



Dark archives or a dark age for reasoning over archives?

Mark Bell¹ · Jenny Bunn¹

Received: 25 November 2020 / Accepted: 23 November 2021 / Published online: 18 January 2022
© Crown 2021, corrected publication 2022

Abstract

This article considers that reasoning over archives is a joint enterprise between archivists and researchers and that both groups are increasingly using machine agents to assist them in it. It starts by considering the processing of archivists, researchers and machine agents separately. Using the different perspectives this brings to highlight different aspects of that processing, as a process of sense-making, as scholarly research activity, as practices that realise and achieve data for the drawing of further inference, it reasserts the argument that archives cannot be regarded as raw data to be reasoned over, but must be seen as the result of multiple representative and interpretive acts, of iterative realisation and activation as ‘data’ potentially involving many, many additional actors. It then goes on to consider how the involvement of machine agents fits into and potentially alters this picture by providing more detail about the basis on which they currently perform such acts.

Keywords Archives · Supervised machine learning · Scholarly process · Representation

1 Introduction

Darkness is not necessarily a bad thing when it comes to archives. Light can damage the organic materials on which ‘archives’ are sometimes stored (Surrey History Centre 2021). Then again, even when archives are stored digitally it is difficult to argue with the wisdom of keeping at least one copy of them ‘dark’—more safely put away as a failsafe in case all other copies are lost (Erickson 2013). The darkness with which this special issue is concerned, however, would seem to be the darkness that exists ‘beyond the reach of scholarship’—the unknown or perhaps even worse the unknowable (Stapleton and Jaillant 2020). There is a general perception then that this darkness is growing. Scholarship (not uniquely) has always sought to push back against the boundaries of the unknown and many scholars have sought to do so by using material collected together in and as archives; repositories of knowledge alongside other such—libraries and museums—collections of objects, books and journals. Scholarship (more distinctively) has also formed itself into long lived collectives—disciplines or fields—to distribute the cognitive load of pushing back against the unknown and to allow it to enlist ever increasing numbers

of scholars to process the ever increasing accumulations or sources of knowledge that become available to them over time. Those involved in such endeavours have sought at times to declare victory—as an Age of Reason or Enlightenment overcame that of the Dark—but now perhaps there is a growing perception that the tide has changed. The war is in danger of being lost.

This article will not tell you how or if that war can be won, but it will identify one way in which it is more likely to be lost. As with any large force, there is always a risk of infighting and factionalism and two factions which have long been prominent within (humanities) scholarship are archivists and academic researchers/scholars. Divided along the lines of gatekeeping—of facilitating legal and ethical access to sources—on the one hand and of wanting access to those sources on the other, these two factions have long scrapped over the bones of the power to decide what is kept, and of data protection and copyright. More recently, and encouragingly, these factions have started to recognise that they face a common problem. This problem concerns the processing capacity and capability they have as groups (albeit large groups) of human beings and is felt and experienced in slightly different ways by the two communities. It arises from both the way in which the stores of the raw materials that are seen as within their remit for potential processing are proliferating and expanding seemingly exponentially (causing issues of capacity), and also in the way in which they can

✉ Jenny Bunn
jenny.bunn@nationalarchives.gov.uk

¹ The National Archives, London, UK

now work with these materials differently (causing issues of capability). A common factor across both these aspects is a new set of tools and technologies which fall under the broad term of artificial intelligence and the narrower term of machine learning.

This article does not offer any solution to the problem of processing capacity and capability, rather it focuses more on exploring and articulating the nature and detail of that processing. Firstly, it will seek to highlight the way in which the work of archivists in preparing material for use by researchers (one way in which archival processing is defined) can be seen as akin to the sort of scholarly research carried out by the researchers who also use that material. Secondly it will, through discussion of attempts by researchers to categorise and conceptualise this scholarly process in the face of new technology (primarily under the umbrella of digital humanities), seek to highlight another aspect of that process that is increasingly gaining recognition amongst the research community, namely data practices. With a focus on data practices established, the article then inserts an explanation of the data practices of the machine agents brought about by the current state of artificial intelligence development. Finally it will build from these separate elements to articulate a shared vision for all three constituencies (archivists, researchers and machine agents) in the joint enterprise of reasoning over archives.

