DELIVERING DISCOVERY

Visualising Infrastructures and Labour in Digitised Library Collections

Timothy Busuttil

Masters by Research [Design]
University of Technology Sydney
Faculty of Design, Architecture and Building

2019

I acknowledge the traditional owners of the land on which this work was produced—the Wangal people, upon whose land I live, and the Gadigal people, upon whose land I study and work. I pay my respects to Elders past, present, and future, as the traditional custodians of knowledge for this land that was never ceded.

ii.

CERTIFICATE OF AUTHORSHIP

I, Timothy Busuttil, declare that this thesis, is submitted in fulfilment of the requirements for the award of Masters by Research (Design), in the School of Design, Architecture, and Building at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise reference or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This document has not been submitted for qualifications at any other academic institution.

This research is supported by the Australian Government Research Training Program.

Signature: Production Note:
Signature removed prior to publication.

Date: 13 September, 2019

ACKNOWLEDGEMENTS

With great admiration, thank you to Rebecca Hall; your care and your support carried me through this project. Thank you to my family; for checking in, for not asking, and for writing down the name of my degree so you could tell your friends.

Thank you to Kristelle De Freitas, for your patience with me and with the thesis document. Thank you also to Thomas Ricciardiello, for sharing pide and various office spaces throughout the past three years.

Thank you to my supervisors Associate Professor Jacqueline Lorber-Kasunic and Professor Kate Sweetapple, for your perseverance through this project, and for the opportunities that you've offered me alongside it.

And finally, thank you to Matt Burgess, Joanna Fleming, and Rachel Franks, for welcoming me at the State Library of New South Wales. You shared your knowledge and your metadata, and allowed me to look behind the interface.

vi.

FORMAT OF THESIS

This thesis includes creative work. It is comprised of both *Delivering Discovery*, a prototype web interface available at www.deliveringdiscovery.com, and this exegesis.

x.

LIST OF FIGURES

ABSTRACT

In undertaking digitisation projects, libraries create large digitised collections that present new opportunities for access and display. For designers and developers, these growing collections drive the development of generous interfaces that reflect the depth and richness of digitised library collections. At the same time, many scholars, in particular from digital humanities, media theory, and information science and design backgrounds, are engaged in discussions regarding transparency—or lack thereof—in the practices, labour, and decisions that factor into the creation of these digitised library collections. Every digitised collection is built on a complex ecosystem of hardware, software, and specialist knowledge. While generous interfaces shed new light on digitised collections, their generosity is rarely extended to the administrative structural metadata which contain evidence of this network of infrastructure and labour.

This project integrates the frameworks of generosity and transparency, creating a productive intersection where media-rich, browsable representations are combined with greater visibility of the infrastructures and labour which construct and maintain our digitised library collections. To achieve this integration, this project uses a Research through Design methodology to develop an interface that materialises a response to calls for transparency in a digitised library collection interface. It deploys diagramming and prototyping as key methods in an iterative design process which culminates in *Delivering Discovery*, a prototype web interface for a subset of the State Library of New South Wales' David Scott Mitchell collection. The interface draws on principles of generous interface design and applies them to the administrative and structural metadata embedded in the collection to create a more transparent interface that affords users a view of the infrastructures and labour which construct the collection. *Delivering Discovery* represents one approach to the integration of generosity and transparency in a digitised library collection interface. It is a propositional prototype that contributes to the ongoing exploration of this intersection, and opens the door to further research on the possible uses of administrative and structural metadata in digitised library collections.

INTRODUCTION

In the State Library of New South Wales' David Scott Mitchell collection, the metadata for the record of *The death of Oenone, Akbar's dream, and other poems* by Alfred Lord Tennyson includes a curious note: 'dead bug p4'. This note remarks on the condition of the book at the time of digitisation, but it also carries with it a sense of the life (and death) of the bug, the life and work of the technician photographing the book, and offers insight into the richness of the David Scott Mitchell collection and the objects housed within it. The metadata attached to the digitised books in the collection varies from unexpected intrusions of bugs and bookmarks to their date of digitisation, copyright information, and checksum data used to prevent file corruption over time. This metadata, however, is frequently omitted when digitised books are made available to patrons. This project demonstrates the possibilities that unfold when the interfaces we use to access digital collections unveil these often-hidden layers of metadata that surround digitised objects. At its core is a consideration of generous interface design and the transparent representation of digitised collections. These concepts, generosity and transparency, work in tandem to offer ways forward for scholars working with digitised collections—particularly in design, information science, digital humanities, and media studies.

Increasingly, our library collections are made up of digitised material. The State Library of New South Wales (SLNSW) is engaged in a ten-year program to digitise '20 million collection images' (State Library of New South Wales, 2019), and the National Library of Australia has digitised '5.5 million images… from the Library's various collections' (National Library of Australia, 2019), in addition to the '128 million digitised newspaper articles' hosted on Trove (Whitelaw, 2015b, para. 2). Outside of Australia, the New York Public Library hosts over 860,000 digitised items (New York Public Library, 2019), while the British Library provides digitisation services on demand for patrons or in bulk for other institutions (British Library, 2017), and hosts a number of digitised collections including over 60,000 digitised

printed books. The digitisation of collections places their contents in electronic storage, physically out of reach from library patrons.

The creation of a digitised library collection involves the creation, storage, and management of digital images of physical collection items. Every digitised collection is built on a complex ecosystem of hardware, software, and specialist knowledge—cameras and scanners and storage media, digital catalogues and metadata management tools, domain experts in cataloguing and data fixity. Evidence of this ecosystem is deeply embedded in the collection items themselves, in the form of administrative and structural metadata [figure 1]. These metadata annotate familiar collection items with meaningful insights into the logistical realities of digitised library artefacts; the labour of digitisation, quality control practices, and challenges of storage and display. While most representations of digitised collections prioritise the digital images of physical objects, these metadata are a rich and multifaceted source of information, and a valuable product of the digitisation process.

Shifting demands from patrons, new technical developments, and experimentation in the fields of user interface and interaction design have resulted in a gradual move away from traditional search interfaces, towards more browsable, engaging platforms that provide users with the tools and encouragement to explore vast and varied collections. However, while interfaces to digital collections offer up new affordances to the patrons of these institutions, the infrastructures that enable and support digital collections remain hidden away, relegated to the domain of archivists and back-end developers. By continuing to obscure these infrastructures, library interfaces act as black boxes which prevent patrons and scholars from seeing the labour and institutional processes that construct our digitised collections. Understanding how digitised library collections are built and maintained—understanding that they are built and maintained by human labour—serves as a reminder of the constructedness of our digitised cultural collections, and cautions us against conceiving of cultural data as *a priori.* This project works towards the inclusion of this information by integrating two key concepts: generosity and transparency.

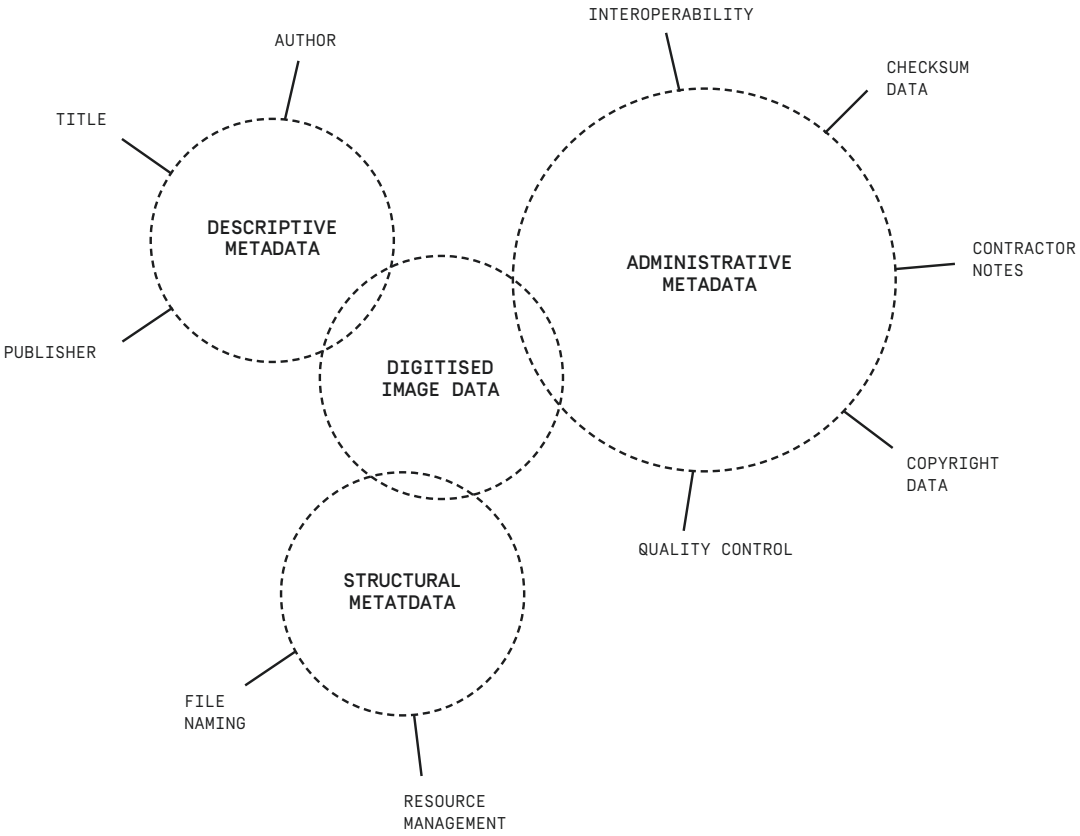


Figure 1:

Digitised library collections are comprised of more than just images of book pages; alongside every digital image is a wealth of metadata, recorded and used for various purposes. Descriptive metadata is most commonly surfaced in searching and browsing. Structural metadata is used to keep composite objects together, linking individual pages to form books. Administrative metadata, the broadest category, includes unique identification numbers necessary for interoperability, checksum data to prevent file corruption, contractor notes to aid resource management and quality control, and more.

Many digitised library collection interfaces exemplify the concept of generosity, first identified by academic and practitioner Mitchell Whitelaw in 2012. Generous interfaces 'represent the scale and richness' of library collections; they 'offer multiple ways in, and support exploration' (2015b, para. 3). Whitelaw suggests generous interfaces as an alternative to the more familiar keyword search interface, in which users enter a query into a search bar, and receive paginated lists of titles (and sometimes small thumbnail images) in return. While search interfaces are effective for users seeking specific information, generous interfaces better support exploratory modes of engagement with library collections.

Generous interfaces come in many forms, but share key features: they are rich in media, encourage exploration, offer 'overviews to establish context', and use 'fragmentary representations' to 'reveal detail at multiple scales' (Whitelaw, 2015b, para. 46). They prioritise the rich representation of collection data, allowing exploratory browsing and serendipitous discovery. In providing these experiences, generous interfaces often omit the logistical complexities of digitisation. For a specialised set of library patrons—researchers and professionals in the GLAM (Galleries, Libraries, Archives and Museums) sector, and curious members of the public—these interfaces continue to withhold the valuable information embedded in administrative and structural metadata. A more transparent interface, however, might reveal this information.

Scholars working with libraries and the digitisation of their collections are already engaged in discussions regarding the transparency of the practices, labour, and decisions that factor into the digitisation of library collections. In this exegesis, transparency centres on making visible the hidden infrastructures, labour, and institutional processes that underpin digitisation projects. A transparent interface reveals digitised library collections as constructed and maintained, shaped by organisational biases and logistical realities—and it may do so by highlighting administrative and structural metadata. In his 2016 talk *Collections as Data: Conditions of Possibility,* digital scholarship librarian Thomas Padilla

addresses the issue with a series of rhetorical questions: 'Who designed the schema? Who did the transcription?… How were the data processed? What was left in what was left out? Why? How representative is the data? What bias does our organisation of the data reflect? Can we admit a bias?' (2016, para. 43). By centering these questions, Padilla asserts that practitioners working with digitised cultural collections must be 'guided by an ethics that focuses on transparency' (2016, para. 96).

Like Padilla, media scholar Shannon Mattern recognises the invisibility of the people and systems that construct our digitised collections; librarians and archivists, hardware and software, each with their own goals and constraints. In a growing interdisciplinary discussion of labour practices that uphold our networked world, Mattern argues that 'acknowledging this complicated logistical network makes visible the labor, equipment, and expertise required to build and maintain our libraries, one of our society's few remaining intellectual and cultural commons' (Mattern, 2015, para. 9). Library patrons have increasing access to collections which, as Padilla notes, necessarily contain biases (2016, para. 92). In such a collection-rich environment, developing a critical literacy of digitised collections, and the institutions and processes that construct them, is paramount if we are to use these collections responsibly.

In response to these provocations, this project makes transparency its priority. It applies the techniques of generous interface design to the often-unseen administrative and structural metadata within digitised library collections. In doing so, the project works towards an interface that offers a rich and multifaceted view of the collection by making visible the infrastructure, processes, and labour that support it. This goal aligns with Whitelaw's framing of generosity as a notion that emphasises the interpretation of complex and diverse digitised collections. Surfacing administrative and structural metadata, and the hidden labour they reveal, has the potential to make interfaces generous in new ways and with new information, opening up library collections to new avenues of research and use for academics and professionals in the GLAM sector.

The culmination of my research is *Delivering Discovery:* a prototype web interface for a subset of SLNSW's David Scott Mitchell collection. The interface uses administrative and structural metadata as a new lens through which to view the collection of roughly 2,600 digitised books. Although this represents a small fraction of the entire collection (only about 15% of digitised books in the DSM collection), the subset remains a significant case study as the product of SLNSW's newly-updated protocol for the digitisation of printed material, making it rich in the administrative and structural metadata around which this project pivots.

Delivering Discovery was developed as the practical outcome of a Research through Design (Frayling, 1994) methodology. Working with digitised collection specialists at SLNSW to access data and metadata that are normally hidden from library patrons, the initial aim of the project was to build upon the foundation of generosity in representing the collection. This process combined approaches from visual communication, data visualisation, and interaction design, as I discuss in the methods section. Diagramming and prototyping were critical methods in revealing the presence and potential of administrative and structural metadata within the collection. While discussions of transparency and logistics in libraries are not new, their implications have been slow to effect change in interface design decisions. By approaching the issue with a Research through Design methodology, this project seeks to realise a practical integration of transparency and generosity through the creation of a digitised library collection interface.

Generous interfaces have made great strides in creating new ways to explore digitised library collections. However, their goal of enabling rich exploration stands to benefit from a greater engagement with ongoing discussions of transparency. While these interfaces openly share collection media and descriptive metadata, this generosity is rarely extended to administrative and structural metadata, and is generally not leveraged to address Mattern and Padilla's concerns. As a result, the ability of library patrons—particularly those specialised academics and professionals—

to investigate and critically interpret the full complexity of a dataset is constrained. Generous or not, library interfaces rarely reveal the administrative and structural metadata which contribute to a digitised collection's breadth and depth.

This exegesis documents the process of developing the *Delivering Discovery* interface, and explores its potential as a prototype that combines both generosity and transparency in its representation of a digitised library collection. I begin with an examination of Mattern and Padilla's positions on transparency in libraries, and survey contemporary interface practice to identify precedents and possible opportunities for the integration of transparency and generosity in digitised collection interfaces. I then detail the methods employed throughout the development of the collection interface, showing how early visual experiments formed the basis for later design decisions. Finally, I discuss the insights generated by this approach to digitised collection interface design—how a focus on transparency builds on the goal of generous interfaces to open digitised library collections to deeper interpretation for particular users.

The interface offers two interactive representations of the collection. The first is an explorable diagram depicting SLNSW's recently updated Statement of Requirements for the Digitisation of Printed Materials, a document which details the Library's technical standards for scanning, metadata management, and file organisation. This view [figure 2] is a visual representation of the complex, interconnected ecosystem of digitisation. It provides context for the second view [figure 3], which represents all 2600 books in the collection in an interactive visualisation. As users move through different administrative and structural metadata categories (date of digitisation, digitisation staff, quality control notes, and more), the collection is arranged and rearranged. Both views are simultaneously linear and exploratory; viewers are guided by annotations through the often-hidden and unfamiliar categories that shape the interface, but are free to dig deeper into the facets that interest them.

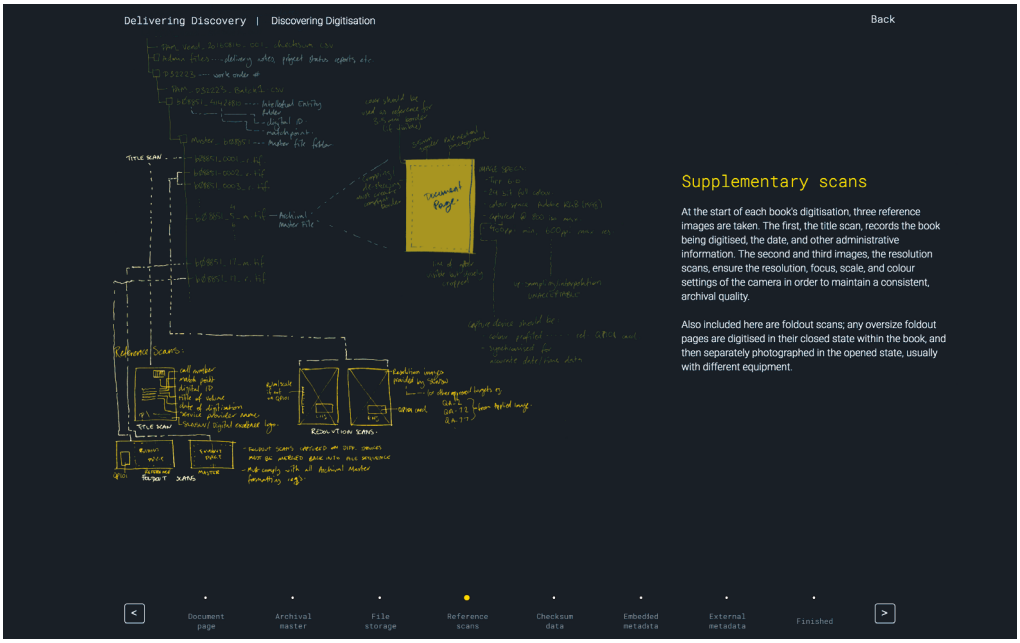


Figure 2:

Delivering Discovery's infrastructure diagram depicts SLNSW's Statement of Requirements as an interconnected network, with annotations describing the uses of various data.

