Remaking Collections

How to Cite: Whitelaw, M 2018 Mashups and Matters of Concern: Generative Approaches to Digital Collections. *Open Library of Humanities*, 4(1): 26, pp. 1–28, DOI: https://doi.org/10.16995/olh.291

Published: 24 April 2018

Open Library of Humanities

Mitchell Whitelaw, 'Mashups and Matters of Concern: Generative Approaches to Digital Collections' (2018) 4(1): 26 *Open Library of Humanities,* DOI: https://doi.org/10.16995/olh.291

REMAKING COLLECTIONS

Mashups and Matters of Concern: Generative Approaches to Digital Collections

Mitchell Whitelaw
School of Art and Design, Australian National University, AU
mitchell.whitelaw@anu.edu.au

This article discusses two practical experiments in remaking collections. *Drifter* (2016) and *Succession* (2014) build on the affordances of machine-readable collections and APIs to harvest large datasets from diverse sources, and show how these sources can be re-deployed to address complex spatiotemporal sites. These projects demonstrate the potential of a mashup-like generative approach based on sampling and recombination. Such approaches generate an expansive range of unforeseeable outcomes, while retaining a highly authored character. Here these projects are analysed through three key constituents: the troublesome trace of data; their extraction of digital samples; and their generative recomposition of samples into emergent outcomes. These techniques remake collections in a way that addresses the intrinsically complex, entangled and heterogeneous nature of what Latour terms 'matters of concern'.

Article

In Bethany Nowviskie's (2015) keynote address at Digital Humanities 2014, she set out a vision of Digital Humanities (DH) in the Anthropocene that is both sobering and hopeful. It calls for digital humanities practice to recognise the gravity of its situation, but also sets out some contributions that the community might offer. These hinge on DH's capacity to address large temporal and spatial scales, and on the need for representational approaches that might support 'picturing histories anew' (Nowviskie, 2015: i12). Nowviskie celebrates the 'forward-facing reorientation' (ibid.) of experimental and makerly humanities approaches that help map out 'possible and positive futures' (ibid.). This article documents two projects directly inspired by Nowviskie's call. They draw on and remake large digital collections around rich spatio-temporal sites, using generative techniques to create expansive and experimental outcomes. They seek to picture histories anew, as well as to convey the openness of future possibilities.

This work also draws on an active community of practice in the remaking of digital collections. As digital collections rapidly grow to post-human scales, these resources pose new technical and creative challenges. How might we engage with the Internet Archive's Flickr stream of 2.4 million images? Or the National Library of Australia's Trove collection of 211 million digitised newspaper articles? One strand of practice takes a poetic and playful approach, offering serendipitous samples and chinks of algorithmic insight. Tim Sherratt's *Trove News Bot* tweets archival news articles based on daily headlines (Sherratt, 2013); the British Library's Mechanical Curator posts random images from the library's digitised books (Baker, 2013). Sherratt's *Eyes on the Past* (2014) harvests faces from digitised newspapers, and has its eyes peer out through the interface, inviting investigation. These approaches reflect an emerging interest in collections as active sites of meaning-making, and experimentation with how we might encounter such collections in an everyday digital environment. They also signal the twin roles of computational techniques and human authorship in remaking digital collections.

The central question of this article is how such makerly, experimental approaches to digital collections can be applied to what Nowviskie (2015) terms 'picturing

histories anew' (i12), and in particular to the tangled and troubled character of the Anthropocene. Two projects are documented here: *Succession* (2014) generates composite images from around two thousand visual sources linked to Newcastle-Upon-Tyne; *Drifter* (2016) renders a digital portrait of the Murrumbidgee River catchment through multiple scientific, archival and spatial datasets. Both projects take on deliberately expansive subjects, and celebrate and evoke complexity rather than positing single or summative points of view. In the following discussion these topics are considered through Latour's (2004) notion of 'matters of concern', in which knowledge is neither purely objective fact nor socially constructed value, and the objects of knowledge are tangled and controversial 'gatherings' rather than stable, discrete units. Before Nowviskie's hopeful call to picture anew, Latour offered a more direct challenge: 'where are the visualization tools that allow the contradictory and controversial nature of matters of concern to be represented' (Latour, 2008a: 13)?

This article outlines a set of techniques that attempt to answer Latour's challenge. Drawing together diverse digital collections and recombining them into unforeseeable outcomes, these projects are best described as *mashups* — playful, speculative constructs that date from a more optimistic phase in our recent digital history. After introducing the projects and their development, it returns to elaborate on Latour's formulation. The following analysis considers three key components of these mashups: their foundation in the troublesome trace of data; their extraction of digital samples; and their generative recomposition of samples into emergent outcomes that prompt interpretation. These techniques offer productive approaches to remaking collections, generating rich, unexpected and engaging new forms; they also propose tools that may help address the ubiquitous complexity of Latour's matters of concern.

Succession

Succession (Whitelaw, 2014) addresses the industrial and pre-industrial heritage of Newcastle-Upon-Tyne and surrounds, reflecting my encounters with the city during an extended visit, and the meditations on energy, history, industrialisation and capital that the place provoked. Newcastle was a key Roman settlement, later one

of the crucibles of the Industrial Revolution, now finding its way in a 'de-industrial' Britain (Chakrabortty, 2011). While the work is highly localised, it reads this place as a cypher with far wider implications. The work's title draws a term from ecology, alluding to continuous and ongoing change and adaptation. *Succession* aims to mine the city's industrial past in order to fuel consideration of our possible futures.

In practice, the work is a web application that draws on a corpus of some two thousand image records harvested from the Flickr Commons, and combines these elements into new composites (or *fossils*). Each fossil is composed of five randomly selected source images arranged, composited and potentially repeated. Sources are cited, so that while composites often radically obscure, transform or juxtapose their elements, the sources and their attendant contexts remain navigable and intact. Composites may be saved, acquiring a permalink to become a new citable online object. The generative process of composition is performed live, in the browser; each viewer will encounter a series of unique composites. The system allows for around 2.5×10^{15} combinations of elements (ignoring spatial and blending variations). At a rate of one combination per second, it would take around eight million years to show all permutations.

The saved composite in **Figure 1** shows how the generated artefacts can operate as what Drucker and Nowviskie call 'aesthetic provocations' (2004: 431) while enlisting the contexts and referents of their source elements in speculative juxtapositions. This composite is dominated by a 1993 photograph of Wearmouth Colliery in its final week of operation, a poignant image of the last days of Newcastle coal. But a spectral water-bird (from Gould's 1837 *Birds of Europe*) seems about to splash down in those desolate puddles. Faintly in the background is the bustling River Tyne itself, circa 1880; and on the left of the frame a carriage destined for Newcastle's Metro system, under development in the mid-1970s. Almost imperceptible at bottom left is the *HMS Opal*, a torpedo destroyer, under construction at the Sunderland shipyards in 1915. So this composite encompasses not only 150 years of urban history, but a latent portrait of twentieth-century capital, the rise and (UK) fall of extractive industry, war, urbanism, and pre-Industrial naturalism and the non-human lives it records.

