Metaphors We Work By

Reframing Digital Objects, Significant Properties, and the Design of Digital Preservation Systems

CHRISTOPH BECKER

ABSTRACT Much digital preservation research has evolved around the idea of authentic *digital objects* and their *significant properties*. However, the nature of digital preservation work continues to be ill-defined. This paper unpacks definitions of *digital objects* and their *significant properties* to deconstruct misleading conflations. I review how the use of the terms *digital object* and *significant properties* has evolved in digital preservation, and I identify conceptual inconsistencies. By critiquing the system boundaries assigned by different writers to the term *digital objects*, I explore the metaphorical nature of the concept and show that the discourse routinely ignores the role of computation in favour of artifact-centred concepts of bits and *records*. As a consequence, I illustrate category errors around what it means to migrate and preserve digital objects. I suggest a reformulation of both terms based on their metaphorical nature and discuss how this reformulation aligns with insights from research on electronic records and digital preservation. The discussion shows that digital objects are best understood as a metaphorical concept that allows us to articulate the emergent properties of computed performances relying on data, software, and hardware. Significant properties are best understood as mechanisms that allow curators to specify shared understandings of what constitutes authentic reproductions of digital objects. At the core of digital preservation is the design of software-based information systems intended to reproduce authentic digital objects. The article thus contributes to a reframing of the nature of digital preservation and emphasizes the importance of the conceptual frame of computing and systems design in archival education and practice.

RÉSUMÉ L'idée des *objets numériques* authentiques et de leurs *propriétés importantes* a fait évoluer une grande quantité de recherche portant sur la préservation numérique. Cependant, la nature du travail en préservation numérique n'est toujours pas bien définie. Cet article décortique les définitions des objets numériques et leurs propriétés importantes afin de déconstruire des amalgames trompeurs. J'examine comment l'utilisation des termes *objets numériques* et *propriétés importantes* ont évolué en préservation numérique et j'identifie les inconsistances conceptuelles. En critiquant l'attribution de limites de système à la notion *d'objet numérique* par divers auteurs, j'explore la nature métaphorique du concept et je montre que ce discours ignore couramment le rôle du traitement informatique en faveur de concepts centrés sur l'artéfact tels les bits et le *document d'archives*. Comme résultat, je dresse les erreurs de catégories dans ce que l'on entend par la migration et la préservation d'objets numériques. Je suggère une reformulation des deux termes en fonction de leur nature métaphorique et je discute de la façon dont cette reformulation s'aligne sur les résultats de la recherche sur les documents numériques et la préservation numérique. Cette discussion démontre que les objets numériques sont davantage compris comme un concept métaphorique permettant d'énoncer les propriétés émergentes des performances de système qui utilisent des données, des logiciels et du matériel informatique. Les propriétés importantes sont davantage comprises comme des mécanismes qui permettent aux conservateurs de préciser les compréhensions communes de ce en quoi consiste la reproduction authentique d'objets numériques. Au cœur de la préservation numérique se trouve la conception de systèmes d'information logiciels qui servent à la reproduction authentique des objets numériques. Cet article contribue alors à la redéfinition de la nature de la préservation numérique et met l'accent sur l'importance d'un cadre conceptuel en matière de traitement informatique et de conception de systèmes d'information dans la formation et la pratique archivistiques.

Introduction

This article aims to address the long-standing debate in digital preservation about the role of significant properties in preserving digital objects. The field of digital preservation has experienced a long-running back and forth between extremes in practice and in theory: first, the debate about whether migration or emulation was the best approach; subsequently, whether the specification of significant properties was something inevitable or impractical, necessary or impossible, measurable or indefensible; and, finally, whether the approach to specify such properties was an expression of unacceptably reductionist objectivist ideas or whether in fact this critique took subjective relativism to absurdity. This article takes the perspective that there is a useful framing that suggests a way forward out of this maze. The linguistic analysis of conceptual frameworks has shown that metaphors (i.e., conceptual mappings)[1] structure what we perceive, how we reason about it, and how we articulate the questions we ask. The contemporary theory of metaphor thus offers a new angle that allows us to unpack the conceptual frameworks used in prior discourse.

To make its argument, this article will discuss three key elements: the frame of systems design, the concept and use of significant properties, and the concept of the digital object. It will proceed to show that the idea of a digital object is a metaphorical construct that has been used with different meanings across different communities. Incongruence between the conceptual mappings that structure these meanings has remained implicit. This has prevented the debate on significant properties from engaging substantively with its core questions.

The pragmatic conclusions we arrive at are the following: 1) a constructive path forward lies in the suggestion that significant properties are best understood as mechanisms that allow curators to specify shared understandings of what should be accepted as authentic reproductions of digital objects; and 2) at the centre of digital preservation work is the design of software-intensive information systems intended to reproduce authentic, usable, and understandable digital objects. The discussion suggests that the conceptual frame of systems design is an important part of archival education and practice.

1 George Lakoff, "The Contemporary Theory of Metaphor," in *Metaphor and Thought,* ed. Andrew Ortony, 2nd ed. (Cambridge, UK: Cambridge University Press, 1993); George Lakoff and Mark Johnson, *Metaphors We Live By* (Chicago: University of Chicago Press, 1980).

The article compares this argument to findings from InterPARES and CASPAR,[2] two collaborative research projects concerned with the challenges of preserving digital content. The discussion shows that the conclusions here correspond to the InterPARES and CASPAR findings, but are based on different concepts and arguments. This article thus provides a new synthesis that resolves previous contradictions in support of a precise discourse.[3]

The Frame of Systems Design

The design of archival information systems raises complex questions about social and technical environments and their systemic interactions. Tackling these requires constructive use of conceptual frameworks that deal with challenging situations in which social, technological, and humanistic questions interact.

Systems thinking comes in diverse forms, which historically grew out of multiple parallel developments in fields as diverse as sociology, biology, mathematics, operations research, anthropology, management, and critical theory. The systems thinking label has thus been applied to very different ways of thinking.[4] This article cannot explore each of these in detail, but we emphasize the plurality of perspectives one can take: systems thinking is as much about *systemic ways of thought* as it is about the idea that systems can be distinguished out there in the world. In the words of Gerald Weinberg, "a system is a way of looking at the world."[5] In the words of Peter Checkland, the *system* concept "embodies the idea of a set of elements connected together which form a whole, this [whole] showing properties which are properties of the whole, rather than

2 See InterPARES Project, http://www.interpares.org/welcome.cfm; and Digital Curation Centre (DCC), Resources: Briefing Papers, CASPAR (Cultural, Artistic and Scientific Knowledge for Preservation, Access and Retrieval), http://www.dcc.ac.uk/resources/briefing-papers/technology-watch-papers/caspar.

3 This may be a necessary step of "theoretical hygiene," as Blackwell calls it in reference to Brian Cantwell Smith in an excellent paper on the role of metaphor in human computer interaction; see Alan F. Blackwell, "The Reification of Metaphor as a Design Tool," *ACM Transactions on Computer-Human Interaction (TOCHI)* 13, no. 4 (2006): 490–530; Brian Cantwell Smith, *On the Origin of Objects* (Cambridge, MA, and London: MIT Press, 1998).

4 Excellent overviews are given in Peter Checkland, *Systems Thinking, Systems Practice: Includes a 30-Year Retrospective* (New York: Wiley, 1999); and Michael C. Jackson, *Systems Thinking: Creative Holism for Managers* (Chichester, UK, and New York: Wiley, 2003).

5 Gerald Weinberg, *An Introduction to General Systems Thinking* (New York: Wiley, 1975), 52.

properties of its component parts."[6] This suggests that models of systems are both representations of systems that are thought to exist as well as cognitive devices and social constructs. Systems thinkers have distinguished types of systems from many angles, be they natural, social, artificial, and socio-technical[7] or goal-seeking, purposive, and purposeful.[8] Concepts such as hierarchy, control, feedback, emergence, environment, and system boundary are important ideas for reasoning about complex systems of any type (all the way from *the climate* to *my team*). Some strands of systems thinking focus on investigating how structure causes behaviour (system dynamics)[9]; some focus on mechanisms of feedback and control in organizations (organizational cybernetics)[10]; and others examine the architecture of complex artificial systems.[11] While it is impossible to do justice to the diversity of approaches here,[12] it is worth highlighting that systems thinking concepts have become commonplace in many disciplines and have long extended beyond the functionalist paradigm of systems analysis and systems engineering to interpretive paradigms and critical approaches.[13] Soft and critical systems thinking places the central emphasis on the constitutive role of the observers and their worldviews and the process of creating boundaries through what Geoffrey Vickers termed *appreciation:* "a combined judgment of value and fact"[14] grounded in the observer's experience. Common to all systems

6 Checkland, *Systems Thinking, Systems Practice,* 3.

7 Eric Trist, "The Evolution of Socio-Technical Systems: A Conceptual Framework and an Action Research Program" (Toronto, ON: Ontario Ministry of Labour, 1981).

