Toward a "Third Order" Archival Interface: Research Notes on Some Theoretical and Practical Implications of Visual Explorations in the Canadian Context of Financial Electronic Records[1]

VICTORIA L. LEMIEUX

RÉSUMÉ Ce texte aborde les défis liés à l'abstraction et aux représentations des documents d'archives et participe au débat sur ce sujet qui existe dans la littérature scientifique par un certain nombre de ses contributions théoriques et pratiques. Présentant les résultats d'un projet de recherche multidisciplinaire visant la création d'un modèle de référence interactif de haut niveau portant sur le contexte canadien des documents d'archives financiers numériques, il avance un cadre théorique au sujet du contexte sociétal comme ontologie de domaine et il fournit une approche pour établir les frontières du contexte sociétal. Il s'inspire également de la théorie sur les systèmes d'information, en particulier sur la théorie des représentations, afin d'élaborer la théorie des documents d'archives comme représentations. Il poursuit en abordant des expériences visant à développer un prototype d'une représentation visuelle interactive d'une ontologie de domaine du contexte canadien des documents d'archives financiers numériques, suggérant que les représentations visuelles interactives qui combinent des caractéristiques des éditeurs et concepteurs d'ontologie avec des caractéristiques d'outils pour l'analyse visuelle peuvent fournir des bases solides pour des interfaces d'archives de « troisième ordre ».

ABSTRACT This paper addresses challenges related to abstraction and representation of archival records and makes a number of theoretical and practical contributions to discussions in the archival literature on this topic. Reporting on an interdisciplinary research project aimed at creating a high-level interactive reference model of the Canadian context of financial electronic records, it contributes a framework for theorizing about societal context as a domain ontology and an approach to establishing the boundaries of societal context. It also draws upon information systems theory, in particular representation theory, to extend the theory of records as representations. It then moves on to discuss experiments in developing a prototype interactive visual representation of a domain ontology of the Canadian context of financial electronic

1 This research was funded by a University of British Columbia Hampton Research Grant. The author would like to thank Giovanni Michetti, School of Library, Archival and Information Studies, the University of British Columbia, and two anonymous reviewers for their suggestions to improve this paper. Any errors and omissions remain the sole responsibility of the author.

records, suggesting that interactive visual representations that combine features of ontology editors and builders with features of tools for visual analysis may provide a good foundation for "third order" archival interfaces.

"…in the third order of order, knowledge doesn't
have *a* shape. There are just too many … ways to make
sense of our world."
– David Weinberger[2]

"New knowledge today materializes at the boundaries
of existing disciplines…."
– Kapil Sibal[3]

Introduction

As a means of engaging with archival discourse on the theme of abstraction and representation of records and archives, this paper reports on aspects of a collaborative interdisciplinary research project, carried out by a research team at the University of British Columbia from 2009 to 2011, that involved experimentation with what might be described as creation of a "third order" archival interface (i.e., a system in which users can easily arrange archival resources into as many different aggregations and sequences as desired according to their needs). The overall goal of the research was to create a high-level, interactive reference model of the Canadian context of financial electronic records.[4] The project was motivated by two questions: What is the current context of financial record creation and recordkeeping in Canada, and how can we best represent it? This led the researchers to integrate archival theory and practice with information systems theory and practice, thereby reconceptualizing the abstraction and representation of societal provenance as an exercise in building and representing a domain ontology of the context of financial electronic records in Canada. The project has led to a number of theoretical and practical contributions in

2 David Weinberger, *Everything is Miscellaneous: The Power of the New Digital Disorder* (New York: Times Books, 2007), 83. Weinberger proposes three orders: The first order is the order of physical things – fonds arranged on shelves in order of their provenance, for example. The second order is the catalogue order, in which it is possible to have multiple descriptive representations of an entity that resides in one physical place. Weinberger's example is the card catalogue. The third order is the digital order, in which there is no limit to the number of possible orderings.

3 Kapil Sibal, "Preface," in *The Road to Academic Excellence: The Making of World-Class Research Universities*, ed. Philip G. Altbach and Jamil Salmi (Washington, DC: The International Bank for Reconstruction and Development/The World Bank, 2011), xiv.

4 For the original inspiration for creating a high-level contextual reference model, see EDRM: Creating Practical Resources to Improve E-Discovery & Information Governance, "Electronic Discovery Reference Model" (2009), under "Frameworks: EDRM Stages," accessed 20 May 2013, http://www.edrm.net/resources/edrm-stages-explained.

relation to questions of abstraction and representation of records and archives and of the context of their creation. These points are developed in Parts II and III and can be summarized as follows:

Theoretical

1. Use and demonstration of upper-level ontologies to provide a theoretical basis for the conceptualizing of societal provenance.
2. Combined application of Bunge's upper-level "substantive" ontology[5] and Searle's upper-level social ontology[6] to provide greater semantic expressiveness as the basis for the development of archival domain ontologies.
3. Application of systems and network theory to provide a theoretical basis for determining the boundaries of a domain ontology and for abstractions and representations of societal provenance based on ontology theory.
4. Use of upper-level ontologies to extend conceptualizations of records as representations to arrive at novel theoretical insights into the "nature" of records and theoretical integration of archival theory with information systems theory.

Practical

5. Demonstration of a manual top-down methodology for building a domain ontology to represent societal provenance, using financial records as an example.
6. Application of theories of visual cognition and perception, combined with ontology theory, to arrive at representations of archival records and their context that potentially strike a balance between parsimoniousness and expressiveness in archival representations.
7. Demonstration of new theoretical understandings of societal provenance and of records as alternate ways of visually representing archival records and their provenance.
8. Development of a prototype third-order representation of the Canadian context of financial electronic records.
9. Discovery that the combination of functionality found in software applications designed for ontology representation and the functionality

5 Yair Wand and Ron Weber, "Mario Bunge's Ontology as a Formal Foundation for Information Systems Concepts," in *Studies on Mario Bunge's Treatise*, ed. Paul Weingartner and Georg J.W. Dorn (Atlanta: Rodopi, 1990), 123–43.