Figure 1: *Succession* saved composite 1413513552860 (Whitelaw, 2014a).

This example shows how formal and visual transformations inflect narrative or historical interpretations, and thus that these generative artefacts are not simply bundles of citations, but speculative visual propositions. Layering emphasises simultaneity and atemporal juxtaposition, rather than chronology: Gould's duck, about to dive into the colliery puddles; or perhaps swimming on the 1880s Tyne? Faded traces evoke the presence of the lost; visual collisions prompt an interpretive search for coherence.

One byproduct of this generative approach is that it is impossible to characterise its results exhaustively. Instead, we can think of the system as defining a large space of potential outcomes. Artist Bill Seaman has developed a number of works using similar techniques: he outlines a 'recombinant poetics' (Seaman, 2001: 424) that seeks 'emergent meaning' (ibid.) within digital arrays of textual and audiovisual elements. Seaman (2001: 426) usefully describes his combinatoric generative system as an 'authored electronic space', emphasising that it is expansive, but not arbitrary. Where Seaman created each source element, *Succession* imports them; but nonetheless its generative space is highly authored — in part through this curation

of its corpus. Sources were selected for content relevance, as well as visual potential. The content-base grew and was pruned around the conceptual focus in a slow process of subjective evaluation, exploratory search and tangential investigation. The Internet Archive's Flickr stream was a key influence: this immense set of bulk-harvested book illustrations offers uncurated insights from a range of sources, including nineteenth-century trade journals and volumes of naturalist scholarship. The resulting corpus is diverse and heterogeneous, but retains enough commonality to generate meaningful connections; in Seaman's words, this authorship seeks a 'resonant unfixity' (Spielmann, 2002).

The rules for compositing elements are also authored, tailored around the idiosyncracies of the sources and the poetic aims of the work. Layered composition is both formal machine and metaphor: to address a city built on coal it seemed necessary to combine and compress, to obscure while hopefully at the same time intensifying the energy latent in those sources. Thus, the image blending modes are biased to overlay dark elements; this treatment resonates with the engraved illustrations in the Internet Archive and British Library collections. Process and corpus (algorithm and data) are interdependent, but equally authored. Rather than a neutral, inert (Cartesian) space, this authored space is co-constituted by an active assemblage of media, concepts, subjective interests and computational processes.

Drifter

One of the unexpected outcomes of *Succession* was the folding in of documents of nonhuman lives, and the poignant juxtaposition of natural science illustrations with the machinery of industrialisation. This prompted an interest in how digital archives might be used to more intentionally address the troubled complexities of landscape. As well as drawing on Nowviskie's conception of digital humanities in the Anthropocene (2015), this approach was informed by John Thackara's (2015) advocacy of a 'bioregional' practice focusing on local ecological systems including food networks and water catchments.

Drifter (Whitelaw, 2016) was developed in early 2016 in response to an invitation from the Land Dialogues conference at Charles Sturt University, Wagga Wagga (Land

Dialogues, 2016). Canberra, where I work and live, shares a river catchment with Wagga: Land Dialogues, a multidisciplinary meeting sited on the Murrumbidgee, offered an ideal venue for a project investigating multi-layered digital representations of our shared river system.

The Murrumbidgee River is a major inland river in south-east Australia. It stretches some 1600 kilometers from its headwaters in the mountains of southern New South Wales (NSW), passing close to Canberra before running west through inland rural centres including Wagga Wagga. Like many of Australia's watercourses, the river has a dense and troubled history of European settlement, exploitation and intervention. It remains a site of significant conflict and controversy, focused on the competing demands of environmental conservation and agriculture. The river's course includes multiple dams and weirs, constructed over the twentieth century to divert water for irrigation and electricity generation. Its catchment also sustains ecologically significant swamps and wetlands, including the Fivebough and Tuckerbil wetlands, recognised under the Ramsar convention. The river's flow is regulated under the 2012 Murray Darling Basin Plan, which sought to reduce agricultural extraction and return enough water to ensure the river's long-term sustainability. The plan has been greeted with open hostility from some communities along the river; in 2011 young men in Griffith, NSW burned copies of the draft plan in protest (Barlow & Doyle, 2011).

Drifter approaches this landscape as an expansive bundle of post-settlement concepts and perspectives on place and site including geography, history, ecology, agriculture, social discourses and cultural representations. These dimensions are formative forces, intervening in the land (building dams, diverting water for irrigation) as well as offering divergent epistemological and ontological frames through which the landscape is represented (as a resource, as habitat, as a figure in a narrative of settlement and colonisation). As the ongoing conflict around this landscape shows, these divergent frames are far from resolved. *Drifter* begins from the premise that traces of these diverse dimensions of the landscape persist in various forms, and are legible as data — digitised photographic records, administrative archives,

newspaper reports, scientific observations, geospatial information. Its core question is how this data can be drawn together to address the tense, complex multiplicity that continues to characterise the river catchment. As in *Succession*, it aims to draw together and recombine digital collections for speculative ends, inviting unforeseen interpretations and unpredictable connections.

In practical terms, *Drifter* is a set of three web-based interfaces or displays. Each of these draws on specific datasets and collections and recombines them in specific visual and temporal structures. This approach acknowledges from the outset the impossibility of a total or comprehensive representation of this landscape or its data traces; these three displays are intended as partial and provisional views, drawing on Drucker's notion of parallax in humanities displays — the revealing divergences between distinct but related representations (Drucker, 2011).

Map

Drifter's Map offers a geospatial view of the river catchment, using familiar web map methods to create a navigable surface for exploration. However rather than use pre-existing cartographic 'base layers' (like the ubiquitous Google Maps) this map selects and overlays a limited set of features. In an effort to defamiliarise the landscape, and in particular to unwind the anthropocentric focus of traditional maps, it omits roads, rail lines, administrative boundaries and other conventional features such as elevation; it includes some towns and cities only as minimal indications. Instead, this map emphasises water in the catchment. It combines and layers geospatial data from a range of sources, including OpenStreetMap, the Digital Chart of the World, the US Geological Service HydroSHEDS dataset and the Australian Department of the Environment and Energy's Australian National Aquatic Ecosystem classification framework, which provides detailed documentation of wetland areas within the catchment. Collectively these datasets map not only the Murrumbidgee River but its entire filigree network of creeks, streams and tributaries, as well as a multitude of other watery structures, intermittent and permanent: swamps, soaks and flood plains as well as dams and weirs.