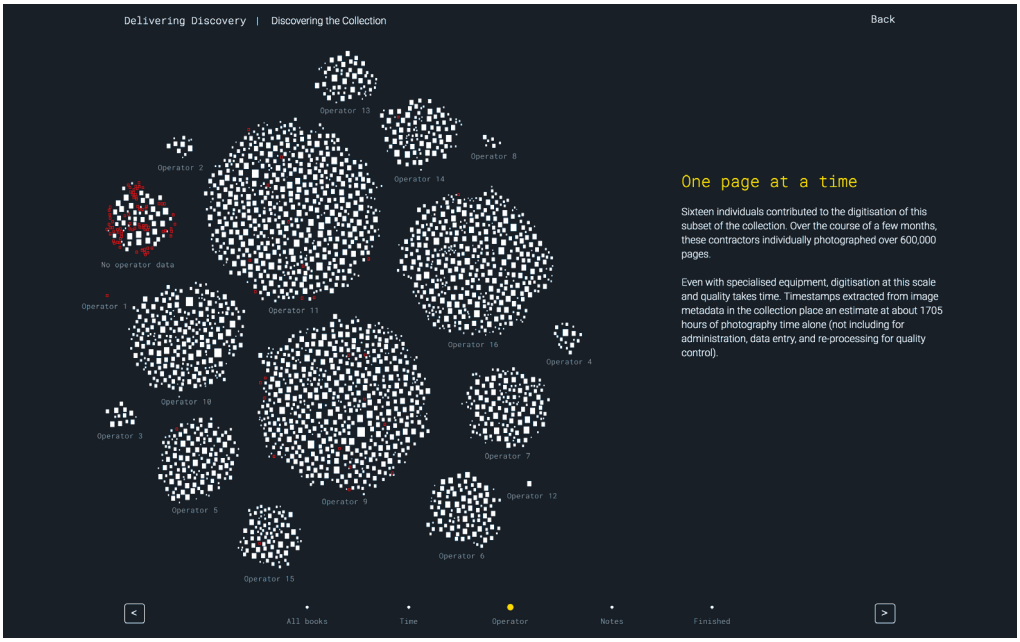


Figure 3:

The collection visualisation view displays all 2,579 books digitised in the David Scott Mitchell collection in 2016–17. The rectangle depicting each book is scaled to reflect the total file count for that book.

The *Delivering Discovery* interface is a culmination of an iterative design process. Drawing from a series of prior prototypes, *Delivering Discovery* provides a more transparent view of the labour and infrastructure of digitisation. It is not intended to replace existing traditional or generous library interfaces. Rather, it is a provocation; a practical response to the lack of transparency in libraries identified by Mattern and Padilla, and an invitation to further consider the possibilities of creating more transparent interfaces to digitised library collections.

LITERATURE REVIEW

1. TRANSPARENCY

In order to work towards a more transparent digitised collection interface, we must understand who might look through such an interface, and what they hope to see. For the majority of library patrons accessing digitised collections, generous interfaces fulfil their goal of revealing 'the complexity and diversity of cultural collections' (Whitelaw, 2015b, para. 35). They make the contents of their collections readily available, and enable users to explore rich displays of text, images, video and audio. However, these interfaces fail to extend that generosity to the administrative and structural metadata embedded in their collections. Seeing digitised library collections through a more transparent interface could provide a view of the complex ecosystem of infrastructures and digital labour that supports every digitised library collection.

Surveying the work of scholars who write in and about libraries from humanities, digital humanities, media studies and design backgrounds reveals a vibrant research space comprised of multiple stakeholders. Each scholar brings a unique perspective on the issue of transparency that informs this project's approach. In this section, I will provide an overview of these key concepts, before investigating them in detail. One of the key factors of transparency in digitised library collections is the acknowledgement that these collections are built and shaped by infrastructures of labour, software, and hardware. Digitised library collections are implicated in a social and technical ecosystem, and an ability to see this ecosystem is one of the drivers of transparency as a framework. Shannon Mattern, a professor of media studies, works towards transparency by unveiling the labour, institutions, and infrastructures which create, store, and maintain media, and unpacking the implications of the forces acting in this 'infrastructural ecology' (2014, para. 22). Mattern highlights the need to acknowledge material logistical networks that support library operations (2014, 2015).

Once these systems of labour and technology are made visible,

transparency also supports us in avoiding the misconception of digitised library collections as a given, a set of *a priori* cultural facts. Johanna Drucker approaches this issue from a design perspective. She is an author, book artist, visual theorist, and cultural critic, whose work engages with the theoretical and the visual. Drucker advocates for a humanist approach to visualisation, asserting that data should be reframed as *capta,* and perceived as 'taken' rather than 'given' (2011, para. 3). Sharing a similar perspective, media historian Lisa Gitelman's writing on the emergence of new media frames data as disciplinarily constructed. She warns against the oxymoron of 'raw data' (2013, p. 3).

The final prevalent idea that permeates discussions of transparency in library interfaces is that acknowledging the labour which constructs these digitised collections is critical for ethically conducting research within them. For digital scholarship librarian Thomas Padilla, the opportunities presented by new perspectives on digitised collections are only viable if the scholars take responsibility for the ethics of working with digitised collections (2016, para. 66). Michelle Moravec is a historian whose body of work spans feminist cultural activism and digital humanities. Her writing also reflects on the implications of 'conducting historical research in digital archives from a feminist perspective' (2017, p. 186), challenging researchers to engage ethically with digital archives.

These scholars work across a number of disciplines, but their work converges on the historical, technical, ethical, and political implications of acknowledging the infrastructures, processes, and labour at work in our digital ecosystems, of which digitised library collections are undoubtedly a part. While not all of them reference transparency explicitly, transparency is the framing concept I use in this exegesis to gather these ideas and create an actionable counterpart to generosity. By working towards transparency, designers and developers are given a framework to engage with the perspectives and concerns of the various stakeholders and scholars for whom they create interfaces. This section will explore the work of these scholars, highlighting what might be seen through a transparent

interface, and identifying opportunities presented by making these infrastructures and labour visible.

Mattern's perspective on transparency exists within the context of her work on library logistics and the labour of maintenance. She has written on a variety of invisible infrastructures: smart cities (2014), cloud computing (2016), artificial intelligence (2017). In her article *Library as Infrastructure,* Mattern offers 'infrastructure' as an alternative to the metaphor of library as 'platform,' highlighting the hardware, software, and labour that allows libraries to perform their myriad roles. She frames libraries as 'sites where spatial, technological, intellectual and social infrastructures shape and inform one another' (2014, para. 10). Much of Mattern's work on libraries leads her to investigate and tour offsite library operations facilities, from the New York and Brooklyn Public Library's shared technical services center BookOps to San Antonio's BiblioTech, a "bookless' library featuring 10,000 e-books' (2014, para. 49). In examining such facilities, Mattern notes the 'increasing recession of these technical infrastructures—and the human labor that supports them—further off-site, behind the interface, deeper inside the black box' (2014, para. 50). Notably, Mattern implies an opaque interface, through which library patrons struggle to see 'the ways in which those structures structure our intellect and sociality' (2014, para. 50). As Mattern identifies interfaces as opaque, perhaps they are a potential site for the development of transparency in libraries.

We must be careful here not to confuse the deployment of transparency with the term's use in media theory, where it denotes an interface that functions so intuitively and provides so little obstruction as to become fully invisible. As media theorist Alexander Galloway describes it, 'the more a dioptric device erases the traces of its own functioning (in actually delivering the thing represented beyond), the more it succeeds in its functional mandate; yet this very achievement undercuts the ultimate goal: the more intuitive a device becomes, the more it risks falling out of media altogether, becoming as naturalized as air or as common as dirt' (2012, p. 25). This tendency for interfaces to erase themselves through functioning is problematic for humanities scholars. Michelle Moravec identifies a clear tension between the recession of archival labour off-site and the

significance of acknowledging this labour in humanist research. She notes that 'Digital archival environments, which can increase access to materials, often simultaneously erase the labour that makes this possible' (2017, p. 191). Citing literature that identifies archives' and libraries' reliance on 'feminised labour', Moravec also stresses that acknowledging this labour 'should be viewed as part of a feminist research practice as well' (2017, p. 191). Humanities scholars working with digitised collections sit at the crux of this issue. Humanities research works within the bias, subjectivity, and complexity that permeate our cultural records. Digitised collections present countless opportunities for this work, but access to the information in these collections is restrained by designers and developers when they seek to build interfaces which seamlessly and intuitively represent these collections. So, how then might we approach a transparency that is productive for these scholars?

The disappearance of the interface occurs when information access is seamless; when the interface hides the connections and mappings between the heterogeneous information systems that are involved in the creation and presentation of a digital collection. From imaging hardware, to storage media, to multiple software catalogues, and finally the user interface itself, each of these systems schematises and describes aspects of the intellectual entity differently, and translation the required at each seam is usually concealed. This paradigm has been critiqued by Matthew Chalmers, who calls instead for a 'seamful' design approach that emphasises those points of mapping (Chalmers and Galani 2004, p. 9). Embedded in those seams lies evidence of the technical structures, institutional knowledge, and feminised labour that construct the very collections users seek to engage with. By broadening the types of metadata that are included in digitised collection interfaces, our representations of collections can benefit from the exploration of these seams. By viewing our collections through the lens of the labour that constructs them, we can work towards a productive transparency for digital humanists.

Writing on the labour of data maintenance, Mattern cites Jean-Christophe

Plantin's work on the invisibility of data processing work in the social sciences, noting that 'internal maintenance work isn't supposed to be visible to end users, who tend to like the idea that they're working with 'raw' data' (Mattern, 2018, para. 48). However, Plantin recognises that 'data never come as raw' (Mattern, 2018, para. 48). Acknowledging data as constructed rather than given—cooked, rather than raw—taps into a broader conversation led by humanists working with data. Media historian Lisa Gitelman's *Raw Data is an Oxymoron* offers the 'friendly reminder' that 'different datasets harbor the interpretive structures of their own imagining' (2013, p. 3). That is, data are necessarily shaped by the circumstances of their collection and presentation. Visual theorist Johanna Drucker encapsulates this constructed nature of data by arguing for their recharacterisation as *capta*. 'Capta is 'taken' actively while *data* is assumed to be a 'given' able to be recorded and observed' (2011, para. 3). This reframing of data as capta acknowledges the subjective and situated nature of the construction of digitised cultural collections.

Drucker's *capta* is a succinct expression that aligns with Gitelman's observation that 'Every discipline and disciplinary institution has its own norms and standards for the imagination of data' (2013, p. 3). Both hold true for digitised library collections. Digitisation and archiving are labour-intensive tasks that require ongoing attention from domain experts, but the labour of image capture, quality control, data cleaning, and administration behind every digitised library collection goes largely unseen by patrons. Libraries present digitised collections as rich, interconnected, and ready to browse, with little to no public acknowledgement of the work done, decisions made, and data excluded to make them so. This seamless presentation may be desirable for library patrons, but such collection interfaces perpetuate the fallacy of raw data, and intentionally or not, avoid acknowledging the biases of the (usually white, colonialist, patriarchal) institutions that produce these data. Obfuscating the labour that constructs digitised collections also frames the work of digitisation as the preservation of existing cultural heritage, rather than an ongoing process of creation of cultural heritage in a new form. For GLAM professionals, researchers in the digital humanities, media theory and information science, and the designers and developers who create these

interfaces, a representation of collections which frames them as raw data is an incomplete representation.

A more transparent interface might represent collection data as a function of the institutional practices and disciplinary assumptions that construct it, making the interface (and the collection itself) a richer resource for scholars. For Mattern, such a commitment to transparency is also a political project; 'Better understanding and investing in the logistical systems that route our power, packages, and—particularly with libraries—knowledge means greater potential to shape not only our built landscapes, but our political and intellectual ones as well' (2015, para. 9).

Padilla works from within libraries as a digital scholarship librarian. His concerns about transparency are built on the foundational concept of collections-as-data. He sees libraries' role in offering collections to patrons in the form of data as an extension of their commitment to '[supporting] the formation of community' by '[building] infrastructure and collections' (2016, para. 17). As libraries adopt 'open data' policies and develop systems to make their collections more computationally accessible (Mattern, 2016, para. 23), cultural data are becoming increasingly available for exploration and interpretation. Such work certainly has benefits; the opening of library collection data has enabled and encouraged 'uses beyond reading, viewing, and listening' (Padilla & Higgins, 2014, p. 324)—investigation, research and play across numerous academic and professional fields and by members of the public. However, Padilla notes that these benefits may be contingent on transparency. 'Will our users expect similar levels of transparency in the documentation of computational process and analytics decisions that were made to generate the collections that we provide access to? They may. Transparency in this vein may be a precondition of viable use' (Padilla, 2016, para. 71). Again, we see an incentive for institutions to more openly disclose the labour and practices used to construct their digitised collections: without some transparency, certain avenues of enquiry or usage of digitised collections, particularly by researchers and GLAM professionals, may be closed off. Beyond these

specialised cases, sharing a greater variety of information broadens the scope of possible uses for digitised collections in general.

Figuring digitised cultural collections as data also grants new perspectives. Padilla exemplifies the potential of this perspective with a single digital image: 'Approaching this image from a technical, materialist angle, as a matrix of pixels, rather than as human viewers, brings new facets of this meaning to light' (Padilla and Higgins, 2014, p. 328). Here, Padilla and Devin Higgins describe the 'facet effect', where the contents of digitised collections (most often text, images, and audio), are understood in terms of 'the facets that are leveraged in the course of mining, visualizing, and generating new objects of inquiry' in digital humanities research (Padilla and Higgins, 2014, p. 335). These facets are comprised of metadata which annotate a digital object, capturing both high-level information and granular details. Conceiving of administrative and structural metadata as annotations calls attention to the constructed nature of the collection; these metadata dispel the notion of collection data as *a priori*, revealing the ecosystem of labour and infrastructures which inflect the process of building a digitised library collection. For digital humanists, these facets present a fertile ground for analysis. Examples of these facets include word frequency, collocation, part of speech and line length in digitised text, and colour data, size, format, and human- or machine-annotated tags describing features or category in digital images. This approach grants access to more of the information stored in a digitised artefact. Rather than a black-box which can only be viewed as a two-dimensional grid of pixels, the digital image becomes a dense, multifaceted object which can be understood and analysed in a variety of ways.

Understanding collections as data—using alternative methods to 'see the image from a new, often estranging angle' (Padilla and Higgins, 2014, p. 328)—is easily extended to the array of administrative and structural metadata associated with objects in digitised library collections. Although not commonly represented in collection interfaces, facets like date of digitisation, quality control notes, numbers of foldout pages, and checksum data are not inherently more esoteric than word frequency in a digitised text, or colour and size data in a digitised image. Including these metadata

in the multifaceted perception of digitised collection objects opens them up to the analysis of researchers both within and beyond the digital humanities. Furthermore, these metadata carry evidence of the collection images' place in a larger ecosystem of digitisation—a connection that enables scholars to apply humanist computational analysis techniques to study the creation and management of digitised collections, as well as the individual collection items themselves. These metadata already exist in collections for 'back-end' purposes of quality control and interoperability, so making them available serves to enhance the value of a digitised collection. By using these metadata as the basis of a new representation of a collection, generous interfaces could become more transparent, drawing attention to these metadata and serving digital humanists and other researchers.

Each scholar presents both challenges and opportunities for transparent interfaces. Mattern notes the invisibility of library infrastructures and the labour of digitisation, and invests in uncovering and understanding these systems. She implies that interfaces are often opaque, opening space for the creation of transparent interfaces which make these infrastructures and labour more visible. Padilla reminds us of the decisions and biases that construct library collections. He sees transparency—in decision-making and computational process—as a necessity for libraries opening their digitised collections to public use, as patrons may expect clarity on the digital provenance of their data. There are also benefits beyond necessity; metadata offer new and productive ways to locate, view, and engage with collection items. Records could be found not only by title and author, but also by image format, date of digitisation, quality control notes, and more. Padilla notes that 'Algorithmically derived technical and descriptive metadata have the potential to create new paths to library collections and items, in ways that not only serve traditional searching and browsing but also rise to meet the research interests of Digital Humanists', whose research relies on both 'the image as *substance* and the image as *picture*' (Padilla and Higgins, 2014, p. 328). Extrapolating from the work of these scholars, we can imagine a transparent interface as one which: makes

visible the labour and infrastructures that create and maintain our digitised collections, and includes broader varieties of collection metadata to offer a richer and more comprehensive view of the items which makes up a digitised collection. These perspectives inform this project's investment in transparency, and form a critical lens through which to build upon generosity in library interfaces.