6 John R. Searle, "Social Ontology: Some Basic Principles," *Anthropological Theory* 6, no. 1 (March 2006): 12–29.

found in software applications designed for visual analysis of data provides helpful clues about how to design third-order archival systems.

Before discussing the theoretical and practical contributions of the research as noted above, it is worthwhile to review some of the challenges in archival abstraction and representation that the research sought to address.

Part I: Challenges in Archival Abstraction and Representation

Daily, archivists face the daunting task of researching and representing the rich and varied context of archival fonds. This is both part and parcel of the process of respecting the fonds and of the process of communicating with end users about particular fonds. As such, we may conceptually situate the archival function of arrangement and description within the broader context of knowledge representation.[7] Indeed, as Katie Shilton and Ramesh Srinivasan note in their article on participatory appraisal and arrangement for multicultural archival collections, "the 'power to represent' has been wielded by information institutions throughout history."[8] What is knowledge representation? Elizabeth Yakel draws on the work of Elin K. Jacob and Deborah Shaw, who describe representation as:

> the process or activity of representing.… The process of representing seeks to establish systematic correspondence between the target domain and the modelling domain and to capture or "re-present," through the medium of the modelling domain, the object, the data, the information in the modelling domain. To the extent that this re-representation corresponds to, or models, the object, data or information in the target domain the two can be thought of as representationally one.[9]

Computer scientists Randall Davis, Howard Shrobe, and Peter Szolovits suggest that the notion of representation can best be understood in terms of the five distinct roles it plays:

- First, a knowledge representation is most fundamentally a *surrogate*, a substitute for the thing itself, that is used to enable an entity to determine consequences by thinking rather than acting, that is, by reasoning about the world rather than taking action in it.
- Second, it is a set of ontological commitments, that is, an answer to the question, In what terms should I think about the world?

7 Elizabeth Yakel, "Archival Representation," *Archival Science* 3, no. 1 (January 2003): 1–25.

8 Katie Shilton and Ramesh Srinivasan, "Participatory Appraisal and Arrangement for Multicultural Archival Collections," *Archivaria* 63 (Spring 2007): 88.

9 Elin K. Jacob and Debora Shaw, "Sociocognitive Perspectives on Representation," *Annual Review of Information Science and Technology* 33 (1998): 146, quoted in Yakel, "Archival Representation," 1–2.

- Third, it is a fragmentary theory of intelligent reasoning expressed in terms of three components: (1) the representation's fundamental conception of intelligent reasoning, (2) the set of inferences that the representation sanctions, and (3) the set of inferences that it recommends.
- Fourth, it is a medium for pragmatically efficient computation, that is, the computational environment in which thinking is accomplished. One contribution to this pragmatic efficiency is supplied by the guidance that a representation provides for organizing information to facilitate making the recommended inferences.
- Fifth, it is a medium of human expression, that is, a language in which we say things about the world.[10]

Knowledge representation entails the process of abstraction. This process may be explained as one that groups specific instances of entities in a domain of analysis into semantically meaningful classes.[11] In software engineering, it may involve conceptual modelling wherein particular entities (e.g., functions or data instances) are grouped into aggregate classes in order to, for example, reduce duplication in the writing of code. In science in general, the process of abstraction allows us to make inferences about things, such as: if instance *x* is part of class *y*, then *x* is a type of *y*. Based on this inference, we may then attribute a number of features of *y* to *x* as well. In relation to records, recordkeepers use this process to infer the semantics of records, the contents of which do not in themselves always provide full information about the context of their creation. For instance, if we know that *x* is a file with a name on it – H. Jenkinson – and a code *y*, which represents the class of functions having to do with Human Resources Management, we may infer that *x* is H. Jenkinson's personnel file.

Within the archival literature, there has been a persistent strand of discussion regarding the limitations of current approaches to abstracting and representing our knowledge of records and archives. The literature on this theme may be grouped into two broad issues: 1) the issue of abstraction, that is, how well a given abstraction represents provenance or the context of creation; and 2) the issue of representation, that is, whether a representation is a good representation of our abstraction of provenance or the context of creation. Though we may treat these as two separate issues for the purposes of analysis, in the literature and in reality they are often intertwined. Our abstractions shape how we represent, and our representations shape and sometimes constrain our abstractions.

10 Randall Davis, Howard Shrobe, and Peter Szolovits, "What Is a Knowledge Representation?" *AI Magazine* 14, no. 1 (Spring 1993): 17.

11 Benjamin C. Pierce, *Types and Programming Languages* (Cambridge, MA: MIT Press, 2002).

As a foundation principle of the field of archives, the concept of provenance needs no introduction to the readers of this journal, but a brief description of its traditional dimensions will serve to frame the discussion that follows. Provenance is often defined as "the organization or individual that created, accumulated and/or maintained and used RECORDS in the conduct of business prior to their transfer to a RECORDS CENTRE or ARCHIVES."[12] The *General International Standard Archival Description (ISAD(G))* and the *International Standard Archival Authority Record for Corporate Bodies, Persons and Families (ISAAR (CPF))* define provenance as the relationship (in *ISAD(G)*) or relationships (in *ISAAR (CPF)*) "between records and the organizations or individuals that created, accumulated and/or maintained and used them in the conduct of personal or corporate activity."[13] The principle of provenance holds that archival documents can only be understood in context, or in relation to their origins and to other documents, not as self-contained, independent items.[14] In order to understand the provenance of a particular body of archival documents, it is necessary to conduct systematic research into the origins of their creators, the purposes for which the records were created, and the organic and interrelated characteristics of the documentation, often constructed as "the context of the records," or contextualization of the records.[15] Within this framework of abstraction, the standard constructs have come to be the fonds or group, the sous-fonds or sub-group, the series, the file, and the item.[16] Following a Weberian logic, archivists have tended to order these constructs hierarchically.[17]