2 The processing of archivists and researchers

The Society of American Archivists' Dictionary of Archives Terminology defines processing as 'preparing archival materials for use' but this preparation is also defined in (narrower) terms of arrangement, description and housing (Society of American Archivists 2021). The process which archivists are trained to follow in order to prepare archival materials for use is well documented in any number of professional text books, and can be seen to involve tasks as intellectually undemanding (but physically difficult) as removing rusty staples and paperclips to more intellectually demanding ones such as (re)ordering, interpreting, creating surrogates and designing architectures for representational systems' (Yakel 2003). This last characterisation comes from an article by Elizabeth Yakel, written in 2003, in which she sought to reframe the archival function 'identified as arrangement and description, processing, and occasionally archival cataloguing' as archival representation, asserting that this was a more precise description for the work archivists were undertaking. Later on in the article she also associated this work with that of structuring, linking this idea back to Anthony Giddens' Theory of Structuration, and with that of categorization, referencing the idea that this 'is not an individual cognitive

process, but rather the result of a complex dynamic of cultural and social forces' (Yakel 2003).

Yakel's article marked the start of a number of attempts by archivists to further consider their own processes of arrangement and description. For example, six years later, Meehan (2009) started her own examination of the subject with the characterisation of such work as 'an analytical process' and MacNeil (2005) sought to draw an analogy between 'the archivist's work in arranging and describing records' and that 'performed by traditional textual critics in preparing a scholarly edition for publication'. The two articles took slightly different paths, but ended up in very similar territory.

Meehan (2009) was motivated by a desire to move beyond a more general understanding of arrangement and description 'as interpretive and representative' towards an elaboration of 'some of the particular acts of interpretation and representation involved in archival arrangement and description'. In this pursuit, she placed an emphasis on evidence and inference, the analytical process she had asserted being further characterised as; 'an active process of using the sources as evidence—that is, as the basis for inferring facts about past events' (Meehan 2009). She acknowledged that this 'process of inferring one thing from another and drawing meaningful conclusions [...] is anything but conclusive' and involved 'a fair amount of speculation' and went on to state that that it is 'impossible to configure archival analysis in arrangement and description as anything other than various ongoing, often overlapping, and ultimately open-ended processes of reasoning about records' (Meehan 2009). In her conclusion, she retrospectively framed her account as being one 'of the archivist's process of *making sense* of the records' (Meehan 2009). MacNeil (2005) on the other hand was interested in drawing her analogy between the processes of archivists and those of textual critics and with exploring 'the theoretical and socio-historical foundation on which the relationship between archival description and authenticity rests'. She does not delve in as much detail as Meehan into particular acts, but does raise questions about 'what can and cannot be reconstructed from the surviving evidence' and shows a concern with 'the relationship between material parts and imaginary wholes' that perhaps carries the possibility that such a relationship is based on or implies inference (MacNeil 2005).

The territory that both articles share is in their suggestion and advocacy for methods by which archivists might 'begin to account for the inferential and speculative nature of archival analysis' or 'make the act of picking our text transparent to our users through description' (Meehan 2009; MacNeil 2005). These suggestions include; using 'conditional phrases such as "perhaps" and "may have been" to qualify statements that would otherwise seem conclusive' (Meehan 2009 after MacNeil 2005), documenting, for example, 'the rationale for

a particular arrangement, the reasoning behind decisions, and the sources of information used in reaching a particular decision’, using footnotes and citations, and adding ‘information about the archivist who has processed the collection’ (Meehan 2009). In making these suggestions both Meehan (2009) and MacNeil (2005) reference an earlier article by Light and Hyry (2002) that suggested the inclusion of colophons and annotations as a means ‘to acknowledge the inherent subjectivity of archival work’. This article did not seek to explicate ‘the archivist’s process of *making sense* of the records’, but it did set out the context, an exposure to post-modern thought, which had led archivists to become more self-conscious about it.

Whatever prompted it though, it is clear that many of the suggestions being made to archivists in respect of the presentation of their analysis (in the form of a finding aid or archival description) are similar to the techniques employed by researchers of many kinds in respect of their presentation of their analysis or research (albeit in different forms of production such as articles in academic journals). Indeed Meehan (2009) also made the suggestion that the archival profession should address ‘specific methods for using sources and formulating appropriate questions to guide analysis’ in a move towards the formalisation of appropriate methods for undertaking their data analysis. Then again, her statement that to undertake such analysis, an archivist required ‘not only well-developed research skills and subject knowledge, but also a nuanced understanding of archival principles, critical and creative thinking, and, perhaps more than anything, an imaginative frame of mind’ could also, bar the reference to archival principles, be considered applicable to researchers in many other disciplines (Meehan 2009).

