interpretation of its contents. Both item-to-collection relationships and item-to-item relationships are important.

5. *Share high quality primary content.* At the core of the ethos of generosity is “sharing liberally.” A generous interface should deliver on its promise and provide seamless access to high quality primary content.

Challenges

Transforming established conventions for digital collection interfaces involves a number of challenges across multiple domains. The technical challenges are relatively easy to define: generous interfaces are more technically demanding, requiring more data and more computation on both the server and client sides. Rich, web-based display and interaction also requires modern browsers compliant with the latest standards (such as CSS3 and HTML5). The need for rich overviews involves new data requirements: in general collection databases are search-oriented, providing batches of results in response to a query. New database and API designs may be required that can provide data in suitable forms for these interfaces.

For interface and web designers, generous interfaces pose interesting new challenges. In our work to date we draw on conventions from both web design and information visualisation - the result is a hybrid form whose design language is not yet known. Implementing the principles outlined here - for example, providing a rich, representative overview of a large collection - is a design challenge in itself. In our work to date we are beginning to develop reusable strategies, but these must be adapted to suit the characteristics of specific collections.

These interfaces also pose potential challenges for collection users. They may be unfamiliar; though with careful design they should not be intimidating. At some point the increased data-density of such interfaces may be “too much” for some users - we don’t yet know where that point lies. Having said that, we propose that these interfaces should complement, rather than replace, existing approaches; they need not eliminate search, for there will always be instances where search is a desirable solution. Instead, the aim here is to do what search cannot - provide rich, exploratory interfaces to digital collections.

Conclusion

In the context of a rapidly changing information landscape, and the recognised limitations of the search paradigm for information retrieval, this paper proposes an ethos of generosity - being “liberal in giving or sharing” - as a guiding principle for the design of digital collection interfaces. Drawing on established literature in information retrieval, and applying the critical lens of generosity itself, I have argued that search alone provides an inadequate (and ungenerous) interface to our digital collections. An analysis of interfaces to archival collections demonstrates both the dominance of search, and the emergence of more generous interface elements such as browsable categories and facets. The experimental interfaces presented here are more radical, combining methodologies from data visualisation with the techniques of web design. They demonstrate practical approaches to creating rich, exploratory interfaces for digital cultural collections, and suggest some principles for more generous interfaces to archival collections.

References

- Belkin, N. J. "Some (what) Grand Challenges for Information Retrieval." *ACM SIGIR Forum*. Vol. 42. 2008. 47–54.
- Belkin, N. J, R. N Oddy, and H. M Brooks. "ASK for Information Retrieval: Part I. Background and Theory." *Journal of documentation* 38.2 (1982): 61–71.
- Chang, S. J, and R. E Rice. "Browsing: A Multidimensional Framework." *Annual review of information science and technology (ARIST)* 28 (1993): 231–76. Print.
- Dörk, M., S. Carpendale, and C. Williamson. "The Information Flaneur: A Fresh Look at Information Seeking." *Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems*. 2011. 1215–1224.
- Hinton, S., and M. Whitelaw. "Exploring the Digital Commons: An Approach to the Visualisation of Large Heritage Datasets." London, 2010. Available: <http://www.bcs.org/content/conWebDoc/36049>
- Lynch, C. "Digital Collections, Digital Libraries and the Digitization of Cultural Heritage Information." *First Monday* 7.5 (2002). Available: <http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/949/870>
- Marchionini, G. "Exploratory Search: From Finding to Understanding." *Communications of the ACM* 49.4 (2006): 41–46.
- Shneiderman, B. "The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations." *Proceedings of the 1996 IEEE Symposium on Visual Languages*. IEEE Computer Society, 1996. 336.
- Taylor, Robert S. "The Process of Asking Questions." *American Documentation* 13.4 (1962): 391–396.
- Whitelaw, M. "Visualising the Digital Archive: The Visible Archive Project." *Archives & Manuscripts* 37.2 (2009): 22–40.

Acknowledgements

The Prints and Printmaking Collection prototypes discussed here were produced as part of research funded by the Gordon Darling Print Fund and the Faculty of Arts and Design at the University of Canberra. The Manly City Council Historic Image collection prototype was produced as part of a research project funded by the State Library of New South Wales.