12 ICA Dictionary of Archival Terminology, "DAT III: Draft, English List," accessed 25 October 2012, http://www.staff.uni-marburg.de/~mennehar/datiii/intro.htm (capitalization in original). It is important to emphasize that provenance was not initially conceived as an abstraction, or intellectual construct, but was understood more concretely as physical aggregations of records (see Jennifer Douglas, "Origins: Evolving Ideas about the Principle of Provenance," in *Currents of Archival Thinking*, ed. Terry Eastwood and Heather MacNeil (Santa Barbara, CA.: Libraries Unlimited, 2010), 23–43.

13 International Council on Archives, *ISAAR (CPF): International Standard Archival Authority Record for Corporate Bodies, Persons and Families*, 2nd ed. (Paris: ICA, 2004), accessed 20 May 2013, http://www.icacds.org.uk/eng/ISAAR(CPF)2ed.pdf; International Council on Archives, *ISAD(G): General International Standard Archival Description*, 2nd ed. (Ottawa: ICA, 2000), accessed 20 May 2013, http://www.ica.org/?lid=10207.

14 See Douglas, "Origins."As Douglas notes, there has been some debate about whether the principle of provenance includes both the principles of respect des fonds and respect for original order as subprinciples. She follows Peter Horsman in stating that provenance is the only principle in archival theory, and this is also the position adopted in this paper for the purposes of discussion.

15 See, for example, Tom Nesmith, "Reopening Archives: Bringing New Contextualities into Archival Theory and Practice," *Archivaria* 60 (Fall 2005): 259–74.

16 Douglas, "Origins."

17 See Giovanni Michetti, "Archives Are Not Trees: Hierarchical Representations in Digital Environments," *Proceedings of the UNESCO Conference The Memory of the World in the Digital Age: Digitization and Preservation*, ed. Luciana Duranti and Elizabeth Shaffer (Vancouver, 26–28 September, 2012), 1002–10, accessed 20 May 2013, http://www.unesco

Records, which we can link to the concept of item-level archival abstraction, traditionally have been viewed (abstracted) as objects existing *within* a context.[18] There is, however, a tradition of also viewing records, or archival items, as *representing* that context. In this conceptualization of records, they are both an abstraction and a representation at one and the same time. Indicative of this viewpoint is the old archival adage "form follows function," which is understood to mean that archival records represent, and provide evidence of, the functions that they are created to support.[19] Diplomatics also views records as representations of their context: diplomatic analysis proceeds on the basis of identifying elements of the content and form of records that signify or reveal aspects of the context of their creation.[20] Recently, Geoffrey Yeo has articulated a view of records as "representations of occurrent," using the term "occurrent" to refer to activities, steps, processes, functions, and other related phenomena.[21] In spite of accepting that the content or form of records can represent the context of their creation, all of these theories acknowledge that there are limitations to what can be known about the context of records from the content or form of the records themselves – hence the need for archival analysis of the provenance of records that draws on other sources as well. This is also why "functional" hierarchical classification has been an important archival activity. Recordkeepers use this process to infer the semantics of records. In essence, archival analysis of provenance and records classification aims to fill a semantic gap.

Archivists have long felt some disquiet with traditional notions of the archival principle of provenance and hierarchical orderings of archival constructs. Jennifer Douglas notes that "much of modern archivists' criticism and discus-

.org/new/en/communication-and-information/events/calendar-of-events/events-websites/the-memory-of-the-world-in-the-digital-age-digitization-and-preservation/.

18 See, for example, Barbara Reed, "Records," in *Archives: Recordkeeping in Society*, ed. Sue McKemmish, Michael Piggott, Barbara Reed, and Frank Upward (Wagga Wagga, NSW: Charles Sturt University, Centre for Information Studies, 2005), 101–30.

19 For a discussion of this, see David Bearman, "Structural Formalisms in Documentation: Reflecting Function and Supporting Meaning," in *Controlling the Past: Documenting Society and Institutions; Essays in Honor of Helen Willa Samuels*, ed. Terry Cook (Chicago: Society of American Archivists, 2011), 241–56.

20 See Luciana Duranti, *Diplomatics: New Uses for an Old Science* (Lanham, MD: Scarecrow Press, 1998).

21 Geoffrey Yeo, "Concepts of Record (1): Evidence, Information, and Persistent Representations, *American Archivist* 70, no. 2 (2007): 315–43; Yeo, "Concepts of Record (2): Prototypes and Boundary Objects," *American Archivist* 71, no. 1 (2008): 118–43. In connection with Yeo's use of the term "occurrent," he writes that defining records as representations of activities may "preclude recognition of the notions that elementary records often represent steps *within* activities and that aggregations of elementary records can constitute records at higher levels." He therefore chooses to use the word "occurrent" to represent not just activities, but steps, processes, functions, and "other such phenomena." (See Yeo, "Concepts of Record (2)," 136). I return to a discussion of Yeo's concept of "record" in Part II of this paper.

