Arch Sci (2015) 15:25–49
DOI 10.1007/s10502-013-9212-y

ORIGINAL PAPER

Visual analytics, cognition and archival arrangement and description: studying archivists' cognitive tasks to leverage visual thinking for a sustainable archival future

Victoria L. Lemieux

Published online: 29 September 2013

Abstract This paper explores the possibility of using visual analysis to support archival functions. It first reviews the current state of experimentation with the use of interactive visual interfaces in the archival domain, noting several gaps including the absence of formal studies of cognitive and perceptual tasks. The paper then reports on preliminary results from a cognitive task analysis of archival arrangement and description, providing a broad brush description of archival analysis during arrangement and description as an example of archival sense-making. It then suggests some possible leverage points where visual analytics technology might be applied to support archival arrangement and description and concludes by reflecting on some theoretical and practical aspects of visualization technologies.

Keywords Arrangement and description · Visual analytics · Visualization · Cognitive task analysis · Protocol analysis · Sense-making · Human–computer interaction

Introduction

Visual analytics (VA)—"the science of analytical reasoning facilitated by interactive visual interfaces" (Thomas and Cook 2005, p. 28)—is being applied in many domains to assist analysts where there is a need to process masses of complex data and blend computational analysis with interactive visualization of the results (VisMaster 2010). For this reason, it is an approach that is well suited to the problems archivists face in arranging and describing vast quantities of digital

V. L. Lemieux (✉)
School of Library, Archival and Information Studies, I.K. Barber Learning Centre, University of British Columbia, 4th Floor, 1961 East Mall, Vancouver, BC V6T 1Z1, Canada
e-mail: vlemieux@mail.ubc.ca

records. Like many other organizations, archival institutions are facing a data tsunami, as data are growing massively and rapidly: about 2.5 exabytes (2.5 million terabytes) of data are created every day, and 90 % of the data in the world today have been created in the last two years (IBM 2012a, p. 1). Not surprisingly, it is a challenge to store and process data that are growing so rapidly. Data are also changing quickly to become more complex and less structured: according to IBM, an estimated "80–90 % of any organization's data is what is referred to as unstructured… [and] that data is growing at 40–60 % per year" (IBM 2012b). The International Data Corporation projects that unstructured data will grow at 60 % or more compounded annually, far higher than transactional (structured) growth rates of 20 % plus (Villars 2012).[1] There are endless new types and sources of data: IBM cites "the rising use of interactive web technologies, such as blogs and social media platforms" as one of the big drivers in the growth of complex data (IBM 2012b). Given these developments, it is increasingly difficult to manipulate the flood of data and information and to make sense of it (Hey 2004). The sheer volume of information becomes overwhelming and the lack of structure renders it decontextualized and meaningless (Hey 2004).

In the light of these changes, traditional approaches to managing archival documents seem unlikely to remain tenable. This is particularly the case for archival arrangement and description, where backlogs of material waiting to be processed are already the norm rather than the exception to the rule. As an example, a 2006 survey in the United States indicated that 37 % of the average university repository holdings were unprocessed, and thus effectively unavailable for use (Prom et al. 2007, p. 159). Given such backlogs in archival arrangement and description, archivists have begun to explore new approaches to ensure that it is a sustainable activity in the era of 'big data'. Greene and Meissner's (2005) 'more product, less process' methodology aims to speed up processing by creating less detailed finding aids. It has attracted the attention in the US archival community over the last few years. But in a survey of Canadian historians in 2004, 37 % of respondents reported that existing finding aids already were not detailed enough to meet their needs (Duff et al. 2004, p. 14). Minimal processing solutions may increase archivists' productivity but seem unlikely to generate finding aids that enhance scholarly research processes or protect the authenticity, security, or integrity of archives. Another approach was suggested by Evans, that instead of constructing detailed descriptions themselves, archivists should rely on web-based user contributions, establishing systems that use "the eyeballs and the intellect of… volunteers… throughout the world" (2007, p. 397). This suggestion chimes with ideas about the need to reflect different voices and multiple perspectives in description, by means of communicating alternate representational possibilities (Meehan 2010) or by introducing version control in finding aids to make explicit their evolution and the corrections and adaptation they have undergone over time (Yakel 2003). Similar to Evans, others have also suggested removal of some descriptive tasks from the

[1] Readers should be aware that this report is behind a pay wall and only accessible at a significant cost; however, the IDC site from which the publication can be accessed does point to a number of sources that cite the report and from which a good amount of information about what is in the report can be obtained.

archivist's work schedule by 'crowd sourcing' the processing of archival materials. Duff and Harris (2002) have suggested incorporation of contributions from users in finding aids as yet another approach to reducing the burden of archival arrangement and description on archivists. In practice, most archivists who have sought ways of involving users in description, typically through web 2.0 interfaces (see Krause and Yakel's (2007) work on next generation finding aids for an example), have found that establishing, maintaining, and monitoring forums for user engagement are themselves very demanding tasks, especially if quality control is to be achieved.

Some archivists have even explored the possibility that traditional approaches to aggregating documents—as is done in arrangement and description after records are transferred into the custody of an archival repository—may be outmoded in the digital era (Bak 2012). Other writers, such as Hurley (1998), dismiss archival arrangement and description as being necessary only for archives that collect records from organizations lacking records systems where descriptive metadata can be defined and appended to records at creation. Coming from a records continuum perspective, Cunningham (2001) has argued for merging records management metadata practices with archival description. Others, however, suggest that archival arrangement and description have a role beyond what can be provided by records management metadata in supporting a more contextual perspective (Gilliland-Swetland 2005; MacNeil 2005; Nesmith 2006) and, as Yeo (2010) notes, recent discourse has moved away from the simple view that all or most descriptive needs can be met by capture of metadata at record creation alone; rather, there is an understanding that descriptive information can be added progressively throughout a record's life and reused and repurposed over time. These debates about archival arrangement and description illustrate how much the digital era is presenting unprecedented challenges to archivists' traditional ways of carrying out one of their core archival functions. At the same time, it is opening up opportunities to transform the way that such work is conducted. Thus, now is the perfect time for the profession to explore new technologies and new approaches to the performance of archival activities. One such new approach may be VA.

VA can be classified as an 'intelligence augmenting' technology that harnesses the power of human visual perception and cognition. Visualization has advantages over other modes of communication and thought because humans have evolved visual and spatial skills that include the ability to detect edges and discontinuities, to recognize patterns, and to retrieve information using visual cues (Kosslyn 2005; Lurie and Mason 2007). Each of these visual and spatial attributes (i.e. edges, discontinuities) can be transformed into a graphical image to provide a rich visual description of data. As these features can be observed with 'pre-attentive processing', that is, they are perceived prior to conscious attention, they are understandable at a glance and much more rapidly than words (Ward et al. 2010). Through encoding of data into graphical images, visualizations can act as a repository of data which allows individuals to offload cognition to the perceptual system, using visuals as a form of virtual memory (Munzner 2009). This can enlarge problem-solving capabilities by enabling the processing of more data without overloading the decision-maker (Tegarden 1999). Finally, because visual cues stand out to human perception more than words, a picture sometimes forces us to notice

what we never expected to see, thereby allowing for greater insight (Ward et al. 2010).

Visual analytics adds a computational component to visualization. In essence, VA uses computation (for example, text mining or machine learning algorithms) to offload cognitively demanding tasks from human users. Other tasks that are performed more effectively by humans are facilitated by careful design of interactive visual interfaces that leverage the human visual system to aid perception and cognition. Human and machine reasoning function together in a VA system to combine the best of both components. Thus, VA could provide a possible solution to future sustainability of archival arrangement and description in the archival function's current form or in some possibly transformed future iteration.

Development of VA technologies requires an understanding of the analytic task that the technology is being designed to support. The VA research cycle therefore includes working with decision-makers in the context of their organizations to characterize data and solutions in the situations in which analysis takes place (Fisher et al. 2011). This work defines research questions for laboratory investigation, the results of which guide the design of new interactive visualization technologies and analytical methods that are evaluated in partnership with the decision-makers and their organizations. This 'translational' research approach (Fisher et al. 2011) follows human-centred design principles and processes (ISO 9241-210 2010), which base technology design on an explicit understanding of users, tasks, and environments; engage with users throughout the design process; and follow an iterative process which is driven by user-centred evaluation of products (mock-ups, prototypes, and other artefacts) throughout.