2. GENEROSITY

Whitelaw developed generosity in response to the critique of keyword search interfaces as 'inadequate' for some applications (Whitelaw, 2015b, para. 8; Hinchcliffe & Whitelaw, 2015, para. 1). Keyword search interfaces excel when a patron seeks specific information they can define in key terms. For more exploratory use cases however, 'search fails to match the ample abundance of our digital collections and the generous ethos of the institutions that hold them' (Whitelaw, 2015b, para. 3). Generosity has proven a significant guiding principle in countering this inadequacy. The key features of generous interfaces—media richness, explorability, overviews, and multifaceted representations—lend themselves to a variety of collections and user experiences, from relatively focused searching to serendipitous browsing. Generous interfaces seek to 'reveal the complexity and diversity of cultural collections, and to privilege the process of interpretation' (Whitelaw, 2015b, para. 46), an objective that shares promising overlap with Mattern and Padilla's calls for transparency. I have identified existing interfaces from galleries, libraries, archives and museums (GLAM institutions) that demonstrate varying degrees of overlap between elements of generosity and transparency. None of these interfaces are solely focused on presenting administrative and structural metadata: some represent only descriptive metadata, while others find ways to integrate the administrative and structural. Through these examples, we begin to see how generosity can align with transparency—bringing a greater variety of collection data to the surface can bring with it metadata which provides evidence of the labour and infrastructures that lay behind an opaque interface. Together, these interfaces stake out a possibility space for the

practical integration of transparency and generosity in digitised collection interfaces.

Manly Images

SLNSW is not unfamiliar with the shift towards generous interfaces; in 2012, they commissioned Whitelaw to develop *Manly Images*, an experimental interface to the Manly Local Studies Image Library (and the first of Whitelaw's case studies presented in *Generous Interfaces for Digital Cultural Collections)*. The tool provides a browsable visual overview of the collection of approximately 7000 images, grouping images together into 'tiles' that are scaled by the number of images they contain [figure 4]. These tiles can be sorted by either subject matter or decade, allowing a level of control over the browsing experience through multiple views. Users are not required to provide a search term before seeing images. Instead, these decades or key terms that would normally be typed into a search bar are presented by the interface to users, immediately displaying what the collection offers through a visual overview. While the interface doesn't exemplify transparency, it presents a clear baseline for generous interfaces. Manly Images offers the whole collection in aggregate, clustered by human readable keywords and dates. It achieves its goal of providing a 'rich,

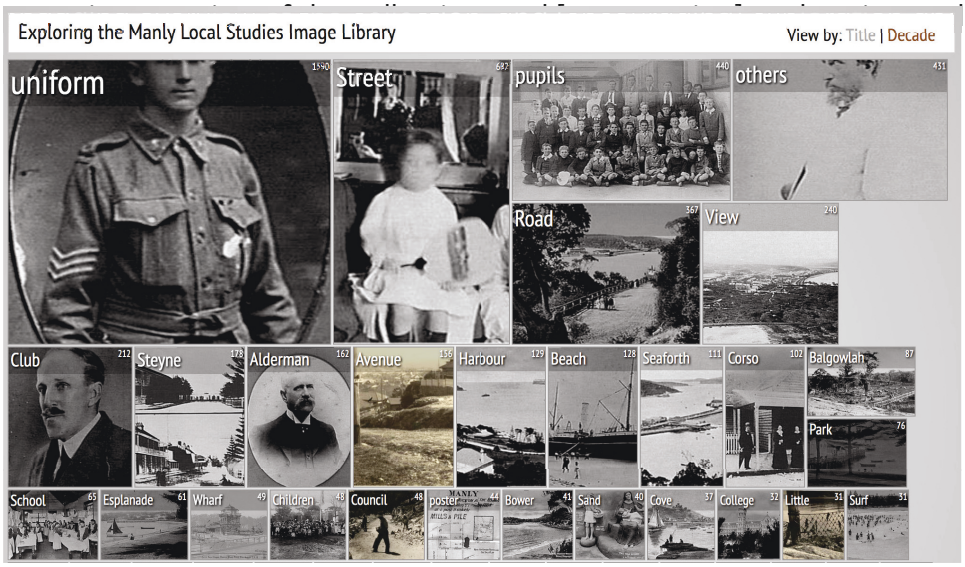


Figure 4:

Whitelaw's Manly Images typifies the generous overview–the whole collection is presented at a glance, grouped by subject matter and scaled by the size of the groups. Users quickly become aware of the distribution of the collection's contents.

Visionaries Explorer

Continuing to explore generosity as a guiding principle for collection interfaces at SLNSW, designer/developer (and frequent collaborator with Whitelaw) Geoff Hinchcliffe designed and developed *Visionaries Explorer* [figure 5] for the DX Lab's Digital Learning Fellowship in 2018. The interface was designed to 'support both strategic and playful exploration of the content' (Hinchcliffe, 2018, para. 20), SLNSW's Visionaries collection: a set of digitised texts, images, maps and objects curated for secondary students. Containing 'over 17,000 images' (Hinchcliffe, 2018, para. 1), the interface is rich in media. Visionaries Explorer offers a number of tools to allow users to filter collection records. Date ranges, keywords, and result types can all be optionally set, allowing users to start with a broad overview, and narrow down to multiple levels of detail. The interface also offers a more playful browsing tool—the colour palette search—building on Hinchcliffe's previous work in *Discover the Queenslander* (2014). Clicking on a colour swatch reveals all images containing that colour, providing an unconventional—but increasingly common (Hinchcliffe & Whitelaw, 2015, para. 4)—facet with which to browse collections. The interface is another clear example of

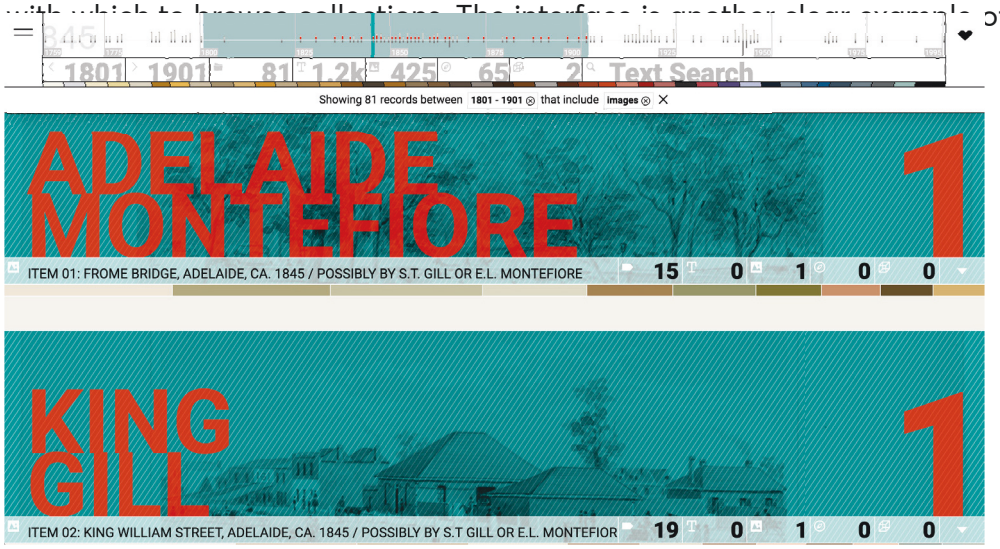


Figure 5:

Visionaries Explorer provides a compact toolbar with a number of filters and visualisations. Records can be filtered by year, type, and keywords. The toolbar lists the number of results of each type, and shows their distribution by date.

Manly Images and *Visionaries Explorer* are exemplary generous interfaces. They are each media-rich and explorable, and give patrons the ability to see an overview of the whole collection, or dive into areas of interest at multiple levels of detail. However, they are not particularly transparent. The generosity of these interfaces focuses on the collection media (digitised images) and descriptive metadata. Administrative and structural metadata, which might give some insight into the processes of digitisation, are absent. Writing on the field of data visualisation, media theorist Lev Manovich points out that when faced with complex datasets (such as digitised library collections), 'designers and their clients have to choose which dimensions to use and which to omit' (2002, p. 3). The decision to omit administrative and structural metadata can be attributed to the intended purpose of these interfaces. For patrons who wish to browse collection images, topic, date, and colour might be relevant dimensions by which to browse. Inversely, administrative metadata such as the date of digitisation, make and model of capture equipment, or quality control scans aren't useful, and are excluded as a result. While this focus on descriptive data is reasonable for the intended purpose, the continued omission of administrative and structural metadata contributes to the opacity of library interfaces.

Loom

Not all generous interfaces are entirely without elements of transparency. In 2015, SLNSW's DX Lab developed *Loom* in collaboration with design studio Grumpy Sailor. *Loom* is a generous interface for a collection of photographs taken in Sydney throughout the last 130 years, using the date, location, and subject of the photographs to sort them in three interactive views. The first view displays photographs in a 3-dimensional 'time tunnel'; the images are arranged chronologically in decade groups, with older images fading into the distance [figure 6]. The second view is location-based, clustering groups of images on a map of Sydney [figure 7]. The third view is tag-based; it reveals trends in subject matter with reference to time and location

[figure 8]. Clearly, *Loom* is generous; not only does it readily offer records to explore, it presents three distinct ways of exploring them. Each of the different views frame the collection's images with different metadata, encouraging the use of all three views to more completely explore the intricacies of the collection it represents.

Integrated into this generous representation of the collection is a subtle but engaging representation of administrative metadata. In the location-based view, clusters of images are represented as 3D geometric forms extending above and below the plane of the map [figure 9]. The volume above the map represents images that are available to view, while the volume below the map represents those which have not yet been digitised. This representation is reminiscent of an iceberg; the presence of more images below the surface gestures at the scale of the collection, and offers a small insight into the ongoing nature of the digitisation process. While it is not the focal point of the interface, this elegant representation provides a model for the unobtrusive inclusion of administrative data in a generous interface. Through *Loom*, the DX Lab demonstrate the potential of generous interfaces to exemplify transparency by integrating

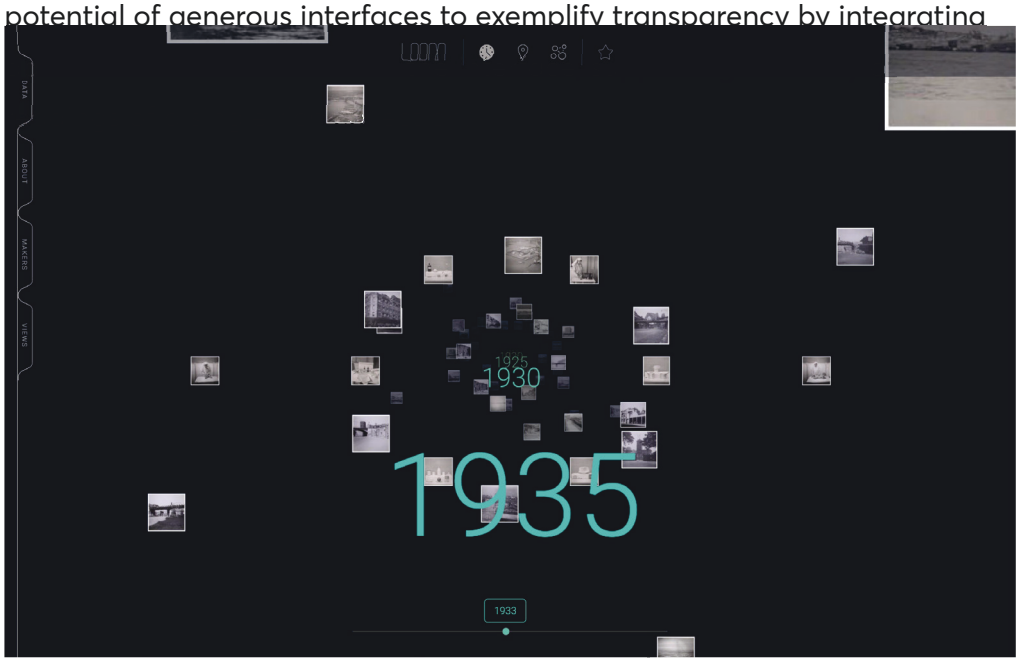


Figure 6:

Loom's 'time tunnel' view sorts records by year, but its spatial presentation incorporates a scrolling interaction and parallax movement to create a more engaging and exploratory browsing experience.

Figure 7:

Loom's location-based view gathers records into regions throughout the Sydney CBD. The height of the three-dimensional forms reflects the number of records available in that region for a given year.

Figure 8:

Loom's complex tag-based view allows users to filter the collection by location and time, and surfaces images in groups organised by topics.

Figure 9:

When viewed from the side, the location-based view reveals that there are more images below the surface. This example shows 21 digitised records for the Botanic Gardens [1880–1885], but shows a further 13 are yet to be digitised. This is represented visually as well as numerically.

The Real Face of White Australia

While many generous interfaces aim to be expansive in their representation of digitised collections, they are also capable of being highly focused. Tim Sherratt is a historian and hacker who employs digital tools to reveal alternative voices within cultural collections. His desire to use digital collections to tell these stories is exemplified in *The Real Face of White Australia,* an interface developed in 2011 as part of the Invisible Australians project with historian Kate Bagnall [figure 10]. The interface, and the Invisible Australians project more broadly, seek to make visible some of the people—'people whose lives were monitored and restricted because of the colour of their skin' (Bagnall and Sherratt, 2011)—living under the White Australia Policy in the early twentieth century. But as Sherratt points out, 'this is not just an exhibition, it's a finding aid' (Sherratt, 2011, para. 15).

Figure 10:

The Real Face of White Australia's image grid is not complex in form, but it successfully encourages browsing with its focused subject matter. By offering a compelling and approachable interface, the website avoids confronting users with complex searching tools, instead generously drawing them into the collection.

Clicking on any of the 7,247 portraits reveals the whole document from which the portrait was cropped, and provides a link to the document in the National Archives of Australia's (NAA) RecordSearch database for further research. With this project, Sherratt demonstrates the great potential of a generous interface that focuses on a specific type of data. *The Real Face of White Australia* gives users access to NAA's Series ST84/1, but it does so through an arresting overview of all-too-often erased faces. By extracting a specific subset of the collection data and presenting it in a separate interface, Sherratt leverages his skill and sensitivity with digital tools to provide a compelling gateway to an existing dataset. *The Real Face of White Australia* provides a model for collection interfaces which focuses on specific types of data or subject matter.

In a blog post linked in the interface's 'About' section, Sherratt shares his methods for the extraction of the portraits. Using a combination of face-detecting code (which is freely available on GitHub) and careful, manual data cleaning, Sherratt extracted the portraits from a range of government documents held by the NAA. This blog post reveals the labour, care, and decision making behind the interface; the guiding principles, trial and error, and multiple stages of cleaning that make such a project possible. This information provides the kind of transparency 'in the documentation of computational process' that Padilla argues is necessary (2016, para. 71), although it is external to the interface. Nonetheless, it offers insight to curious viewers, guidance to other practitioners, and important context to researchers who might be interested in working with the dataset.

Coins

Digitised collection interfaces take a variety of approaches to the apparent physicality of their contents. *Coins,* developed by Flavio Gortana, Franziska von Tenspolde, Daniela Guhlmann, Marian Dörk at the Urban Complexity Lab in 2017–2018, is an interface which leverages physical metaphors of sorting and rearranging to create a dynamic and engaging presentation of 26,000 digitised coins (Gortana et al., 2018).

As in many generous interfaces—although perhaps most literally in this

example—the resting state of the interface is one large pile of coins [figure 11]. This choice is framed a response to the tendency of numismatic collections to 'end up in tabular grids, regardless of whether they are in the storage drawers of the depot, the glass cabinets in the exhibition, or thumbnail pages in the online interface' (Gortana et al., 2018, p. 2). By initially displaying the collection as a disorganised pile, the interface encourages users to begin sorting the coins on their own terms. This sorting is enabled along a number of descriptive metadata facets: country and region of origin, minting place, period and date, material, weight, and size. As facets are selected, the coins are rearranged into various layouts, chosen by the type and number of facets selected [figure 12]. The collection allows users to quickly rearrange the collection to make comparisons and identify trends between the metadata facets; the usage of certain materials over time, the size of coins in certain regions. Clicking on a single coin in the collection reveals its details along the facets provided in the interface. Like *The Real Face of White Australia*, a link is

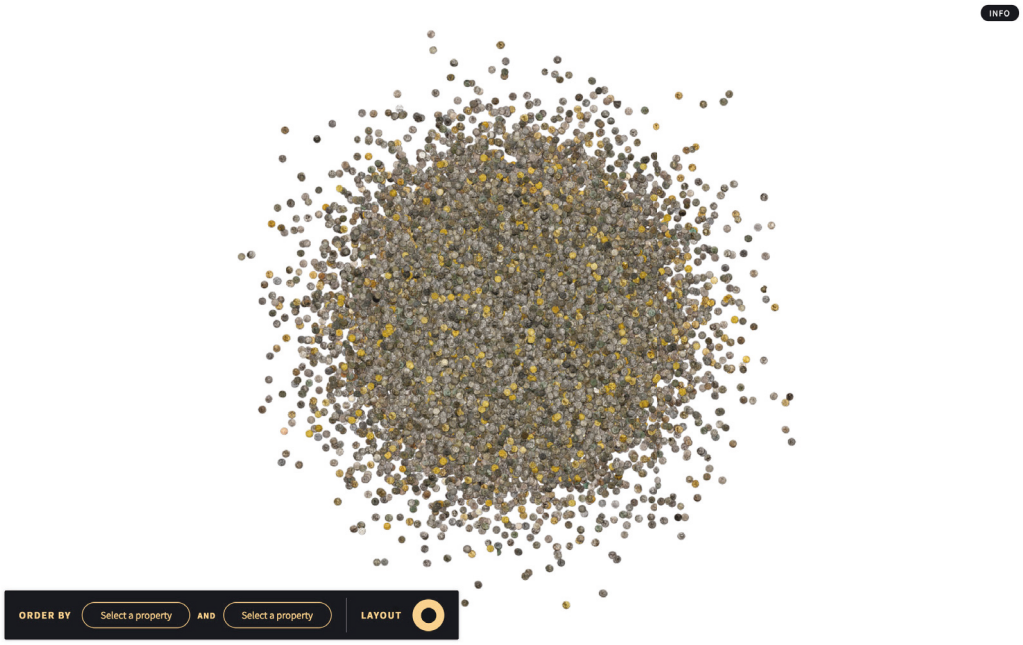


Figure 11:

Upon entering the Coins interface, users are presented with the interface's entire collection. From here, users can begin to sort the collection along up to two facets of their choice, and watch the pile re-organise in response.

provided to the coin's page in the Münzkabinett Berlin digitised collection for further research in a traditional collection interface. These filters are easily removed, allowing users to zoom back out to the broader collection before further exploration.