sion of the principle of provenance (and, therefore, of its subprinciples respect des fonds and respect for original order) has focused on how to effectively represent the fluid and changing nature of both the external and internal structure of provenance of archival aggregations."[22] As early as the 1950s, Peter J. Scott advanced the notion of what came to be called the "series system" as a reaction to the limitations he found in traditional approaches to expressing the rich complexity of archival provenance.[23] Scott's series system advocated the adoption of the series as the primary locus of intellectual control and description and the use of authority records to link series with as many records creators as warranted. Though not abandoning the traditional archival constructs, Michel Duchein, in his 1983 article "Theoretical Principles and Practical Problems of *Respect des fonds* in Archival Science," noted that the practical application of the principle gave rise to many theoretical difficulties (e.g., how to ascribe provenance in cases when records creators contributed collectively to the creation of records in a shared database system).[24] At the time that the Canadian *Rules for Archival Description* were under development, Debra Barr emphasized that an archival fonds should be viewed as an abstraction rather than a physical entity, as it had been up to that point, and she also criticized the conceptualization of the fonds proposed for the rules as being too reductionist, stating that "Respecting provenance means reflecting more than one aspect of the complex history of many records."[25] Terry Cook also advised archivists to view archival fonds as primarily intellectual constructs, thereby clearly establishing a distinction between aggregations as groupings of physical records and aggregations as abstractions of records.[26] Cook further observed that the fonds is created through description of *relationships* (e.g., between records and creators and records and functions) and that provenance lay "at the heart" of these relationships.[27]

Greg Bak has argued recently that the item is the most important level of abstraction for archival representations and that we should abandon efforts to

22 Douglas, "Origins," 29.

23 See Peter J. Scott, "The Record Group Concept: A Case for Abandonment," *American Archivist* 29, no. 4 (October 1966): 493–504; *The Arrangement and Description of Archives amid Administrative and Technological Change: Essays and Reflections by and about Peter J. Scott*, ed. Adrian Cunningham (Brisbane, QLD: Australian Society of Archivists, 2010).

24 Michel Duchein, "Theoretical Principles and Practical Problems of *Respect des Fonds* in Archival Science," *Archivaria* 16 (Summer 1983): 64–82.

25 Debra Barr, "The *Fonds* Concept in the Working Group on Archival Descriptive Standards Report," *Archivaria* 25 (Winter 1987–88): 168.

26 Terry Cook, "The Concept of the Archival Fonds: Theory, Description, and Provenance in the Post-Custodial Era," in *The Archival Fonds: From Theory to Practice*, ed. Terry Eastwood (Ottawa: Bureau of Canadian Archivists, 1992), 52–64.

27 Terry Cook, "Mind Over Matter: Towards a New Theory of Archival Appraisal," in *The Archival Imagination: Essays in Honour of Hugh A. Taylor*, ed. Barbara L. Craig (Ottawa: Association of Canadian Archivists, 1992), 66.

group items into functional classes, which, he argues, better serve recordkeepers than end users.[28] Bak also strongly rejects the ordering of archival constructs (e.g., items, series, sous-fonds) into hierarchies. Bak's views reflect a growing criticism among archival theorists of the hierarchy. Chris Hurley also has been a critic of hierarchical arrangements, writing that the "hierarchical structure fixes the portrayal of the records into a single view, strangling the depiction of changing relationships between descriptive entities."[29]

In "Bringing Things Together: Aggregate Records in a Digital Age," Geoffrey Yeo asks whether the folder model and the classification scheme still deserve the primacy they have long enjoyed in records management and whether archivists can still cling to the beliefs that fixed hierarchies of files and series can represent provenance adequately.[30] In a thoughtful exegesis, Yeo ultimately concludes that the answer to these questions is no. And, more recently still, Giovanni Michetti attempts to drive a stake through the heart of hierarchical representation.[31] Michetti observes that recordkeepers have so internalized the tree-like hierarchical approach that they have come to view archives *as* trees (here we are reminded of Bruno Latour's "black box," in which the social relations that construct the artifact become so embedded as to be made invisible[32]). Michetti reminds us that "archives are not trees."[33]

The acts of abstracting and representing are tightly coupled; while abstraction is the process of abstracting away from particular instances of things into classes according to some logic, representation is the form in which we choose to represent the abstraction. In software engineering, that form may be a Uniform Modelling Language (UML) diagram or some other form of notation. In the field of archives, it takes the form of archival description. Many archivists have criticized the way we represent the archives as reductionist. Among them

28 Greg Bak, "Continuous Classification: Capturing Dynamic Relationships among Information Resources," *Archival Science* 12, no. 3 (September 2012): 287–318.

29 Chris Hurley, "The Making and Keeping of Records: (2) The Tyranny of Listing" (Chris Hurley, 17 Jan. 2000): 3, accessed 20 May 2013, http://www.descriptionguy.com/images/WEBSITE/the-making-and-the-keeping-of-records-part2.pdf.

30 Geoffrey Yeo, "Bringing Things Together: Aggregate Records in a Digital Age," *Archivaria* 74 (Fall 2012): 43–91.

31 Michetti, "Archives Are Not Trees."

32 Bruno Latour, *Science in Action: How to Follow Scientists and Engineers Through Society* (Cambridge, MA: Harvard University Press, 1987).