8 Russell L. Ackoff, "Towards a System of Systems Concepts," *Management Science* 17, no. 11 (July 1971): 661–71.

9 See, for example, Peter M. Senge, *The Fifth Discipline: The Art and Practice of the Learning Organization* (New York: Doubleday/Currency, 1990).

10 See, for example, Stafford Beer, *Diagnosing the System for Organizations* (Chichester, UK: Wiley, 1985).

11 See, for example, Herbert A. Simon, "The Architecture of Complexity," in *Facets of Systems Science,* International Federation for Systems Research International Series on Systems Science and Engineering (Boston: Springer, 1991), 457–76, accessed 1 July 2017, https://doi.org/10.1007/978-1-4899-0718-9_31.

12 See Magnus Ramage and Karen Shipp, *Systems Thinkers* (Dordrecht, NE, and New York: Springer, 2009).

13 Jackson, *Systems Thinking.*

14 Geoffrey Vickers, *Value Systems and Social Process,* The International Behavioural and Social Sciences Library: Sociology & Social Policy, IX (Abingdon, UK: Routledge, 2001), 191; Peter Checkland, "Webs of Significance: The Work of Geoffrey Vickers," *Systems Research and Behavioral Science* 22, no. 4 (2005): 285–90.

approaches, however, is a holistic approach to the structure of systems and the organization of their elements.

Systems concepts play a central role in the design of information systems. While a comprehensive review of the multitude of views and conceptions of *systems design* is beyond the scope of this article, a few aspects should be highlighted.[15] Terry Winograd and Fernando Flores characterize *design* as "the interaction between understanding and creation,"[16] a succinct but opaque description. In unpacking it, we find a dual nature of design. It encompasses an appreciation of an as-is situation and the concerns of stakeholders, and it uses this as a starting point for creating desired change through an intentional intervention. In many but not all design professions, that intervention involves a tangible artifact. In the words of John M. Carroll, "the objective of design is to produce or specify an artifact that satisfies needs identified in the current situation."[17] *Design* (as a verb) thus refers to the intentional specification or production of an object that is to fulfill an identified purpose and is meant to satisfy an identified need of identified stakeholders in a given environment through the application of certain techniques and the use of resources. The process of design leads to a plan for a product based on the designers' intent. This is not commonly straightforward, however: "the possible moves that designers may take in reasoning from a description of the current situation in the world toward an improved version of that situation are not specified."[18] Those who design must identify the environment, stakeholders, and needs to be considered, and specify *a design* for the object to be placed into this environment using available means. In doing so, those who design must address the concerns of those affected by the change. Paul Ralph and Yair Wand formalize *design* (as a noun) as "a specification of an object, manifested by an agent, intended to accomplish goals, in a particular environment …

15 A particularly comprehensive discussion on information systems design is given in Peter Checkland and Sue Holwell, *Information, Systems and Information Systems: Making Sense of the Field* (New York: Wiley, 1998).

16 Terry Winograd and Fernando Flores, *Understanding Computers and Cognition: A New Foundation for Design* (Norwood, NJ: Ablex, 1986), 4.

17 John M. Carroll, *Making Use: Scenario-Based Design of Human–Computer Interactions,* 1st ed. (Cambridge, MA: MIT Press, 2000), 29.

18 Ibid., 20.

satisfying a set of requirements, subject to constraints."[19] In discussing six properties that differentiate design from simple problem-solving, Carroll highlights such characteristics as the incompleteness of initial knowledge about the problem to be addressed, the difficulty of anticipating what effects design choices will have, the impossibility of knowing the state of the world resulting from a design intervention, and the need to incorporate diverse perspectives and knowledge.

When the object of design itself is a system, the structure of that system's elements, the scoping of its boundary, and the relationships to its environment become central parts of the design process. Since the introduction of artificial components such as software systems into a socio-technical system such as an information system will inevitably change the overall system's structure and behaviour and affect the individuals acting within this context, the design of information systems must take a holistic view of the resulting overall system. Models of systems in design are used both indicatively, to represent and examine a current state of a system or the environment of the system under design, and optatively, to explore desired states of the resulting system.

Central to this notion of design is the distinction of means and ends and the importance of trade-offs. While *ends* refer to the problem domain in which the stakeholders articulate a problem for which a design should be developed (the *what*), *means* belong to the solution domain that comprises the various techniques and resources that can be employed to realize the object under design (the *how*). Since there are inevitably multiple ways to achieve the same functional purpose, designers must identify which to employ. Means–ends relationships allow designers to evaluate and describe *how well* a combination of means employed in a certain context can achieve desired ends, and they may express this evaluation as *quality*. In the design of systems, means and ends are interlinked horizontally (choices of *what* and *how* are often mutually dependent) and vertically (in a functional hierarchy of systems and subsystems, a higher level's *how* often becomes the lower level's *what*). In software-intensive information systems, means and ends manifest most visibly as *requirements* (statements of needs) and *architecture* (the structure of a system or, more formally, the "fundamental concepts or properties of a system in its environment, embodied in its

19 Paul Ralph and Yair Wand, "A Proposal for a Formal Definition of the Design Concept," *Design Requirements Engineering: A Ten-Year Perspective,* LNBIP, 14 (2009): 103–36.

elements, relationships, and in the principles of its design and evolution"[20]). The discipline of requirements engineering has developed sophisticated conceptual frameworks for eliciting, specifying, validating, and modelling requirements. Requirements are considered well articulated if the statements are expressed independently of how they are to be achieved. Yet dependencies between requirements and architecture necessitate that design progresses iteratively between them.[21]

Success criteria for design depend on the assumptions of the paradigm from which design is evaluated. In an environment with a functionalist worldview, the role of the designer is to ensure efficacy and efficiency, while the identification of the problem and the purpose of the system are considered unproblematic. Others outside of this paradigm highlight the political and constructed nature of social reality and emphasize the plurality of valid stakeholder concerns. The concept of *wicked problems*, often misunderstood by proponents of a functionalist paradigm in engineering as simply "difficult" problems, highlights the disjunct worldviews and concerns, legitimately divergent needs, and potentially irreconcilable aspirations of stakeholders in a complex problem situation.[22] Others emphasize the political and coercive nature of stakeholder relationships and highlight the need for *fairness* in the processes of boundary judgments and systems design.[23] In either perspective, however, design is commonly conceived as excluding construction: "When the solution space is specified, the design work per se is complete."[24] The heterogenous but interconnected nature of problem situations that call for systems design by necessity also drives the interdisciplinarity inherent in both the idea of design and systems thinking. Problems worth addressing are not bounded by disciplinary epistemologies, so systems design is always interdisciplinary and collaborative.

20 International Organization for Standardization, "ISO/IEC/IEEE 42010:2011 – Systems and Software Engineering – Architecture Description," 2011, accessed 1 July 2017, https://www.iso.org/standard/50508.html.

21 Bashar Nuseibeh, "Weaving Together Requirements and Architectures," *Computer* 34, no. 3 (2001): 115–19.

22 Richard Buchanan, "Wicked Problems in Design Thinking," *Design Issues* 8, no. 2 (April 1992): 5–21; Horst W.J. Rittel and Melvin M. Webber, "Dilemmas in a General Theory of Planning," *Policy Sciences* 4, no. 2 (1973): 155–69; C. West Churchman, "Guest Editorial: Wicked Problems," *Management Science* 14, no. 4 (1967): B141–42.