This paper discusses the first phase of a design study aimed at the application of visual analytics technologies in support of the traditional model of archival arrangement and description. Following a typical design methodology, the aim of the study was to understand the analytic tasks that the visual analytics system is being designed to support; that is, to understand how archivists think as they undertake archival arrangement and description.

Related literature

Literature on archival arrangement and description

Surprisingly, a review of the archival literature on arrangement and description has revealed that explanations of the archival reasoning process, by and large, are absent. Most of the literature on arrangement and description can be grouped into two categories in relation to how it reflects upon arrangement. The first discusses the main concepts involved in arrangement. Provenance, *respect des fonds* and original order are the key concepts examined. In order to evaluate whether, when, and to what extent these concepts are pertinent to arrangement and description, authors explain how they were developed within specific recordkeeping traditions, and how they have been used and adapted in different countries. Recognizing the difficulty of achieving consensus on such key concepts, and the challenges involved when

applying them to concrete situations, authors sometimes suggest a few pragmatic rules on how to deal with conceptual inconsistencies in practical terms (Duchein 1986; Eastwood 2000; Horsman 2002). MacNeil (2005) notes, for example, that in recent years the literature on archival arrangement and description has been dominated by the discussion of the relative merits of fonds versus record group and fonds versus series-based arrangement and description, on issues associated with the development and implementation of national and international descriptive standards as a means of enhancing intellectual retrieval, and on the needs of researchers in both traditional and web-based environments. When discussing concepts, however, authors do not make clear how archivists actually reason about their concrete application of the concepts when arranging and describing different fonds.

The second category of literature about archival arrangement and description tends to be normative or prescriptive in nature and deals with the practicalities archivists have to address when actually performing arrangement and description (i.e. by providing procedural steps). Targeting professionals looking for guidance concerning specific routines, this group of works sees arrangement and description as process. It also stresses the steps involved in their execution. Authors suggest preparatory tasks, such as looking for secondary sources related to the context of records creation and having questions in mind during the first contact with the records, as well as commenting on general organizational issues, such as how to set up a proper workspace. These works also give hints and present examples to help archivists establish levels of control (Millar 1988; Miller 1990; Roe 2005). Instructions provided in this literature mostly correspond to a sequence of required or suggested activities regarding arrangement and description, but do not explore the reasoning implied before or during the execution of the suggested steps.

There is also literature related to technology development projects designed to support archival arrangement and description in a digital context and which have involved some study of users' task requirements as part of the development process. The Archivopteryx Project at Simmons College (Anderson et al. 2010) uses web-based technology to teach students about the archival analysis process involved in arranging and describing archival records. Hypatia (AIMS 2012) is an initiative to create a Hydra application that supports the accessioning, arrangement and description, delivery, and long-term preservation of born digital archival collections.

Since the 2000s, another category of studies has been slowly emerging. By explicitly or implicitly acknowledging the role of archivists in shaping the meaning of records and archives, these works offer an important step forward in understanding how archivists think as they conduct arrangement and description. In an examination of various archival representational forms (finding aids, EAD descriptions, and MARC records), Yakel (2003) sees arrangement and description as a fluid, evolving and socially constructed practice, which she names 'archival representation'. MacNeil (2005), in comparing the work of archivists during arrangement and description to that of textual critics, calls attention to the selectivity involved in archival description, as well as to its incomplete character. Meehan (2009, p. 72) sees arrangement as "identifying and creating the contextual relationships of a body of records", and points to the historical standpoint of the

archivist, the use of evidence, and the role of inference as central factors in the analytical process in arrangement and description. The works of Yakel, MacNeil, and Meehan provide a solid starting point for research into how archivists think; yet, these works also demonstrate the pertinence and necessity of a thorough empirical investigation of the cognitive and perceptual tasks involved in archival arrangement and description.

Though it is not based on empirical research, Meehan's work comes closest to the objective of this study, which is to describe and develop a conceptual model of the process of archival analysis during arrangement and description which can be used to design a VA tool for archivists to use in processing archival documents. She observes that:

> For the processing archivist, speculation is key to inferring what may be known about context from what is known in the form of gathered contextual information. In other words, it is a key means of making the leap from the parts available – the bits and pieces of information gleaned from the records and other documentation – to an imagined whole – an understanding of the contexts of records creation, maintenance, transmittal and use. This in turn lays the groundwork for transforming the physical parts in hand – rarely, if ever, "the whole of the records" – into a meaningful, if imaginary, whole – a collection or fonds, organized and represented in a finding aid (2009, p. 84).

Meehan describes this process as "the archivist's making sense of the records en route to contextualizing them" (2009, p. 85).

Literature on human information behaviour, sense-making, and information processing

Given Meehan's reference to sense-making, it is useful to consider the analytic process of archival arrangement and description in the context of different theories of sense-making, particularly theories of experts' sense-making grounded in systems engineering and human factors research. Sense-making has been defined as how people make sense of their experience in the world using pieces of information or data, often under uncertain and complex conditions (Duffy 1995).

Brenda Dervin's theory and methodology of sense-making is probably best known in the world of information science. Dervin (2003) has studied sense-making in the context of human information seeking behaviour, emphasizing the 'cognitive gap' that people experience as they try to make sense of observed data. In the context of this study, theories of individual expert sense-making are especially relevant, and two stand out in particular. Klein et al. (2006) researching sense-making from the perspective of expert decision-making and situational awareness, discuss sense-making as the interrelation of concepts such as creativity, curiosity, comprehensive mental modelling, and situational awareness. They view sense-making as a series of processes that are initiated when an individual or organization recognizes the inadequacy of their current understanding of events. While any one of the concepts can help an individual to make sense of their situation, the combination creates a continuous effort to understand connections (which can be

among people, places, and events) in order to anticipate their trajectories and act effectively. These efforts to understand events start from a personal perspective, viewpoint, or understanding, which Klein et al. call a frame (or what also could be otherwise described as a mental model). Sense-making, according to these theorists, is an active two-way abductive reasoning process of fitting data into a frame and fitting a frame around the data. Abductive reasoning typically begins with an incomplete set of observations and proceeds to the likeliest possible explanation for the set. It is the kind of daily decision-making that does its best with the often incomplete information at hand. This differs from an inductive reasoning process wherein generalizations are made from specific observations. In the Klein et al. model, neither data nor frame comes first: data evoke the frame and the frame selects and connects data. When there is no adequate fit, the data may be reconsidered or an existing frame may be revised. This describes the metacognitive processes that are used by individuals to build, verify, and modify working models to account for an unrecognized situation (Fig. 1).

The second relevant model derives from Pirolli and Card's study of sense-making in intelligence analysis. Their model places emphasis on the importance of expert schemas (a set of patterns around the important elements of their tasks) (Pirolli and Card 2005). Pirolli and Card's interviews with analysts revealed that schemas were used to organize and represent incoming information. Analysts developed insights through manipulation of the schema using pen and paper or a computer-based system, and then finally created some knowledge product or decision based on their insights. Pirolli and Card organize this process into two major loops of activities: (1)

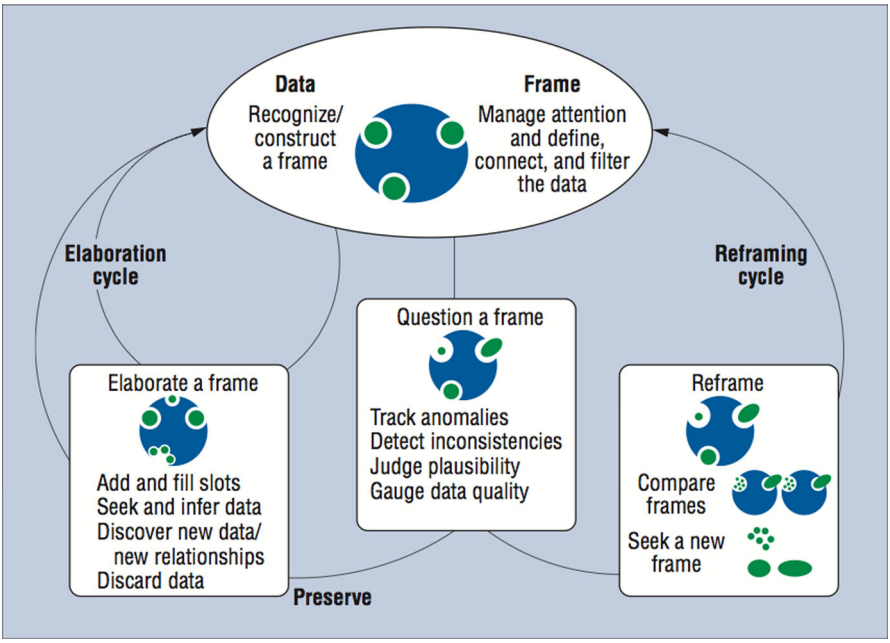

Fig. 1 Klein et al.'s (2006) data-frame model of sense-making

a foraging loop that involves processes aimed at seeking information, searching and filtering it, and reading and extracting information, possibly into some schema and (2) a sense-making loop that involves iterative development of a mental model (a conceptualization) from the schema that best fits the evidence. They note that information processing can be driven by bottom-up processes (from data to theory) or top-down processes (from theory to data), which are used in an opportunistic mix (Fig. 2).