33 Michetti's thesis reflects, and perhaps draws upon, the classic argument by Alfred Korzybski that "the map is not the territory," as a metaphor for the notion that an abstraction of something is not the thing itself and a warning not to confuse models with reality. See Alfred Korzybski, "A Non-Aristotelian System and Its Necessity for Rigour in Mathematics and Physics," presented at the American Mathematical Society meeting of the American Association for the Advancement of Science, New Orleans, 28 December 1931; reprinted in Alfred Korzybski, *Science and Sanity: An Introduction to Non-Aristotelian Systems and General Semantics*, 5th ed. (Brooklyn, NY: Institute of General Semantics, 1995), 747–61.

is Laura Millar, who has argued that the *Rules for Archival Description* establish an approach to the fonds that does not allow for the description of virtual bodies of records, accumulated over time and scattered over space.[34] Tom Nesmith is also critical of the reductionism inherent in current forms of archival representation. In his paper "Reopening Archives: Bringing New Contextualities into Archival Theory and Practice," Nesmith calls for archival description to be thought of as "the action mediated by archivists of researching and representing the multi-faceted contextuality (or history of records or 'archival narrative' about them) which enables records and knowledge to be made through archiving."[35] He conceives of this as being a "general overlay to any descriptive system, a series of essays on the approach to description taken by the system/archives and the nature of the contextual information found in it, *and* not necessarily found there."[36] He illustrates this approach in his article "The Concept of Societal Provenance and Records of Nineteenth-Century Aboriginal-European Relations in Western Canada: Implications for Archival Theory and Practice," which is, itself, an essay exploring the different contextualities of records documenting Aboriginal-European relations in Western Canada.[37]

Where abstraction and representation are concerned, there is often a trade-off between parsimoniousness and expressiveness.[38] Parsimoniousness is the characteristic of being able to produce a model of (perceived) reality that represents the classes in such a way that relevant details about particular instances are not lost, yet not providing so much detail as to confuse or lose sight of the essence of the domain of interest. To elaborate, it is not necessary to know the distance between Canary Wharf and Notting Hill to take the tube across London. Tourists use a flat Underground map, a model that is just complex enough for the problem at hand. Expressiveness, on the other hand, is the inclusion of enough detail and richness in the model to fully represent a domain of interest without reductionism. For example, if an artist paints a portrait of someone as a red blob, the blob can serve as a representation of the person, but there *may not* be enough expressiveness in the blob for someone

34 Laura Millar, "The Death of the Fonds and the Resurrection of Provenance: Archival Context in Space and Time," *Archivaria* 53 (Spring 2002): 1–15. See also Peter Botticelli, "Records Appraisal in Network Organizations," *Archivaria* 49 (Spring 2000): 161–91.

35 Nesmith, "Reopening Archives," 270.

36 Ibid., 271.

37 Tom Nesmith, "The Concept of Societal Provenance and Records of Nineteenth-Century Aboriginal-European Relations in Western Canada: Implications for Archival Theory and Practice," *Archival Science* 6, no. 3–4 (December 2006): 351–60.

38 Xavier Gabaix and David Laibson, "The Seven Properties of Good Models" (paper presented at NYU Methodology Conference, accessed 25 October 2012, http://scholar.harvard.edu/laibson/publications/seven-properties-good-models), published in *The Foundations of Positive and Normative Economics: A Handbook,* ed. A. Caplin and A. Schotter (Oxford: Oxford University Press, 2008).

else to identify the person (assuming that that is what is important about the painting in the first place).[39] Returning to archival representations, while Nesmith's approach would yield high expressiveness, it is doubtless lacking in parsimoniousness.[40] Given this limitation, how might we represent archival context in a manner that strikes the right balance between parsimoniousness and expressiveness? Many initiatives, such as using metadata to represent the complex reality of the context of records, have been pursued,[41] yet the volume of metadata needed to capture even a shadow of the complex reality is daunting. Some, like Heather MacNeil, have suggested that such approaches would result in systems collapsing under their own weight.[42] Drawing upon reports by Margaret Hedstrom[43] and David A. Wallace,[44] Nesmith notes that the "2000 metadata forum in Holland drove home this type of observation by concluding that such metadata work is so complex, seemingly open-ended, and impeded by cost and other practical questions … that it requires further major research initiatives to pursue responses to these problems."[45] Add to this the problem of representing records-creator metadata, and the issue of achieving a parsimonious representation is further compounded. At the other end of the spectrum, some have sought parsimoniousness by way of representing the fonds as a hierarchical node-link diagram that visually conveys a hierarchically ordered relationship between fonds, sous-fonds, series, files, and items.[46] Although expressively more efficient than a series of essays, the diagram still suffers from the same

39 I italicize the words "may not" in this example because use of a red blob could be very expressive if a red blob is always used to signify a particular person. In this case, we would be able to identify the person easily.

40 To fairly characterize Nesmith, he does suggest that the proposed essays on the contextualities of records should be linked to higher-level and more summarized descriptions of archival context. See Nesmith, "Reopening Archives."

41 Wendy M. Duff, "Evaluating Metadata on a Metalevel," *Archival Science* 1, no. 3 (January 2001): 285–94. For more on the limits of means of representation, see Wendy M. Duff and Verne Harris, "Stories and Names: Archival Description as Narrating Records and Constructing Meanings," *Archival Science* 2, no. 3–4 (January 2002): 263–85.

42 See Heather MacNeil, review of *Holding On to Reality: The Nature of Information at the Turn of the Millennium*, by Albert Borgmann, *Archival Science* 3, no. 1 (January 2003): 67–74. For her discussion of these insights in relation to the findings of the InterPARES Project, see also Heather MacNeil, "Providing Grounds for Trust II: The Findings of the Authenticity Task Force of InterPARES," *Archivaria* 54 (Fall 2002): 24–58.

43 Margaret Hedstrom, "Recordkeeping Metadata: Presenting the Results of a Working Meeting," *Archival Science* 1, no. 3 (January 2001): 243–51.

44 David A. Wallace, "Archiving Metadata Forum: Report from the Recordkeeping Metadata Working Meeting, June 2000," *Archival Science* 1, no. 3 (January 2001): 253–69.

45 Nesmith, "Reopening Archives," 269.

46 For an example of visual representations of fonds, see Library and Archives Canada, "Archives Search" (in this example, Department of Indian Affairs and Northern Development Fonds), "Arrangement Structure," accessed 19 March 2012, http://collectionscanada.gc.ca/pam_archives/index.php?fuseaction=genitem.displayHierarchy&lang=eng&rec_nbr=135001&back_url=()&&back_url=().

reductionism that critics of hierarchical representations have drawn to our attention. Parsimonious, yet sufficiently expressive, approaches to representing archival contextual complexity remain elusive.