In the work of these authors to ‘make some of the archivist’s implicit processes more explicit’, it does become possible to see at the abstract level, a parallel between the processing being undertaken by researchers and that being undertaken by archivists (Meehan 2009). And yet, these parallels are easily lost and a division appears, when the work of archivists in preparing materials for use is regarded as only pre-preparation to the proper processing then undertaken using those materials by proper researchers. That archivists at least, do feel this shift in regard is perhaps apparent in Light and Hyry’s (2002) article where it is stated that the mechanisms of colophons and annotations ‘would also force researchers to acknowledge the value we [the archivists] add to collections’.

3 Facilitating our processing

As has already been made clear, this article is not interested in perpetuating in-fighting and factionalism. It therefore asserts the proposition (in the above) that the processing of

archivists and researchers—in sense making, in undertaking analysis drawing on a range of source material, in drawing inferences on a balance of evidence can be seen as essentially the same basic action or operation; one of drawing conclusions on the basis of data, of undertaking research. The two communities may not research to the same ends, but both communities have articulated problems with continuing to accomplish this aspect of their work. These problems are seen to arise as the possible volume, variety and velocity of what they have to work with has increased exponentially in an era defined by ‘big’ data.

For example, from the community of archival practice, Bearman (1989) published an essay in which he undertook a comparison between the magnitude of the tasks archivists had set themselves and the magnitude of their capabilities revealing ‘substantial discrepancies’. This mismatch has been raised subsequently by others including Greene and Meissner (2005) and a number of solutions have been suggested, e.g., co-opting others to help with the task and accepting that ‘the golden minimum [...] is all we can realistically accomplish’. That technology might be able to carry out some of these tasks has also long been recognised. For example, listening to a paper at the Conference of the Society of Archivists in 1964 is reported to have inspired those in attendance to ‘happy visions of indexes being compiled and sorted by machine’ (Jones 1964). More recently, however, these two currents of discourse have become increasingly and more urgently intertwined with Victoria Lemieux (2015) writing in the early 2010s of how ‘Traditional archival pre-processing [owning the pre-labelling of the archivist’s work] 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 analytical tasks associated with their research’. Lemieux has also been involved with attempts to establish a new interdisciplinary field of computational archival science concerned with ‘the application of computational methods and resources to large scale records/archives processing, analysis, storage, long-term preservation, and access’ (University of British Columbia 2016).

Humanities researchers on the other hand, seem to have articulated and been able to experience the problem not so much as an impending tsunami of material threatening to drown them, and more as an opportunity for (and project of) reinvention. Certainly, they have the longer history of recognising the potential of using computers as assistive tools. Father Roberto Busa was inspired by the idea of machine supported indexing a number of years before 1964 and humanities researchers were also slightly quicker in realising the possibilities of creating an encoding standard tailored to their needs; with the origins of the Text Encoding Initiative predating those of Encoded Archival Description by a matter of a few years (Hockey 2004; Barry et al. 2013).

Then again, the reinvention of humanities computing into digital humanities arrived before and has (perhaps as a consequence) reached a more mature stage of development than attempts at reinvention (into computational archival science or perhaps more successfully into digital curation) within the professional archival community.

John Unsworth, who is seen, along with his fellow editors of 2004's *A Companion to Digital Humanities*, as one of the architects of the reinvention, is also well known for introducing the influential idea of Scholarly Primitives (Schreibman et al. 2004). Starting from a position that focused on the operation of undertaking analysis, particularly analysis based on textual materials, Unsworth (2000) attempted to break down this higher-order operation associated with arguments, statements and interpretations into seven more basic functions or primitives. These he defined as; discovering, annotating, comparing, referring, sampling, illustrating and representing (Unsworth 2000). Unsworth's work has proven very influential and has inspired a whole host of subsequent work. More recently, for example, those responsible for the so-called Taxonomy of Digital Research Activities in the Humanities have acknowledged that 'classifying and categorizing the activities that comprise "digital humanities" has been a longstanding area of interest for many practitioners in this field' (Borek et al. 2016). That practitioners in the process of reinventing their practice might desire a way to define what it is that they do should come as no surprise, but this does not appear to have been the only motivation that has prompted this work.