Both of these models can be contrasted with the classic data, information, knowledge, wisdom (DIKW) model—a model widely cited in the fields of computer science, management, and information science representing purported structural and/or functional relationships between data, information, knowledge, and wisdom—which posits that human information processing proceeds in a linear and hierarchical fashion transforming data to information, information to knowledge, and knowledge to wisdom (Ackoff 1989; Frické 2009). A number of variations of the model exist. Data were not in the original model suggested by Eliot or Harland but were added later on (Hey 2004). Since then, others have also proposed extensions to the hierarchy; Ackoff includes understanding (and some use intelligence) as its own level before attaining wisdom, and Zeleny proposes enlightenment as the final stage beyond wisdom (Hey 2004). Others, such as Rowley (2007), leave wisdom out of the model altogether. Frické (2009) has criticized the DIKW model as being reductionist and dependent upon outdated and discredited ideas such as operationalism. Hey's (2004) study of the model uses metaphorical analysis to explore the relationships between data, information and

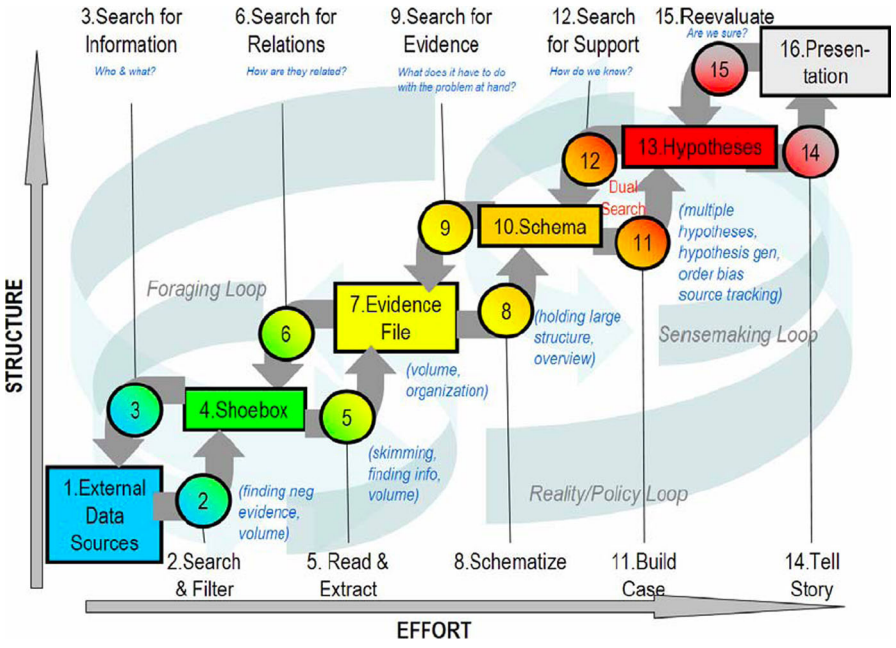

Fig. 2 Pirolli and Card's (2005) model of sense-making

knowledge. Hey discusses, in particular, the metaphor of 'knowing as seeing' in which individuals make use of their physical experiences to help structure their thinking on more complex abstract concepts (2004, p. 4).

Literature on archival use of information visualization and visual analytics

In terms of VA and archival arrangement and description, there are, at present, no studies that apply VA strictly to support this archival function, though there are a number of studies that use visualization to represent the outcome of archival analysis or which use VA to support access to archival collections and archival preservation decision-making. For example, an early and common visualization is a hierarchical node link graph, which is also typically used in representing organizational charts. These graphs have been used to represent the relationship between fonds, *sous-fonds*, series, and items in an archival fonds. The introduction of XML encoded-finding aids, particularly EAD (encoded archival description), and widespread implementation of descriptive standards such as ISAD(G) (general international standard archival description), has created opportunities for the introduction of new visualization tools that leverage archival descriptive metadata. Anderson (2009) has explored the use of visualizations to represent multiple 'dimensions' (for example, relationships among persons, committees, and so on) in online finding aids. His project sought to visualize archival information by applying Ted Nelson's ZigZag™ structure to two existing EAD finding aids.[2] Allen (2005) has explored the possibility of visually expressing hierarchies, networks, processes, and timelines in US government archival holdings using EAD to extract data and structure from source documents. Similar in purpose to Anderson's work, Allen proposes an interactive approach to visualizing links between archival materials and has created a prototype interface that uses a 'mass-spring model' to spatialize the relationship among the concepts (Provot 1995). ArchivesZ, developed by Jeanne Kramer-Smyth, is a prototype of an information visualization tool that, in the same vein as Anderson and Allen's work, leverages the structured data available in EAD encoded-finding aids. By representing the distribution of subjects and time periods using the metric of total aggregate linear feet, Kramer-Smyth argues that ArchivesZ enables tool users to view total available research materials more quickly than they would by viewing a standard search result list (Kramer-Smyth et al. 2010). More recently, researchers at the University of Texas have developed a prototype VA system to aid archivists in identifying files requiring digital preservation and as a possible VA tool in support of archival research (Esteva et al. 2011). They developed a visualization based on a space-filling treemap (Johnson and Shneiderman 1991) to present digital file-related metadata extracted from the collection at different levels of aggregation and abstraction. Mitchell Whitelaw has also developed a number of interactive visual interfaces to explore archival collections (Whitelaw 2009, 2012; Hinton and Whitelaw 2010). Ribarsky et al. (2011) have begun to explore the application of VA technologies to extract and visually

[2] For more information about ZigZag™, see http://users.ecs.soton.ac.uk/lac/zigzag/, accessed on 10 June 2013.

represent spatiotemporal data from historical archives to develop the whole story of place.

In summary, the application of visualization and VA in the archival domain primarily has relied upon using archival descriptive metadata; that is, it has focussed on creating visualizations using post-archival analysis data (EAD encoded-finding aids). Some researchers have begun to explore the application of visualization and visual analytics approaches directly to the content of archival documents and associated pre-archival descriptive metadata. To date, visualizations have, by and large, been targeted at end-users, rather than focused on aiding archivists to perform archival functions. None of the work so far has relied upon a formal analysis of researchers' or archivists' cognitive tasks in the design of tools supporting interactive visual analysis. It is this necessary, but overlooked, aspect of interactive visual system design that the research discussed in the next section of the paper seeks to address.

Methodology

As mentioned, this study sought to explore archivists' cognitive processes as they arrange and describe archival material as part of a broader design study on applying VA to the arrangement and description of archival material. In essence, the research questions were, 'How do archivists think as they undertake arrangement and description?' and 'What are the implications of how archivists think for the design of a VA tool to support archival arrangement and description?' To seek answers to these questions, the researchers used cognitive task analysis and verbal protocols (Trickett et al. 2007; Crandall et al. 2006; Schraagen et al. 2000) to observe archivists arranging and describing archival records in an archival fonds.