Visualised, these structures generate a satisfying intensity of detail, but also prompt a foundational question: what is a river? The HydroSHEDS dataset prompted this question during the project development. This global hydrological database is based on land elevation data, and thus records watercourses at extreme spatial detail: a psychedelic overload of capillaries known as stream lines (**Figure 2**). This water-landscape is both recognisable and unfamiliar. When overlaid with water areas and wetlands in the *Drifter* map, the Murrumbidgee no longer seems a single, discrete thing, but a dappled field of ponds, lagoons, creeks and rivulets, fans and meanders; many complex things, not one simple blue line on a map (**Figure 3**). Visualised in this way, the data suggests that this river may not be what we imagine it to be.

Two other datasets are overlaid on the map's base layer. One is a corpus of newspaper articles linked to 37 different locations within the river catchment: towns and other European settlements as well as lakes, dams and weirs. For each location, a set of up to 200 article references were harvested from the National Library of Australia's Trove digitised newspapers service. Along with place names a set of query terms ('river', 'creek', 'dam', 'irrigation', etc) was used to extract articles referring to

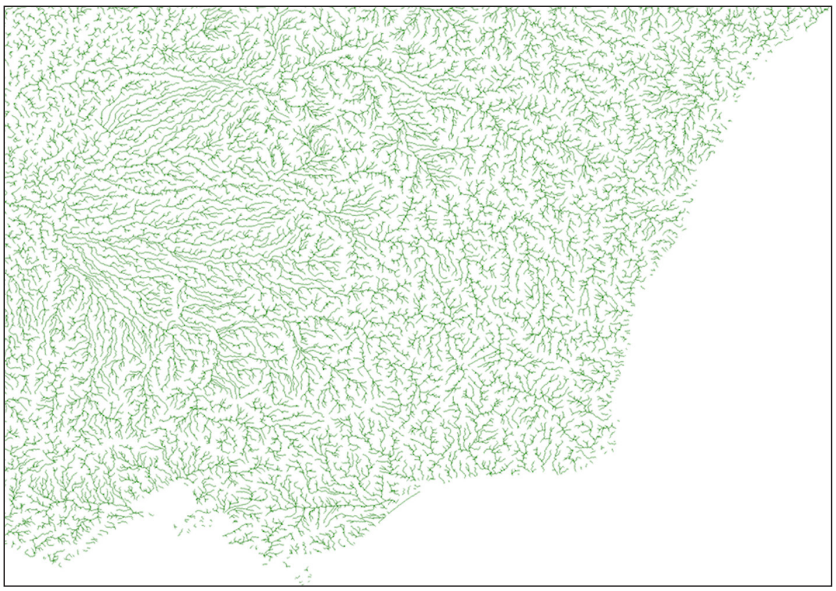

Figure 2: HydroSHEDS stream line data, south-eastern Australia. Image by the author.

Figure 3: *Drifter* Map (from Whitelaw, 2016). Murrumbidgee catchment between Griffith and Hay, New South Wales.

the river system. Article excerpts appear spontaneously on the map, anchored to their related location, showing snippets of the digitised text, along with links to the full article on Trove. In line with Trove's overall coverage, articles are largely from the period 1880 to 1945, a period that coincides with many of the major post-settlement interventions in the river system, including the establishment of the Murrumbidgee Irrigation Area.

The second dataset is a corpus of scientific observations, drawn from the Atlas of Living Australia (ALA), an aggregator of biodiversity data. The ALA includes detailed taxonomic data structures and geospatial search functions, along with a well-documented API. Across the whole Murrumbidgee catchment, the Atlas lists millions of individual observations. Here a slice of that data is used, focusing on a single taxonomic family. Frogs (*Amphibia*) are tightly bound to watery habitats and heavily impacted by the degradation of the river. Several frog species endemic to this region are critically endangered, such as the Southern Bell Frog, the Yellow-spotted Tree Frog and the Southern Corroboree Frog. The ALA lists some 6700 individual geolocated observations of 37 frog species within the catchment. Observations date as far back as 1891, based on specimens held by the Australian Museum (Anon, 1891), but most observations date from after 1960. Like the

newspaper articles, the frog observations appear on the map spontaneously, randomly selected based on the current field of view, and linked to the full ALA record. Audio of the frog's call is linked with each observation wherever possible, so we both see and hear these data points. Their appearance is sequenced in order to overlap, so that the observations are artificially reconstituted into a chorus, an audible trace that is also a spatialised sonification of frog occurrences across the river catchment.

Through extracting, combining and layering datasets, the map interface forms a stage, a designed space constituted through juxtapositions. These juxtapositions are random and in part unpredictable, much like the image overlays of *Succession.* A key difference here is that the data sources are intentionally heterogeneous; orthogonal rather than parallel; and in this sense the collisions between layers are designed (or staged). The juxtaposition of frog observations and newspaper excerpts is a staged conversation whose moment-by-moment details are uncertain, but whose broad range of concerns are, to use Seaman's term, resonant rather than disparate.

Sifter

The Sifter interface attempts to probe similar questions — of coexistence in the landscape, and the divergent relationships between European settler culture and the living world it occupies. Once again it uses a corpus of digitised newspaper articles, and juxtaposes fragments that provide small insights and points of connection. Here the Trove API was used to harvest articles linked to the Murrumbidgee and its major tributaries: the Bredbo, Cotter, Goodradigbee, Gudgenby, Molonglo, Numeralla, Tumut and Yass rivers. The original intention in gathering this material — over 15,000 articles in total — was to winnow out traces of the living systems of the river, identifying points of contact between the affairs of humans and nonhumans. However, mentions of the Murrumbidgee and its tributary rivers during this period are overwhelmingly human-centred. These are sites of accidental drownings, picnics and football games, the construction of bridges and roads and the discovery and extraction of gold, as well as the extraction and stockpiling of water. Animals such as fish — often trout, introduced from Europe — appear only as prey or food. One

of a handful of mentions of frogs in this corpus illustrates the point. In 1924, the Adelaide Register reported on the commission into establishing the federal capital of Canberra:

> Mr. Yates (S.A.) was willing to vote for Canberra if he could be shown how it was to benefit Australia. He had seen a little humpy they called a hostel. The water supply was ridiculous and extravagant. He had been told of the wonders of the Molonglo River, but he found it to be simply a frogs concert room. (Anon, 1924: 13)

In response, the Sifter interface works more actively against the grain of this corpus. In the absence of traces of the nonhuman histories of the river, it essentially mines the text of these newspaper snippets to reassemble what is missing. As the newspaper excerpts unfold, a simple search process slowly assembles the name of an endangered species endemic to that river, matching characters in the species' scientific name with those in the newspaper text. Once it is complete, the more familiar common name is also revealed, as the text excerpts fade into the background. This sifting cycle continues; the interface loads another corpus and begins to search for another name.