The origination of Unsworth's scholarly primitives lay, as he himself made clear in two funding proposals directed towards research into text analysis tools, and when illustrating many of the primitives what he actually presented were interfaces, means and ways for scholars to interact with scholarly materials digitally rather than in physical person. Considering what sort of basic operations scholars carried out acted for Unsworth (2020) therefore, as a precursor to and a way of organising his efforts towards the facilitating of scholarly activity through the design and production of tools. Similar motivations seem to have lain behind OCLC's Scholarly Information Practices Project which started from the perspective of research libraries as existing 'to support scholarly work' and aimed to provide 'an empirical basis [...] for development of digital information services to support and advance scholarship' (Palmer et al. 2009). And finally the Taxonomy of Digital Research Activities in the Humanities took place in the context of the DARIAH project, an attempt to build a pan-European Digital Research Infrastructure for the Arts and Humanities (DARIAH-EU 2021). A motivation in many of these efforts does therefore seem to have been the facilitation of scholarly work.

The taxonomic approach is not the only one that has been taken within digital humanities. As well as classifying and

categorising, attempts have also been made at mapping and modelling. These approaches are less concerned with breaking things down, and as such are perhaps better suited to seeing the whole as a whole, rather than an amalgam of constituent parts. In the project that laid the foundations for DARIAH, Benardou et al. (2010) produced 'A Conceptual Model for Scholarly Research Activity' which confined and hid 'the detailed structure of the research process, and way of working for each step' within one part of their model—procedure. In this model Research Activity is considered to involve following a particular procedure, being directed towards a particular goal, and also as both developing and referring to Propositions and involving interaction with Information Objects. Within this model the scholarly primitives previously developed 'can be interpreted as specific operations on conceptual or information objects' and can therefore 'be represented as specializations of properties relating Research Activity to Proposition, Concept and Information Object' (Benardou et al. 2010). This model then would seem to conceptualise, as scholarly research activity, processing that is similar to that discussed in the previous section. Reading the descriptions of archival processing, as sense-making, as the 'process of inferring one thing from another and drawing meaningful conclusions' alongside this conceptualisation, it becomes even harder to argue against a proposition that some of the work of archivists is also a form of scholarly research activity (Meehan 2009).

Less conceptually, McCarty and Short (2002) have led efforts to map out the field of digital humanities at the intersection of various disciplinary groupings and 'clouds of knowing'. The intersection is conceived of as a 'methodological commons.' In 2002, when the map was first drawn what lay within the methodological commons was a number of different forms of data, e.g. narrative text, tabular alpha-numeric, numbers, music, images and the phrase 'communications, hypermedia and the digital library.' In a later version, types of data were still distinguished, as; text, image, 3D Vis, sound and numbers, but the phrase 'communications, hypermedia and the digital library' was replaced with a two way relationship connecting 'analytical tools and data structures' with 'formal methods.' This comparison can be made in a article by Siemens (2016) in which the maps are considered side by side and what lies in the methodological commons is characterised as 'those things central to the practices of our community: data and data structures modeling core materials, and tools modeling formal methods.' Data practices are also highlighted in the OCLC's Scholarly Information Practices Project, which defined its own taxonomy of scholarly primitives, included one group of so-called cross-cutting primitives. Within this last group a further distinction was made between monitoring, notetaking and translating as 'of interest because of their significance in the research process' and data practices which stood out

from the others because it was ‘not a primitive in its own right, but a set of activities around which a growing body of discourse and new research is emerging’ (Palmer et al. 2009).

Researchers have always worked with data of course, but they have perhaps not previously paid so much attention to their data practices. Just then, as archivists increasingly assert their role as scholarly researchers, so too have scholarly researchers started to assert their role as data practitioners. In the latter case, this assertion seems to have arisen in part from an increasing engagement with and desire to employ information technologies to build interfaces and services for serving up and interacting with the materials they use within their work. Data are therefore no longer just something that researchers collect and use, rather data is something that they now do. As mentioned previously, archivists have sometimes felt ignored by researchers in their doing of data and have felt that the value which their doing adds to collections (data) often goes unacknowledged. With the advent of artificial intelligence however, both archivists and researchers are increasingly employing machine agents to do more of their data work, and in the next section we seek to explore and explain how it is that our new partners are actually doing that work.

4 Machine reasoning

The field of Artificial Intelligence is broad but archivists and researchers are increasing finding application for and employing methods from the sub-field of Machine Learning. For the purposes of this paper we will concentrate on Supervised Machine Learning, a family of algorithms which learn by example.

Supervised machine learning is a method for identifying patterns in data in order to predict a value or label for an input record. Predicting a label (e.g. ‘Sensitive’ or ‘Non-sensitive’ for archived documents) is also known as classification. The input data, termed a feature vector, could represent a row of numbers in a table, the pixels of an image, or the text in a document. The output is a vector of one or more numbers. In the case of regression (predicting a continuous value such as height, or an amount of money) the number could be any floating point value while in classification the numbers often total to 1, each being interpreted as the probability of a record being representative of a possible class.