Two archivists participated in the study, both of whom were recent graduates and relatively new to the archival field. The participating archivists were asked to conduct their standard archival procedure on the records, while being observed using the 'concurrent' think aloud verbal protocol (Trickett et al. 2007; Kuusela and Paul 2000; Lewis and Rieman 1993; Ericsson and Simon 1993; Capon and Davis 1984). This particular verbal protocol is a method commonly used in interface design and evaluation, and entails asking participants to describe what they are doing as they perform a task. This might include descriptions of what they are thinking, looking at, feeling, any unexpected problems they might be having, and so on. This allows any internal information that may be apparent to the subject but not to the researchers to be made externally evident, thus allowing for insight into the archivists' experience. In spite of some limitations (Payne 1994; Kuusela and Paul 2000), this method can help to reveal how an analyst resolves uncertainties, to determine the implicit heuristics or qualitative mental models that an analyst might be employing, or to identify common conceptual simulations that arise during an analysis (Trickett et al. 2007). There were five separate sessions with the archivists, each lasting between 80 and 260 min depending on the participants' pace and the range of activities being performed. Sessions were conducted in the participants' normal archival location, so as to achieve a more naturalistic recording of the

methods involved in archival processing. Participants were audio-recorded with their consent, and text transcriptions were made of each session. A brief interview also followed each session, to gather retrospective data to complement data gathered during the arrangement and description process (Kuusela and Paul 2000). Observational notes on the environment and process (such as how the archivists interacted with the archival material and with one another) were taken by the observers.

During each data-gathering session, the archivists worked as a pair, articulating their thought processes as they arranged and described the material. This modified approach to the use of the standard think aloud protocol, inspired by pair analytics (Arias-Hernández et al. 2011), overcame one of the weaknesses of the think aloud method, which is that as individual participants become more cognitively engaged in their work, they articulate their thoughts less frequently (Randsell 1995; Kuusela and Paul 2000). The research team found that working in pairs encouraged the archivists to continue to articulate and communicate their thoughts to each other throughout the process.

The archival records used in the study—99 cm of textual records, 2 floppy disks, and 1 film reel of private papers—had not been previously processed, so as to obtain a clear understanding of the entire process rather than repeating arrangement and description on a known set. This approach also has allowed uncertainties to arise as normal during the process, since with previously processed material such problems may have been expected or already dealt with.

Text transcriptions were made of each session and related observational notes were analysed using a grounded theory methodology (GTM). GTM involves the generation of theory from the data collected through the conduct of research and, as proposed by Corbin and Strauss (2008), starts with 'open coding' of the data by breaking down, examining, comparing, conceptualizing, and categorizing data. The data were coded and categorized into major task-related segments and these segments were, in turn, used to infer a model of the archivists' analytic process during arrangement and description (Kuusela and Paul 2000).

Findings

The research team observed that archival processing followed a three-stage sense-making process: in phase one, from first contact with the archival records, the archivists were concerned with gaining an overview of the structure of the archival fonds and creating a draft arrangement; in phase two, the archivists confirmed and refined the arrangement structure and re-ordered archival documents, as necessary; and in phase three, they described the final arrangement, documenting their description in an archival finding aid (Fig. 3).

The first phase, consisting of analyzing the archival material and developing a mental model of its contents, represents a significant cognitive load. The subjects reported being much more fatigued in the first session than they did in any other session. Researchers observed that there was less joking during the process and there were more recorded instances of archivists' use of language expressing uncertainty (such as the use of 'hmm' and 'um') in verbalizing their thought

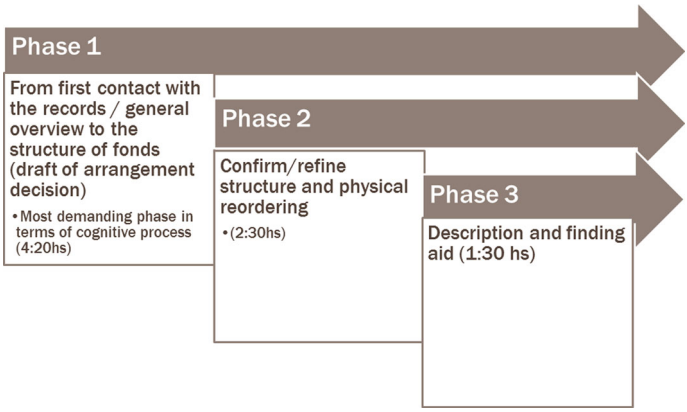

Fig. 3 Making visual sense of records: three-stage model of archival sense-making during arrangement and description

processes. Part of this may be accounted for by the fact that the participants were still adjusting to being observed and performing the think aloud protocol, but observers remained silent throughout, and both subjects worked in their standard location and using their own methods. More time was devoted to figuring out the timeline and context of the fonds; indeed, this was the longest of all the sessions (4 h and 20 min). The following extract illustrates this sense-making process:

1. S1: from when?
2. S2: 1984
3. S1: that seems awfully late
4. S2: His vitae
5. S1: unless it's for a promotion?
6. S2: Maybe. Well he was … here, 84–85 … he's got records here up to 92, maybe he has funds that he knows are incoming…or maybe this is outdated?
7. S2: I'm just crosschecking dates with a couple of my copies…

Observers noted that the physical act of rehousing documents in new acid-free folders facilitated archivists' review of the material in the fonds and the process of forming a high-level mental overview of the fonds, as suggested by this statement:

> S2: okay so I put the accession file aside and I'm going to be starting going through my first box, um, focusing first on uh simple rehousing a lot of these materials because they are quite deteriorated um, and coming up with potential file titles because the first there are three ledgers and four notebooks. It looks to be all on research and different chemical compounds…
> And
> S1: So I just finished all my refoldering, hurray, so now I can start seeing I can come up with some basic series with my stuff.

After the mental model was formed, particularly in the second and third sessions, both participants were able to proceed more quickly.

Further contextual information was easy to integrate once they had determined an overall mental model of the total arrangement, and fewer uncertainties were exhibited when new information became evident. The archivists were even able to infer the date and purpose of various files once this general mental framework was established. A more solid understanding of the fonds appears to have been established in the second session, when the subjects began deciding on the series structure despite not having completed arrangement on all of the files and having focused mostly on rehousing. Even with these unknowns, their final decisions on series closely mirrored the ones they considered at this point. Similarly, initial predictions on the content of the individual folders or containers were more accurate once they had established more information about the creator of the files. The points of confusion that arose in the later sessions were raised because of inconsistencies in what they had established in their interpretation of the timeline and the context of the files they were working with. While neither archivist explicitly mentioned the development of a mental model, the cognitive task analysis carried out for this study suggests that creating a mental model is a major part of the process of arrangement and description.

There was not a strong focus on any specific template or archival style throughout the arrangement. Both subjects followed their own method without significant difficulties, and were able to integrate their work with one another. While there was a little debate on the number of series or the manner in which to categorize them, both archivists noted that their respective choices were based on their own experience rather than a specific method. This supports the idea that each archival arrangement has its own element of uniqueness, in spite of standardized approaches to arrangement and description, as each archivist may make different choices and the process itself may differ depending on the archivist. By allowing the archivist to make note of their unique 'take' on the fonds and having them make known the context that they considered important, the process of arrangement and description was made more transparent.

The actual form of the data was very important for both subjects as this exchange between the archivists reveals:

1. S2: I'm starting with box three. I'm just going through some preliminary observation to kind of get a sense of what's here
2. S1: So I've got a full box of file folders and it is slightly easier. And they're all titled
3. S2: and um it looks like I've got some ledgers to begin with that are in some, um, they have various states of deterioration so they're going to have to be handled with care
4. S1: one thing about file folders is that although I like file folders normally, often times um people like to reuse file folders I've found so we always have to look inside them and this person seems to have liked to write on folders which is always an interesting challenge because often times we rehouse files in the folders but if the file folder is a part of the file then you have to keep it. I've got a file of letters that appear to relate to the title they're all the same date
5. S2: It looks like I've got some … notes here

6. S1: Aww I got personal correspondence! Often times it's really helpful to have um distinctions between different kinds of correspondence because correspondence often times becomes a big miscellaneous file of correspondences all over the place so it's nice to know there's correspondence from different people
7. S2: Um I've got a couple more notebooks in here and I'm going to have to look at more material before I decide kinda where they fit into the scheme of things, um, but at the bottom of this box there's another box so I'm gonna have a look and see what's in there …

While the importance of form is not in itself surprising given the old archival adage of 'form following function', it is of interest from a design perspective. The archival fonds that the subjects worked on was relatively small, comprising only three major divisions, but it still contained a large amount of information. It would have been impractical to have attempted to read every single piece of data or document, or to explore every film reel or floppy disk. The amount of content in even this relatively small fonds meant that trying to create series by reading all content would have been a time-consuming proposition. However, both archivists were able to quickly examine and categorize hundreds of files into their various series by paying attention to the form of the material.