23 Jackson, *Systems Thinking*.

24 Carroll, *Making Use,* 20.

In the design, operation and maintenance of archival information systems, digital preservation is a concern that is shared by archivists, librarians, and many other information professionals. The computational nature and the exploding volumes of born-digital objects result in a tension between the desire to interact with individual resources and the drive to reduce individual interaction in the process of preservation to allow that process to scale economically to large-scale operations.[25] This tension manifests differently across professional, institutional, and disciplinary contexts. Since the individual treatment of digital resources is simply infeasible for large collections of resources of any kind, attention must shift to the design and configuration of systems that in turn operate on resources at scale. The design of such systems needs to focus on identifying and working with commonalities if we are to bring the ability to enact curatorial responsibility to the age of large-scale networked computing. In the "age of algorithms," as Clifford Lynch[26] calls it, this need is more pressing than ever.

The argument in this article arose as a by-product of a decade of empirical work with preservation planning and the design of digital preservation systems across those different professional contexts. Preservation plans are a translation of preservation policy into a specification of how an identified set of digital objects (records) in a given setting should be preserved.[27] The preservation planning approach and system developed in PLANETS and SCAPE, projects concerned with building services and tools in the field, afforded insightful perspectives on core questions that surround the preservation of digital objects through the lenses of practice, teaching, and research.[28]

25 Christoph Becker, Luis Faria, and Kresimir Duretec, "Scalable Decision Support for Digital Preservation," *OCLC Systems & Services: International Digital Library Perspectives* 30, no. 4 (2014): 249–84.

26 Clifford Lynch, "Stewardship in the 'Age of Algorithms,'" *First Monday* 22, no. 12 (2 December 2017), http://firstmonday.org/ojs/index.php/fm/article/view/8097.

27 Christoph Becker, Hannes Kulovits, Mark Guttenbrunner, Stephan Strodl, Andreas Rauber, and Hans Hofman, "Systematic Planning for Digital Preservation: Evaluating Potential Strategies and Building Preservation Plans," *International Journal on Digital Libraries* 10, no. 4 (2009): 133–57.

28 See PLANETS (Preservation and Long-term Access through NETworked Services), http://www.planets-project.eu, and SCAPE (SCAlable Preservation Environments), http://scape-project.eu, both accessed 28 February 2018. See also Adam Farquhar and Helen Hockx-Yu, "Planets: Integrated Services for Digital Preservation," *International Journal of Digital Curation* 2, no. 2 (2007), http://www.ijdc.net/index.php/ijdc/article/view/45. For information about the preservation planning tool, see PLATO, accessed 28 February 2018, http://www.ifs.tuwien.ac.at/dp/plato/intro; and Christoph Becker, Hannes Kulovits, Andreas Rauber, and Hans Hofman, "Plato: A Service Oriented Decision Support System for Preservation

The Debate over the Concept of Significant Properties

The emergence of digital preservation more than 20 years ago was marked by a number of milestone publications that set forth the cornerstones of the nascent field. One, the task force report "Preserving Digital Information," set the stage by explaining that "for digital objects, no less than for objects of other kinds, knowing *how* operationally to preserve them depends, at least in part, on being able to discriminate the essential features of *what* needs to be preserved."[29]

Subsequently, the topic of significant properties became a central question in digital preservation research.[30] Five key examples of attempts to define the concept of significant properties illustrate the conceptual frameworks deployed in this discussion (emphasis added):

1. Margaret Hedstrom and Christopher A. Lee's well-cited paper defined significant properties as "those properties of digital objects that affect their quality, usability, rendering, and behavior."[31]

2. The Cedars project suggested two definitions for significant properties. One defined the concept as "those components of a digital object deemed necessary for its long-term preservation."[32]

Planning," in *Proceedings of the 8th ACM/IEEE-CS Joint Conference on Digital Libraries,* 16–20 June 2008, Pittsburgh, PA (New York: ACM, 2008), 367–70.

29 Donald Waters and John Garrett, "Preserving Digital Information: Report of the Task Force on Archiving of Digital Information," commissioned by the Commission on Preservation and Access and the Research Libraries Group (1 May 1996), 17.

30 Fairly comprehensive overviews are given in Andrew Wilson, *Significant Properties Report,* InSPECT Work Package 2.2 (Arts and Humanities Data Service, 2007); Helen Hockx-Yu and Gareth Knight, "What to Preserve?: Significant Properties of Digital Objects," *International Journal of Digital Curation* 3, no. 1 (2008): 141–53; and, in particular, Geoffrey Yeo, "Nothing Is the Same as Something Else: Significant Properties and Notions of Identity and Originality," *Archival Science* 10, no. 2 (2010): 85–116.

31 Margaret Hedstrom and Christopher A. Lee, "Significant Properties of Digital Objects: Definitions, Applications, Implications," in *Proceedings of the DLM-Forum* 200 (2002): 218–27, accessed 1 July 2017, https://www.ils.unc.edu/callee/sigprops_dlm2002.pdf.

32 Cedars Project, "Cedars Guide to Digital Collection Management," Cedars Project (2002), https://web.archive.org/web/20060227202356/http://www.leeds.ac.uk/cedarsguideto/dpstrategies .og/10.1145/1378889.1378954

3. Elsewhere, Cedars specified that significant properties are "those technical characteristics agreed by the archive … to be most important for preserving the digital object over time.… A digital object's significant properties are not assumed to be empirical; archives will make judgments at levels appropriate to fulfil their preservation responsibilities and meet the needs of the archive's user communities."[33]

4. Paul Wheatley, in a CAMiLEON paper of the same period, refers to "the significant properties (i.e., the intellectual content of the raw textual data) of word processor files."[34]

5. Like its predecessors, the InSPECT project defined the concept as "the characteristics of digital objects that must be preserved over time in order to ensure the continued accessibility, usability, and meaning of the objects, and their capacity to be accepted as evidence of what they purport to record."[35]

It is worth noting that each of these definitions establishes a causal means–ends relationship between *components* or *characteristics* (the means) that are positioned as necessary *for* preservation to succeed (the ends). In each, significant properties are positioned as the means to the ends of digital preservation and are described as the required elements that must be present *in order* to achieve successful preservation. The debate in all these articles asked several questions simultaneously, without always distinguishing them explicitly, a symptom of evolving discourse. On reflection, we can clarify these questions as the following:

1. Do digital objects have distinguishable properties?

2. Do sets of digital objects have *common* distinguishable properties?

33 Ibid. See also CAMiLEON (Creative Archiving at Michigan and Leeds: Emulating the Old on the New), accessed 28 February 2018, http://www.dcc.ac.uk/resources/external/camileon-creative-archiving-michigan-and-leeds-emulating-old-new.

34 Paul Wheatley, "Migration: A CAMilEON Discussion Paper," *Ariadne* 29 (2001), http://www.ariadne.ac.uk/issue29/camileon. See also InSPECT (Investigating the Significant Properties of Electronic Content over Time), accessed 28 February 2018, https://web.archive.org/web/20160520082501/http://www.significantproperties.org.uk.

35 Wilson, "Significant Properties Report," 8 (emphasis added).

3. Is the question whether any of these properties are significant a decidable question (a) a priori, (b) in a given context, (c) in an absolute sense, or (d) not at all?

4. Who gets to decide?

Multiple projects have addressed the desire to operationalize the use of significant properties in digital preservation practice. The InSPECT project aimed to "determine sets of significant properties for a specified group of digital object types," and a set of reports was produced to that end. The report of the Significant Properties workshop, which marked the culmination of several projects in 2008, in turn shifts the attention to the role of the concept itself and describes it as "a mechanism for determining the characteristics that must be preserved for digital objects to remain accessible and meaningful."[36]

Backlash came in part from the difficulty of operationalizing this mechanism in practice, and the role of significant properties turned into a contested question in digital preservation.[37] The debate is well illustrated by Geoffrey Yeo, who discusses what he refers to as "variously called" significant properties or essential characteristics of "objects."[38] In discussing projects such as those above, Yeo takes a critical stance on efforts to operationalize the concept of significant properties in digital preservation:

> The search for methods of quantifying significant properties of digital objects … has usually allowed little space for recognising subjectivity or fuzziness. Classification systems for data about significant properties are predicated on beliefs that aspects of experience can be definitively fixed and formally codified, and attempts to construct canonical lists of such

36 Hockx-Yu and Knight, "What to Preserve?" 142 (emphasis added).

37 For examples of this debate, see Kevin Bradley, "Defining Digital Sustainability," *Library Trends* 56, no. 1 (2007): 148–63; Yeo, "Nothing Is the Same"; Jyue Tyan Low, "A Literature Review: What Exactly Should We Preserve? How Scholars Address This Question and Where Is the Gap" (paper written for the digital preservation class, Master of Library and Information Science program, University of Pittsburgh, Spring 2011), accessed 1 July 2017, https://arxiv.org/ftp/arxiv/papers/1112/1112.1681.pdf; and Colin Web, David Pearson, and Paul Koerben, "'Oh, You Wanted Us to Preserve That?!' Statements of Preservation Intent for the National Library of Australia's Digital Collections," *D-Lib Magazine* 19, no. 1/2 (January/February 2013), doi:10.1045/january2013-webb.