The selection of a machine learning algorithm depends on the data (both amount and type), the application, and pragmatism. The process begins with modelling. Numerical data is often modelled as having a linear relationship with the output variable moving up or down in proportion to some combination of the input variables. The method of Least Squares was used as early as 1795 by Gauss and is

still an important tool in regression analysis. Least squares is a mathematical process but it has much in common with many machine learning algorithms. It takes an input matrix representing a column for each feature and a row for each data item, and an output vector. In general there is not a single function which can map any input row accurately to its output value. What the method returns instead is a function which maps the inputs to an approximation of the output vector, such that the total error between approximations and true output is minimised at an aggregate level. In statistics regression is used to understand data and the relationships between input features, whilst in machine learning it is a tool for prediction. The input data is known as training data and the aim is to find a function that will best predict previously unseen data. As Efron and Hastie (2016) have observed, the field of machine learning has moved further from the statistician’s aim of understanding and more in the direction of prediction. This has resulted in great technical innovation but as the models become more complex they become less understandable in terms of how they do what they do. Evaluation of algorithms is achieved by assessing them against benchmark datasets which means incentives for researchers are based on accuracy scores not theoretical understanding.

Cultural heritage applications tend to use machine learning for classification, and the records being classified tend to be either images or text, rather than tabular numeric data. The modeller has two tasks, one to select a numerical representation of the data, and another to model how features within the data interact or relate to each other. In image classification tasks the data is already in a numerical form, a digital image being a matrix of pixel values. The modelling decision is whether colour is important to the task, or whether greyscale, or binary (black/white) is sufficient. In computational terms this makes a difference of 256^3 possible pixel values in the colour case, 256 shades of grey for greyscale, or 2 for binary images. Colour images increase the dimensionality of the problem thereby increasing the number of parameters the machine learning algorithm needs to optimise to perform well, and this in turn increases the amount of examples needed to train the model. Classical approaches to computer vision involve feature engineering approaches such as SIFT and modelling images as a Visual Bag-of-Words (a collection of visual shapes, for example, corners). While these perform well for many applications attention in recent years has turned to neural network architectures. The 2012 ImageNet LSVRC, an object categorisation and detection contest, was described as “a turning point for large-scale object recognition, when large-scale deep neural networks entered the scene” (Russakovsky et al. 2015). The clear winner was a 228 layer “deep” neural network (Krizhevsky et al. 2012) which significantly outperformed the other entrants in the competition. As the designers state, “the immense complexity of the object recognition

task means that this problem cannot be specified even by a dataset as large as ImageNet, so our model should also have lots of prior knowledge to compensate for all the data we don't have." They used an architecture called a Convolutional Neural Network which has an advantage over the bag-of-words model since it is able to model how features relate spatially.

Similar modelling choices are made in Natural Language Processing. Text needs to be converted into a numerical form and there are a range of techniques available, depending on the application. The simplest form of modelling words is to choose a vocabulary (the top N most frequent words in the corpus), and to represent each document as a vector indicating (with 1 s and 0 s) which words from the vocabulary it contains. A more common technique is to use TF-IDF which weights each word according to both its frequency within the document (term frequency, TF) and how many other documents it appears in (inverse document frequency, IDF) (Spärck Jones 1972). This applies some measure of importance a word has to a document relative to the remainder of the corpus. These approaches don't incorporate any semantic meaning to words, they are treated as unrelated tokens. The development of Word2vec (Mikolov et al. 2013) changed this by creating a representation of words as vectors in a many (50,100,300) dimensional space. These vector representations are derived by analysis of words in context over millions of documents. While in TF-IDF the words "cat" and "dog" will be represented as floating point numbers in two columns of a table, the word2vec representation (a word embedding) for each is a long vector of floating point numbers (same length as the number of dimensions). These two vectors can be compared (by measuring the angle between them) and will be to some degree similar because cats and dogs appear in similar contexts. The actual embeddings and similarities will depend on the corpus they were trained against.