The pace of this process is intentionally slow and almost completely passive. Newspaper excerpts are animated to unfold languorously, braiding a stream of text that disintegrates back into its constituent characters. Reading is both slowed and abstracted, as excerpts collide and intermingle, creating incidental poetry from the pieces. In the background, a loop of audio runs continuously: a recording of the Molonglo River, trickling through a gorge on the western fringes of Canberra. The aim is to divert and delay our habitual pace of information seeking, inviting a more reflective frame for interpretation. The only interactions available during this process are in the form of citation links leading to the Trove sources for each text excerpt.

Compositor

The third and final view takes a contrasting approach; like *Succession*, it draws on visual records and constructs composites, though there they are simple crossfades between pairs of images drawn from different historical domains. *Drifter* was initially presented

in a gallery exhibition context, as a non-interactive display, cycling between its three views. In this presentation, the compositor acted as an interstitial or transition mode, between the longer form, text-centric Sifter and Map visualisations. The collection data here consists of two harvested and curated image sets. The first is a set of some two hundred historic images linked to the river, drawn from collections including the National Library of Australia, State Archives of NSW and the Powerhouse Museum. These date largely from the late 19th century to the mid 20th: major interventions such as Burrinjuck Dam, the Irrigation Area and its newly-planned settlements are documented in maps, plans, diagrams and photographs. Scenic views, sketches and mementoes of river settlement also feature. The second set is a collection of 49 more recent photographs. Provided by freshwater ecologist Dr. Skye Wassens and the Institute for Land Water and Society, Charles Sturt University, these images document scientific fieldwork in the wetlands of the Lowbidgee and the mid-Murrumbidgee between 2008 and 2015. This set includes close-up specimen photographs recording specific frog, insect and fish species, as well as wider views of the wetlands, verdant lagoons and reedy marshes. Although gathered as documentation of ecological fieldwork, these images are often strikingly beautiful, implying a subjective point of view with a strong attachment to these places.

The compositor randomly selects a pair of images, one from each set, and stages a gradual cross-fade between them. As in *Succession*, blend modes are used to generate unpredictable interactions between the two layers such that one will selectively reveal or obscure the other; the intention is to generate and prolong a liminal state where both documents (and both domains) coexist. Here too a simple generative, combinatorial process multiplies two small datasets into an expansive set of possible outcomes — almost 10,000 different composites. At times these juxtapositions are relatively blunt (a frog, cupped in a human hand/Burrinjuck Dam under construction); at others they can be more subtle and ambivalent. In either case, much like *Succession* this view uses selective harvesting and curation combined with generative unpredictability to seek out resonant collisions between collection items (**Figure 4**).

Mashups and Matters of Concern

In 2004, Bruno Latour asked, anxiously: 'why has critique run out of steam'? He was referring to the way in which the critical project (in part his own) of revealing the social construction of scientific knowledge seemed to lead to wholesale skepticism, artificial controversies, conspiracy thinking and an ideological erosion of the status of 'facts', particularly after September 2001. While not 'reversing course' (Latour, 2004: 231) Latour argues that the critical spirit has 'sent us down the wrong path': 'the question was never to get away from facts but closer to them' (ibid.). Rather than a retreat, Latour (2004) calls for a reorientation, a change of target: for 'a realism dealing with … *matters of concern*, not *matters of fact*' (ibid., emphases in original). Through Heidegger, Latour characterises matters of concern as complex 'gatherings', folded entanglements that are never separate objects, but which involve and include us. In a later lecture he writes: 'A matter of concern is what happens to a matter of fact when you add to it its whole scenography, much like you would do by shifting

Figure 4: *Drifter* compositor (from Whitelaw, 2016). In this composite are: (1) a 1907 image of a cable tram trailer, constructed for the Burrinjuck Dam tramway (NSW State Archives, 2010); and (2) a 2011 photograph of Avalon Swamp in the Lowbidgee Wetlands, provided by Dr. Skye Wassens, Charles Sturt University.

your attention from the stage to the whole machinery of a theatre'. 'Matters of fact were indisputable, obstinate, simply there; matters of concern are disputable, and their obstinacy seems to be of an entirely different sort: they move, they carry you away, and, yes, they too *matter*' (Latour, 2008: 38–9, emphasis in original).

Latour's formulation is a useful encapsulation of the approach that both *Drifter* and *Succession* take to the sites they address. Both projects seek to render a proliferation, a gathering; an empirical stance that uses digital collections as resource to restore to these topics some of their 'scenography': the webs of context, human and nonhuman actors, systems and narratives that constitute them. Even as they deploy speculative and imaginative methods both projects share Latour's aim to 'get closer' to the complexities of matters of concern. And if matters of concern *matter* — if they move us or carry us away — similarly these projects are motivated by a personal mattering, and they seek to propagate that concern through shared circuits of feeling and thinking.

But if what these projects seek is a way to reveal or represent matters of concern, Latour offers a challenge. Addressing a conference of design historians, he sets it out:

> In its long history, design practice has done a marvellous job of inventing the practical skills for drawing objects, from architectural drawing, mechanic blueprints, scale models, prototyping etc. But what has always been missing from those marvellous drawings … are an impression of the controversies and the many contradicting stake holders that are born within with these … [T]hree hundred years after projective geometry, fifty years after the development of CAD computer screens, we are still utterly unable to draw together, to simulate, to materialize, to approximate, to fully model to scale, what a thing in all of its complexity, is. (Latour, 2008a: 12)

Put more directly:

> Where are the visualization tools that allow the contradictory and controversial nature of matters of concern to be represented? … Critique, deconstruction and iconoclasm, once again, will simply not do the job

> of finding an alternative design. … What I am pressing for is a means for drawing *things* together — gods, non humans and mortals included. (Latour, 2008a: 13, emphasis in original)

The response from design and visualisation practice has been limited to date. For example, Murray (2016) recognises Latour's challenge but proposes in response an emphasis on the physicalisation of data representation. Davila et al. (2014: 60) describe an artistic visualisation project dealing with homelessness, and drawing on Latour assert the value of attempting to render it richly: 'The complexity of a problem such as homeless [sic] renders it unrepresentable. What can be represented is complexity itself, as well as a sense of the immensity of the issue, its material reality, and the care and concern associated with it'.