38 Yeo, "Nothing Is the Same."

> properties can be seen as motivated by a wish to find an objective basis for decision-making and selectivity.[39]

Yeo suggests that projects such as PLANETS attempted to define "the significant properties" of classes of objects free of their organizational and curatorial context:

> Undaunted by InterPARES's failure to construct a usable typology, digital preservationists often lay similar emphasis on determining which properties 'should be considered significant for different classes of record' or 'for a range of object types' (InSPECT 2006; 2008, p. 10). For example, a project in Portugal (Ferreira et al. 2007) has attempted to compile lists of significant properties for preserving still images, text documents and relational databases, and a researcher in Austria has set out to define the significant properties of dynamic documents and digital art (Guttenbrunner 2008). The InSPECT project in the United Kingdom initially followed a similar route, seeking to identify the properties of raster images, e-mails and other formats, and suggesting that e-mails, for example, have '28 properties … that must be preserved' (InSPECT 2008, p. 25).

Yeo suggests that Mark Guttenbrunner's focus was the development of a definitive set of significant properties of broad classes of objects ("dynamic documents and digital art") with objective validity; that his approach was similar to that taken in the InSPECT project; and that these projects are misguided in assuming what Yeo later calls a "seemingly objective basis." However, a few inaccuracies cause a conflation of multiple different approaches. Unlike the InSPECT project team, Guttenbrunner and his colleagues were specifically applying a preservation planning approach to a given set of objects to be preserved and developed specifications of properties considered significant in a given context.[40] The

39 Ibid.

40 Mark Guttenbrunner, Christoph Becker, Andreas Rauber, and Carmen Kehrberg, "Evaluating Strategies for the Preservation of Console Video Games," in *Proceedings of the 5th International Conference on Preservation of Digital Objects – iPRES, London, 29–30 September 2008,* accessed 1 July 2017, https://phaidra.univie.ac.at/view/o:294190: 115–21; Mark Guttenbrunner, Christoph Becker, and Andreas Rauber, "Keeping the Game Alive: Evaluating Strategies for the Preservation of Console Video

method was replicated across organizational contexts. In the method of preservation planning,[41] the concept of significant properties was operationalized, but not on a "seemingly objective basis," as alluded to by Yeo, but through a contextualized perspective based on an analysis of stakeholders situated within an organizational context. The InSPECT project followed a very different route and aimed to develop a comprehensive catalogue of properties of object types with broad validity across organizations, based on ideas of object types, file formats, and technical instantiations. Contrary to preservation planning, the ambition was indeed to develop comprehensive specifications across different situations. The PLANETS characterization registry also mentioned by Yeo was like InSPECT in this aim, but it focused on technically measurable characteristics, not on significant properties. The situated specification of significant properties *for specific sets of digital objects* was, and remains, a core element of preservation planning.[42]

As Yeo's discussion illustrates, the significant properties debate evolved simultaneously around questions about what the concept meant, whether it was feasible to apply the concept in digital preservation practice (in specific contexts, which ranged from research data and library resources to electronic art and electronic records), whether it would be ethical to do so, how the concept could be operationalized, or why this could never work. The debate took place in parallel across multiple professional and academic disciplines with different terminologies, including archives,[43] digital libraries,[44] library and information science,[45] computer science,[46] and digital curation.[47] It was typified by acrimonious disputes

Games," *International Journal of Digital Curation* 5, no. 1 (2010): 64–90.

41 Becker et al., "Systematic Planning."

42 Ibid.

43 Yeo, "Nothing Is the Same."

44 Angela Dappert and Adam Farquhar, "Significance Is in the Eye of the Stakeholder," in *Proceedings of the 13th European Conference on Digital Libraries, 27 September to 2 October 2009,* 297–308, accessed 1 July 2017, http://www.planets-project.eu/docs/papers/Dappert_SignificantCharacteristics_ECDL2009.pdf; Web, Pearson and Koerben, "Oh, You Wanted Us to Preserve That?!"

45 Hedstrom and Lee, "Significant Properties."

46 Jeannette M. Wing and John Ockerbloom, "Respectful Type Converters," *IEEE Transactions on Software Engineering* 26, no. 7 (2000): 579–93.

47 Hockx-Yu and Knight, "What to Preserve?"; Gareth Knight and Maureen Pennock, "Data without Meaning: Establishing the Significant Properties of Digital Research," *International Journal of Digital Curation* 4, no. 1 (2009): 159–74.

about necessity, complications, feasibility, inevitability, and pointlessness. Early on, Kevin Bradley dismissed the idea of significant properties because it "takes us no further if we cannot define its meaning in such a way that we understand what properties are under consideration, and describe them in a way that is machine-readable and automatically actionable."[48] Ultimately, the debate did not lead to a resolution. In hindsight, this is not surprising because the conceptual frameworks underpinning the debate seem flawed.

Multiple conflations overlap in the definitions and in the discourse summarized and critiqued by Yeo. At least three of these conflations are quite significant, and it is essential to address them to move on. The first concerns *classes and instances*, the distinction between properties that apply to sets of objects (classes), and the features of specific instances or specific objects. Typically, the former is referred to as properties, while the latter is called characteristics. The second concerns *essence and attribution*, the distinction between the belief in intrinsic essence and the perspective of properties as *attributed* by an observer. *Essential features* point to the former, while *significant properties* focus on the latter. Finally, the third concerns *means and ends*, the distinction between the ends to be achieved through preservation and the various means (actions, instruments, and interventions) through which these ends are to be achieved.

The first two conflations were identified, addressed, and resolved effectively by Dappert and Farquhar. First, properties apply to sets, while characteristics connect an element of that set to a value. For example, people have names (a property of the class *Person*), and this author's first name is given at the head of this article (a characteristic of this person). Second, Dappert and Farquhar highlight that "significance" by definition is "in the eye of the stakeholder" and always attributed.[49] I take the pragmatic perspective that there is no practical need to resolve the deep philosophical dispute between intrinsic essence and attributed significance – significant properties are all we must agree on. We shall return to the third distinction soon. However, a fourth challenge remains: Where are those properties to which we could attribute significance supposed to be located? The intuitive answer is, of course, "in the digital object," the object that is to be preserved.

48 Bradley, "Defining Digital Sustainability."

49 Dappert and Farquhar, "Significance Is in the Eye of the Stakeholder."

The Concept of the "Digital Object"

Archivists and other information professionals like to think of a digital object as something they can protect, safeguard, and keep. For the Digital Curation Centre (DCC), digital objects are discussed under the heading of data:

> Data, any information in binary digital form, is at the centre of the Curation Lifecycle. This includes Digital Objects, simple digital objects (discrete digital items such as text files, image files or sound files, along with their related identifiers and metadata) or complex digital objects (discrete digital objects made by combining a number of other digital objects, such as websites; and Databases, structured collections of records or data stored in a computer system.[50]

This points to a very particular conceptual frame predicated on data, as highlighted by Costis Dallas.[51]

Definitions of digital objects abound. The definition by Sandra Payette et al. serves as an illustration: "a digital object model … enables the aggregation of distributed, heterogeneous elements or streams of data to create complex multimedia objects."[52] It is consistent with the widely cited definition of Robert Kahn and Robert Wilensky: "A digital object is a data structure whose principal components are digital material, or data, plus a unique identifier for this material."[53]

It also corresponds to the OAIS definition of a digital object: "An object composed of a set of bit sequences."[54]

50 Digital Curation Centre (DCC), "DCC Curation Lifecycle Model," accessed 1 June 2017, http://www.dcc.ac.uk/resources/curation-lifecycle-model.

51 Costis Dallas, "Digital Curation beyond the 'Wild Frontier': A Pragmatic Approach," *Archival Science* 16, no. 4 (2016): 421–57.

52 Sandra Payette, Christophe Blanchi, Carl Lagoze, and Edward A. Overly, "Interoperability for Digital Objects and Repositories," *D-Lib* 5, no. 5 (May 1999), http://www.dlib.org/dlib/may99/payette/05payette.html (emphasis added).