In light of these issues with abstraction and representation in the archival domain, several archivists have begun to call for different abstractions and new, flexible, and fluid forms of archival representation. Wendy Duff and Verne Harris call for "a user-friendly descriptive architecture – or at least interface with it – that eloquently represents relationships and contextual information in a clear, understandable fashion."[47] Greg Bak suggests, in his recent article "Continuous Classification: Capturing Dynamic Relationships among Information Resources," that records users should be given the freedom to use their own tools to determine how they aggregate records and conduct "resource discovery."[48] Elizabeth Yakel says that "archivists should begin to think less in terms of a single, definitive, static arrangement and description process, but rather in terms of continuous, relative, fluid arrangements and descriptions…"[49] Yeo, in turn, calls for the creation of third-order systems,[50] following Weinberger,[51] in which resources can be arranged into as many sequences as desired by users, independently of the limitations imposed by analog systems.

Part II: Abstraction

Given the challenges associated with archival abstraction and representation discussed in the previous section, we decided to explore outside the field of archives in our search for different and more expressive ways to abstract and represent archival records. Specifically, we chose to experiment with the use of ontology theory in developing abstractions of societal provenance. An ontology is a form of knowledge representation about a domain of interest – in our case, the Canadian context of financial electronic records. Natalaya F. Noy and Deborah L. McGuinness, following Gruber, define ontology as an explicit formal specification of the terms in the domain and relations among them.[52] Similarly, Studer, Benjamins, and Fensel define ontology as "a formal, explicit

47 Duff and Harris, "Stories and Names," 274.
48 Bak, "Continuous Classification," 314–15.
49 Yakel, "Archival Representation," 4.
50 Yeo, "Bringing Things Together."
51 Weinberger, *Everything is Miscellaneous*.
52 Natalya F. Noy and Deborah L. McGuinness, "Ontology Development 101: A Guide to Creating Your First Ontology" (Stanford, CA: Stanford University Knowledge Systems Laboratory, 2001), accessed 25 October 2012, http://www.ksl.stanford.edu/people/dlm/papers/ontology101/ontology101-noy-mcguinness.html; Thomas R. Gruber, "A Translation Approach to Portable Ontology Specifications," Knowledge Acquisition 5, no. 2 (June 1993): 199–220.

specification of a shared conceptualisation."[53] Uschold and Gruninger explain that ontology is a term used to refer to the shared understanding of some domain of interest that may be used as a unifying framework, and that it necessarily entails or embodies some sort of world view with respect to a given domain.[54] The World Wide Web Consortium (W3C) defines an ontology as a technical artifact that is used to describe and represent an area of knowledge; ontologies are used by people, databases, and applications that need to share domain information.[55] Ontologies include computer-usable definitions of basic concepts in the domain and the relationships among them. They encode knowledge in a domain as well as knowledge that spans domains.[56] In this way, they make that knowledge reusable. In discussing ontologies, we may distinguish, as Nicola Guarino does, between capital "O" ontology (i.e., a formal upper-level ontology) and lowercase "o" ontology.[57] The former refers to a formal philosophical discipline that concerns itself with developing a system of categories that account for a certain conceptualization of the world, while the latter refers to an engineering artifact constituted by a certain vocabulary (based on a set of explicit assumptions) that describes a certain reality or domain. The focus of our project was closer to the object of building a small "o" ontology of the Canadian context of financial electronic records, though we ultimately linked our ontology to the abstractions expressed in two upper-level ontologies.

Research on ontologies is becoming increasingly widespread: it is found in fields as diverse as knowledge engineering, knowledge representation, qualitative modelling, language engineering, database design, information modelling, information integration, object-oriented analysis, information retrieval and extraction, knowledge management and organization, and agent-based systems design.[58] In recent years, ontology development has been moving from laboratories to the desktops of domain experts and has become common across the World Wide Web, from the categorization of websites (Yahoo!) to the categorization of products for sale (Amazon).[59] Ontology research and usage have found

53 Rudi Studer, V. Richard Benjamins, and Dieter Fensel, "Knowledge Engineering: Principles and Methods," *Data & Knowledge Engineering* 25, no. 1–2 (March 1998): 184.

54 Mike Uschold and Michael Gruninger, "Ontologies: Principles, Methods and Applications," *Knowledge Engineering Review* 11, no. 2 (June 1996): 93–136.

55 "Frequently Asked Questions on W3C's Web Ontology Language (OWL)," The World Wide Web Consortium, accessed 21 May 2008, http://www.w3.org/2003/08/owlfaq.

56 Kevin Bradley, ed., *Guidelines on the Production and Preservation of Digital Audio Objects*, 2nd ed. (Auckland Park, South Africa: International Association of Sound and Audio Visual Archives, 2009), accessed 25 October 2012, http://www.iasa-web.org/tc04/publication-information.

57 Nicola Guarino, ed., "Formal Ontology and Information Systems," *Proceedings of the International Conference on Formal Ontology in Information Systems, 6–8 June 1998, Trento, Italy* (Amsterdam: IOS Press, 1998), 3–15.