After selecting a representation of vocabulary, the modeller must choose one for language. Bag-of-words assumes words in a sentence or document are independent of each other. While this sounds like an over-simplification, it works well in practice for many applications. A marginally more nuanced approach is to consider a word to be only dependent on the word which precedes it, the Markov assumption. These approaches are computationally simple and do not require a lot of data. With the deep learning era came more sophisticated models of language which treated text as a sequence of related tokens. So called sequence to sequence models are able to map a sequence of tokens to another output sequence. This has application in machine translation, mapping one language to another, and handwriting recognition, mapping a sequence of visual features to a string of text. The most recent innovation is the Transformer network which is able to model relationships between words in a

whole document, and generate embeddings which change according to context (Devlin et al. 2019). An additional modelling choice is defining how a document will be divided into units of text, i.e. is each document left whole, or segmented into paragraphs or sentences. Nguyen et al. (2020) suggest that "From a computational perspective, the unit of text can also make a huge difference, especially when we are using bag-of-words models, where word order within a unit does not matter".

The models discussed so far are generalised conceptual solutions designed to interpret data and then represented in computer code. Without data they are incapable of any inference. There is another definition of 'model' which is the trained model used in real world applications for prediction. The training process uses a set of input feature vectors, with associated output vectors (known as ground truth), and an algorithm adjusts the internal parameters of a model until it attains an optimal mapping between inputs and outputs. Optimal implies minimising the error between predictions made by the model using the input features against the ground truth values. This optimal final state is the trained model. It is effectively a mathematical function which maps a numerical input to a numerical output, and is defined by a collection of weightings and a set of rules for applying them. Ultimately it is the data which defines the parameters of the function. If the training data provides a good representation of the world being modelled, and an appropriate conceptual model has been chosen, then the model should perform well. Neural network models are initialised randomly meaning that the learned parameter values will be different if the exercise is repeated. The trained model, however, is deterministic when mapping input to output. The output values are informative of what machine learning algorithms are doing. Although a classification model is trained using labels, no labels are returned when the model makes predictions. Instead we, the user, interpret the floating point values it returns as classes, often assigning the class which has the highest value in the output vector. The algorithm is unaware of our interpretation of these as classes, they are just numbers.

We must therefore be careful not to project any notions of awareness or intelligence on the ML agents we employ to perform a task on our behalf. Ultimately they are mathematical constructs optimised to get as few wrong answers as possible on a given task. As the data grows bigger and the models more sophisticated the system increasingly approximates the appearance of intelligence, but the numerical, vectorised, representations of our documents, images or videos, which it takes as inputs are divorced from any prior knowledge or understanding of the world. In terms of representative and interpretive acts then, the forms of representation employed (numerical and vectorised) are deliberately designed for machine, rather than human processing and

those machines interpret these representations on the basis of algebraic measures, rather than on that of an advanced understanding and experience of the ‘real’ world being represented or interpreted.

5 Conclusion

This article has so far treated—for the most part—the processing of archivists, researchers and machine agents as separate elements, but it now seeks to bring them together within a joint enterprise of reasoning over archives. Looking from different angles often leads to a more holistic view leaving less room for blind spots. From the archivists’ perspective, acknowledging that their processing is a form of scholarly research, that it does involve inference from one thing to another and the drawing of conclusions on the basis of evidence has been accompanied by another, namely that they have always, in thereby shaping the data on which further conclusions are drawn, been more complicit in shaping the knowledge we have of the world around us and our place and agency in it than they might like to think (Schwartz and Cook 2002; Duff and Harris 2002; Wood et al. 2014). Then again, from the researchers’ perspective, more direct involvement with designing and maintaining the interfaces and services through which they access or achieve their data, has led to a renewed understanding that accounting for all the multiple interpretive and representative acts through which that data has been realised is becoming even more complicated, especially now that some of those acts are being carried out by machine agents on the basis of unfamiliar (numerical and vectorised) means of representation and interpretation (using algebraic measures). Just as in the past humanities scholars have not always taken into account these acts when they have been carried out by archivists, so too both archivists and humanities scholars should in future take care to continue to account for these acts when they are carried out by what we might prefer to consider and subsume as neutral tools of automation.

Artificial Intelligence as a field includes branches focused on logical and probabilistic reasoning, and decision making, but the type of machine learning most commonly being applied in the data work of archivists and researchers at the current time is more basic than that. As Pearl states, “Current machine learning systems operate, almost exclusively, in a statistical, or model-free mode,....” And “To achieve human level intelligence, learning machines need the guidance of a model of reality, similar to the ones used in causal inference tasks” (Pearl 2018). With access to such a model, the interpretative acts carried out by machines stand some chance of being conducted on the same basis as those of human agents. Looking towards the future then, we are already using machine learning to extract entities

and relationships from data in order to create knowledge bases and these may one day inform such a model of reality. Then again, graph neural networks allow us to create vectorised representations of knowledge graphs, similar to word embeddings (Bianchi et al. 2020). Combining these knowledge representations with text embeddings may advance our models beyond semantic similarity of words to utilising the real world events the words describe.