53 Sandra Payette, Christophe Blanchi, Carl Lagoze, and Edward A. Overly, "Interoperability for Digital Objects and Repositories," D-Lib 5, no. 5 (May 1999), http://www.dlib.org/dlib/may99/payette/05payette.html (emphasis added).

54 Consultative Committee for Space Data Systems (CCSDS), *Reference Model for an Open Archival Information System (OAIS): Recommended Practice, CCSDS 650.0-M-2* (June 2012), 1–11, accessed 1 July 2017, https://public.ccsds.org/pubs/650x0m2.pdf.

Similar definitions have been brought forward in digital libraries and in information science and other fields. Variations exist, of course, but the principal notion they share points to the conception that a digital object is an aggregate of data streams, i.e., bits. Even Yuk Hui's extensive exploration of the philosophical basis of digital objects approaches and defines the book's subject matter as follows:

> These objects are basically data, sharable and controllable.… By digital objects, I mean objects that take shape on a screen or hide in the back end of a computer program, composed of data and metadata regulated by structures or schemas.[55]

This focus on the tangible mirrors a logical way of thinking of tangible artifacts that can be stored, ordered, arranged, described, duplicated, and sent around. However, this quickly becomes an untenable position when we consider that actions for preservation – whether migration or emulation – in fact destroy what these definitions call "digital objects."[56] From the viewpoint of computer science, Guttenbrunner and Rauber explain that there is little difference between migration and emulation: emulation replaces parts of the software in the computing stack; migration will in fact replace both the data *and* the software stack, since the common purpose is to allow the "digital object" to be reproduced in a different software environment.[57]

The fact that the data itself (i.e., the bits) does not need to survive has been acknowledged since the beginning of digital preservation research. For example, the InterPARES Authenticity Task Force's *Requirements for Assessing and Maintaining the Authenticity of Electronic Records* explicitly declares that an "electronic record … [is] essentially complete and uncorrupted if the message that it is

55 Yuk Hui, *On the Existence of Digital Objects* (University of Minnesota Press, 2016), e-book accessed 1 July 2017, https://muse.jhu.edu/book/45372.

56 See also Simone Sacchi, "What Do We Mean by 'Preserving Digital Information'? Towards Sound Conceptual Foundations for Digital Stewardship" (PhD dissertation, University of Illinois at Urbana-Champaign, 2015), 46–50, accessed 1 July 2017, https://academiccommons.columbia.edu/catalog/ac:185846.

57 Mark Guttenbrunner and Andreas Rauber, "Evaluating Emulation and Migration: Birds of a Feather?," in *The Outreach of Digital Libraries: A Globalized Resource Network, Proceedings of the 14th International Conference on Asia-Pacific Digital Libraries, ICADL 2012, Taipei, Taiwan, 12–15 November 2012,* ed. H. H. Chen and G. Chowdhury, Lecture Notes in Computer Science, vol. 7634 (Berlin: Springer, 2012): 158–67, accessed 1 July 2017, doi:10.1007/978-3-642-34752-8_22.

meant to communicate in order to achieve its purpose is unaltered. This implies that its physical integrity, such as the proper number of bit strings, may be compromised."[58] The performance model of the National Archives of Australia articulates this through the idea that the record is no longer experienced directly: instead, a source (data file) is used in a process (hardware and software) to produce a performance (such as a rendering on screen). The OAIS model distinguishes between the Data Object (composed of bits) and the Information Object ("a Data Object together with its Representation Information"[59]). It remains silent, however, on the nature of computation; by stating that software "is used to access the Information Object,"[60] it does not acknowledge the constitutive role of computation in (re)producing the Information Object. By suggesting that the Information Object is merely an *aggregate* of bits and representation information (including software), the model denies that it really is a product of the interactions between these pieces in a computational process.

The conclusions are clear, then, and have been for a while:[61] if the data and the computation are replaceable, they cannot be significant. Instead, significance lies in the performance. Digital preservation intervenes on the level of means – bits, metadata, formats, software technology – and the criteria for success must be defined on the level of outcomes: meaningful engagement of members of the designated community with authentic reproductions of understandable digital objects. However, all definitions mentioned above are focused on the *components of* a digital object, the elements *necessary for* preservation, and the *properties* of files. These are concepts of the solution space. All focus on the means, not the ends, of digital preservation.

Because they conflate problem domain and solution domain, the use of these definitions as a basis for the design of digital preservation systems is fraught

58 InterPARES Authenticity Task Force, "Requirements for Assessing and Maintaining the Authenticity of Electronic Records" (March 2002), accessed 1 July 2017, http://www.interpares.org/book/interpares_book_k_app02.pdf.

59 CCSDS, "Reference Model for an Open Archival Information System (OAIS)," 1–12.

60 Ibid., 2–5.

61 Helen Heslop, Simon Davis, and Andrew Wilson, "An Approach to the Preservation of Digital Records," National Archives green paper (2002), accessed 1 July 2017, http://trove.nla.gov.au/version/38579727; Luciana Duranti and Kenneth Thibodeau, "The Concept of Record in Interactive, Experiential and Dynamic Environments: The View of InterPARES," *Archival Science* 6, no. 1 (2006): 13–68; Becker et al., "Systematic Planning"; Guttenbrunner and Rauber, "Evaluating Emulation and Migration."

with difficulties and confusion. Various contributors to the debate of significant properties have all used terms such as *digital object* but have attributed different boundaries, and have often done so implicitly. Sometimes, the term refers to data[62] contained in files or streams; sometimes, to conceptual objects in the mind of an observer or a designated community, perhaps referred to through their media type (document, image) or genre (letter, portrait photograph); sometimes, to the concept of the record or another kind of information object, or, explicitly, to an OAIS Information Object.[63] *Digital object*, as the term used to gesture at all of this, served as a false friend that led to a false sense of discourse when, in fact, many of these arguments for and against the ideas of significant properties of these supposed objects do not connect substantively. For example, Yeo suggests that the InterPARES definition of an "authentic record"[64] as "a record that is what it purports to be and is free from tampering or corruption" provides grounds to "argue that the authenticity of digital objects stands or falls by whether they remain unchanged at bit level." However, as he points out, the InterPARES team also asserted that migration creates "new records," which can be authentic under certain conditions.[65]

The definitional inconsistency, incoherence, and lack of clarity of the *digital object* concept is also highlighted by Stephen Abrams,[66] who compares previous models, including those discussed above, from a semiotic perspective and finds that none distinguish effectively and comprehensively between related but distinct concerns. Based on the suggestion that semiotics provides a coherent

62 The term *data* itself is of course overloaded and contested as well. In this article, it is used in consistence with the OAIS information model, i.e., it refers to bits unless qualified as in, for example, "research data" or "metadata."

63 See Ken Thibodeau, "Overview of Technological Approaches to Digital Preservation and Challenges in Coming Years," in *The State of Digital Preservation: An International Perspective* (2002). Thibodeau highlights that digital objects are simultaneously "physical, logical, and conceptual" and emphasizes some of these boundaries. As Sacchi points out, however, Thibodeau's use of multiple inheritance to explain this multiplicity is logically inconsistent and leaves many questions of identity and composition unresolved; see Sacchi, "What Do We Mean?," 48–50.

64 Heather MacNeil et al., *Authenticity Task Force Report* (Vancouver: InterPARES Project, 2001), 2, accessed 1 July 2017, http://www.interpares.org/book/interpares_book_d_part1.pdf.

65 Yeo, "Nothing Is the Same," 4.

66 Stephen Abrams, "A Foundational Framework for Digital Curation: The Sept Domain Model" (paper presented at IPRES 2015–12th International Conference on Digital Preservation, 2–6 November 2015, Chapel Hill, NC), 6, accessed 1 July 2017, http://escholarship.org/uc/item/75v3z67n.pdf.

and comprehensive framework, Abrams has developed the Sept model. It distinguishes between seven types of objects of increasing utility, differentiated by a set of analytic dimensions and aspects that apply at each level. These types are (1) *blobs*, such as unidentified bit sequences; (2) *artifacts*, such as identified demarcated bit sequences; (3) *exemplars*, i.e., characterized artifacts with known internal structure (such as a JPEG file); (4) *products*, such as a JPEG file with associated description; (5) *assets*, i.e., performed products (such as a rendering of the JPEG image); (6) *records*, defined as "trustworthy object[s] resulting from an evaluative act of verification,"[67] and (7) *heirlooms*, distinguished by the expectation that they have been made viable and useful through interventions. While the terminology does not easily translate to archives, the model effectively distinguishes which class of curation activities applies to what type of object – for example, forensic disk imaging applies to blobs, verification of fixity to artifacts, and migration to products.[68] These distinctions are used to develop a conceptual model of digital curation practice. The model remains as silent as the OAIS model, however, on the role of computation in constituting the asset through a performance, and thus remains within an artifact-centric worldview.