Grant applications, for example, were very common throughout the files that the subjects received. Different grants had different application forms, but they tended to follow a general spatial structure, and at the very least were simple to quickly identify and recognize as being grants as opposed to correspondence. After a fairly in-depth examination of a few of these grants, the archivists were able to recognize files that fit the visual description of a grant, and were then quickly able to pick out the relevant content, such as the size of the grant, the year, or the organization providing the grant:

> S1: Hmm. I've got another grant application here that's three so far, and this one appears to show he was collaborating with I think biologists? and to do um, experimental tests um of his theories …

This then allowed the archivists to place the grant in a timeline of the creator's life, and to draw inferences as to what the creator may have been doing with regard to their career, or even their geographical location at the time. With time information, the archivists could then further guess as to the contents of the folder, and ascertain the original order of the material. This then expanded to other files and folders; more grant applications allowed them to apply the same analytic procedure and quickly categorize other folders. The archivists were able to pull even more context-specific information by examining things like resumes, which they did not obtain until arranging the last division of files. Naturally, they could then determine that one series would likely be grant applications, and apply similar methods when determining their other series. This is not exclusive to types of files or documents that reveal personal accomplishments. One subject noted that looking for e-mail correspondence was a good way to identify timelines, as these types of records are always dated—a simple way to help obtain more context about the data.

The subjects also used a number of other programs or resources to help their process, as illustrated by this extract from the transcripts:

> S1: I'm here on my laptop which is quite nice because it has a dictionary on it so I can look up words I have no clue what they mean, like, morphogenesis. Morphogenesis. Just to find out um…what I'm working with here. More grant applications here…okay…well that's really not helpful…heh, morphological…form of biology that deals with the relationships between living structures. More grants,…

Dictionaries and an online encyclopedia were used to define or elaborate on terms they were not familiar with, while physical implements like rulers were used to help mark their current location in a box as they arranged it. Using a text editor, they made note of the original organization of the archival data. In some cases, there were no available aids, such as when comparing two photographs to determine whether they were duplicates, or trying to find whether an article had been published in a journal.

Discussion

Archivists' cognitive processes

The research findings reported in the previous section provide empirical evidence about archivists' cognitive processes during arrangement and description that can be compared to reflections on this process in the archival literature. Meehan (2009) describes archival arrangement and description as a two-step analytic process: in the first step, the archivist gathers contextual information and in the second step, the archivist uses contextual information to generate an understanding of the various contexts of the records. Meehan states that 'reading for context' aptly describes certain interactions between the archivist and the records. With respect to the archivist's interaction with records, she observes that:

> The top-down mode of analysis typically involves reading documentation by and about the creator. The archivist might read mandates, annual reports, organizational charts, and the like, or CVs, reference works, and printed material to take note of functions, activities, dates, names and places… Likewise, the bottom-up mode of analysis typically involves reading the records themselves, individually and collectively, to take note of names, dates, record formats, and any existing organization, seeking to gather any available facts that provide specific clues about the internal relationships, or original order, of a body of records (Meehan 2009, p. 80).

In contrast, the research team in this study did not find evidence of a two-stage process nor of a sharp distinction between a top-down and bottom-up approach. Rather, as mentioned previously, the team found that archival processing followed a three-stage sense-making model in which gathering and using contextual information occurred throughout. As in the study by Klein et al. (2006), the processes of

reasoning from the evidence in the records and searching for evidence from other sources took place in parallel and in an opportunistic mix.

The importance of form in the arrangement and description process is of interest in that it sheds light on reasoning in the context of the arrangement and description process. Drawing upon the thinking of David Schum, Meehan emphasizes the importance of inference in the process of archival reasoning and cites Schum as saying that "evidence and inference are of concern to any discipline and practical activity in which conclusions are reached and decisions are made on the basis of incomplete information" (Meehan 2010, p. 78). In this study, the research team found evidence to support the assertion that archivists use inferential reasoning, especially in inferring information about function from the form of records. The team also found support, however, for Klein et al.'s (2006) view that sense-making does not follow the supposed progression of data to information to knowledge to wisdom, wherein primitive data or isolated cues are successively massaged by inferential operations until they emerge from the other end as knowledge or wisdom. In other words, reasoning in the archival process of arrangement and description is, more often than not, abductive.

Pirolli and Card, citing a study by Ericsson and Lehmann, have observed that experts do not simply review the data and match them to patterns directly from memory; instead, they select the relevant information and encode it in special representations (2005, p. 2), which seems very close to how archivists described their analytic process in this study. This finding supports Meehan's statement about the constructed nature of the records: "Rather than merely identifying these relationships on the basis of gathered information, the archivist for all intents and purposes constructs these relationships … drawn from the gathered information" (Meehan 2009, p. 82). If further research verifies this finding, the implications could be far-reaching for more traditional archival theories–including appraisal theories—purporting that archivists uncover and preserve relationships in the records rather than construct them.

A surprising finding was how important the physical process of refoldering or rehousing archival documents is in the analysis that goes on during archival arrangement and description. Archivists used this time to gather a significant amount of contextual information in order to build up their mental model without disturbing the physical order in which they found the records. Archivists' interaction with the physicality of records harkens back to the aforementioned work by Jonathan Hey. Hey's (2004) metaphorical analysis of the DIKW model notes that language, in particular words that express materiality such as 'flow', 'explosion', and 'mine', encapsulates individual's use of physical experiences to help structure their thoughts on more complex abstract concepts, such as data. The archivists who participated in this study were not using language, but were using the actual materiality of the records to create mental 'space' to think about more abstract concepts of series and the intellectual order of the archival documents with which they were interacting.

Design implications

While the observations were done only on physical archives as opposed to digital ones, relevant information can still be drawn out and applied to the design of a VA

tool for archivists to use in working with digital archives. Overall the findings suggest the need for a VA system that will visually support archivists' thinking by helping them to extract (by using computational methods) and use contextual information (such as events, people, places, forms of records) throughout the process of arrangement and description. The tool must support visual exploration of the data space to aid the archivist in determining relationships among such entities and a suitable arrangement and description of material in an archival fonds.

As form is a key mechanism by which archivists infer contextual information, it should be incorporated into the design of a VA system to support archival arrangement and description. Though form is more apparent in physical, analogue documents, it is still present in digital media. File type alone can reveal if something is a text document, an image, a sound clip, a movie, and so on. However, a preview of a word-processed document may further reveal to an archivist that it pertains to something like grant applications by virtue of the spatial layout of elements of form within the document itself. Our design concept therefore provides support for archivists to use form to infer context as part of the arrangement and description process. This approach also allows for cutting down the cognitive load for archivists in manually comparing duplicates.

To leverage archivists' use of form to infer context, we propose to develop a computer algorithm to cluster archival records by their form. File type is the most obvious attribute to use in the clustering of material of similar form, but file type alone is insufficient. Our study reveals that archivists also use the spatial layout of certain types of documents (for example, emails, word-processed documents, forms, pdfs, and so on) to identify form. Thus, we propose to enhance our algorithm by incorporating a calculation of the spatial layout of documents (i.e. representing the documents as a matrix of spatial regions which are assigned a computed value). We can further enhance the design of our algorithm by studying the spatial regions of particular documents to which archivists pay most attention as they categorize them during arrangement and description. In order to understand these visual perception mechanisms better, we propose to carry out an eye-tracking study of visual attention during arrangement and description. This will provide us with data about the spatial regions that should be given greater weight in the design of our clustering algorithm. Using this approach, we will be able to cluster documents by file type as well as by their spatial layout to reduce cognitive demand on archivists faced with grouping large volumes of archival material by form. In order to leverage the human capability to visually detect patterns and outliers, we propose to visually represent archival documents of similar form as clusters, with individual documents semantically mapped to a single dot in a cluster and projected onto a 2D space. This approach is consistent with Shneiderman's (1996) advice on "visual information seeking" to provide an overview of datasets first in order to support visual exploratory data analysis (Fig. 4).