Succession and *Drifter* share this interest in revealing complexity, but offer a distinctive approach in drawing on digital heritage collections, and combining generative techniques with structured juxtapositions to draw together the multiple layers and actors at play. In doing so they adopt many of the characteristics of the *mashup*.

The term mashup often refers to musical remixes that seamlessly blend multiple sources into new composites. This term was adapted by the IT industry around the mid-2000s, to describe web applications that built on, combined and visualised multiple diverse data sources. This practice grew out of the rise of Web 2.0 social media and e-commerce platforms, and the public availability of structured data via APIs. Feiler's (2007) instructional volume *How to Do Everything with Web 2.0 Mashups* offers a cultural and technical snapshot of this era. Web 2.0 mashups were cultural products of their times. While some developed into serious commercial applications, mashups were typically motivated by a spirit of playful experimentation. This experimentation was supported by well-documented, accessible technologies and standards such as JavaScript, HTML, XML and RSS. Mashups were *non-destructive* and *non-exclusive*: combining and building on disparate sources without changing or superseding those sources.

A decade later, with the increasing centralisation and commercialisation of web services and data, and rising evidence of the social risks of large digital platforms,

mashups seem *passé*. In many cases, they are simply no longer possible, as the public APIs that they relied on have been shuttered or commercialised. One notable exception to this trend is the growth in the availability of public-sector data, including digital heritage collections. In this domain, mashups remain not only possible but potentially valuable in investigating how cultural and scientific data can be transformed and redeployed in widely accessible public platforms. Moreover, given the challenges of dealing with our entangled present, and Latour's ubiquitous matters of concern, generative mashups offer a productive way to draw things together.

Drawing Things Together: Data, Sampling and Recomposition

We can consider this proposal through a stack of characteristic elements and techniques used in both *Succession* and *Drifter*. Interrogating these components and their interactions sheds light on how generative techniques to digital heritage play out, and how such approaches might address Latour's challenge.

1. Data as Troublesome Trace

The foundational constituent here is data, for to be blunt these works draw *data*, not things, together. They harvest tens of thousands of records, taking the form of metadata, textual data and digital images. The value and scope of any mashup is inevitably conditioned by its source data, and here we have reason to be critical. As Drucker argues, the etymological underpinnings of data as 'given' is problematic for humanities thought, bringing with it a realist epistemology of an observer-independent world (Drucker, 2011). Drucker proposes as an alternative *capta*, which is 'taken' rather than given; that is, actively shaped through selection, observation, interpretation and structuring. The heritage records that these projects build on are indeed capta, typically constituted through multiple stages of 'taking', including the acquisition of a physical object (photograph, newspaper article or biological specimen), its registration within an institutional collection, according to specific norms and practices, and the eventual digitisation of metadata and/or object. At every stage along this chain of provenance, data is shaped and conditioned through a stack of social, cultural, organisational and technological processes. Computation

plays an increasingly significant role. The Internet Archive's Flickr stream, a key source in *Succession*, is the product of an automated computational process that harvests image content from the Archive's digitised books (Miller, 2014), enriching the collection while also radically amplifying its scale. Optical character recognition, used to automatically transcribe Trove newspaper articles, is similarly invaluable in making article content digitally legible even though its output is riddled with errors. As Miriam Posner (2015) argues, humanities data is messy, unwieldy, and demands careful interpretation. In this sense data is *troublesome*.

This character prompts caution, but it does not negate data's value or agency as a material in these works. These projects maintain faith in the connection, however mediated and conditioned, that data offers to diverse worlds and actors. Data here is a *troublesome trace*. It is undeniably constrained in its scope; it indexes a tiny portion of the complex histories and landscapes and systems at issue. In documenting the pasts of these places and landscapes it shows only what was legible, relevant or significant to the record keepers. Those partial, conditioned records are further mediated through the socio-technical processes of digitisation. These factors would trouble any summative, generalising conclusions — any move to reduce or collapse data into definitive findings. By instead using generative techniques that retain the fragmentary specificity of their data sources and privilege open-ended interpretation, these works seek to expose and build on the trace that remains, and to draw things (not only data) together.

2. Sampling

In working with large datasets, these projects face a question of representational scale. In previous work with digital collections, I have used data visualisation techniques to generate compact summative representations of whole collections (Whitelaw, 2015). These approaches rely on abstracting details of individual records, grouping and clustering them according to their metadata. *Succession* and *Drifter* demonstrate a complementary approach based on sampling. Rather than generating abstracted, summative representations, a sampling approach represents the collection through its individual constituents. This technique has several distinct dimensions. Firstly, it

is a form of systematic statistical representation. Given a repeated, random selection method, a set of samples offers a compact but statistically representative subset of the whole corpus. In these works, samples often unfold over time, developing a cumulative impression of the dataset. This approach reflects the empirical use of random sampling in the sciences; a pragmatic approach to dealing with scale and complexity.

This dimension has nothing to say about the content of the sample, however; in fact it entails a statistical averaging-out that blurs specific details. The metonymic dimension of the sample begins to account for the way that specific content operates, as extracted fragments often signify through part-for-whole relations. A photograph from the Newcastle-Upon-Tyne shipyards is a documentary fragment that also signifies an overlapping set of larger contexts and concerns: shipbuilding and shipyards in general; Newcastle's industrial heritage; Northern industry and its decline; the de-industrialisation of the UK. Similarly, an ecologist's image of a frog from a Murrumbidgee swamp, cupped in a human hand, stands as part of a complex network including other frogs, river habitats and systems, other wildlife, scientists and institutions. Samples also act as metonyms in relation to their sources, and the disciplines and domains they represent. A diagram from the *Transactions of the Institution of Mining Engineers* showing the forces on a suspended cage in a mine shaft, speaks to a specific professional domain and its associated epistemology (Internet Archive Book Images, 1898). The metonymic function of these samples depends heavily on an interpretive understanding, on the knowledge that links these nested and associated contexts and concerns. As such, it too involves a degree of abstraction; the part-for-whole function takes the sample to represent something greater than, but also other than, itself.

The third and final dimension is the sample as an evocative fragment in and of itself. These samples are never self-sufficient, but shards of larger narratives, discourses, institutions and contexts. In both *Succession* and *Drifter*, links to the public source of each sample encourage the audience to pursue and parse these connections. On 14 November 1966, the *Canberra Times* reported: 'Four bulldozers changed the course of

the Cotter River in dense bush country south-west of Canberra on Saturday morning, in the first major step of the Corin Dam wall construction' (Anon, 1966) (**Figure 5**). This fragment points to highly localised, specific events in time and space. That specificity is in part indexical — it cues the function of the data as an indexical trace. At the same time it is evocative, and in a sense incomplete: its terse description is a prompt for imagination, an invitation to fill in the details of the story.