If we consider that the elements that supposedly make up the digital object are a system, and that the whole includes not just the files but also the computational process,[69] then Russell Ackoff, a systems thinker famous for his aphorisms, reminds us that a system is never just the sum of its parts: it must be understood as the product of their interactions.[70] These interactions produce emergent properties that we cannot attribute to the parts, only to the whole. A simple technical example illustrates this point. The number of pages in an editable electronic document is typically not stored in the file. This number is also not stored in the software. Instead, it is computed in the performance, and humans really have no unmediated access to this process. Characterization tools provide measures for these characteristics, but any characterization relies on an algorithm, and for many characteristics, the algorithm amounts to a partial or full performance. Identifying typologies of common standardized properties is highly complex, but

67 Abrams, "A Foundational Framework," 6.

68 Ibid., 9.

69 Guttenbrunner and Rauber, "Evaluating Emulation and Migration."

70 Russell Ackoff, *Re-Creating the Corporation: A Design of Organizations for the 21st Century* (New York: Oxford University Press, 1999).

of course formal approaches have been suggested and implemented.[71] Extensible characterization languages (XCL)[72] aim to provide canonical specifications of characteristics of object types such as "raster images" and to supply extractor functions for these characteristics across different formats. Their implementation, however, uses algorithms contained in off-the-shelf components and scripts to perform the characterization process. The Bitstream Segment Graphs approach provides a formal declarative bottom-up mapping of file format structures to recognizable information elements, but was only implemented as a proof-of-concept for limited simple formats.[73] Algorithmic characterization remains the common approach, but while it is considered useful, most algorithms remain ultimately unverified.[74]

Do digital objects exist? Theory and practice in digital preservation will continue to talk about preserving digital objects, but we cannot preserve digital objects, because what we preserve does not actually have an existence in a tangible, manifest way that continues across time. This, of course, essentially corresponds to a key finding of InterPARES 1, that "empirically, it is not possible to preserve an electronic record: it is only possible to preserve the ability to reproduce the record."[75] However, two comments on this are in order. First, insufficient attention to the finer mechanisms of how these objects emerge through computational processes, and that some of their properties cannot be attributed solely to the components that are kept and safeguarded, has caused persistent misunderstandings in the archival and digital preservation literature

71 Wing and Ockerbloom, "Respectful Type Converters."

72 Manfred Thaller, *The EXtensible Characterisation Languages – XCL,* 1st ed. (Hamburg: Kovac, 2009).

73 Michael Hartle, Friedrich-Daniel Möller, Slaven Travar, Benno Kröger, and Max Mülhäuser, "Using Bitstream Segment Graphs for Complete Description of Data Format Instances," in *ICSOFT 2008 – Proceedings of the Third International Conference on Software and Data Technologies,* vol. ISDM/ABF, Porto, Portugal, 5–8 July, 2008 (Setubal, PT: INSTICC Press, 2008): 198–205, accessed 1 July 2017, http://dblp.org/db/conf/icsoft/icsoft2008-3; Michael Hartle, Max Mülhäuser, Daniel Schumann, Arsene Botchak, "A Logic-Based Approach to the Formal Specification of Data Formats," in *iPRES 2008 – Proceedings of the 5th International Conference on Preservation of Digital Objects* (London: iPRES 2008), 292, accessed 1 July 2017, https://phaidra.univie.ac.at/view/o:294183.

74 Christoph Becker, Kresimir Duretec, and Andreas Rauber, "The Challenge of Test Data Quality in Data Processing," *ACM Journal of Data and Information Quality (JDIQ)* 8, no. 2 (February 2017).

75 InterPARES 1 Preservation Task Force, "Preservation Task Force Report" (2002), accessed 1 July 2017, http://www.interpares.org/book/interpares_book_f_part3.pdf.

and practice. Second, we do not preserve the "ability" mentioned above either. It is in constant change, designed and redesigned, not preserved.[76]

Digital preservation is often conceived in terms of *using* software, but *to preserve digital objects* really means to *design sustainable systems* intended to *reproduce* what we then call digital objects. And we call these phenomena digital objects because "our ordinary conceptual system, in terms of which we both think and act, is fundamentally metaphorical in nature."[77]

Metaphors We Work By

As Lakoff and Johnson explain in their influential book *Metaphors We Live By*, "the primary function of metaphor is to provide a partial understanding of one kind of experience in terms of another kind of experience."[78] Consider that the argument laid out in this article was first articulated in a talk this author gave to an audience, whose members then had to decide whether there was something useful in the talk, perhaps an idea they could take away from it. The previous sentence contains multiple metaphors: ideas are objects, a presentation is an object, and language is a container of ideas. The takeaway, an idea, is conceived in the previous sentence as if it were something we can physically carry out of the room.

Metaphors thus constitute a "cross-domain mapping in the conceptual system"[79] that links concepts from a source domain to a target domain. Using many examples, Lakoff and Johnson illustrate their argument that through these mappings, the metaphorical structures that ground our conceptual system are ultimately based in experience:

76 A similar conclusion is drawn by Simone Sacchi in "What Do We Mean by 'Preserving Digital Information'? Towards Sound Conceptual Foundations for Digital Stewardship" (PhD dissertation, University of Illinois at Urbana-Champaign, 2015), 60, accessed 1 July 2017, https://academiccommons.columbia.edu/catalog/ac:185846. Sacchi argues that preservation is "a misleading metaphor" since everything involved in it is either an abstract concept ("for which preservation does not apply") or a material object (which will eventually be replaced in the process).

77 Lakoff and Johnson, *Metaphors We Live By,* 1.

78 Ibid., 154.

79 Lakoff, "The Contemporary Theory of Metaphor," 203.

> Ontological metaphors are grounded by virtue of *systematic correlates within our experience*. As we saw, for example, the metaphor THE VISUAL FIELD IS A CONTAINER is grounded in the correlation between what we see and a bounded physical space. The TIME IS A MOVING OBJECT metaphor is based on the correlation between an object moving toward us and the time it takes to get to us. The same correlation is a basis for the TIME IS A CONTAINER metaphor (as in "He did it in ten minutes"), with the bounded space traversed by the object correlated with the time the object takes to traverse it. Events and actions are correlated with bounded time spans, and this makes them CONTAINER OBJECTS.… we typically conceptualize the nonphysical in terms of the physical – that is, we conceptualize the less clearly delineated in terms of the more clearly delineated.[80]

Similarly, "spatialization metaphors are rooted in physical and cultural experience; they are not randomly assigned. A metaphor can serve as a vehicle for understanding a concept only by virtue of its experiential basis."[81] Note how this explanation in turn relies on the metaphor LEARNING AS A JOURNEY, which in turn relies on the metaphor PROCESS AS MOVEMENT. In learning, we walk the path of understanding, and a vehicle can provide external support for transportation, which we use to ease the journey. We thus comprehend one intangible experience – learning – through a tangible concept "rooted in physical and cultural experience,"[82] in this case, transport. The mapping comprises "sets of conceptual correspondences"[83] between source and target domain that allow us to leverage our knowledge about one to comprehend the other. We are generally unaware of the transparent nature of metaphor in this process of communication.

Lakoff and Johnson argue that far beyond daily life, "so-called purely intellectual concepts, e.g., the concepts in a scientific theory, are often – perhaps always – based on metaphors that have a physical and/or cultural basis."[84] This

80 Ibid., 58–59.

81 Ibid., 18.

82 Ibid.

83 Ibid., 207.

84 Ibid., 18–19.

cultural basis is in turn contingent on the norms of disciplines and communities of practice, and the correspondences and entailments brought forward using metaphorical concepts will vary accordingly. As Lakoff and Johnson explain, metaphors highlight parts of our experience in the cognitive and communicative process while masking others: "metaphors … have the power to define reality. They do this through a coherent network of entailments that highlight some features of reality and hide others."[85] Of course, metaphor is not the only aspect of language that introduces cultural contingency. Due to their nature as conceptual mappings across domains, however, metaphors play a particularly important role here.