58 Guarino, "Formal Ontology and Information Systems," 3.

59 Noy and McGuinness, "Ontology Development 101," 1.

their way into the domain of cultural heritage as well. Ontologies are being used to aid access to museum artifacts, such as the University of Gothenburg's project to support multilingual access to artifacts through the use of semantic web ontologies, and as the basis for metadata schemas for digital preservation.[60] Srinivasan and Huang have used "fluid, community-chosen ontologies" in their design of the *Village Voice* online agora.[61] Recently, the UK National Archives and Ontotext have used ontologies to improve transparency and accessibility to holdings.[62] Katifori et al. have used ontologies to support researchers' exploration of historical news archives,[63] while Hunter and Yu have explored the use of ontologies in annotating 3D museum objects.[64] Many other examples could be cited, though a full survey of such initiatives is beyond the scope of this work.[65]

Ontologies are closely related to taxonomies, though there is a good deal of debate about the precise nature of the differences between them. Patrick Lambe defines the term "taxonomy" as any system that can be used to group, arrange, and describe items according to meaningful principles, and which provides

60 Dana Dannélls, Mariana Damova, Ramona Enache, and Milen Chechev, "Multilingual Online Generation from Semantic Web Technologies," *Proceedings of the 21st International Conference Companion on World Wide Web, Lyon, France, 16–20 April 2012 (WWW '12 Companion)* (New York: ACM, 2012), 239–42.

61 Ramesh Srinivasan and Jeffrey Huang, "Fluid Ontologies for Digital Museums," *International Journal on Digital Libraries* 5, no. 3 (May 2005): 193–204, cited in Shilton and Srinivasan, "Participatory Appraisal and Arrangement," 98.

62 See Ontotext, "The National Archives: Semantic Knowledge Base," under "Clients & Stories: Success Stories," accessed 25 October 2012, http://www.ontotext.com/case/nationalArchives-skb. This was achieved by creating an ontology capturing various aspects of the government structure and functions from archived websites and populating the ontology with instances from existing government data sets. Through information extraction methods, the project team identified key entities and facts in the documents and linked them to the knowledge base, at the same time handling ambiguity and duplication issues. They were able to create an ultimate scale multi-paradigm semantic index powering the final archive search system.

63 A. Katifori, C. Nikolaou, M. Platakis, Y. Ioannidis, A. Tympas, M. Koubarakis, N. Sarris, V. Tountopoulos, E. Tzoannous, S. Bykau, N. Kiyavitskaya, C. Tsinaraki, and Y. Velegrakis, "The Papyrus Digital Library: Discovering History in the News," *TPDL '11 Proceedings of the 15th International Conference on Theory and Practice of Digital Libraries: Research and Advanced Technology for Digital Libraries* (Berlin: Springer-Verlag, 2011), 465–68.

64 Jane Hunter and Chih-hao Yu, "Assessing the Value of Semantic Annotation Services for 3D Museum Artefacts," in *Sustainable Data from Digital Research*, ed. N. Thieberger, L. Barwick, R. Billington, and J. Vaughan (Melbourne: University of Melbourne, 2011), 137–57.

65 For examples, see Museums and the Web, "Museums and the Web 2012, 11–14 April 2012, San Diego, CA," under "Sessions," accessed 25 October 2012, http://www.museumsandtheweb.com/mw2012/sessions.html; Kevin Bradley, International Association of Sound and Audiovisual Archives, IASA Technical Committee, *Guidelines on the Production and Preservation of Digital Audio Objects* (Auckland Park, South Africa: IASA, 2009), accessed 25 October 2012, http://www.iasa-web.org/tc04/publication-information.

users with an overview of the domain being organized.[66] Using Lambe's definition, an ontology would be a type of taxonomy. On the other hand, in the field of knowledge management, taxonomies are usually conceived of as being narrower than ontologies, and sometimes derived from them, as ontologies describe a larger range of relationship types.[67]

Debates about the precise difference between ontologies and taxonomies aside, there is general agreement that *hierarchical* taxonomies aggregate items along a continuum, where one end of that continuum consists of fewer items (i.e., constructs, classes, types) than the other end. This notion of hierarchies as continuums explains why taxonomies take on a tree-like shape. Here, it is worthwhile to remark that hierarchies, or trees, are not in themselves bad. They are merely one shape that we can assign to represent archival abstractions, which reveal some features of archives and hide others, as is true of any representational form. So while it is important to note that archives are not trees, neither should recordkeepers avoid planting trees, so to speak.

Hierarchical taxonomies support inferences using a single generic relationship type (e.g., supertype-subtype or parent-child relationships, or "is a") in a consistent, systematic manner to bring together categories.[68] Ontologies sup-

66 Patrick Lambe, *Organising Knowledge: Taxonomies, Knowledge and Organisational Effectiveness* (Oxford: Chandos Publishing, 2007).

67 For an example, see Hendra Suryanto and Paul Compton, "Learning Classification Taxonomies from a Classification Knowledge Based System," in *Proceedings of the First Workshop on Ontology Learning OL '2000, Berlin, Germany, 25 August 2000,* ed. Steffen Staab, Alexander Maedche, Claire Nedellec, and Peter M. Wiemer-Hastings (CEUR Workshop Proceedings, 2000), 31.