In fact, this fragment can equally be read under all three dimensions of the sample. It is specific, but also metonymic: these bulldozers stand for all bulldozers, building dams, bridges, weirs, channels and other river interventions; and in turn for the whole modern paradigm that reshapes the landscape in this way. It is metonymic, but also a random sample; just one of over 5,000 harvested snippets on the Cotter River used in *Drifter*. Here these three dimensions of the sample operate concurrently (**Figure 6**).

3. Generative Recomposition

In *Drifter* and *Succession*, fragmentary samples are recomposed into complex, unstable composites. This recomposition makes use of simple computational techniques, but builds on the resonance of its content to generate expansive outcomes. The works

overall effect on the sion.

el cell trouble developed he spacecraft earlier to-

esterday Aldrin stood in pen hatch of the space- for a record two hours minutes and took photo- hs. of star fields and ducted several experi- ts.

e covered 42,000 miles the hatch open, at 00 mph.

Pictures of rare scene

arlier Lovell had flown mini-12 into rendezvous a total solar eclipse South America, re- ling the first pictures of rare event ever taken n above the earth's dis- ng atmosphere.

he four pictures taken ng the scant eight nds in which the astro- ts had the total eclipse view are expected to scientists discover the ets of the solar corona, fiery halo of gases ch reaches out thous- s of miles from the s surface.

he corona, which ap- red like a rim of vivid w pearls as the moon ked the sun, is believed e a major influence on earth's weather and on -range radio com- ications.

irely satisfactory", a Space Agency spokesman said.

From **Moscow** comes the information that the Soviet Union had launched an unmanned artificial earth satellite, number 131 in its Cosmos series.

It was believed the satellite might have been sent into orbit to study a massive belt of sunspots reported yesterday by the Bochum Observatory in West Germany.

The Bochum laboratory said the belt of sunspots, 75,000 miles long, could affect the upper atmosphere and endanger US astronaut Aldrin in his planned manoeuvres outside the Gemini-12 spacecraft.

US scientists did not feel the effects would be so strong.

Picture — Page 5

died tonight in a fire which destroyed his brake van on a run from Port Kembla to Goulburn.

Late tonight police were still on the scene at Robertson where the death was discovered.

Police believe the van was on fire, with smoke and flames billowing behind the train, for the greater part of its journey from Port Kembla.

The guard's name has not been released.

Cotter River diverted

Four bulldozers changed the course of the Cotter River in dense bush country south-west of Canberra on Saturday morning in the first major step of the Corin Dam wall construction.

RANGERS FIRE ON U.S. CLUB

SAIGON, Sunday (AAP). — South Vietnamese Government Ranger troops shot up an American recreational club after reports of friction between some Vietnamese rangers and US troops in the rubber-growing region for several weeks.

The agency quoted its sources as saying the

the variation of Crown leases was inhibiting investors in general and investors in accommodation in particular.

For years the Government had released sites on the outskirts of Canberra specifically for hotels and motels, but now was varying the conditions of residential leases in the heart of the city to allow private houses to operate as guest houses.

The outer, commercial sites had been bought at high cost, on the assurance that it was government policy to place such development on the periphery of the city.

"Now we find the Government is releasing residential leases in the centre of the city, to be used for commercial accommodation

to crack open a wall of the recreation club, then fired into the building with automatic weapons.

Meanwhile the US command in Saigon announced today that Vietcong infil-

the main entry to the capital", the statement said.

A Commonwealth representative had asked the Supreme Court last June to vary residential leases, simply on the grounds that private houses had been operating as guest houses, although illegally.

The courts were being asked to cure a situation caused by bad planning and ineffective administration.

40mph gusts hit city

Maximum wind gusts of 40mph were recorded by the Bureau of Meteorology yesterday afternoon at Fairbairn.

The forecast for today is for clearing showers in the early morning, and cloudy periods with fresh to strong and gusty west to south-westerly winds for the remainder of the day.

A southerly buster which swept up the NSW coast yesterday with wind gusts of up to 45mph passed through Sydney shortly before noon.

In Queensland, a thick layer of dust today blank-

He is new C. Cody G'Neill (A (Lib).

OXLEY held by M (Lab). He by T. J. W. H. Fle J. Logan.

CANN Mr J. M. is being o Edwards (DLP).

MOORE Mr D. W. is opposed (ALP), H. G. P. O'C

Ad ca

A spo group, Mr Sydney, s ment that concentrate the electo mite, Indi ning, and His g mainly pe primary in

In addi

Co D

The Leader

Figure 5: Cotter River Diverted, *Canberra Times*, 14 November 1966.

show how simple generative processes can be developed to deal with digital heritage content, and begin to outline the space of similar applications.

In *Succession*, each fossil is composed of five unique, randomly selected images from a single corpus of around two thousand. The selection of elements here is independent of their content, so the diversity of the corpus, curated through the project's development, is a key factor in the character of the results. A second layer of generative logic controls the appearance of each composite. A set of defined presentation modes position an image as either covering the whole frame, contained within it, or repeated in a vertical or horizontal strip; and a set of compositing rules control the visual layering of the images within the composite. In combination with the random selection of sources, these rules describe an immense range of variation, but at the same time they constrain that variation within an authored space of potential outcomes that is constituted through both the corpus and its recombinant logic.

Where *Succession* uses a single heterogeneous corpus, *Drifter* uses multiple corpora, and in this case generative techniques juxtapose elements between corpora as well as within them. The set of frog observations in the map view share

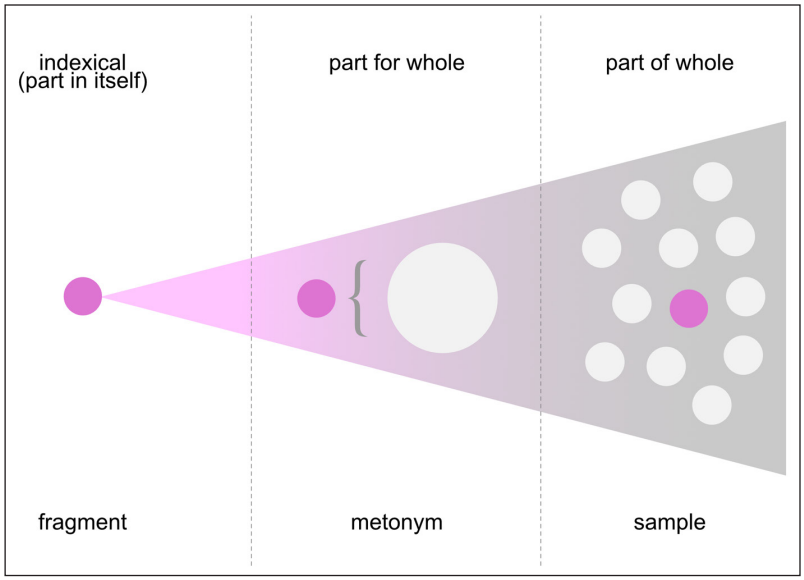

Figure 6: Three concurrent dimensions of the sample. Image by the author.