Reasoning over archives can be seen as a form of scholarly research activity, whether that activity is conceptualised, in Benardou et al.’s terms, as one of developing and referring to Propositions through interaction with Information Objects, or more broadly as one of sense-making. Archives cannot be (and never should have been) regarded as raw data to be reasoned over, but must be seen as the result of multiple representative and interpretive acts, of iterative realisation and activation as data, potentially involving many, many additional actors. In reasoning over Archives, we must also reason over, acknowledge and sometimes account for these acts. Particularly as we start to work with techniques that employ unfamiliar forms of representation and interpretive logic, it is vital that we properly understand and incorporate their working within our own.

Funding Not applicable.

Availability of data and materials Not applicable.

Code availability Not applicable.

Declarations

Conflict of interest Not applicable.

Ethics approval Not applicable.

Consent to participate Not applicable.

Consent for publication Not applicable.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Barry R, Gibbs Thibodeau S, Kiesling K, Pitti D, Ransom H, Spindler R (2013) Development of the encoded archival description DTD. <https://www.loc.gov/ead/eaddev.html>. Accessed 30 Mar 2021
- Bearman D (1989) Archival methods. https://www.archimuse.com/publishing/archival_methods/. Accessed 30 Mar 2021
- Benardou A, Constantopoulos P, Dallas C, Gavrilis D (2010) A conceptual model for scholarly research activity. <http://hdl.handle.net/2142/14945>. Accessed 30 Mar 2021
- Bianchi F, Rossiello G, Costabello L, Palmonari M, Minervini P (2020) Knowledge graph embeddings and explainable AI. In: Knowledge graphs for explainable artificial intelligence: foundations, applications and challenges. IOS Press, Amsterdam, pp 49–72
- Borek L, Dombrowski Q, Perkins J, Schöch C (2016) TaDiRAH: a Case Study in Pragmatic Classification. *Digit Humanit Q* 10.1. <http://www.digitalhumanities.org/dhq/vol/10/1/000235/000235.html>
- DARIAH-EU (2021) Taxonomy of Digital Research Activities in the Humanities. <https://www.dariah.eu/tools-services/tools-and-services/tools/tadirah-taxonomy-of-digital-research-activities-in-the-humanities/>. Accessed 30 Mar 2021
- Devlin J, Chang M-W, Lee K, Toutanova K (2019) BERT: pre-training of deep bidirectional transformers for language understanding. In: Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, vol 1 (long and short papers). Association for Computational Linguistics, Minneapolis, pp 4171–4186
- Duff WM, Harris V (2002) Stories and names: Archival description as narrating records and constructing meanings. *Arch Sci* 2:263–285. <https://doi.org/10.1007/BF02435625>
- Efron B, Hastie T (2016) Computer age statistical inference: algorithms, evidence, and data science. Cambridge University Press, Cambridge, p 446
- Erickson C (2013) Light, dark and dim archives: what are they? <http://preservationmatters.blogspot.com/2013/05/light-dark-and-dim-archives-what-are.html>. Accessed 30 Mar 2021
- Greene M, Meissner D (2005) More product, less process: revamping traditional archival processing. *Am Arch* 68:208–263. <https://doi.org/10.17723/aarc.68.2.c741823776k65863>
- Hockey S (2004) The history of humanities computing. In: Schreibman S, Siemens R, Unsworth J (eds) *A companion to digital humanities*. Blackwell, Oxford
- Jones P (1964) The society's chronicle: annual conference, Newcastle-upon-Tyne 10th and 11th June 1964. *J Soc Arch* 2:490–491. <https://doi.org/10.1080/00379816009513799>
- Krizhevsky A, Sutskever I, Hinton G (2012) ImageNet Classification with Deep Convolutional Neural Networks. *Communications of the ACM* 60, no. 6 (24 May 2017) pp 84–90. <https://doi.org/10.1145/3065386>
- Lemieux V (2015) Visual analytics, cognition and archival arrangement and description: studying archivists' cognitive tasks to leverage visual thinking for a sustainable archival future. *Arch Sci* 15:25–49. <https://doi.org/10.1007/s10502-013-9212-y>