The flexibility of the interface achieves the project's stated aim to 'do justice to the multidimensional and material abundance that the collection exhibits while providing a playful mode of access that is inviting to a lay audience' (Gortana et al., 2018, p. 2). The interface is certainly playful; the ease of rearranging the collection allows for a great many visualisations, which the designers hoped 'could generate unique observations and insights into the collection' (Gortana et al., 2018, p. 10). However, it may be difficult for such a lay audience to draw meaningful insights from these arrangements. The interface allows the comparison of coins in the collection along their descriptive metadata, but provides no guidance or numismatic context for the potential significance of such arrangements.

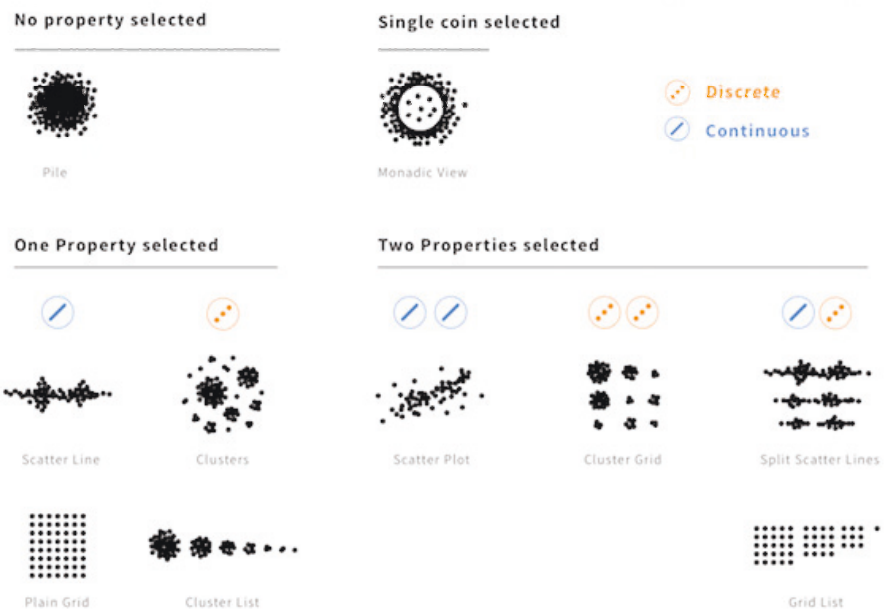


Figure 12:

Coins' dynamic layouts [Gortana et al., 2018] are selected automatically depending on the properties selected. While the layouts themselves support browsing and exploration, the constant change in structure can become disorienting.

Repcol

Generous interfaces are often easily recognised by their richness in media, usually digital images of collection items. *Repcol*, a digitised collection interface prototype developed by now defunct digital consultancy and product development studio Bengler, explores a more abstract representation of collection items. Designed to give an overview of a whole collection in one interactive visualisation, *Repcol* allows users to investigate a snapshot of The Norwegian National Museum's (NNM) entire archive of artworks. Like many generous interface, Repcol clusters the items it represents. Artworks are grouped by artists, and these artists are represented as prisms in 3-dimensional space [figure 13]. Such an

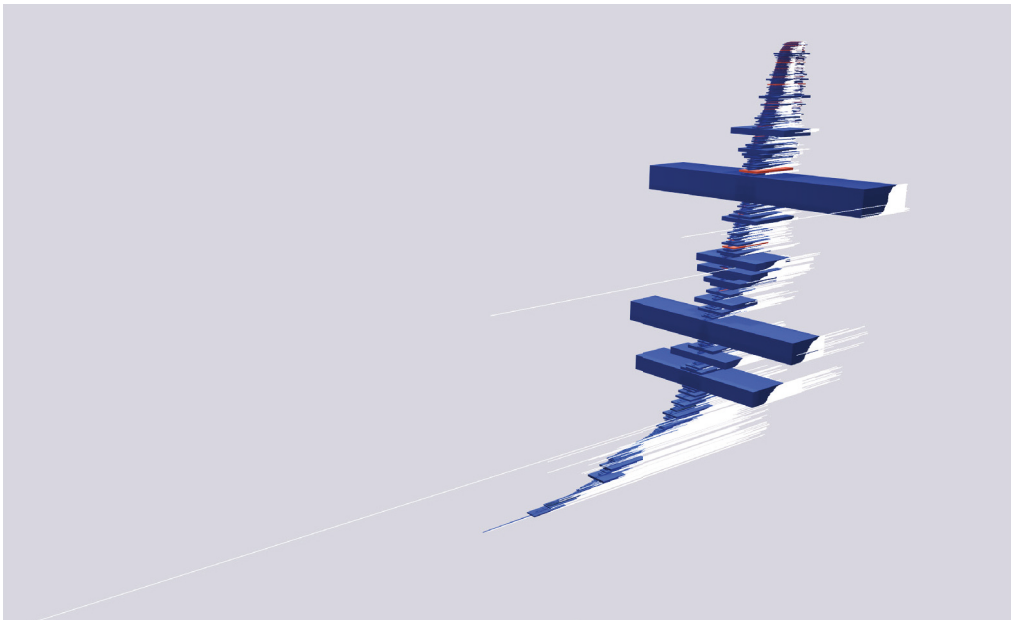


Figure 13:

Repcol's use of size, colour, and position create immediate questions for the user: which artists are most prominent in the collection? Are there any women represented substantially in the collection? What are the collection's oldest and newest acquisitions?

abstract, geometric representation of a collection is relatively uncommon in GLAM interfaces; the arrangement of these prisms makes up a data visualisation. The artists are sorted chronologically by birth, scaled according to the number of their works in the collection, and colour coded to represent gender.

This visualisation/overview reveals a number of interesting institutional patterns; the Museum's most prominent artists are easy to identify, as is the collection's uneven gender representation. Clicking on a prism reveals some of the individual artworks by that artist in the collection, but this function seems secondary to the overview itself—the visualisation offers information that the items themselves cannot. A closer inspection of the prisms continues this pattern. On the side of each prism is another layer of the data visualisation; a timeline, marking the creation of each artwork throughout the artist's life and linking it to the date of acquisition [figure 14]. Read in aggregate, these smaller timelines make visible the institution's differing acquisition habits over time; in a video introducing the interface, Bengler demonstrate with *Repcol* that 'a hundred years ago,

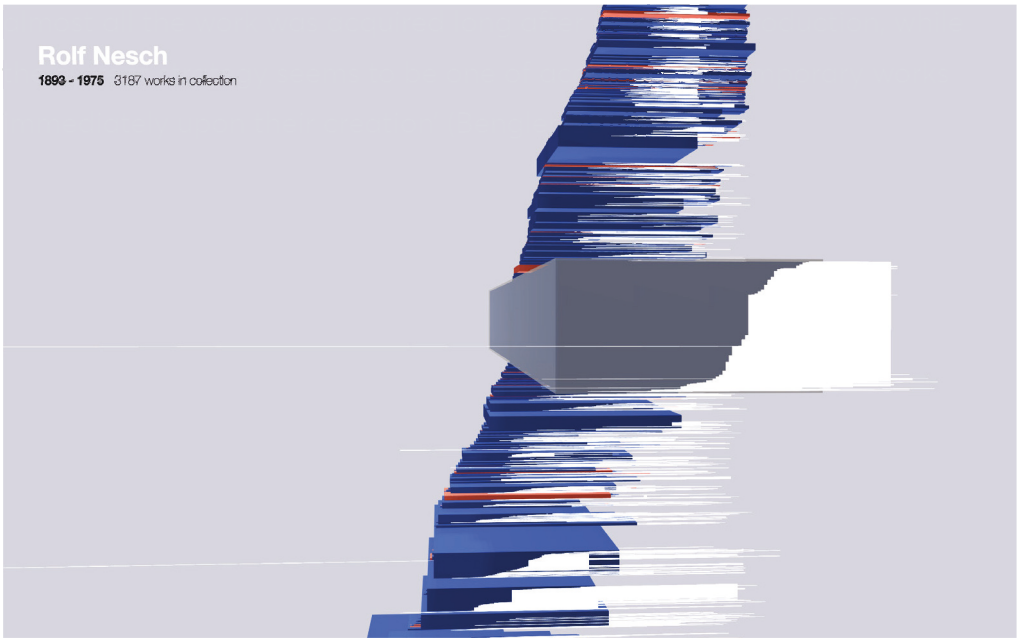


Figure 14:

A closer look at Rolf Nesch's prism in the interface reveals one aspect of NNM's collecting process; the white graph on the side of the prism shows the bulk of Nesch's work was acquired within a few years, with some older works [seen at the bottom of the white graph] collected some years later to build a more complete collection.

Repcol includes administrative metadata which are not often available in digitised collection interfaces. By foregrounding the prominence of certain artists within the collection, their genders, and the artworks' dates of acquisition, Bengler create a level of transparency through which patterns in the collection can be seen. This revelation of new information affords patrons the ability to ask different questions of the collection. Of all the artwork metadata included in Repcol, date of acquisition is perhaps the least commonly encountered in other interfaces. *What year was this item created?* can now be paired with *what year was this item acquired by the collection?* With this in mind, patrons can then ask: *what are the institutional practices and biases governing the construction of this collection?* Such questions are enabled by Repcol's inclusion of administrative metadata, which begin to reveal the labour and decision-making that is required to build a collection. This transparency affords a different mode of engagement with the collection—a line of questioning that begins to unveil the systems and labour behind digitised collections.

Together, these examples represent a spectrum of transparency in digitised collection interfaces. Whitelaw's *Manly Images* and Hinchcliffe's *Visionaries Explorer* provide a foundation of generosity. Their overviews and multiple levels of detail readily offer rich collections for browsing. *Loom*, from SLNSW's own DX Lab, demonstrates a small but elegant inclusion of administrative data which hints at the ongoing nature of digitisation. It is a strong example of how transparency might be worked into existing generous interfaces. Sherratt's *Real Face of White Australia* presents a focused subset of its collection data, and is transparent about how this subset was extracted. It forms a compelling alternate interface which

encourages further exploration of the NAA's database. Finally, Bengler's *Repcol* deliberately de-emphasises images of artworks, creating space for administrative data to shape the interface. It employs visualisation techniques to highlight these facets, and creates transparency by applying the generous overview administrative metadata.

Clearly, there is potential to integrate transparency into generous interfaces. Generous interfaces, by design, prioritise the rich representation of collection data. As a result, they often omit the logistical complexities of digitisation—but the expansive representation of digitised collections and a view into the ecosystem that supports them are not mutually exclusive. To draw this potential out of generous interfaces, this project makes transparency its priority. It draws upon lessons learned from these successful generous interfaces: the significance of overviews and the ability to rearrange a collection, the viability of an interface focused on a particular subset of a collection, and the potential to enable new readings of a collection by including administrative data. This project then applies these techniques of generous interface design to administrative and structural metadata, working towards an interface that integrates transparency and generosity. It offers a rich and multifaceted view of the collection, and materialises academic and professional discussions of transparency in the GLAM sector by making visible the infrastructure, processes, and labour that support digitised collections.

METHODS

1. METHODOLOGY

In this project, I approach the exploration of transparency and generosity in digitised library collection interfaces with a Research through Design methodology. My approach exemplifies what Christopher Frayling has defined as Research through Design, where design practice is the initiator of, and vehicle for, research into questions outside design (Durrant and Price, 2015). In a Research through Design project, design methods are employed as tools of knowledge production; designerly knowledge of materials, technologies and practices are applied in the investigation of a research question (Frayling 1994, p. 5).

In an article discussing Research through Design in the context of a research project focusing on digitised archives, Tom Schofield, Mitchell Whitelaw, and David Kirk note that 'In RtD work... there is precedent for experimental design work which has a commitment to uncovering new kinds of knowledge through the production of artefacts' (2017, para. 6). As well as enabling the production of knowledge, these artefacts also embody that knowledge, giving tangible presence to complex research findings. This approach is well-suited to the exploration of new ways to represent digitised collections; the production of artefacts through the research process parallels the production of possible representations of digitised collections. Schofield et al. cite Johanna Drucker's assertion that 'all forms of design share a propositional orientation that is well-suited to the challenges that come with designing new structures for design asks 'What if?' (Schofield et al, 2017, para. 7). Research through Design adopts this propositional orientation, applying it to research outside design to provide new methods of enquiry and modes of communication within these fields.

The Research through Design methodology enables the production of artefacts as practical embodiments of research findings. In this project, the design artefacts developed throughout the research process reflect discussions of transparency in libraries that are already taking place.

The work of Mattern and Padilla demonstrates an existing attention to the labour, infrastructures, and implications of library digitisation in an academic context. Digital humanities research by Florian Kräutli and Stephen Boyd Davis is built on the observation of 'an increased awareness of the possible merits of metadata', citing a number of institutions (including tcvhe Cooper-Hewitt, Tate Britain, and MoMA) making their collections metadata available for download (2016, p. 10) through application programming interfaces (APIs), public repositories on GitHub (a platform for sharing open-source code), and direct downloads. However, despite clear concerns regarding transparency and metadata in both academic and professional contexts, these discussions have been slow to make a discernable impact on library collection interface design. This project combines these existing perspectives and concerns with critical analysis of existing interface design, and employs Research through Design methods—primarily diagramming and prototyping—in order to materialise these concerns in an artefact.

2. DESIGN PROCESS

Statement of Requirements for the Digitisation of Printed Materials

After initial consultation with digitisation staff, the State Library of New South Wales provided me with their Statement of Requirements for the Digitisation of Printed Materials. The document describes the specifications for digitisation and related tasks at SLNSW. As much of the library's digitisation is completed by third-party contractors, the document (updated in 2017) ensures consistent completion of digitisation tasks, ranging from photography and image processing to file storage and metadata management. The Statement of Requirements provides valuable information on the processes of creating and maintaining digitised library collections. It is also a dense technical document that is difficult to digest. As part of a designerly process of close reading and visual note-taking, I created a diagram of the document's contents [figure 15]. The diagram encapsulates much of the information in the Statement of

Requirements in a form that affords quick reference, and foregrounds the interconnectedness of the document. Archival master images are created in line with strict specifications, and are stored under specified file naming conventions. Specifications for image capture, image storage, file naming, administrative and technical metadata, and data fixity; these discrete but related pieces of technical information are represented in relation to each other by proximity and linear connection.

Figure 15:

The full infrastructure diagram, as it appears in Delivering Discovery. The diagram is a visual performance of the complex, entangled connections between collection data and descriptive, structural, and administrative metadata.

While it is tempting to argue that this is a natural representation of an interrelated information ecosystem, it's imperative to maintain an understanding of the discursive quality of diagrams. As Drucker reminds us, 'every graphic representation is a rhetorical device. Every presentation structures arguments—it doesn't "reveal" facts in all their purity through the fallible, flawed system of graphical expressions' (2010, p. 23). The Statement of Requirements diagram may appear to be a logical or self-evident display of information, but its presentation alongside a digitised library collection interface is an intentional visual argument that the records in the DSM collection—and indeed, in all library collections—should be understood in terms of their relationship to such an ecosystem. This is especially true in its final presentation in *Delivering Discovery*, where interaction and animation involve the user in the performance of unfolding the complex network of connections.

The production of diagrams allows design practitioners to explore and represent complex information systems through graphical means. In her foundational study of visual knowledge production, Johanna Drucker asserts that 'Graphics are economical, efficient, and offer a way to grasp complex relations in an immediately legible form' (2010, p. 12). Drucker describes three ways in which images embody information: 'They can work 1) through offering a visual analogy or morphological resemblance, 2) through providing a visual image of non-visible phenomena, or 3) by providing visual conventions to structure operations or procedures' (2010, p. 5). This variety of modes of containing information in graphical forms makes diagrams useful across multiple fields, from mathematics, sciences and engineering to the humanities. Of particular interest for this project is the second mode; the ability to provide a visual representation of—and thus the ability to communicate economically about—complex, invisible systems. This aspect of diagramming offers a strong foundation from which to approach the representation of the largely invisible labour and infrastructures of digitisation in library collection interfaces.