both subject matter and institutional domain, forming a coherent data layer. The newspaper excerpts are more diverse in content, but remain bound by the shared features of their sources. As a result, the character of the generative composite is shaped more by the character of each corpus than random selections within it. Each sample functions metonymically to refer to its corpus, and its associated domain. Here generative recombination juxtaposes divergent perspectives and ontologies, as well as data-points; mashing up domains is one way that these techniques can draw things together, without erasing their differences. As Latour's 'compositionist manifesto' states, 'things have to be put together … while retaining their heterogeneity' (Latour, 2010: 473–4).

Computational autonomy is a key attribute of these generative techniques. Automated selection and juxtaposition generate astronomical spaces of potential, and thus unforeseeable collisions between samples. As in other uses of generative techniques in design, the payoffs for this unpredictability are in part hedonistic: the constant variety of outcomes promises endless potential; sparks of surprise offer small pleasurable rewards. Like other mashups, these techniques also entail a sense of liveness: as fragments are continually sampled and recombined, we encounter something that *is happening*, rather than something that *has happened*. In working with digital heritage content this is particularly relevant. While the samples refer to mixed and heterogeneous pasts, their continual reassembly enacts a restless and ongoing renewal, a call for reconsideration.

Ross Gibson recognises something similar at work in his project with Kate Richards, *Life After Wartime* (*LAW*). *LAW* is a 'story-engine' that recomposes an archive of crime scene photographs together with sound and text fragments in the service of an affective, emergent understanding (Richards, 2006: 453). For Gibson (2006) the process of re-connecting and re-animating fragments is a response to the 'aftermath cultures' of Australia (and arguably other western nations), where landscapes, ecologies and knowledge are 'ailing or out of balance'. In response, he proposes:

> Our parlous states need imagination. We need to propose "what if" scenarios that help us account for what has happened in our habitat so that we can then better envisage what might happen. We need to apprehend the past. Otherwise … we won't be able to divine the continuous tendencies that are making us as they persist out of the past into the present. (Gibson, 2006)

Gibson further links this aim to the restlessness of generative recomposition and its interpretation: 'It's the *restlessness* that's crucial, the way the artwork … prompts the imagination by artful imbalances and implied possibilities for completion or patterning' (Gibson, 2006: emphasis in original). For Gibson the computational juxtaposition of fragments is ultimately an engine for human interpretation. Incompleteness and implied pattern drive an active process, a stitching-together and imaginative reconstruction. The same processes are at play in *Drifter* and *Succession*. As Gibson argues, the computational scale of generative techniques is not an end in itself, but a means to invite complex understandings of a fragmented terrain.

Conclusion

Where Gibson outlines their aspirations, these projects also demonstrate some of the limits of this approach. To draw things together through the proxy of data relies on the existence of data; mashups may simply reinforce or reproduce gaps or blind spots in their source collections. *Drifter's* attempts to address Australia's troubled European history are constrained by the absence of Aboriginal history — a critical blind spot, in relation to concepts and models of landscape. As Tim Sherratt observes:

> The contents of Trove's newspaper zone, like any online collection, is constructed — shaped by many competing priorities. The consequences of this process are not always obvious … There's a danger that the sheer scale of aggregation services like Trove will reinforce existing prejudices. People already struggling for visibility and recognition within our cultural record might be lost. (Sherratt, 2015)

This risk also applies to the remaking of collections such as Trove. It may be balanced in part by the opportunity that *re*making provides to author, curate and enrich its sources, as Bagnall and Sheratt's *Invisible Australians* (2012) shows.

Drifter and *Succession* demonstrate a generative approach to the remaking of digital cultural collections. They build on the affordances of machine-readable collections and APIs to harvest large datasets from diverse sources, and they show how these sources can be curated and redeployed to address specific topics and concerns. These digital collections are large in scale and broad in scope, while also offering intimate, immediate traces of complex domains. How might we work with such collections? These projects demonstrate the potential of a generative approach based on sampling and recombination. Juxtaposition reveals connections and contrasts, combining different domains and historical moments while retaining a resonance based on place. In multiplying complexity and privileging interpretation, generative techniques provide a means to address complex topics and heterogeneous domains. This is significant because, as Latour would argue, many of our urgent challenges are intrinsically complex, entangled and multi-domain: matters of concern. Latour seeks a way to draw things together in order to represent the complexity of matters of concern. *Drifter* and *Succession* show how mashup techniques can work with the troublesome traces of digital heritage to meet that challenge.

Acknowledgements

I acknowledge the support of CultureLab, Newcastle University, where I was a visiting fellow during the production of *Succession* in 2014; and the Land Dialogues conference and exhibition, Charles Sturt University, Wagga Wagga (2016), which was instrumental in the development and exhibition of *Drifter*.

Competing Interests

The author is a co-editor for the Special Collection "Remaking Collections." This piece was handled by another editor and underwent a standard double-blind peer-review process.

References

Anon 1891 Record: Herpetology:R.19544. *Atlas of Living Australia*. Available at: https://biocache.ala.org.au/occurrences/26c3bcc5-5c64-48ff-8964-7ed146a0545d (Last accessed 26 October 2017).

Anon 1924 Canberra Commission. *The Register*, 9 May, pp. 13. Available at: http://nla.gov.au/nla.news-article57397035 (Last accessed 26 October 2017).

Anon 1966 Cotter River Diverted. *Canberra Times,* 14 November, pp. 1. Available at: http://nla.gov.au/nla.news-article106946104 (Last accessed 27 October 2017).

Bagnall, K and **Sherratt, T** 2012 *Invisible Australians: Living under the White Australia Policy*. Available at: http://www.invisibleaustralians.org (Last accessed 31 October 2017).

Baker, J 2013 The Mechanical Curator. *Digital Scholarship Blog*, 30 September. Available at: http://britishlibrary.typepad.co.uk/digital-scholarship/2013/09/the-mechanical-curator.html (Last accessed 3 November 2014).

Barlow, K and **Doyle, M** 2011 Leaked Murray-Darling Plan Stokes Job Loss Fears. *ABC News*, 18 November. Available at: http://www.abc.net.au/news/2011-11-18/irrigators-predict-anger-over-murray-darling-plan/3678350 (Last accessed 26 October 2017).