Each document in a document repository will be divided into regions (the boxes in the middle of the above figure). Spatial features will be extracted (e.g. numbers of lines in a region, the ratio of lines to space, etc.) and used to calculate a similarity metric (the numbers in the boxes in the middle of the above figure) to cluster together those documents with similar spatial features. Clusters of documents,

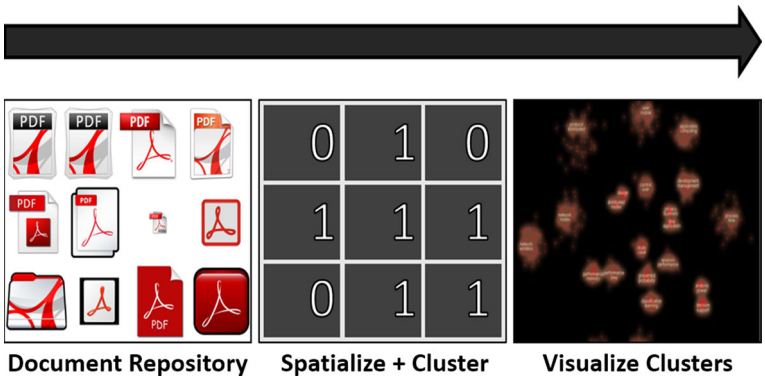

Fig. 4 Proposed approach to clustering archival material by elements of form

represented as dots, will then be projected as a 2D visualization in *n*-dimensional space.

Automated extraction of dates also will assist archivists to create timelines to structure their mental models of the context of records creation. This approach will provide computational support to help archivists as they work to gain a sense of the fonds as a whole, as well as being of possible use to future researchers using that archive. Following Ribarsky et al. (2011), the timeline could be interactive with key events expanding to indicate a detailed history as well as overlaying displays of events occurring at the same time to reveal interesting patterns of development or evolution to help the archivist in understanding important documentary relationships. Though a number of researchers have experimented with combining information retrieval, text mining and natural language processing approaches for entity extraction and feature analysis with interactive visual interfaces (Yu et al. 2010; Ribarsky et al. 2011; Dou et al. 2011; Luo et al. 2012), more extensive research and experimentation is required to combine these techniques with visual technologies in support of archival tasks.

While we can assist archivists in the arrangement and description process by leveraging computational methods to cluster archival material by form or by applying text mining to extract key dates, determination of series still requires human 'in the loop' exploration as part of the sense-making process. Multiple views support moving from an overview (e.g. from visual clusters) to 'details on demand' (e.g. from thumbnail views of contents of clusters to the digital documents themselves) to aid archivists' sense-making (see Fig. 5). By providing even quick visual glimpses of files, archivists should be able to draw more information and categorize digital material in the same manner as they do in the analogue case. A cluster-style archival visualization cannot by itself divulge all this extra, available information. While cluster visualizations can facilitate the process of determining what series exist by providing a high-level overview of a corpus of archival material, archivists may not be able to determine relevant form-related content as intuitively from the clusters as they would by browsing through visual surrogates of the material. This finding is consistent with Greene et al. (2000), who observe that

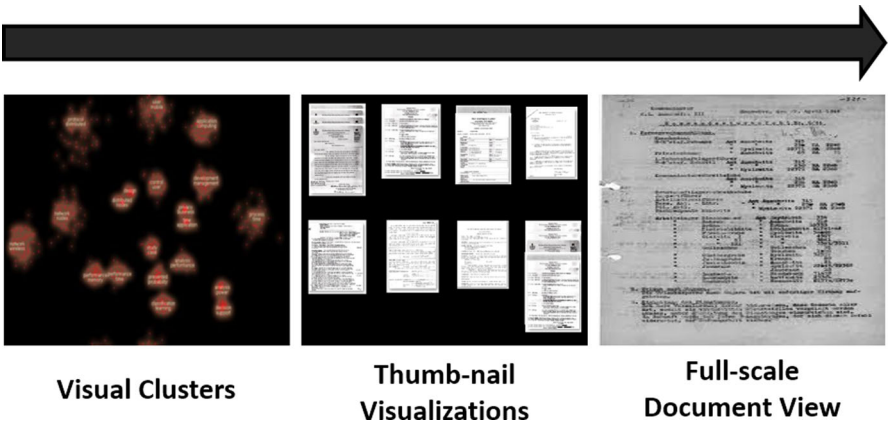

Fig. 5 Multiple views to aid sense-making

designers often fail to provide appropriate preview materials to aid users in gaining a sense of the overall structure of materials. Thus, incorporation of additional thumbnail style visualizations into the design of a VA tool for archival arrangement and description may provide better support for archivists' spatial reasoning about different documentary forms and aid in their ability to hone in on or 'mine' for variable elements within particular documentary forms as they read documents for context. Thus, we propose a design that permits archivists to switch from cluster visualizations, which provide high-level overviews of material related by form, to thumbnail style visualizations of the contents of each cluster. The design could allow archivists to further interact with the thumbnails (for example, by clicking on them) to open documents to obtain 'details on demand' (Shneiderman 1996).

Another finding of the study is how important the materiality of the records is to archivists as they determine an appropriate arrangement and description for an archival fonds. This finding is consistent with theories of embodied cognition purporting that the way the human sensorimotor capacities interact with their environment—in this case archivists' interaction with the materiality of records—enables cognition (see, for example, Clark (2012)). The embodied nature of archival cognition suggests that materiality should remain an important feature of a VA tool to support archival arrangement and description. Where the records that the archivist may be arranging and describing are entirely digital and where the archivist's work takes place in a digital environment, the cognitive aid of materiality may be lost. In this case, a system that replicates materiality in some form may help archivists to analyse the material that they are arranging and describing. The question of the form in which this might be achieved remains to be answered. There are a number of options. Hey (2004) points to the use of metaphoric language on some level, such as files, folders, and refoldering, that maps to the familiar material world but that supports functionality in the digital environment. Beyond just retaining familiar language, a common approach is to emulate the physical world graphically, as in the case of the standard graphical user interface (GUI). Hiroshi Ishii (2008) notes that with the commercial success of Apple Macintosh and

Microsoft Windows systems, the GUI has become the standard paradigm for human–computer interaction. The existing 'windows, icons, mouse, pointer' (WIMP) model of the computer environment was designed to emulate—or digitally represent—real-world tangible objects (such as files and folders) in order to improve ease of use. Our proposal to integrate thumbnail visualizations of archival documents into the design of a VA system to support archival arrangement and description is in line with this tradition of graphically emulating real-world physical objects. However, Ishii (2008) also observes that when interacting with the GUI world, we cannot take advantage of our evolved skills in manipulating physical objects to aid cognition. Additionally, because conventional GUI interfaces duplicate the material world, there is a danger that they will import the limitations of the material world into a digital one. Recent research on tangible user interfaces (TUIs), which aim to make use of haptic interaction skills in addition to graphical interaction, may point the way to overcoming the limitations of the standard GUI and provide an enhanced approach to incorporating materiality into the design of VA systems to support archival arrangement and description processes. With TUIs, users are able to directly manipulate digital information with their hands and use their peripheral senses to perceive digital information through its physical embodiment (Ishii 2008). Thus, one possible approach may be to use large format tabletop displays in the design of an interactive visual interface for archival arrangement and description so that archivists can physically interact with graphical representations of archival data. A continuing research challenge is to determine how to combine the graphical and the haptic components of the interface to achieve best results. To address this challenge, it would be useful to study the archival arrangement and description process of digital materials in the typical desktop computing environment to compare it with the observations obtained from this study. Many other aids used by the archivists in this study could be integrated into a tool to assist the archivist's visual analysis: dictionaries and an online encyclopedia, which were used to define or elaborate on terms the archivists were not familiar with, physical implements like rulers that were used to help mark their current location in a box as they arranged it, and a text editor, which they used to make note of the original organization of the archival data.

Greene et al. (2000) discuss systems that combine multiple coordinated views for presenting visualizations. Allen (2005) urges moving beyond identifying isolated visualization tools to consider how they may be combined into a system. He suggests that by putting several visualization tools together along with other data management tools we could create an archival-researcher's or a record-manager's workbench for supporting complex searches and combining materials from many sources. Both of these approaches are likely to be necessary to support even a single archival task, such as archival arrangement and description, and should therefore be part of the design of the proposed VA tool.