The Statement of Requirements diagram exemplifies the dual purpose of diagramming within the project: investigating the interconnected ecosystem of digitisation, and communicating that complex system to users of the interface. Drucker's discussion of information graphics describes this duality: '[information graphics] are expressions of procedures for generating knowledge through the act of visualization and ways of displaying knowledge embodied in visual imagery' (2010, p. 31). The diagram is designed to communicate that archival master images, which many digitised collection interfaces focus on, are surrounded by a large and complex network of processes and infrastructure—to intentionally visualise a network which is often unseen. Initially created as a form of visual note-taking, the diagram came to be included in *Delivering Discovery* to act as a rhetorical device which demonstrates the presence of these infrastructures in libraries, and calls into question their invisibility.

Metadata definitions

One of the most productive insights into the process of digitisation provided by the Statement of Requirements was found in the glossary: a definition of metadata with sub-categories describing the varied roles of metadata in digitised library collections. SLNSW's Statement of Requirements describes a metadata taxonomy established by Robin Wendler (1999), which includes three main categories (State Library of New South Wales, 2017, p. 29):

1. '**Descriptive** metadata describes a resource for purposes such as discovery and identification. It can include elements such as title, abstract, author, and keywords. This is critical for Discovery.'

2. '**Structural metadata** indicates how compound objects are put together, for example, how pages are ordered to form chapters. This is critical for Delivery.'

3. '**Administrative metadata** provides information to help manage a resource, such as when and how it was created, file type and other technical information, and who can access it. There are several subsets of administrative data; two that are sometimes listed as separate metadata types are rights management metadata, which deals with intellectual property rights, and preservation metadata, which contains information needed to archive and preserve a resource.'

As an example, *The Dream of Pythagoras and other poems* is one recently digitised book from the DSM collection. In the library's database, it is represented by an intellectual entity folder containing archival quality images of its pages, and metadata are stored alongside and within these images. The descriptive metadata annotate the book in familiar terms; written by Emma Tatham in 1854, published by Longman and Co. in London, 216 pages, 18 centimetres tall. This description emphasises a continuity between the analogue book and its digital surrogate—unlike the

other categories of metadata. Structural metadata largely take the form of file naming and storage; each archival master image is named in sequence, prefaced with the book's unique digital identification number, and stored together in a folder under the same number. This number (b06961 for *The Dream of Pythagoras*) along with other information, makes up the administrative metadata; the book is part of digitisation batch D47357, match point 1861919, call number DSM/821.89/T, barcode 1990627, out of copyright, captured March 27th 2017, on device 2, by operator 16 (whose name, usually included in this dataset for quality assurance purposes, has been anonymised for this project).

These definitions provide a useful set of terminology, and reveal the logic behind the exclusion of certain types of metadata from library interfaces. Descriptive metadata contain information that is 'critical for Discovery'; they annotate a digitised book with familiar descriptors that can form key search terms. Structural and administrative metadata, although vital, are unfamiliar to library patrons. They are treated as information that, while 'critical for Delivery' of library data, are not themselves significant data to be delivered. This view of structural and administrative metadata relegates them to the 'back end' of library operations; 'behind the interface, deeper inside the black box,' as Mattern describes it (2014, para. 50).

Metadata are often understood as supplementary, data-about-data, but metadata schemas exert agency over the collections they describe. In her writing on meta-languages for online markup, Johanna Drucker calls attention to the way in which metadata systems 'prescribe and delimit domains, and set out the parameters on which knowledge is shaped and bounded. Information, interpretation and knowledge have to fit the model encoded in the metatext' (2007, p. 179). Establishing a metadata framework means establishing a model for the objects it describes, and the rigidity of these digital metadata systems enforces the bounds of that model. In this way, metadata systems impact how we can describe an intellectual entity, and as a result, what we can know of it. For instance, SLNSW's digital records will never tell us the weight and texture of the paper stock upon

which *The Dream of Pythagoras* and other poems was printed, or which of its poems was David Scott Mitchell's personal favourite—these dimensions of the book are not accounted for.

With this in mind, we can understand the different types of metadata as models for conceiving of an intellectual entity in multiple ways. Descriptive metadata understands the item as a unit of knowledge, interconnected with subject areas, people, places, and times. Structural metadata understands the item as being comprised of various discrete pieces that need to be kept together and in sequence. Finally, administrative metadata understands the item as an object to be quality controlled, tracked, located across multiple systems, and preserved over time. These metadata definitions provide three distinct perspectives on digitised books, and present a productive foundation for experimentation with alternative representations for library collections.

Prototyping representations of digitised library collections

Soon after receiving the Statement of Requirements document, SLNSW provided access to collection data, recently digitised according to the specifications of the then newly-updated specifications. That data was a subset of SLNSW's David Scott Mitchell collection, a significant collection of Australiana containing 'books, maps, works of art, photographs, coins, stamps, magazines, ephemera, sound recordings and electronic resources' (State Library of New South Wales, 2015a). The subset, digitised in the 2016–2017 financial year, contains approximately 2600 digitised books: SLNSW provided access to the external metadata CSV for this batch of digitised records, containing a wealth of descriptive, structural, and administrative metadata. Working with the metadata definitions provided in the Statement of Requirements as a starting point, I began to explore possible representations of the DSM data through an iterative prototyping process.

Prototyping is a significant method in Research through Design. In the context of this project, a prototype can be understood as a propositional design outcome; an experimental artefact that can be used to explore a research space, test a hypothesis. Rather than producing fully functioning design outcomes, prototyping enables designers to quickly develop semi-functional provocations, seek feedback from peers, and continue the research process. Schofield et al. frame design prototypes as expressions of the 'concern for the relationship between artefacts and knowledge' that characterises Research through Design (2017, para. 27). Prototypes present 'propositional or provocative stances' (2017, para. 3), enabling discussion of the artefact's implications and embedded knowledge. Working with prototypes often also entails an iterative design approach, which is well-suited to a research context. Multiple prototypes are often developed throughout the research process; each one provokes a response from 'experimental users' (2017, para. 3), driving the development of the next prototype. Together, the propositional and iterative aspects of prototyping make it a productive method for developing and presenting new approaches to the representation of digitised library collections.

Over the course of this project, I developed five prototypes, the last of which is the *Delivering Discovery* interface. Each prototype tested one possible approach to the representation of digitised collections through the lens of their metadata. The prototyping process facilitated investigation of a variety of forms; video installations, books, and finally interactive web interfaces. The prototypes reflect the project's developing approach to the value of metadata, particularly structural and administrative metadata, as a means to creating more transparent representations of digitised collections.

The project's first prototype, the Document Reader, is a 3-channel video installation which displays DSM collection data and metadata. The installation separates the data into three categories: descriptive metadata, administrative metadata, and the archival master images. Each category is displayed on a separate screen. The installation randomly cycles through

books in the collection, with each of the three screens synchronised. For a given book, one screen displays a grid of digitised pages [figure 16], another screen displays descriptive metadata such as title and author [figure 17], while the third screen presents the often-unseen administrative metadata such as capture date, batch number, and copyright information [figure 18]. The installation uses the different screens as different windows into the collection items, each one providing its own insights. It is an initial response to the metadata definitions in the Statement of Requirements, presenting the categories as distinct but equal schemas with which a digitised collection item is described.

Figure 16:
Empty rectangles act as placeholders for digitised collection images in this prototype. Images referred to by contractor notes are highlighted, calling attention to images which may be of interest.

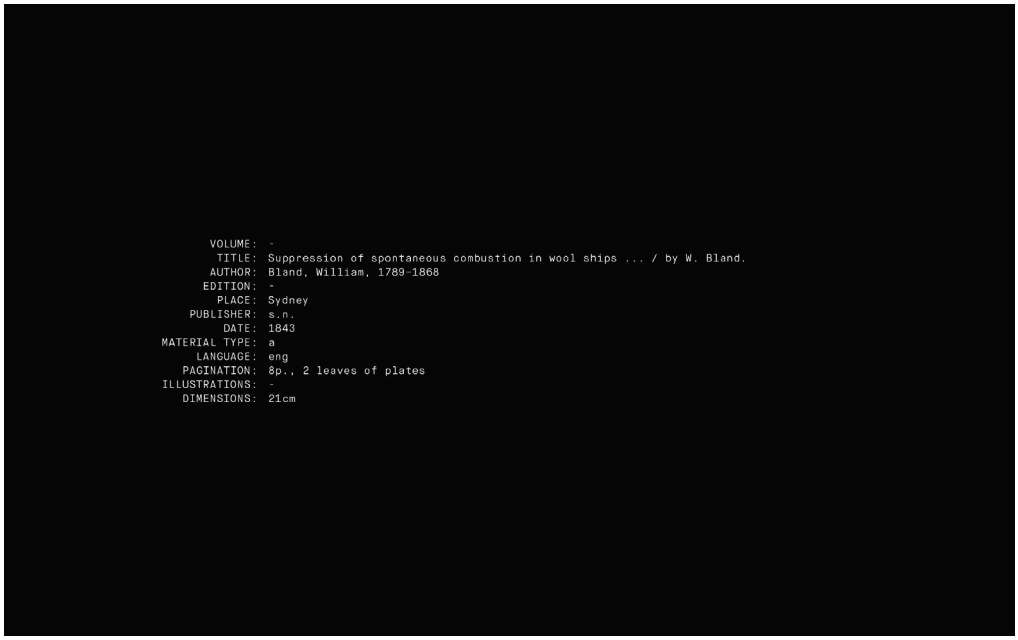


Figure 17:

Descriptive metadata primarily contains human-readable fields like title, author and place. These metadata sketch a recognisable image of a book; a title and author, twenty one centimetres tall, eight pages, with illustrations.

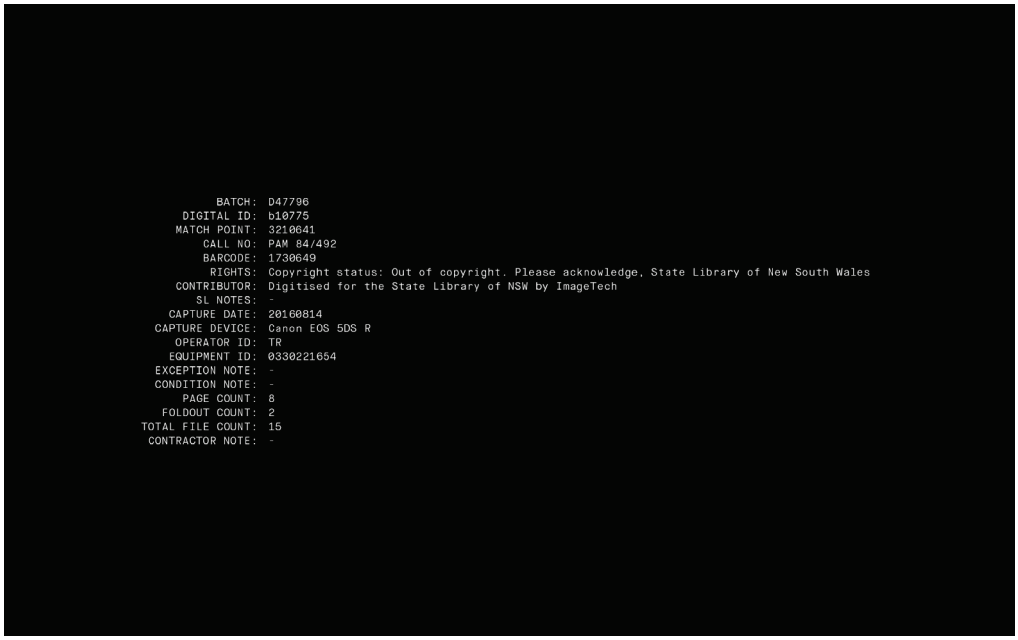
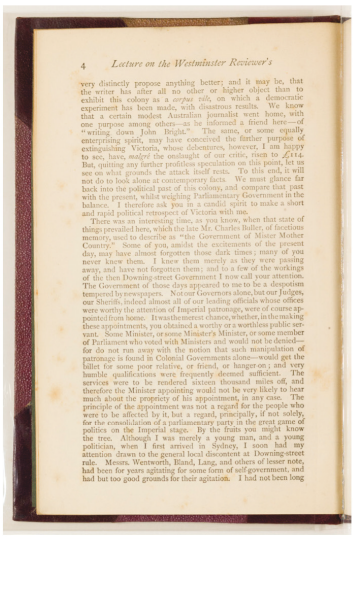


Figure 18:

Administrative metadata leans towards machine-readability. Digital ID, match point, and call number each uniquely identify the book across various library interfaces, while other fields like capture date, batch, operator ID and contractor note are required for quality control and resource management.

The next prototype continued to experiment with the side-by-side display of master image and metadata, this time in a print form. The Metadata Book prototype [figure 19] is a speculative design object which interleaves images of digitised book pages with a printout of the metadata embedded in their TIFF tags. These tags include identifying information such as record name and batch number, date and time of image capture, and technical details pertaining to the reading of the image file itself. Throughout the development of the prototype, exploration of this metadata led to the extraction of image creation timestamps, and the subsequent estimation of time taken to digitise books in the DSM collection. According to these estimates, the digitisation of books in this subset of the DSM collection took anywhere from a few minutes to four hours. These figures are



4 *Lecture on the Westminster Reviewer's*

very distinctly propose anything better; and it may be, that the writer has after all no other or higher object than to exhibit this colony as a *corpse rule*, on which a democratic experiment has been made, with disastrous results. We know that a certain modest Australian journalist went home, with one purpose among others—as he informed a friend here—of "writing down John Bright." The same, or some equally enterprising spirit, may have conceived the further purpose of extinguishing Victoria, whose debentures, however, I am happy to see, have, *malgré* the onslaught of our critic, risen to £114. But, quitting any further profitless speculation on this poser, let us see on what grounds the attack itself rests. To this end, it will not do to look alone at contemporary facts. We must glance far back into the political past of this colony, and compare that past with the present, whilst weighing Parliamentary Government in the balance. I therefore ask you in a candid spirit to make a short and rapid political retrospect of Victoria with me.

There was an interesting time, as you know, when that state of things prevailed here, which the late Mr. Charles Buller, of facetious memory, used to describe as "the Government of Mister Mother Country." Some of you, amidst the excitements of the present day, may have almost forgotten those dark times; many if you never knew them. I knew them merely as they were passing away, and have not forgotten them; and to a few of the workings of the then Downing-street Government I now call your attention. The Government of those days appeared to me to be a despotism tempered by newspapers. Not our Governors alone, but our Judges, our Sheriffs, indeed almost all of our leading officials whose offices were worthy the attention of Imperial patronage, were of course appointed from home. It was the merest chance, whether, in the making these appointments, you obtained a worthy or a worthless public servant. Some Minister, or some Minister's Minister, or some member of Parliament who voted with Ministers and would not be denied—for do not run away with the notion that such manipulation of patronage is found in Colonial Governments alone—would get the billet for some poor relative, or friend, or hanger-on; and very humble qualifications were frequently deemed sufficient. The services were to be rendered sixteen thousand miles off, and therefore the Minister appointing would not be very likely to hear much about the propriety of his appointment, in any case. The principle of the appointment was not a regard for the people who were to be affected by it, but a regard, principally, if not solely, for the consolidation of a parliamentary party in the great game of politics on the Imperial stage. By the fruits you might know the tree. Although I was merely a young man, and a young politician, when I first arrived in Sydney, I soon had my attention drawn to the general local discontent at Downing-street rule. Messrs. Wentworth, Bland, Lang, and others of lesser note, had been for years agitating for some form of self-government, and had but too good grounds for their agitation. I had not been long

Figure 19:

TIFF images, the format used by SLNSW, contain both pixel data and tags which annotate the image with metadata. The Metadata Book prototype presents images and metadata side-by-side, highlighting in pink the fields that are unique to this particular image. The non-unique fields are those which remain consistent between images, including image formatting and camera settings.

especially interesting given that this subset of the collection is comprised of pages with text only, which are less time-consuming to photograph than pictorial content. These estimations of digitisation time frame a digitised book as the product of human labour, rather than a folder of digital images that simply exists. This prototype ultimately called more attention to evidence of the labour of digitisation embedded in administrative metadata, helping to identify that category of metadata as having potential in finding new perspectives on digitised collections, and spurring on the development of web-based prototypes which could allow users to actively engage with such metadata.

Following these experiments in video installation and books as potential forms for alternative representations of digitised collections, the next prototypes followed the lead of existing library interfaces by transplanting the focus on metadata exploration into an interactive, web-based context. The first web interface prototype [figure 20] presents digitised library records in a scatter plot, arranged by date of publication on the horizontal axis and date of digitisation on the vertical axis. The plot can also be filtered by digitisation operator and total file count. Enabling interactive exploration and filtering with collection metadata is engaging, but the interface lacks clarity and purpose. Without context for these administrative facets, it was difficult to filter the records in a way that enabled intentional searching or interpretation. Nevertheless, this prototype was successful as a proof-of-concept interactive web interface centered on administrative and structural metadata; what began as a simple scatter plot revealed a curious narrative embedded in digitisation dates, which enables broader questions to be asked of libraries through their collections. How is the labour of digitisation organised? What are the conditions for digitisation workers? This focus on labour as it is revealed by collection metadata is key to transparency, and was carried forward into *Delivering Discovery*.