Chakrabortty, A 2011 Why Doesn't Britain Make Things Any More? *The Guardian*, 17 November. Available at: http://www.theguardian.com/business/2011/nov/16/why-britain-doesnt-make-things-manufacturing (Last accessed 2 November 2014).

Davila, P, Colangelo, D, Chan, M and **Tu, R** 2014 Expressive Cartography and the Aesthetics of Public Visualization. In: *Proceedings of the IEEE VIS 2014 Arts Program (VISAP'14)*, pp. 56–61. Paris, France, 9–14 November.

Drucker, J 2011 Humanities Approaches to Graphical Display. *Digital Humanities Quarterly*, 5(1). Available at: http://digitalhumanities.org/dhq/vol/5/1/000091/000091.html (Last accessed 16 October 2017).

Drucker, J and **Nowviskie, B** 2004 Speculative Computing: Aesthetic Provocations in Humanities Computing. In: Schreibman, S, Siemens, R and Unsworth, J (Eds.),

A Companion to Digital humanities, pp. 431–47. Oxford: Blackwell. DOI: https://doi.org/10.1002/9780470999875.ch29

Feiler, J 2007 *How to Do Everything with Web 2.0 Mashups*. New York: McGraw-Hill Osborne Media.

Gibson, R 2006 Places Past Disappearance. *Transformations*, 13. Available at: http://www.transformationsjournal.org/wp-content/uploads/2017/01/Gibson_Transformations13.pdf (Last accessed 16 October 2017).

Internet Archive Book Images 1898 Transactions of the Institution of Mining Engineers (image from p. 339). Available at: https://www.flickr.com/photos/internetarchivebookimages/14761419364/ (Last accessed 15 March 2018).

Land Dialogues 2016 *Land Dialogues Conference*. Available at: http://scci.csu.edu.au/landdialogues/ (Last accessed 15 March 2018).

Latour, B 2004 Why Has Critique Run out of Steam? From Matters of Fact to Matters of concern. *Critical Inquiry*, 30(2): 225–48. DOI: https://doi.org/10.1086/421123

Latour, B 2008 *What Is the Style of Matters of Concern. Two Lectures in Empirical Philosophy. Department of Philosophy of the University of Amsterdam*. Assen: Van Gorcum. Available at: http://bruno-latour.fr/sites/default/files/97-SPINOZA-GB.pdf (Last accessed 7 March 2018).

Latour, B 2008a A Cautious Prometheus? A Few Steps Toward A Philosophy of Design (with Special Attention to Peter Sloterdijk). In: Hackney, F, Glynne, J and Minton, V (Eds.), *Networks of Design: Proceedings of the 2008 Annual International Conference of the Design History Society (UK)*, pp. 2–10. Boca Raton: Universal Publishers.

Latour, B 2010 An Attempt at a 'Compositionist Manifesto'. *New Literary History*, 41(3): 471–90. Available at: http://newliteraryhistory.org/articles/41-3-Latour.pdf (Last accessed 7 March 2018).

Miller, R 2014 Millions of Historic Images Posted to Flickr. *Internet Archive Blogs*, 29 August. Available at: http://blog.archive.org/2014/08/29/millions-of-historic-images-posted-to-flickr (Last accessed 2 November 2014).

Murray, R M 2016 A Manifesto for Transdisciplinary Designers. *PLOT(S)*, 3: 108–120. Available at: http://adht.parsons.edu/designstudies/plot/a-manifesto-for-transdisciplinary-designers (Last accessed 31 October 2017).

Nowviskie, B 2015 Digital Humanities in the Anthropocene. *Digital Scholarship in the Humanities*, 30(suppl_1): i4–i15. DOI: https://doi.org/10.1093/llc/fqv015

NSW State Archives 2010 *Cable Tram Trailer*. Available at: https://www.flickr.com/photos/state-records-nsw/4312109973 (Last accessed 31 October 2017).

Posner, M 2015 Humanities Data: A Necessary Contradiction. *Miriam Posner's Blog*, 25 June. Available at: http://miriamposner.com/blog/humanities-data-a-necessary-contradiction (Last accessed 31 October 2017).

Richards, K 2006 Life After Wartime. *Canadian Journal of Communication*, 31(2): 447–58. DOI: https://doi.org/10.22230/cjc.2006v31n2a1729

Seaman, B 2001 Oulipo | vs | Recombinant Poetics. *Leonardo*, 34(5): 423–30. DOI: https://doi.org/10.1162/002409401753521548

Sherratt, T 2013 Conversations with Collections. *Discontents*, 24 June. Available at: http://discontents.com.au/conversations-with-collections (Last accessed 3 November 2014).

Sherratt, T 2014 *Eyes on the Past*. Available at: http://eyespast.herokuapp.com (Last accessed 2 November 2014).

Sherratt, T 2015 Seams and Edges: Dreams of Aggregation, Access & Discovery in a Broken World. *Discontents*. Available at: http://discontents.com.au/seams-and-edges-dreams-of-aggregation-access-discovery-in-a-broken-world (Last accessed 31 October 2017).

Spielmann, Y 2002 Interview with Bill Seaman. *The Daniel Langlois Foundation*, n. dat. Available at: http://www.fondation-langlois.org/html/e/page.php?NumPage=386 (Last accessed 27 October 2014).

Thackara, J 2015 *Bioregions: Notes on A Design Agenda*. John Thackara Personal Website, 23 April. Available at: http://thackara.com/place-bioregion/bioregions-notes-on-a-design-agenda (Last accessed 23 October 2017).

Whitelaw, M 2014 *Succession.* Available at: http://mtchl.net/succession (Last accessed 15 March 2018).

Whitelaw, M 2014a *Succession* Saved Composite 1413513552860. Available at: http://mtchl.net/succession/#/saved/1413513552860 (Last accessed 15 March 2018).

Whitelaw, M 2015 Generous Interfaces for Digital Cultural Collections. *DHQ: Digital Humanities Quarterly*, 9(1). Available at: http://www.digitalhumanities.org/dhq/vol/9/1/000205/000205.html (Last accessed 22 December 2016).

Whitelaw, M 2016 *Drifter.* Available at: http://mtchl.net/drifter (Last accessed 15 March 2018).

How to cite this article: Whitelaw, M 2018 Mashups and Matters of Concern: Generative Approaches to Digital Collections. *Open Library of Humanities,* 4(1): 26, pp. 1–28, DOI: https://doi.org/10.16995/olh.291

Published: 24 April 2018

Open Library of Humanities is a peer-reviewed open access journal published by Open Library of Humanities.

OPEN ACCESS