Their role has been explored for a related area – digital curation – in a paper that initially prompted part of the argument expounded here. In it, M.A. Parsons and P.A. Fox discuss the use of the concept of *publication* to describe research data management and sharing practices and to question its adequacy.[86] The extensive entailments for the conceptual mappings inherent in the concept *digital objects* remain to be explored beyond this article, but a few suggestions arise immediately:

1. Digital objects "can be seen and touched."[87]
2. Digital objects can be damaged and thus need to be physically protected.
3. Digital objects can be edited.

Each of these suggestions contains misleading simplifications. For the first, even though information encoded in digital form of course requires a physical manifestation, in almost all cases it is in no way directly accessible to the human senses. The second focuses attention on the bits, but as noted above, physical integrity is not a necessary requirement for authentic records, and preservation actions such as migration do not preserve it. On the other hand, damage occurs often not as a loss of physical integrity, but as a loss of relationships between elements, whether through link rot, obsolescence, or lack of metadata.

85 Ibid., 157.

86 M.A. Parsons and P. A. Fox, "Is Data Publication the Right Metaphor?" *Data Science Journal* 12 (2013): WDS32–WDS46.

87 *Oxford English Dictionary,* s.v. "object," accessed 1 July 2017, https://en.oxforddictionaries.com/definition/object.

Finally, in a succinct argument, Allen H. Renear and Karen M. Wickett deconstruct the last of these entailments for the case of documents and show it to be false. We shall see that their conclusions hold for the digital object definitions discussed above. Consider these sentences:

- This is a sentence.
- This is not a version of that sentence.

Based on the assumption that documents are sets of strings, Renear and Wickett demonstrate that the editing of a string does not change the original string, an abstract object which still exists, nor does it transform it into the new string, which is an independent object that has no identity relationship to the first sentence.[88] Consider now the standard processes of computing, in which files are read and written by software such as word-processing software. We may "save" a document (note the metaphor) "as" a new file, but the bits in the new file are not a version of the bits in the file that we first "opened" (note the metaphor). These bits are not being "edited"; rather, each string is a new string of bits. Everything else is algorithmic. The document is "the same" by virtue of characteristics such as the text perceived by those who edit it through the on-screen projection. As discussed above, there is typically no readily expressible relationship that can formally verify whether an arbitrary stream of bits in one format is representing the same "object" as an arbitrary stream of bits in another format.

As Lakoff and Johnson explain, "the very systematicity that allows us to comprehend one aspect of a concept in terms of another … will necessarily hide other aspects of the concept. In allowing us to focus on one aspect of a concept … a metaphorical concept can keep us from focusing on other aspects of the concept that are inconsistent with that metaphor."[89]

It seems that this is what happened in the digital preservation discourse described above. Prior archival practice made it natural to overlook the computational aspects that we cannot easily understand based on our cultural experience. This is because computers are not just machines; they are universal

88 Allen H. Renear and Karen M. Wickett, "Documents Cannot Be Edited," in *Proceedings of Balisage: The Markup Conference, 11–14 August 2009, Montreal,* accessed 1 July 2017, http://www.balisage.net/Proceedings/vol3/print/Renear01/BalisageVol3-Renear01.html.

89 Lakoff and Johnson, *Metaphors We Live By,* 10.

machines. Many of the aspects of computational thinking are non-intuitive. Of course, they can be learned, above all through reflective and productive engagement with the practice of computing and systems design.[90] This opportunity also suggests an addition of computing and systems design principles, skills, and knowledge to the digital curation skills matrix.[91]

Conclusions

This article suggests that the conceptual mappings and correspondences employed in the metaphor *digital object* allow us to articulate boundaries around emergent properties of performances computed using data, software, and hardware. These performances and their properties are emergent, so it is simplistic and misleading to assume that individual properties can always be attributed to individual parts. This fact does not suggest, however, that it is futile to articulate what, in a given context, will make a digital object count as authentic, or that it is futile to try to develop automated measures based on such an articulation. In all this, the metaphorical entailments created by seeming correspondences to artifact-centric worldviews mask the computational nature of these "objects." What else is masked may be worth exploring in more detail. Such an exploration could be based in part on semiotic approaches such as the Sept model,[92] but it should incorporate a fuller appreciation of the role of computation in the process of (re)production and must distinguish the performance resulting from computation from the data it uses. The age of algorithms may bring to the fore the dynamic, distributed, computed nature of electronic records, but that nature is not fundamentally new. Electronic records were always computational products of algorithms that used data as input to create a performance. Early algorithms and formats were sometimes simple enough that the approximate conceptual mapping A FILE IS A CONTAINER (implying that the content would be solely contained in the file) was sufficiently accurate to work, and that the simplistic assumption THE FILE IS THE RECORD worked well enough (most of the time).

90 Patricia Galloway, "Educating for Digital Archiving through Studio Pedagogy, Sequential Case Studies, and Reflective Practice," *Archivaria* 72 (Fall 2011): 169–96.

91 Christopher A. Lee and Helen Tibbo, "Where's the Archivist in Digital Curation? Exploring the Possibilities through a Matrix of Knowledge and Skills," *Archivaria* 72 (Fall 2011): 123–68.

92 Abrams, "A Foundational Framework."

The development of standardized archival format specifications such as the PDF/A standard is a strategy to reduce the variability in computational processes so that the approximation can continue to hold. Ultimately, however, it is crucial to distinguish between (1) the reproduced performance, on the one hand, and how stakeholders attribute significance to its various facets and elements; and (2) the various means by which that performance is reproduced, and their relationships.[93] The terms *digital object* and *electronic record* cannot be used for both without risking continued confusion. This aligns with the Sept model's argument that consumers interact not with an *artifact*, exemplar or *product*, but with the *asset:* the product's realized performance.

It follows that *significant properties* are best understood as mechanisms that allow curators to specify shared understandings of what constitutes authentic reproductions of digital objects. A large amount of the disagreement in the debate above can be attributed to the fact that the participants in the debate have not effectively distinguished between the ends that are to be achieved and the means through which to achieve them – in conflating the two, the discussion self-destructs.

The role of significant properties is thus the articulation of requirements, what Joseph Goguen would call the "reconciliation between the social and the technical."[94] Digital archivists and digital curators need to be comfortable with this grey area where the technical and the social meet. Like an effective requirements statement, an effective specification of a significant property describes it independently of the solution space of bitstreams, formats, files, and software packages. This conclusion is consistent with David Giaretta's suggestion "that the function of Significant Properties … is the identification of 'those characteristics [technical, intellectual, and aesthetic] agreed by the archive or by the collection manager to be the most important features to preserve over time.'"[95]

In preservation planning, significant properties are a basis for the specification of criteria on which one can decide whether "a digital object is authentic" or, more precisely, whether a performance designed through a preservation action

93 See also Duranti and Thibodeau, "The Concept of Record."

94 Joseph Goguen, "Requirements Engineering as the Reconciliation of Technical and Social Issues," in *Requirements Engineering: Social and Technical Issues,* ed. Marina Jirotka and Joseph Goguen (San Diego, CA: Academic Press, 1994), 165–99.

95 David Giaretta, *Advanced Digital Preservation* (Berlin: Springer, 2011), 232. Note that this is the only definition that focuses on outcomes.

produces emergent phenomena that can be accepted as an authentic reproduction. The determination of specific properties amounts to a specification of success criteria that are used to evaluate how effective different preservation actions (means) are in enabling the reproduction of authentic digital objects (ends). This specification can be combined with criteria referring to the process by which these means are employed in order to facilitate a structured evaluation of available means to achieve the goals of preservation.[96]

Based on this curatorial attribution of significance, software can sometimes go to great lengths in supporting the tedious task of verification. For example, Stephan Bauer and Christoph Becker showed that for a given set of objects (born-digital photographs in raw formats), the entire set of properties deemed significant in a given context was expressible in measurable terms and measurable by algorithms.[97] It should also be noted that, in archival practice, some institutions have laid down policies to that effect. For example, the Parliamentary Archive of the UK's digital preservation policy states that the archive "will define the properties of any given record type which are significant to its authenticity"; it adds that the "authenticity of an electronic record derives from a set of quantifiable properties of that record" and states that "these properties are independent from any given technical representation of that record."[98] Through this statement, the policy effectively distinguishes success criteria that refer to significant properties from the various actions that operate on technical representations in the processes that preserve electronic records.