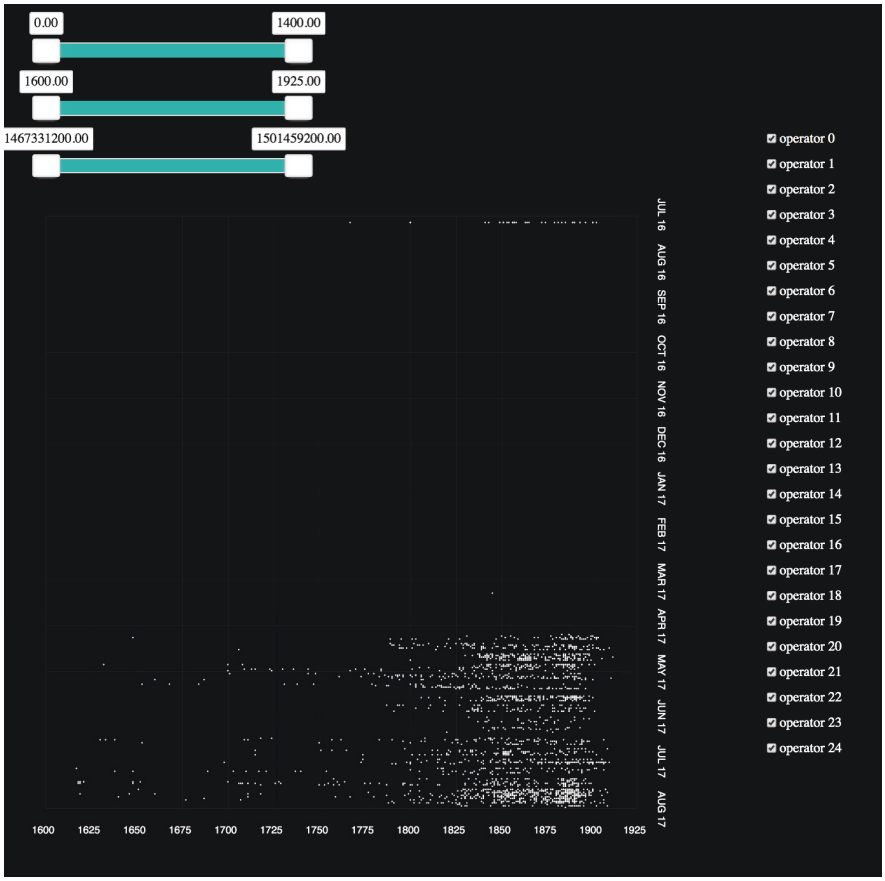


Figure 20:

Creating this interface led to the immediate realisation that date of digitisation might provide interesting insights. The complete lack of records between July 2016 and March 2017, which initially appeared to be an error, is actually evidence of the schedule of the digitisation project. This accidental discovery, brought about by plotting records by administrative metadata, was carried forward into the final Delivering Discovery interface.

To respond to the previous prototype's lack of guidance, a second web interface prototype was designed [figure 21] with a linear structure and annotations, guiding exploration of administrative metadata facets.
In order to use this metadata as the organising factor of the interface, collection records are laid out in a two-dimensional plot, generated using a t-SNE algorithm which 'visualizes high-dimensional data by giving each datapoint a location in a two or three-dimensional map' (van der Maaten and Hinton, 2008, p. 2579). The process was applied to arrange digitised books on a two-dimensional map by their similarity in many dimensions—in this case, by their administrative metadata facets. Books with similar page counts, digitisation operators, digitisation dates and contractor notes appear in proximity to each other, presenting an alternative arrangement of a digitised collection which foregrounds administrative metadata.
While this experimental approach to layout provided a visually interesting interface, it proved difficult to interpret, detracting attention from the records themselves. As a result, these prototypes were not developed past the mockup stage. However, this prototype cemented the potential of structuring a digitised collection interface around administrative metadata. Where this prototype attempted to compress the many facets of administrative data into a single two-dimensional representation, the final prototype embraces the potential of multiple viewpoints.

The iterative prototyping process enabled exploration of different material forms and focuses of a variety of representations of digitised collections. The Document Reader and Metadata Book demonstrated the value of administrative metadata in providing insight into the labour and infrastructures behind digitised collections. The initial web interface prototypes provided direction for how to present this information in the common and highly accessible form of a web-based, interactive interface. This process culminated in one final prototype, Delivering Discovery. While more complete in design and functionality than the previous prototypes upon which it builds, its purpose is similar: to act as a provocation that argues for the potential of extending the notion of generosity in digitised library collection interfaces to administrative metadata in order to create more transparent interfaces.

Figure 21:

These mockups depict a possible web interface. Each point in the visualisation represents one of the 2,579 books, and the organic layout is the result of the t-SNE algorithm. Such a layout can be difficult to interpret, but is based entirely on the books' administrative metadata, and reveals unexpected relationships between records.

DELIVERING DISCOVERY

Delivering Discovery is a prototype web interface for a subset of SLNSW's David Scott Mitchell (DSM) collection, containing approximately 2600 digitised books. The interface applies characteristics and conventions of generous interface design to the collection's structural and administrative metadata. By focusing on these metadata, the interface highlights the labour and infrastructures that create and support digitised collections, offering a prototype for a more transparent digitised library collection interface. This focus on infrastructure and metadata is articulated through the design of its two interactive views: the infrastructure diagram, and the collection visualisation.

Infrastructure Diagram

The infrastructure diagram view is an interactive, animated diagram of SLNSW's recently updated Statement of Requirements for the Digitisation of Printed Materials. As previously discussed, the Statement of Requirements document was written to provide third-party digitisation contractors and internal staff specific, technical instructions for the capture, storage, and maintenance of digitised printed material at SLNSW. This wealth of information ranges from instructions on how printed material should be photographed (eg. file format, resolution, cropping, background, width of borders), to how those images are quality-controlled and stored (eg. file naming, folder organisation, reference images), and what metadata is included in and around the digital images (eg. unique identifiers for multiple databases, date of digitisation, digitisation operator). By translating the document into a diagram, the dense technical information can be presented in a more approachable form; as Drucker argues, the diagram leverages its graphical economy and efficiency to 'offer a way to grasp complex relations in an immediately legible form' (2010, p. 12), and to act as a rhetorical device which positions collection media within a complex ecosystem.

DELIVERING DISCOVERY

When users first enter the diagram view, only the archival master image is revealed in the diagram [figure 22]. These archival master images are often the highest priority in a digitised collection interface (often implicitly presenting them as the sole product of the process of digitisation), which makes them a familiar starting point for the diagram. As users progress through the linear structure of the diagram view, more sections of the diagram are revealed; digitisation specifications, file storage, reference images and quality control, checksum data, and embedded and external metadata [figure 23]. These sections are accompanied by annotations which explain the significance of this web of information; not only their role to the digitisation process, but also their relevance to researchers and other library patrons. As these interconnected areas of the diagram are revealed, the diagram begins to position archival master images within what Mattern describes as an 'infrastructural ecology': a complex material ecosystem of labour, hardware, and software, all of which are active in the images' creation, retrieval, and use (2014, para. 22).

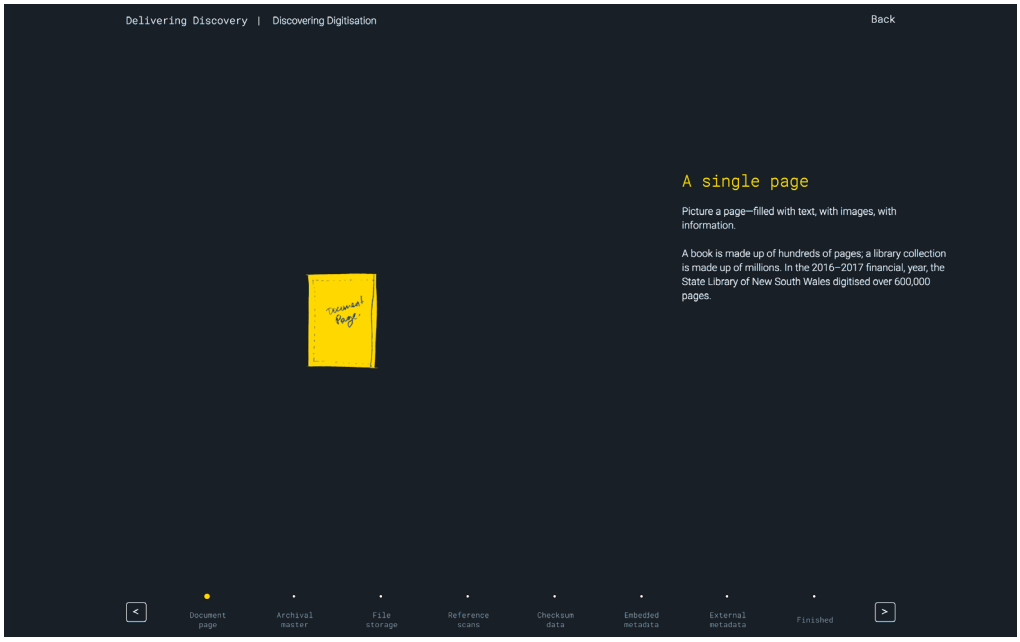


Figure 22:

Like all digitised books, the infrastructure diagram begins with a single page. This one digitised image represents a familiar starting point for users of digitised library collections. As the diagram slowly unfolds, the page is recontextualised in a complex ecology of hardware, software, metadata and labour.

This infrastructure diagram exemplifies *Delivering Discovery*'s shift in focus from the digitised master image to the infrastructure that supports it. Visually depicting this often-unseen infrastructure allows library patrons a view into the extensive ecosystem that surrounds and produces digitised collections, and offers a resource for researchers and GLAM professionals investigating these systems. It creates a more transparent interface by displaying, in direct connection to archival master images, the infrastructures and labour that creates digitised collections. Visualising archival master images—and by extension, digitised library collections—as products of a complex network of labour, hardware, and software provides a new lens through which we can view, use and analyse them. The diagram view also provides context and introduction for the website's visualisation view. It serves to call attention to the absence of this infrastructure in many collection interfaces, and primes viewers for the collection visualisation view, in which they will explore the DSM collection through the lens of this infrastructure.

Figure 23:

The development of the diagram is performative, slowly revealing the complex ecosystem of digitisation. Rather than overwhelming users with the full diagram, it is constructed over time. This makes the connections between sections clearer, and creates space for curiosity.

DELIVERING DISCOVERY

Collection Visualisation

Delivering Discovery's second view is a guided, interactive visualisation of the 2600 books in the 2017–18 subset of SLNSW's DSM collection. Each book in the collection is represented by a small rectangle, scaled according to its number of pages [figure 24]. Like the diagram, this view has a linear structure that guides viewers through the visualisation. As viewers explore, the books are rearranged, clustered into different groups based on facets (date of digitisation, digitisation operator, and quality control notes taken by operators) of structural and administrative metadata embedded in the books. The collection visualisation builds on the argument of the infrastructure diagram, using the administrative and structural metadata embedded in the collection as evidence of, and a lens through which to view, the ecosystem of labour and infrastructure that produces digitised library collections.

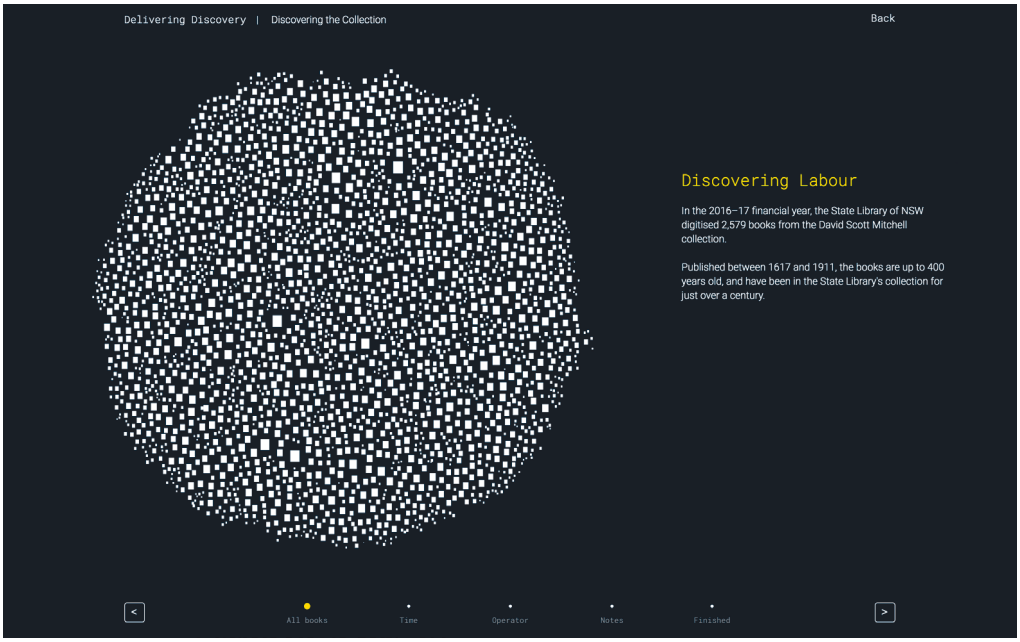


Figure 24:
The first stage of the visualisation introduces the collection as an unorganised cluster. It begins the process of familiarising users with the format of the visualisation, and creates space for the annotation to provide a general introduction before users dive into the collection's administrative metadata.

The individual images of pages in the DSM books are not included in the interface as they were unavailable at the time of development. Although the items were photographed, the images had not yet been ingested into SLNSW's public-facing digital collection, a separate process which involves the creation of lower resolution copies of the archival master images, and the indexing of these files in the appropriate databases. However, this logistical requirement presented an opportunity to break from the conventions of digitised library collection interfaces. A collection interface without collection media de-emphasises the digital image as the only significant product of digitisation. Removing these images may obscure our ability to see the collection—the content of the text, its formatting and layout, the presence of marginalia are all hidden. However, by departing from this naturalised and familiar representation, the interface provides an opportunity to engage with Padilla and Higgins' 'facet effect'; where digital images have a 'potentially alternate reality as data' (Padilla & Higgins, 2014, p. 328). The constellation of data and metadata that surrounds and describes the digitised books is a rich and varied set of information; it offers compelling evidence of the labour and infrastructures that enable library digitisation, and forms a significant resource for library patrons and researchers to browse, read, and utilise digitised collections in new ways.

It is this metadata-focused, multifaceted approach to representing digitised collections that informs the structure of the visualisation view. As users progress through the visualisation, the collection is reorganised according to different metadata facets, each one offering a different perspective on the collection as a whole. Here, the interface draws on the principle of providing multiple views of a dataset exemplified by generous interfaces like *Loom*. *Delivering Discovery* replicates this offering of multiple views, but extends the generosity to the administrative and structural metadata which reveal the organisation and processes which construct the collection.

DELIVERING DISCOVERY

Viewing the collection by digitisation date [figure 25] casts light on the logistical complexity of SLNSW's digitisation project, and presents digitisation as an ongoing process at SLNSW (and other cultural institutions). As we see with *Repcol,* this inclusion of administrative metadata opens the collection—and more broadly, the institution—to questions regarding what collection items (and what collections) are prioritised for digitisation. From curious patrons who wish to better understand the increasingly prevalent process of digitisation, to digital humanists and media scholars researching the development of digitised cultural collections, the inclusion of this data enables broader readings of the collection, and inquiry into the institutional processes that brought it into being.

Reorganising the collection by digitisation operator [figure 26] gives an insight into the human labour of digitisation; third-party contractors photographing individual pages with specialised book photography equipment. Although the contractors are anonymised, it becomes clear that the process of creating a digitised library collection is labour-intensive, requiring skill and care. As seen in the previous stage, more than half of the 12-month digitisation project was devoted to planning and training, and the Statement of Requirements diagram demonstrates the complexity of digitising printed material to an archival standard. These sections of the interface work to dispel any notions of digitisation as a simple or automated process. When users hover over individual books, the interface reveals an estimate of time taken to digitise books, which was extrapolated from image creation timestamps in the collection images [figure 27], as discussed in the methods section. Combined with an understanding of the time taken to plan, the technical standards set by the Statement of Requirements, and annotations describing contractors diligently recording quality control notes, the interface gives viewers a sense of the time and labour required to digitise library collections.