Limitations and future directions

The observations in this study encompass only a single pair of archivists processing one collection and therefore it is uncertain to what extent the preliminary findings in

this study are definitive or generalizable. Moreover, the study used physical archives, and it is still unknown to what extent the processing of entirely digital collections changes the analytic process. For example, how might archivists compensate in their cognitive processes for the loss of the material? In spite of these limitations, it is still possible to use the findings to produce a prototype VA system for archival arrangement and description which can then be evaluated and field tested by users to refine the design and test the understanding gained from this study about archivists' cognitive processes during these activities. This approach is consistent with the application of design thinking or the 'thinking-in-action' approach (Cross 2011), in which solutions to problems are sought and built. The research team, therefore, is now working on the development of the prototype. Following design thinking, once the prototype is developed, the resulting VA tool can be used as a test bed to study archivists' interaction with the tool as they arrange and describe digital records in order to yield additional insights into the analytic processes involved in archival arrangement and description in addition to testing how well the solution works in the built environment. These insights then can be used to support refinement of the model resulting from this research and generalizability of the research findings. The researchers also plan to expand the study of the analytic processes involved in archival arrangement and description across a larger group of study participants and range of archival collections in order to test emergent theoretical ideas about archivists' cognitive processes using a more traditional scientific approach.

Conclusion

Visualization and VA hold much promise as tools to support a sustainable future for archives, particularly as cognitive aid to archivists in undertaking complex analytic tasks, like arrangement and description, which cannot yet fully be achieved using computational methods. However, in spite of the great potential that VA holds for archives, much more research is needed in order to understand how this technology can be applied effectively to support archival functions and to transfer novel tools from the laboratory to production systems in archives. Work on visualization and VA approaches to archival endeavours would benefit from focusing more on developing technologies to help archivists perform archival analysis on unprocessed archival documents rather than on creating tools that visually re-represent the results of completed archival analysis. As useful as tools that support visual interaction with archival descriptive metadata are, they still require much archival pre-processing before they can be used unless the metadata are created upstream by users before transfer to an archival repository and as a natural part of other processes. Traditional archival pre-processing is likely to become increasingly untenable in the era of big data without new tools that cognitively aid archivists in the task of archival analysis or which aid researchers in the analytic tasks associated with their research. As such, information visualization and VA benefits from the study of different archival functions and their associated cognitive and perceptual tasks, together with the range of visual metaphors or mappings and representation design alternatives that could be used

to support such tasks. With such an approach, there is the potential for a better understanding of how archivists actually think or undertake archival analysis. This understanding is a critical precursor for developing effective tools to support visual analysis. It would also help to better understand the ways in which archival analysis and the analysis undertaken by archival researchers differ at a cognitive and perceptual level so as to build tools that reflect cognitive and perceptual differences as opposed to designing 'one size fits all' tools. Whitelaw (2012) has called for 'generous' interfaces that promote exploration of archival material in a manner that simple search cannot do. However, we also need 'snug' interfaces; that is, interfaces that fit cognitively with the tasks that archivists and researchers want to perform.

Even with additional research, will VA supersede other solutions aimed at addressing growing archival backlogs, such as crowd-sourcing descriptions or making wider use of pre-archival descriptive metadata? Likely, the answer is no. VA will probably complement these other approaches as an additional innovative technological tool in the archivist's professional toolkit. This is not only because there is still a long way to go in researching how interactive visual technologies can be developed and applied in the archives, but also just as much owing to how archivists engage with the visual. Though the wider society has been undergoing a visual revolution, the archival community is only just moving beyond treating iconic materials as decorative adjuncts to textual documents (Schwartz 2004). For a profession that, according to theorists like Schwartz, still lacks due appreciation for the visual (although this has perhaps changed in recent years), the serious use of visual images for analytic purposes may be the first hurdle to be crossed in applying this technology to support archival arrangement and description or any other archival reasoning processes. In short, there is a number of ways in which information visualization and VA can be used in archives in future, but there is still some way to go before these technologies can have a major impact on daily archival work. Nevertheless, experimentation with information visualization and visual analytics in the archival domain is developing into an interesting new theme in archival discourse, signalling that an archival future in which interactive visual tools help archivists perform archival analysis and assist researchers to explore archival documents is both achievable and likely to be at least one answer to realizing a more sustainable archival future.

Acknowledgments This research has been carried out with funding from the Boeing Corporation and the Natural Sciences and Engineering Research Council of Canada. The author would like to thank her research assistants Lara Mancuso, Justin Chan, and Kevin Owen for their contributions to this project. A debt of gratitude is also owed to the two anonymous archivists who participated in the arrangement and description project analysed for this study. Without them, and the support of their managers, this study would not have been possible. The author would also like to thank the anonymous reviewers of earlier versions of this paper whose constructive criticisms greatly improved the final version. Any errors and omissions remain the sole responsibility of the author.

References

Ackoff R (1989) From data to wisdom. J Appl Syst Anal 16:3–9

AIMS Work Group (2012) AIMS born-digital collections: an inter-institutional model for stewardship. http://www2.lib.virginia.edu/aims/whitepaper/AIMS_final.pdf. Accessed 16 Jul 2013

Allen R (2005) Using information visualization to support access to archival records. J Arch Organ 3(1):37–49
Anderson IG (2009) From ZigZag ™ to BigBag: seeing the wood and the trees in online archive finding aids. In: Vitali F, Di Iorio A, Blustein J (eds) Proceedings of the 1st workshop on new forms of xanalogical storage and function, 2009, Turin, Italy http://ceur-ws.org/Vol-508/paper3.pdf. Accessed 22 May 2013
Anderson K, Bastian J, Harvey R, Plum T, Samuelsson G (2010) Teaching to trust: how a virtual archives and preservation curriculum laboratory creates a global education community. In: Paper presented at Questions of trust? Archives records, and identities, FARMER: UK and Ireland Forum for Archives and Records Management Education and Research, Wolfson College, Oxford, England. July 5–6, 2010
Arias-Hernández R, Kaastra LT, Green TM, Fisher B (2011) Pair analytics: capturing reasoning processes in collaborative VA. In: Sprague, R (ed) Proceedings of the 44th annual Hawaii international conference on system sciences, 4–7 January 2011, Koloa, Kauai, Hawaii. IEEE Computer Society Press, Washington, DC
Bak G (2012) Continuous classification: capturing dynamic relationships among information resources. Arch Sci 12:287–318
Capon N, Davis R (1984) Basic cognitive ability measures as predictors of consumer information processing strategies. J Consum Res 11:551–563
Clark A (2012) Embodied, embedded, and extended cognition. In: Frankish K, Ramsey W (eds) The Cambridge handbook of cognitive science. Cambridge University Press, Cambridge, pp 274–291
Corbin J, Strauss AL (2008) Basics of qualitative research: grounded theory procedures and techniques, 3rd edn. Sage, Thousand Oaks, CA
Crandall B, Klein G, Hoffman RR (2006) Working minds: a practitioner's guide to cognitive task analysis. Bradford, Denver, CO
Cross N (2011) Design thinking: understanding how designers think and work. Berg, Oxford
Cunningham A (2001) Six degrees of separation: Australian metadata initiatives and their relationships with international standards. Arch Sci 1(3):271–283
Dervin B (2003) Audience as listener and learner, teacher and confidante. In: Dervin B, Foreman-Wernet L, Lauterbach E (eds) The sense-making methodology reader: selected writings of Brenda Dervin. Hampton Press, Cresskill, pp 215–231
Dou W, Chang R, Wang X, Ribarsky W (2011) Parallel Topics: a probabilistic approach to exploring document collections. In: Miksch S, Ward M (eds) IEEE VAST, 23–28 October 2011. Providence, RI, pp 231–240
Duchein M (1986) Theoretical principles and practical problems of respect des fonds in archival science. Archivaria 16:64–82
Duff W, Harris V (2002) Stories and names: archival description as narrating records and constructing meanings. Arch Sci 2(3–4):263–285
Duff W, Craig B, Cherry J (2004) Historians' use of archival sources. Publ Historian 26:7–22
Duffy M (1995) Sensemaking in classroom conversations. In: Maso I et al (eds) Openness in research: the tension between self and other. Van Gorcum, Assen, NL, pp 119–132
Eastwood T (2000) Putting the parts of the whole together: systematic arrangement of archives. Archivaria 50:93–116
Ericsson AK, Simon HA (1993) Protocol analysis: verbal reports as data, revised edn. MIT Press, London
Esteva M, Xu W, Dutt Jain S, Lee JL, Martin WK (2011) Assessing the preservation condition of large and heterogeneous electronic records collections with visualization. Int J Digit Curation 6(1):45–57
Evans M (2007) Archives of the people, by the people, for the people. Am Arch 70:387–400
Fisher B, Green TM, Arias-Hernández R (2011) Visual analytics as a translational cognitive science. Top Cogn Sci 3(3):609–625
Frické M (2009) The knowledge pyramid: a critique of the DIKW hierarchy. J Inform Sci 35(2):131–142
Gilliland-Swetland A (2005) Electronic records management. Ann Rev Info Sci Tech 39(1):219–253
Greene MA, Meissner D (2005) More product, less process: revamping traditional archival processing. Am Arch 68:208–263
Greene S, Marchionini G, Plaisant C, Shneiderman B (2000) Previews and overviews in digital libraries: designing surrogates to support visual information seeking. J Am Soc Inform Sci 51:380–393
Hey J (2004) The data, information, knowledge, wisdom chain: the metaphorical link http://www.dataschemata.com/uploads/7/4/8/7/7487334/dikwchain.pdf. Accessed 8 July 2013