> "Metaphorical thought is unavoidable, ubiquitous, and mostly unconscious."[99]

96 Christoph Becker and Andreas Rauber, "Decision Criteria in Digital Preservation: What to Measure and How," *Journal of the American Society for Information Science and Technology* 62, no. 6 (1 June 2011): 1009–28.

97 Stephan Bauer and Christoph Becker, "Automated Preservation: The Case of Digital Raw Photographs," in Digital Libraries: *For Cultural Heritage, Knowledge Dissemination, and Future Creation,* ed. Chunxiao Xing, Fabio Crestani, and Andreas Rauber (Berlin: Springer, 2011): 39–49, doi:10.1007/978-3-642-24826-9_9.

98 Parliamentary Archives, Houses of Parliament (UK), "A Digital Preservation Policy for Parliament," 1st ed. (March 2009), accessed 1 July 2017, http://www.parliament.uk/documents/upload/digitalpreservationpolicy1.0.pdf.

99 Lakoff and Johnson, *Metaphors We Live By,* 272.

By unpacking the conceptual mapping created by the term *digital object*, this article has clarified the term's limits and highlighted aspects of the concept that it brings to the foreground and some aspects that the metaphor masks. The argument serves to illustrate that at the centre of digital preservation work is the design of software-intensive information systems intended to reproduce authentic digital objects. This means that the conceptual frame of *systems design* is an important part of archival education. The discussion above suggests the following initial set of focal areas:

Algorithms and *computational thinking*. Algorithms are constitutive of the nature of digital objects and electronic records. To enable archivists and digital curators to appreciate this nature, computational thinking[100] needs to be part of archival education and the digital curation curriculum. This is also highlighted by the emergent focus area of "computational archival science."[101]

Information systems architecture. Archivists and digital curators are keen to retain ownership and control of the software systems and services that implement their archival information system, rather than become disempowered subscribers to commercial services. To do so, they must possess a high-level understanding of their systems' architecture and the processes and issues involved in their design, maintenance, and evolution.

Systems thinking and systems design. In designing such socio-technical information systems, the designers must always bridge the social and the technical domains. Since neither can be understood in separation, integrative thinking is essential. The means–ends relationships in the design of such systems are complex, incompletely known and subject to delayed systemic interactions between the components of the resulting system and those in its environment. This is a key factor in the considerable difficulty that lies in formalizing, codifying

100 Jeannette M. Wing, "Computational Thinking," *Communications of the ACM* 49, no. 3 (March 2006): 33–35.

101 Richard Marciano, Victoria Louise Lemieux, Mark Hedges, Maria Esteva, William Underwood, Michael Kurtz, and Mark Conrad, "Archival Records and Training in the Age of Big Data," in *Advances in Librarianship – Re-Envisioning the MLIS: Perspectives on the Future of Library and Information Science Education,* ed. Lindsay C. Sarin, Johnna Percell, Paul T. Jaeger, and John Carlo Bertot, accessed 1 July 2017, http://dcicblog.umd.edu/cas/wp-content/uploads/sites/13/2016/05/Marciano_Kurtz_et-al-Archival-Records-and-Training-in-the-Age-of-Big-Data-final-1.pdf.

and teaching design expertise in general.[102] An awareness of the nature of collaborative design work is best acquired through experiential learning and reflective practice, recently highlighted as a key component of archival education in the digital age.[103]

Requirements are the key conceptual area in systems design that provide the conceptual framework and methods to elicit, model, negotiate, and specify concerns, needs, goals, objectives, and success criteria for the development of archival information systems.

> "A systems approach begins when first you see the world through the eyes of another."[104]

Systems thinking and systems design perspectives are not functional skills that should be seen as an addition to a set of elective courses; they provide coherent foundations for archival information systems. Archival educators need to teach systems design and requirements analysis to archivists from the ground up to ensure that the next generation of professionals are well versed in the conceptual frameworks and methods they can use to express relationships between means and ends, reconcile the social with the technical, and connect their curatorial, archival perspective with specialist computing knowledge in the design of archival information systems for the age of algorithms.

102 Donald A. Schön, *Educating the Reflective Practitioner: Toward a New Design for Teaching and Learning in the Professions* (San Francisco: Jossey-Bass, 1987).

103 Galloway, "Educating for Digital Archiving."

104 C. West Churchman, *The Systems Approach* (New York: Dell Publishing, 1968), 231.

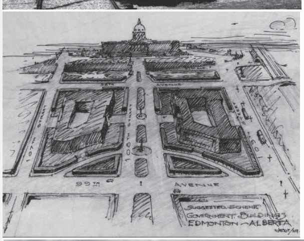
ACNOWLEDGEMENTS I am grateful for discussions on significant properties with collaborators in the PLANETS and SCAPE projects and with students of digital preservation at the University of Toronto. Conversations with Natasa Milic-Frayling, Stephen Abrams, Simone Sacchi, and Emily Maemura have kept the discussion alive. Thanks to several participants of the Archival Education and Research Institute (AERI 2017) for the encouragement to develop an AERI talk into this article. Special thanks to Jess Whyte for her excellent support in preparing the manuscript; to Heather MacNeil and Fiorella Foscarini for insightful comments; and to the anonymous reviewers for their thought-provoking critique and helpful suggestions.

BIOGRAPHY Christoph Becker is an assistant professor at the Faculty of Information at the University of Toronto where he teaches digital preservation to future archivists and other information professionals. He completed his Doctorate in Computer Science at the Faculty of Informatics of the Vienna University of Technology in Austria in 2010, with a thesis on trustworthy preservation planning. He has published widely in the fields of digital curation and preservation, software systems, and digital libraries. As part of the PLANETS and SCAPE projects funded by the European Commission, he led the development of the PLATO preservation planning tool, an award-winning decision support system for digital preservation, in collaboration with an international consortium of universities, national archives, cultural heritage organizations, and commercial partners. Through his research, he further developed open methods to evaluate digital preservation processes and capabilities. As Director of the Digital Curation Institute at the University of Toronto (dci.ischool.utoronto.ca), he brings together graduate students, appointed fellows, faculty colleagues, and partners to conduct research at the intersection of digital curation and systems design, supported by grant-funded state-of-the-art facilities. As a co-founder of the initiative that produced the Karlskrona Manifesto for Sustainability Design (www.sustainabilitydesign.org), he advocates a new interdisciplinary approach to software systems research that emphasizes long-term perspectives on socio-technical systems design. For more information, see https://orcid.org/0000-0002-8364-0593.

223 Colonnade Road South, Ottawa, ON K2E 7K3 phone 613-234-3977 fax 613 234-8500
email aca@archivists.ca www.archivists.ca

Association of Canadian Archivists
Association canadienne des archivistes

Truths, Trust and Technology
June 6 – 9, 2018
Edmonton, AB



The Association of Canadian Archivists invites you to attend the 43rd Annual Conference "Truths, Trust and Technology," June 6th – 9th, 2018 in Edmonton, Alberta. Archivists from far and wide will engage with the theme truths, trust and technology. The truth – whatever that is – is messy these days. In the "post-truth" era, lies can seem more convincing than facts, and in an era of discourse on decolonization, Canadians are reflecting on new truths about historical facts.
"Truths" become "lies" depending on context and interpretative lens. The very existence of fact may even be open to question. Contributing to the messy truth is technology. Amidst this technological and social backdrop, we can be certain that archivists must sustain their role as trusted custodians of the historical record. How might archives and archivists need to adapt? What new approaches, competencies and skills will be needed? How can archivists help people to realize the enriched insights into historical truths that knowledge of context and provenance offers?

Where does "truth" lie for archives and for humanity?

Conference sessions and activities are based in and around the **Chateau Lacombe Hotel**, centrally located in Edmonton.
While in Edmonton participants will be able to:

- Take a ride an historic Edmonton street car over the iconic High Level Bridge
- Tour BioWare–Electronic Arts
- Be part of the East–West softball game
- Enjoy a walking tour of historic 104 Street and the Neon Museum

Edmonton extends a warm welcome to all archivists, national and international. Pre-conference activities, workshops, SIS meetings, conference information and registration are available at the ACA's 2018 website:
https://archivists.ca/content/annual-conference