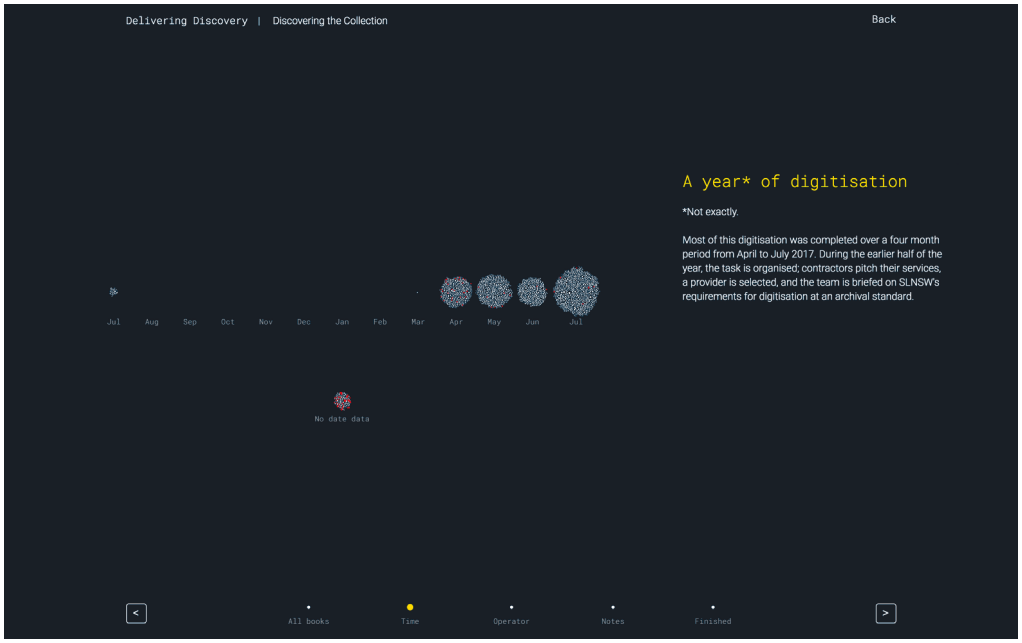


Figure 25:

Date of digitisation may not sound like interesting metadata, but using it to arrange the collection reveals the complexity of large-scale digitisation projects. The small group in July 2016 reflects a test batch of digitisation, after which external digitisation providers pitch their services, are reviewed and selected, and briefed on SLNSW's requirements. No digitisation work is completed until March 2017, at which point the year's project begins in earnest. All this information, reflected in one administrative metadata field.

DELIVERING DISCOVERY

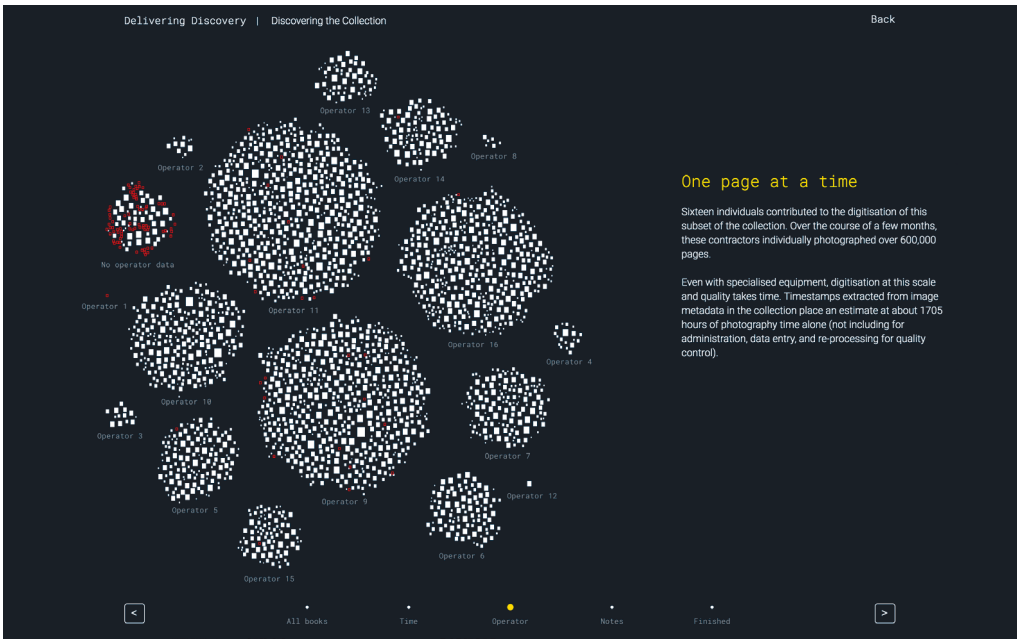


Figure 26:

Sixteen digitisation contractors were responsible for photographing the books in this subset of the collection.

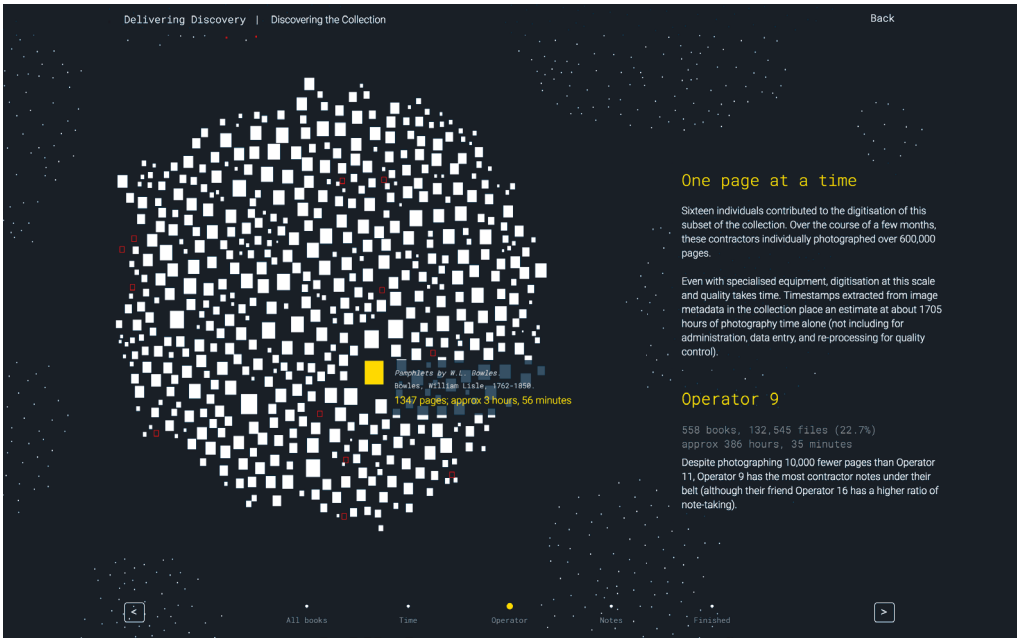


Figure 27:

Users can click on the clusters in each stage of the visualisation to zoom in for more detail, revealing a second text box with the data underpinning the visualisation. For some clusters, more detailed annotations are also provided, highlighting interesting details hidden in the dataset.

Finally, the collection is organised into groups based on notes taken by digitisation contractors [figure 28]. These notes record instances of damaged or fragile books, blank pages, inconsistencies in page numbering, the presence of foldout pages or images (for separate photographing), and so on. While these notes are intended for quality control purposes, for a browsing patron they might offer fascinating nucleation points for inquiry. From torn pages and crooked text, to bookmarks, bus tickets, and pencil illustrations, these notes have great potential to cut through the DSM collection and provide library patrons a new way to browse [figure 29]. Providing ways to browse digitised collections that engage with the physical materiality of the items they represent creates opportunities for compelling exploration of the collection items, avoiding the compression of rich, physical items into flat digital representations.

Each reorganisation presents a new perspective on both the collection, and the labour and infrastructure that creates and supports it. The collection items and the network that surrounds them are given equal presence, combatting the 'increasing recession of these technical infrastructures—and the human labor that supports them… behind the interface' that Mattern observes (2014, para. 50). In her writing on feminist approaches to conducting research in digital archives, Michelle Moravec cites Katherine Rawson's suggestion that 'the transfer of materials to a digital archival environment should entail leaving visible traces, the 'fingerprints' of individuals involved in the earlier collection and processing of materials' (Moravec, 2017, p. 191). From the institutional scale of organising a ten-year digitisation project, to the human scale of a note which reads: 'IMAGE NUMBERS MISLABELLED - THEY BEGIN AT #5. APOLOGIES!', these administrative and technical metadata contain the fingerprints of the individuals who construct and maintain our digitised library collections.

As with the infrastructure diagram, each arrangement of the collection visualisation are connected by an annotated linear structure. This linearity is relatively uncommon for digitised library collection interfaces, which reasonably prioritise open-ended searching and browsing. *Delivering*

DELIVERING DISCOVERY

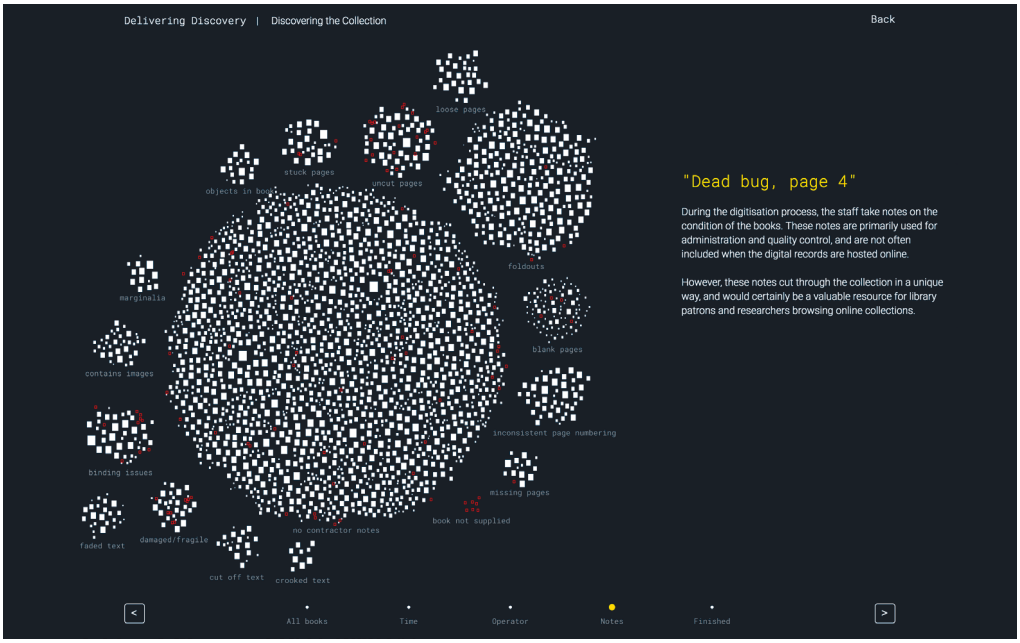


Figure 28:

Records are clustered based on the category of contractor notes taken. Providing these groupings encourages browsing by creating manageable subgroups that can be easily explored. At the overview level, it also reveals the types of abnormalities considered noteworthy by contractors during the digitisation process.

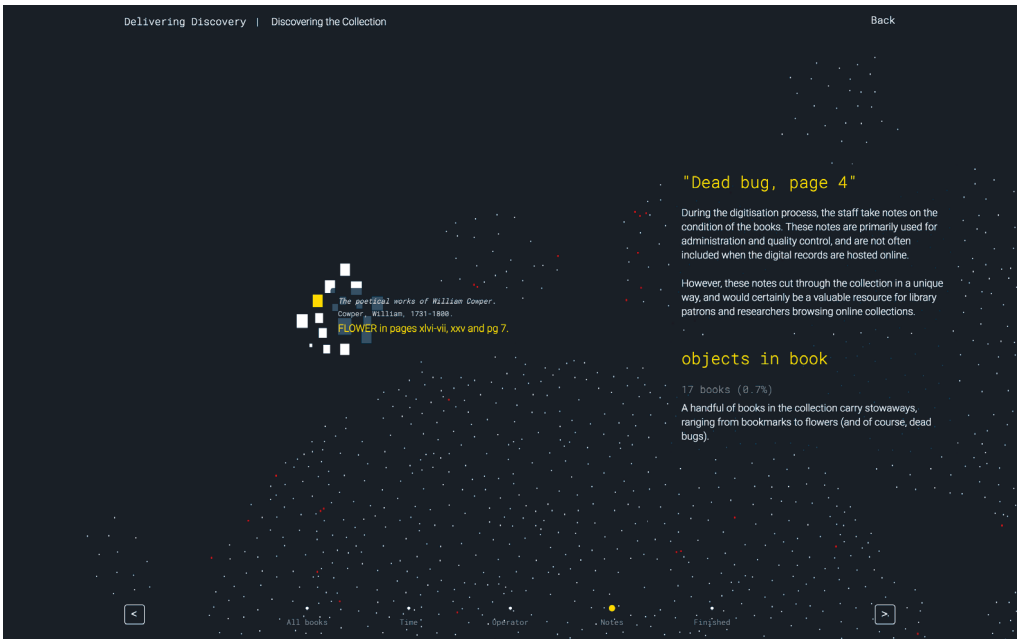


Figure 29:

Throughout the entire visualisation view, users can hover over every book to see its title, author, and the relevant facet from the current stage. Books in other clusters shrink down to single pixels which allows users to focus on the current records, while the other clusters remain clickable to support intuitive navigation of the visualisation.

Discovery's linear structure was developed in response to the often overwhelming array of choices made available in filterable collection interfaces. Interfaces which allow the filtering of a database often create what N. Katherine Hayles describes as 'relational juxtapositions' (2007, p. 1603) in digitised library interfaces, this usually equates to the presentation of a subset collection items that are related across one or more facets. This cluster of items can be interpreted by a user, often prompting further search/filter operations. However, this interpretation relies on the prior knowledge and experiences of the user.

Delivering Discovery does the same, but it represents collection items through unfamiliar metadata. Without familiarity with these data, interpretation is difficult. A filtering interface focusing on administrative metadata might allow filtering by digitisation date. This may be a useful search for an academic or GLAM professional researching patterns of library digitisation, but other patrons may not find that information interesting or useful in isolation. Hayles asserts that 'database can construct relational juxtapositions but is helpless to interpret or explain them, it needs narrative to make its results meaningful' (2007, p. 1603). By providing a linear structure and annotations, *Delivering Discovery* supports the interpretive process by offering a contextualising narrative for these unfamiliar metadata. In doing so, the interface builds familiarity with these metadata facets, and helps to construct them as a new lens through which to view the collection.

DELIVERING DISCOVERY

OUTCOMES

The design and development of *Delivering Discovery* was the primary vehicle for the project's research outcomes, in keeping with its Research through Design methodology. The project's main contribution is the practical integration of generosity and transparency in a digitised library collection interface. By extending our generosity to include broader varieties of data and metadata, generous interfaces can be designed to enable new interactions and support more users, particularly digital humanities and media studies scholars working in the GLAM sector. This outcome is substantiated by the *Delivering Discovery* prototype in three key ways. First, I argue that collection media aren't the only valuable products of digitisation for library patrons. Second, I challenge the naturalised omission of administrative and structural metadata in digitised library collection interfaces. Finally, I explore the possibility of more supporting structure in collection interfaces that provide context and guidance during searching and browsing.

This project asserts that administrative and structural metadata are valuable resources created during the process of digitisation, rather than by-products of the creation of digital images of book pages. Most existing digitised library collection interfaces prioritise collection media, usually images, in their representations; these media are made searchable, browsable, and downloadable. This accessibility is valuable to a wide variety of library patrons, and the ongoing development of generous interfaces is testament to the breadth of possible representations and applications of these images and other collection media.

However, the digitisation of library collection creates more than just images of book pages. During the digitisation process, a wealth of other data is generated—including, but not limited to, administrative and structural metadata. Some of these data are ripe for computational analysis and visualisation; dates of acquisition and digitisation, or embedded TIFF tags

containing metadata about resolution, compression, and photometric interpolation. Others offer rich, human-readable additions to a digitised collection—the carefully recorded notes on damage, fragility, marginalia and more—recorded by archivists and digitisation contractors—highlight existing pencil drawings, newspaper clippings and fabric scraps used as bookmarks, misprints in page numbering, and even dead bugs. Whatever their content, these metadata represent valuable archival objects of enquiry in their own right, and enable broader readings of digitised collections by researchers in digital humanities and media studies.

Through the *Delivering Discovery* interface, I foreground administrative and structural metadata. The interface's infrastructure diagram contextualises digitised collection records within a dense ecosystem of hardware, software, and labour. Furthermore, as a result of both logistical limitations and its function as a propositional prototype, *Delivering Discovery* completely omits collection images in its representation of the DSM collection, inverting the usual availability of digitised collection data. These factors result in an expansive and engaging representation of a digitised library collection. As the interfaces demonstrates, these other products of library digitisation can and should be included in collection interfaces—doing so would better reflect the 'ample abundance of our digital collections and the generous ethos of the institutions that hold them' (2015b, para. 3).

Following this assertion of the value of administrative and structural metadata, the project's focus on the infrastructural aspects of digitised collections challenges their naturalised omission from digitised collection interfaces. *Delivering Discovery's* visualisation view represents the DSM collection through the lens of its administrative and structural metadata. In constructing this interface, I present a prototype for generous interfaces which more visibly and centrally include these metadata. Administrative, structural, and other technical metadata are already present in digitised collections. While these metadata are necessary for quality control, resource management, and interoperability, they also enhance the richness

of a digitised collection, offering a multitude of new perspectives on not only the content of these collections, but the institutions, practices, and labour that digitise and maintain them. These insights into the conditions of the creation and maintenance of our cultural collections directly respond to Mattern and Padilla's concerns regarding transparency in digitised library collections.

This project foregrounds these metadata in our understanding of generous interfaces, making them accessible to library patrons and scholars who may benefit from the supplementary perspectives they provide. Understanding this approach to interface design as an extension of generosity towards collection metadata demonstrates the compatibility of this project's approach with existing generous interfaces. While *Delivering Discovery* centres on these metadata to the exclusion other modes of browsing, it represents only one possible iteration of an interface at the intersection of generosity and transparency. As Whitelaw states, 'other representations are always possible' (2015a, para. 1). This project acts as a provocation to continue working towards more transparent interfaces by broadening the varieties of collection data that are represented generously.