Hinton S, Whitelaw M (2010) Exploring the digital commons: an approach to the visualisation of large heritage datasets. In: Seal A, Bowen J, Ng K (eds) Proceedings of EVA London 2010: electronic visualization & the Arts, 6–8 July 2010, London. British Computer Society http://www.bcs.org/upload/pdf/ewic_ev10_s3paper2.pdf. Accessed 8 July 2013

Horsman P (2002) The last dance of the phoenix, or the de-discovery of the archival fonds. Archivaria 54:1–23

Hurley C (1998) The making and keeping of records (1): what are findings aids for? Arch Manuscr 21(1):58–77

IBM (2012a) IBM PowerLinux big data analytics solutions. http://public.dhe.ibm.com/common/ssi/ecm/en/pos03099usen/POS03099USEN.PDF. Accessed 30 July 2013

IBM (2012b) TAKMI: bringing order to unstructured data http://www-03.ibm.com/ibm/history/ibm100/us/en/icons/takmi/. Accessed 10 July 2013

Ishii H (2008) The tangible user interface and its evolution. Comm ACM 51(6):32–36

ISO 9241-210 (2010) Ergonomics of human-system interaction. Part 210: human-centered design for interactive systems

Johnson B, Shneiderman B (1991) Tree-maps: a space-filling approach to the visualization of hierarchical information structures. In: Proceedings of visualization '91, IEEE information visualization, 22–25 Oct 1991, San Diego, CA, IEEE, pp 275–282

Klein G, Moon B, Hoffman RR (2006) Making sense of sensemaking 1: alternative perspectives. IEEE Intell Syst 21(4):70–73

Kosslyn SM (2005) Mental images and the brain. Cogn Neuropsychol 22(3–4):333–347

Kramer-Smyth J, Nishigaki M, Anglade T (2010) ArchivesZ: visualizing archival collections. archivesz.com/ArchivesZ.pdf. Accessed 10 July 2013

Krause MG, Yakel E (2007) Interaction in virtual archives: the polar bear expedition digital collections next generation finding aid. Am Arch 70:282–314

Kuusela H, Paul P (2000) A comparison of concurrent and retrospective verbal protocol analysis. Am J Psychol 113(3):387–404

Lewis C, Rieman J (1993) Task-centred user interface design. www.hcibib.org/tcuid/index.html. Accessed 8 July 2013

Luo D, Yang J, Krstajic M, Fan J, Ribarsky W, Keim D (2012) EventRiver: interactive visual exploration of constantly evolving text collections. IEEE Trans Visual Comput Graph 18(1):93–105

Lurie NH, Mason C (2007) Visual representation: implications for decision making. J Market 71:160–177

MacNeil H (2005) Picking our text: archival description, authenticity and the archivist as editor. Am Arch 68:264–278

Meehan J (2009) Making the leap from parts to whole: evidence and inference in archival arrangement and description. Am Arch 72:72–90

Meehan J (2010) Rethinking original order and personal records. Archivaria 70:27–44

Millar L (1988) A manual for small archives. Archives Association of British Columbia, Vancouver, BC

Miller F (1990) Arranging and describing archives and manuscripts. Society of American Archivists, Chicago, IL

Munzner T (2009) Visualization. In: Shirley P, Marschner S (eds) Fundamentals of computer graphics, 3rd edn. Peters/CRC Press, Boca Raton, FL, pp 675–708

Nesmith T (2006) The concept of societal provenance and records of nineteenth-century Aboriginal-European relations in Western Canada: implications for archival theory and practice. Arch Sci 6:351–360

Payne JW (1994) Review of thinking aloud: insights into information processing. Psychol Sci 5(5):241, 245–248

Pirolli P, Card SK (2005) The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. http://vadl.cc.gatech.edu/documents/2__card-sensemaking.pdf. Accessed 10 July 2013

Prom C, Schwartz J, Scott W, Rishel CA, Fox KJ (2007) A unified platform for archival description and access. In: Rasmussen E, Larsen RR, Toms E, Sugimoto S (eds) Proceedings of ACM/IEEE joint conference on digital libraries, JCDL '07, 17–22 June 2007, Vancouver, BC. ACM, New York, NY, pp 157–166

Provot X (1995) Deformation constraints in a mass spring model to describe rigid cloth behaviour. Graph Interface 95:147–154

Randsell S (1995) Generating thinking-aloud protocols: impact on the narrative writing of college students. Am J Psychol 108(1):89–98

Ribarsky W, Sauda E, Wartell Z, Balmer J (2011) The whole story: building the complete history of a place. In: Sprague R (ed) Proceedings of the 45th annual Hawaii international conference on systems science, 4–7 January 2012, Maui, Hawaii. IEEE Computer Society, Piscataway, NJ, pp 1864–1873

Roe KD (2005) Arranging and describing archives and manuscripts. Society of American Archivists, Chicago, IL

Rowley J (2007) The wisdom hierarchy: representations of the DIKW hierarchy. J Inform Sci 33(2):163–180

Schraagen JM, Chipman SF, Shalin VL (eds) (2000) Cognitive task analysis. Lawrence Erlbaum Associates, Mahwah, NJ

Schwartz J (2004) Negotiating the visual turn: new perspectives on images and archives. Am Arch 67(1):107–122

Shneiderman B (1996) The eyes have it: a task by data type taxonomy for information visualizations. In: Proceedings of the IEEE symposium on visual languages, 3–6 September 1996, Boulder, CO, pp 336–343

Tegarden DP (1999) Business infoVis. Commun Assoc Inf Syst 1:2–38

Thomas JJ, Cook KA (eds) (2005) Illuminating the path: the research and development agenda for visual analytics. IEEE Computer Society Press, Los Alamitos, CA

Trickett SB, Trafton JG, Saner L, Schunn CD (2007) I don't know what's going on there: the use of spatial transformations to deal with and resolve uncertainty in complex visualizations. In: Lovett MC, Shah P (eds) Thinking with data. Lawrence Erlbaum Associates, Mahwah, NJ, pp 65–86

Villars RL (2012) Worldwide archival storage solutions 2012–2016 forecasts: clouds, big data, and digitization drive preservation expansion. International Data Corporation http://www.idc.com/research/viewtoc.jsp?containerId=235861. Accessed 10 Sept 2012

VisMaster (2010) Mastering the information age http://www.youtube.com/watch?v=5i3xbitEVfs. Accessed 8 July 2013

Ward M, Grinstein G, Keim D (2010) Interactive data visualization: foundations, techniques, and applications. AK Peters, Natick, MA

Whitelaw M (2009) Visualising archival collections: the visible archive project. Arch Manuscr 37(2):22–40

Whitelaw M (2012) Towards generous interfaces for archival collections. In: Paper presented at the international council on archives congress, 20–24 August 2012, Brisbane, Australia. http://ica2012.ica.org/files/pdf/Full%20papers%20upload/ica12Final00423.pdf

Yakel E (2003) Archival representation. Arch Sci 3:1–25

Yeo G (2010) Debates about description. In: Eastwood T, MacNeil H (eds) Currents of archival thinking. Libraries Unlimited, Santa Barbara, CA, pp 89–114

Yu L, Lu A, Ribarsky W, Chen W (2010) Digital storytelling: automatic animation for time-varying data visualization. Comput Graph Forum 29(7):2271–2280

Author Biography

Victoria L. Lemieux is an Assistant Professor at the University of British Columbia (UBC), School of Library, Archival and Information Studies (SLAIS) and Director of UBC's Media and Graphics Interdisciplinary Centre (MAGIC).