- Light M, Hyry T (2002) Colophons and annotations: new directions for the finding aid. *Am Arch* 65:216–230. <https://doi.org/10.17723/aarc.65.2.13h27j5x8716586q>
- MacNeil H (2005) Picking our text: archival description, authenticity, and the archivist as Editor. *Am Arch* 68:264–278. <https://doi.org/10.17723/aarc.68.2.01u65t6435700337>
- McCarty W, Short H (2002) Mapping the field: a report to the Association for Literary and Linguistic Computing. <https://eadh.org/publications/mapping-field>. Accessed 30 Mar 2021
- Meehan J (2009) Making the leap from parts to whole: evidence and inference in archival arrangement and description. *Am Arch* 72:72–90. <https://doi.org/10.17723/aarc.72.1.kj672v4907m11x66>
- Mikolov T, Ilya S, Chen K, Corrado G, Dean J (2013) Distributed representations of words and phrases and their compositionality. In: *Advances in neural information processing systems*, vol 26. Curran Associates, Inc. <https://proceedings.neurips.cc/paper/2013/file/9aa42b31882ec039965f3c4923ce901b-Paper.pdf>
- Nguyen D, Liakata M, DeDeo S, Eisenstein J, Mimno D, Tromble R, Winters J (2020) How we do things with words: analyzing text as social and cultural data. *Front Artif Intell* 3(2020):62. <https://doi.org/10.3389/frai.2020.00062>
- Palmer C, Tefreau L, Pirmann C (2009) Scholarly information practices in the online environment: themes from the literature and implications for library service development. OCLC Research: Dublin, Ohio. <https://www.oclc.org/content/dam/research/publications/library/2009/2009-02.pdf>
- Pearl J (2018) Theoretical impediments to machine learning with seven sparks from the causal revolution. In: Proceedings of the eleventh ACM international conference on web search and data mining (WSDM '18). Association for Computing Machinery, New York, 3. <https://doi.org/10.1145/3159652.3176182>
- Russakovsky O, Deng J, Hao S, Krause J, Satheesh S, Ma S, Huang Z, Karpathy A, Khosla A, Bernstein M, Berg A, Fei-Fei L (2015) ImageNet large scale visual recognition challenge. <http://arxiv.org/abs/1409.0575>
- Schreibman S, Siemens R, Unsworth J (eds) (2004) *A companion to digital humanities*. Blackwell, Oxford
- Schwartz J, Cook T (2002) Archives, records and power: the making of modern memory. *Arch Sci* 2:1–19. <https://doi.org/10.1007/BF02435628>
- Siemens R (2016) Communities of practice, the methodological commons, and digital self-determination in the Humanities. *Digit Stud* 5.3. <https://doi.org/10.16995/dscn.31>
- Society of American Archivists (2021) Dictionary of Archives Terminology, s.v. “processing”. <https://dictionary.archivists.org/entry/processing.html>. Accessed 30 Mar 2021
- Spärck Jones K (1972) A statistical interpretation of term specificity and its application in retrieval. *J Doc* 28:11–21
- Stapleton L, Jaillant L (2020) Special Issue on “Born Digital”—shedding light into the darkness of digital culture. <https://www.springer.com/journal/146/updates/18507690>. Accessed 30 Mar 2021
- Surrey History Centre (2021) Seeing the Light: protection from light damage. <https://www.surreycc.gov.uk/culture-and-leisure/history-centre/depositors/conservation/light-protection>. Accessed 30 Mar 2021
- University of British Columbia (2016) SLAIS faculty member Dr Victoria Lemieux collaborates with University of Maryland iSchool on Computational Archival Science. <https://ischool.ubc.ca/2016/06/slais-faculty-member-dr-victoria-lemieux-collaborates-with-university-of-maryland-ischool-on-computational-archival-science/>. Accessed 30 Mar 2021
- Unsworth J (2000) Scholarly Primitives: what methods do humanities researchers have in common, and how might our tools reflect this? <https://johnunsworth.name/Kings.5-00/primitives.html>. Accessed 30 Mar 2021
- Unsworth J (2020) Scholarly primitives 20 years later. <https://www.youtube.com/watch?v=XyruWILDvIc>. Accessed 30 Mar 2021
- Wood S, Cifor M, Gilliland A, Carbone K, Punzalan R (2014) Mobilizing records: re-framing archival description to support human rights. *Arch Sci* 14:397–419. <https://doi.org/10.1007/s10502-014-9233-1>
- Yakel E (2003) Archival representation. *Arch Sci* 3:1–25. <https://doi.org/10.1007/BF02438926>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.