Finally, I explore the potential of a more structured collection interface to provide context and guidance during browsing. Most collection interfaces have non-linear structures; that is, exploration of the collection is open-ended and not contained in a sequential order. Some interfaces, like *Visionaries Explorer* [figure 30] and *Coins* [figure 31], provide all their tools for searching and browsing (eg. keyword search, tags, filtering) simultaneously, through one complex interface. Interfaces like *Loom* [figure 32] separate their modes of browsing (eg. by time, by location) into different screens or views. In each case, these different modes of searching and browsing are unordered, and users are encouraged to exploration the collection at will. This unguided structure benefits from a user's familiarity with the tools and the collection in order to find relevant collection items.

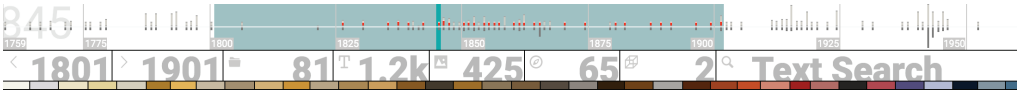


Figure 30:

Visionaries Explorer clusters all its searching/filtering tools in a complex but powerful toolbar.

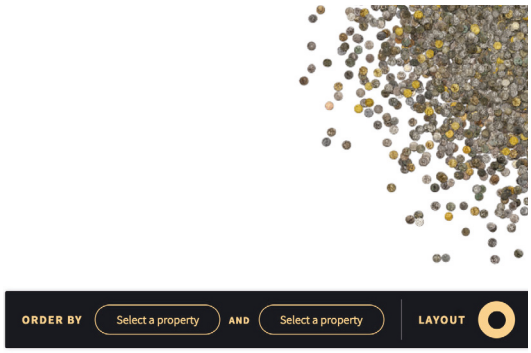


Figure 31:

Filtering options in Coins are made available contextually, reducing the number of options on screen at a given time.

Figure 32:

Loom's multi-faceted representation of the collection prioritises different metadata in each view, allowing users to focus in on a mode of browsing that best suits their needs.

Because *Delivering Discovery* represents collections through unfamiliar metadata, depending on users' prior knowledge of administrative and structural metadata and their significance was an unreliable approach. As a result, I experimented with a more linear interface which provides guidance through its structure and annotations. Both the infrastructure diagram and collection visualisation have a sequential, annotated structure which guides users through the content. The interface still allows for exploration and browsing, but places this within a guided context to support the user's interpretation of the representation of the collection.

This approach has both benefits and drawbacks. By providing a linear structure and explanatory annotations, generous interfaces may be able to better support searching and browsing in specialised collections and interfaces by offering a contextualising narrative. This can help library patrons to develop familiarity with these representations and metadata, and can be achieved without undermining searching and browsing as the primary modes of interaction for these interfaces. On the other hand, constraining searching and browsing behaviour to a linear structure may feel limiting to library patrons, and could reduce the reusability of an interface. Such guided structures may also rely on a particularly focused collection interface, like *Delivering Discovery* or *The Real Face of White Australia*. It would likely be more difficult to create such annotations for more generalised and open-ended collections. Nonetheless, this exploration of guided collection interfaces makes up part of *Delivering Discovery's* contribution to the integration of generosity and transparency in collection interfaces.

Whitelaw's *Generous Interfaces* establishes a strong theoretical framework for the continued development of interfaces to digitised cultural collections. This framework describes (and encourages the development of) interfaces which openly share the content of their collections to 'better match both the ethos of collecting institutions, and the opportunities of the contemporary web' (Whitelaw, 2015b, para. 46). *Delivering Discovery* operates within this framework. It utilises techniques of generous interface design—contextualising overviews, multifaceted representations, and multiple levels of detail—and applies them to unfamiliar facets of digitised library collection data to present a more transparent interface to a digitised library collection. *Delivering Discovery* centers on administrative and structural metadata in the DSM collection, highlighting valuable, pre-existing data and challenging their omission from collection interfaces. It avoids familiar, image-based representations, instead foregrounding these hidden metadata to leverage the 'facet effect' described by Padilla and Higgins, where

digitised collection items can be viewed through other facets of their data to create a new perspective, to 'encourage uses beyond reading, viewing, and listening' (Padilla & Higgins, 2014, p. 324). These data are presented in a linear structure that provides context to support the browsing experience and helps patrons make sense of an unfamiliar representation.

With this project, I seek to expand notions of generosity by combining generous design approaches with a more transparent representation of the infrastructures that support digitised library collections. Alongside existing generous representations of collection media such as images, audio and text, *Delivering Discovery* demonstrates the potential of extending generosity to administrative and structural metadata. It affords broader use cases for researchers and the wider audience of library patrons, and provides new insight into digitised library collections and the institutions that maintain them.

CONCLUSION

The digitisation of library collections presents unprecedented opportunities for access, display, and research. As the number of digitised collections worldwide increases, so too does the breadth of digitised collection interfaces. Designers and developers continue to explore the framework of generosity, creating interfaces that represent digitised collections openly and vibrantly. These interfaces create approachable and explorable representations that reflect the depth and richness of digitised cultural collections. At the same time, many digital humanities and media theories scholars are calling attention to a lack of transparency that problematises the paradigm of open access to digitised library collections. Without a critical understanding of how library collections are digitised and disseminated, how can library patrons access and re-use digitised collections ethically? The integration of these two frameworks, generosity and transparency, presents great potential for interfaces that combine media-rich, browsable representations with greater visibility of the infrastructures and labour which construct and maintain our digitised library collections. Such an interface could offer a transparent view of the ecosystem of digitisation and provide a new lens through which to reveal the depths of digitised collections.

In this exegesis, generosity and transparency are deployed with specific definitions. Generosity is concerned with the ways in which collection data are made available to patrons, and is productive in developing interfaces which create expansive representations of digitised collections. Transparency calls into question which data are made available through these interfaces, framing the (in)visibility of the labour and institutional processes that construct digitised library collections as an ethical concern. This framing of transparency aligns it with ongoing discourses surrounding the ethics of developing and accessing GLAM collections.

Nonetheless, some of the most pressing ethical concerns for contemporary Western GLAM institutions are bound up in the relationships between historic collections and the processes of colonisation. There are 'geographic, political, and even linguistic boundaries' involved in accessing archival material, especially when researching histories of colonisation or state oppression, as pointed out by Indigenous Canadian scholars Crystal Fraser and Zoe Todd (2016, p. 34). Further to this, the authors remind us that as Indigenous scholars, 'a fundamental challenge lies in the fact that the majority of archival documents in Canadian archives have been produced by non-Indigenous people,' (Fraser & Todd, 2016, p. 35). This is also true of collections housed in Australian GLAM institutions, and is a reality that impacts how marginalised groups are understood within and beyond digitised cultural collections. If we are to be truly generous with library collections, these concerns must be reckoned with—the potential generosity of digitised collection interfaces is limited by the silences in the collections they represent. It is beyond the scope of this project to address these challenges. However, further research could explore the ways in which the concepts of generosity and transparency could be extended to address these challenges, particularly in digitised collections constructed and maintained in colonised states like Australia. Design research provides a productive ground for reimagining the possibilities of representing, accessing, and engaging with digitised collections.

This project approaches the design of a digitised library collection interface through a Research through Design methodology, in order to create a prototype that utilises conventions and techniques of generous interface design to materially respond to digital humanities scholars' calls for transparency in digitised collections. The result, *Delivering Discovery*, is a digitised collection interface prototype which demonstrates the value of integrating generosity and transparency. This project positions administrative and structural metadata as valuable products of the digitisation process, as objects of enquiry in their own right. To do this, it foregrounds administrative and structural metadata by de-centering

collection images, creating an unfamiliar representation that challenges naturalised conceptions of what digitised collection interfaces can and should include.

Delivering Discovery demonstrates pathways for designers and developers to address the concerns of digital humanists in digitised collection interface design. Without engaging with the issue of transparency, designers and developers will continue to replicate the ommission of valuable data from digitised library collection interfaces. To combat the recession of libraries' inner workings into digital black boxes identified by Shannon Mattern, the interface's infrastructure diagram reveals the complex network of disciplinary practices, institutional interoperability, and infrastructures of hardware and software that surrounds digitised library collections. It serves as a visual reminder that digital images of book pages are only one part of the ecosystem of digitisation. Drawing on Johanna Drucker and Lisa Gitelman's cautions against notions of raw data, the interface's visualisation view highlights the human labour—from the planning of a year-long digitisation project, to quality control notes written for individual pages—that shape digitised collections. In this way, the project works towards the framing of digitised library collections as capta rather than data. These insights may benefit a variety of library patrons, but all researchers working with digitised library collections sit in this space. This project advocates for a greater variety of data in our interfaces, which, as Thomas Padilla notes, opens library collections to a broader spectrum of use. Furthermore, the inclusion of administrative and structural metadata represents traces of the labour of digitisation and collection maintenance, which Michelle Moravec argues can enable digital humanists to engage more ethically with the opportunities represented by digitised library collections.

Having identified researchers working with digitised collections as key stakeholders, further research should be undertaken in conjunction with scholars from digital humanities, media theory, information science, and design to continue to iterate on this integration of generosity

and transparency; how might such interfaces more directly cater to their needs as researchers? While it makes sense to direct this information towards scholars who would benefit from it directly, Loom's representation of in-progress digitisation demonstrates the potential for interest for a wider population of library patrons. Further research could also explore how elements of transparency could be integrated into more general interfaces, where they might be of use to a variety of users without precluding more familiar browsing and searching behaviours.

*Delivering Discover*y is only one possible iteration of a generous interface which foregrounds administrative and structural metadata. It is a propositional prototype, intended to encourage further exploration of the intersection of generosity and transparency. It opens the door to further exploration of the possible use-cases of administrative and structural metadata in digitised library collections, and continues to explore the representations and browsing behaviours that transparent, generous interfaces can enable in digitised library collections.

APPENDIX

Although *Delivering Discovery* was published in September 2019, this Masters project began in February 2017, and the majority of the final *Delivering Discovery* prototype's development was completed between January and September 2018. As with much design work, iteration is a key aspect of the process, and traces of old designs still exist in the project's code.

For example, a number of different arrangements were tested for the layout of books in the collection visualisation. The file names of these unused past arrangements offer insights into the process of developing a digitised collection interface. These names sometimes echo the personal and idiosyncratic notes taken by digitisation contractors in the State Library of New South Wales' David Scott Mitchell collection. Stored in JSON files, there are eight iterations of layout for the contractor notes stage of the collection visualisation view. The first two iterations, initially created in May 2018, are named *p5.json* and *p5_scaled.json*. This layout was updated in July 2018, and the following six iterations that resulted are named *p5_scaled_jul_wrong.json*, *p5_scaled_jul_right.json*, *p5_scaled_jul_right?.json*, *p5_scaled_jul_right_maybe.json*, *p5_scaled_jul_needsfixing.json*, before finding the final arrangement with *p5_scaled_jul_new.json*.

More evidence of the iterative nature of web development can be found within the full source code, which has been uploaded to GitHub in the interests of transparency and generosity. It can be found at github.com/timbusuttil/delivering-discovery.

BIBLIOGRAPHY

Bagnall , K., & Sherratt, T. 2011, *The real face of white Australia*, viewed 16 August 2019, <http://invisibleaustralians.org/faces/>

Bengler 2014, *Repcol*, video recording, Vimeo, viewed 16 August 2019, <https://vimeo.com/85755986>

British Library 2017, *Digitisation services*, viewed 16 August 2019, <https://www.bl.uk/digitisation-services>

British Library 2018, *Digitised printed books (18th–19th century)*, viewed 16 August 2019, <https://www.bl.uk/collection-guides/digitised-printed-books>

Chalmers, M., & Galani, A. 2004, 'Seamful Interweaving: Heterogeneity in the Theory and Design of Interactive Systems', Proceedings of the 5th conference on Designing interactive systems: processes, practices, methods and techniques

Drucker, J. 2007, 'Performative metatexts in metadata, and mark-up', *European Journal of English Studies*, vol. 11, no. 2

Drucker, J. 2010, 'Graphesis: Visual Knowledge Production and Representation', *Poetess Archive Journal*, vol. 2, no. 1

Drucker, J. 2011, 'Humanities Approaches to Graphical Display', *Digital Humanities Quarterly*, vol. 5, no. 1

Durrant, A. & Price, H. 2015, 'RTD 2015 Provocation by Sir Christopher Frayling Part 1: Research Through Design Evolution,' video recording, Vimeo, viewed 16 August 2019, <https://vimeo.com/129775325>

Fraser, C. & Todd, Z. 2016, 'Decolonial Sensibilities: Indigenous Research and Engaging with Archives in Contemporary Colonial Canada' in R. Ištok, *Decolonising Archives*, L'Internationale Online, pp. 32–39

Frayling, C. 1994, 'Research in Art and Design', *Royal College of Art Research Papers*, vol. 1, no. 1

Galloway, A. 2012, *The Interface Effect*, Polity Press, Cambridge

Gitelman, L. (ed.) 2013, *"Raw Data" is an Oxymoron*, The MIT Press, Cambridge, Massachusetts

Gortana, F., von Tenspolde, F., Guhlmann, D. & Dörk, M. 2018, 'Off the Grid: Visualizing a Numismatic Collection as Dynamic Piles and Streams', *Open Library of Humanities*, vol. 4, no. 2, pp. 1–25

Hayles, N.K. 2007, 'Narrative and Database: Natural Symbionts', *PMLA*, vol. 122, no. 5, pp. 1603–1608

Hinchcliffe, G. & Whitelaw, M. 2015, 'Colouring digital collections: Challenges and opportunities for the use of colour metadata in cultural collections,' *MWA2015: Museums and the Web Asia 2015,* viewed 16 August 2019, <https://mwa2015.museumsandtheweb.com/paper/colouring-digital-collections-challenges-and-opportunities-for-the-use-of-colour-metadata-in-cultural-collections/>

Hinchcliffe, G. 2018, 'Visionaries Explorer', *DX Lab blog*, weblog, State Library of New South Wales, Sydney, 26 February 2018, viewed 16 August 2019, <https://dxlab.sl.nsw.gov.au/blog/visionaries-explorer>

Kräutli, F. & Boyd Davis, S. 2016, 'Digital Humanities Research Through Design', *DH Early Career Conference 'Mapping the Scope & Reach of the Digital Humanities'*, King's College London

Manovich, L. 2002, 'Data Visualization as New Abstraction and Anti-Sublime,' viewed 16 August 2019, <http://manovich.net/index.php/projects/data-visualisation-as-new-abstraction-and-anti-sublime>

Mattern, S. 2014, 'Library as Infrastructure', Places Journal, viewed 16 August 2019, <https://placesjournal.org/article/library-as-infrastructure/>

Mattern, S. 2015, 'Middlewhere: Landscapes of Library Logistics,' Urban Omnibus, viewed 16 August 2019, <http://urbanomnibus.net/2015/06/middlewhere-landscapes-of-library-logistics/>

Mattern, S. 2016, 'Public In/Formation', Places Journal, viewed 16 August 2019, <https://placesjournal.org/article/infrastructural-tourism/>

Mattern, S. 2018, 'Maintenance and Care', Places Journal, viewed 16 August 2019, <https://placesjournal.org/article/maintenance-and-care/>

Moravec, M. 2017, 'Feminist Research Practices and Digital Archives', *Australian Feminist Studies*, vol. 32, nos. 91–92

National Library of Australia 2019, *Digitisation of Library collections*, Canberra, viewed 16 August 2019, <https://www.nla.gov.au/digitisation>

New York Public Library 2019, *New York Public Library Digital Collections*, viewed 16 August 2019, <https://digitalcollections.nypl.org/>

Padilla, T., & Higgins, D. 2014, 'Library Collections as Humanities Data: The Facet Effect', *Public Services Quarterly*, vol. 10, no. 4

Padilla, T. 2016, 'Collections as Data: Conditions of Possibility', 29 September, viewed 16 August, <https://www.thomaspadilla.org/2016/09/29/possibility/>

Schofield, T., Whitelaw, M. & Kirk, D. 2017, 'Research through design and digital humanities in practice: What, how and who in an archive research project', *Digital Scholarship in the Humanities*, vol. 32, no. 1

State Library of New South Wales, 2015(a), 'David Scott Mitchell Library,' State Library of New South Wales, viewed 16 August 2019, <http://www.sl.nsw.gov.au/research-and-collections-about-our-collections/david-scott-mitchell-library>

State Library of New South Wales, 2017, 'Digitisation of Printed Materials: Statement of Requirements,' unpublished internal document

State Library of New South Wales 2019, *Digitisation Strategy,* Sydney, viewed 16 August 2019, <https://www.sl.nsw.gov.au/research-and-collections building-our-collections/digitising-librarys-collections/digitisation>

van der Maaten, L. & Hinton, G. 2008, 'Visualizing Data using t-SNE', *Journal of Machine Learning Research*, vol. 9, pp. 2579–2605

Wendler, R. 1999, 'LDI Update: Metadata in the Library' *Library Notes*, no. 1286, pp. 4–5

Whitelaw, M. 2012, *Manly Images*, viewed 16 August 2019, <http://mtchl.net/manly-images>

Whitelaw, M. 2015(a), 'Representing Digital Collections,' viewed 16 August 2019, <http://mtchl.net/assets/RepresentingDigitalColls-compact.pdf>

Whitelaw, M. 2015(b), 'Generous Interfaces for Digital Cultural Collections,' *Digital Humanities Quarterly*, vol. 9, no. 1