68 Different theories provide slightly different accounts of the nature of generic relationships. Some theories and models describe generic relationships as a series of inclusion sets in which a category is divided into more specific categories. Other approaches describe generic relationships as parent/child relationships, broader term/narrower term relationships, or relationships between a category and a species or member of that category. In all likelihood, each of these theories conceives of the generic relationship in a broadly similar manner. However, none of the theories is sufficient for a precise formal understanding of what the generic relationship entails. See Barbara H. Kwasnik, "The Role of Classification in Knowledge Representation and Discovery," *Library Trends* 48, no. 1 (Summer 1999): 22–47; Stella G. Dextre Clarke, "Organising Access to Information by Subject," in *Handbook of Information Management*, ed. Alison Scammell (London: Aslib-IMI, 2001); Stella G. Dextre Clarke, "Thesaural Relationships," in *Relationships in the Organization of Knowledge*, ed. Carol A. Bean and Rebecca Green (Dordrecht: Kluwer Academic Publishers, 2001); "Resource Description Framework (RDF) Schema Specification 1.0," W3C Metadata Activity, accessed 27 March 2012, http://www.w3.org/TR/2000/CR-rdf-schema-20000327/; David K. Farkas and Jean B. Farkas, *Principles of Web Design* (New York: Longman, 2002); Louis Rosenfeld and Peter Morville, *Information Architecture for the World Wide Web*, 3rd ed. (Sebastopol, CA: O'Reilly Media, 2006); Heather Hedden, "Taxonomies and Controlled Vocabularies: Best Practices for Metadata," *Journal of Digital Asset Management* 6, no. 5 (October 2010): 279–84; Heather Hedden, "Taxonomies and the Information User," *Information Outlook* 14, no. 8 (2010): 10–13; Jean Aitchison, David Bawden, and Alan Gilchrist, *Thesaurus Construction and Use: A Practical Manual*, 4th ed. (London: Aslib-IMI, 2000); Elin K.

port inferences by explicitly defining the precise nature of each category in the system and the precise relationships among those categories. In differentiating between hierarchical taxonomies and ontologies, Gruber points out that ontologies make use of a much wider range of relationships than those in hierarchical nesting; for example, relationships such as owns, is part of, secured by, and restricts.[69] This explains why ontologies often take on network structures or shapes with configurations more like forests than trees. As such, ontologies offer archivists great expressiveness in representing the rich complexity of archival provenance and may help to achieve the elusive balance between parsimoniousness and expressiveness in building archival interfaces.

Establishing the Boundaries of Our Domain Ontology

The literature on ontology development is relatively sparse, and Noy and McGuinness, and Uschold and Gruninger note that there is no "correct" way to develop an ontology.[70] Whether done manually or using semi-automatic or automatic techniques, it is typically an iterative process, beginning with a rough first pass, followed by revision, and then filling in detail.[71] However, unless one is developing an upper-level ontology, an effort must be made to establish the boundary of the domain of interest.[72] As noted previously, we equated the notion of a domain in the development of a domain ontology to the notion of "societal provenance" in the field of archives. Nesmith[73] and others have suggested that

Jacob, "Classification and Categorization: A Difference that Makes a Difference," *Library Trends* 52, no. 3 (Winter 2004): 515–40; American National Standards Institute and the National Information Standards Organization, *Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies* (Bethesda, MD: National Information Standards Organization, 2005).

69 *Encyclopedia of Database Systems*, s.v. "Ontology," by Tom Gruber, accessed 25 October 2012, http://www-ksl.stanford.edu/kst/what-is-an-ontology.html.

70 Noy and McGuinness, "Ontology Development 101," 4; Uschold and Gruninger, "Ontologies," 14. These writers suggest a skeletal method of ontology development as consisting of identifying the ontology's purpose and scope; building the ontology (ontology capture, coding, and integration with existing ontologies); evaluation; documentation; and developing guidelines for each phase. Others who have written on the development of ontologies include: Gruber, "Toward Principles for the Design of Ontologies Used for Knowledge Sharing?" *International Journal of Human Computer Studies* 43, no. 5–6 (November 1995): 907–28; Mike Uschold, Martin King, Stuart Moralee, and Yannis Zorgios, "The Enterprise Ontology," *The Knowledge Engineering Review* 13, no. 1 (March 1998): 31–89.

71 On the subject of domain ontology development, see, for example, Gruber, "Toward Principles"; Uschold et al., "The Enterprise Ontology"; Asunción Gómez-Pérez and David Manzano-Macho, "An Overview of Methods and Tools for Ontology Learning from Texts," *The Knowledge Engineering Review* 19, no. 3 (September 2004): 187–212.

72 Noy and McGuinness, "Ontology Development 101," 4.

73 Nesmith, "Reopening Archives."

societal context is limitless and constantly evolving. Limitless and evolving as it may be, we faced the practical challenge of defining boundaries so that we could represent the Canadian context of financial electronic records in our high-level reference model. This was not an insignificant challenge; in contrast to modelling a physical entity wherein the boundaries are well defined, we had no pre-existing theoretically valid basis to guide us in defining the boundaries of our domain of interest. The field of archives, as noted above, is very vague on this point.

Since the term "financial system" is commonly used to describe the collectivity of entities comprising Canadian financial markets and market participants,[74] we initially chose the metaphor of the "system" – which in turn draws upon on systems theory – as a theoretical basis for defining the boundaries of our model. In systems theory, a system consists of interacting components that are interconnected through a web of relationships.[75] What is common among the parts is that they are working toward achieving a single purpose. Effective operation of the parts in relation to the whole leads to the achievement of the system's purpose.[76] In recent years, economic sociologists have tended to view markets as embedded in social relations and social networks. In these writings, the notion of a "network," as opposed to the system, becomes the defining metaphor.[77] According to this view, networks comprise a collection of nodes tied together by relationships of resources, communication, and other coordinating instances.[78] Both conceptualizations share the notion of a range of interacting relationships as the fabric of the system or network, a view consistent with Terry Cook's notion of archival context.[79]

Financial systems, or networks, work at a global level, but they also operate at the level of countries and other geopolitical jurisdictions. At the global level, Knorr Cetina observes:

> a global market … has been in evidence for some time. This form of globality is not based upon the penetration of countries or of individual behavior. Instead, it rests on the establishment of bridgehead centers of institutional trading in the financial hubs

74 See, for example, Bank of Canada, "What Is the 'Financial System'?" under "About the Bank: Financial System," accessed 25 October 2012, http://www.bankofcanada.ca/about/what-we-do/what-is-the-financial-system/.

75 Ludwig Von Bertalanffy, *General System Theory: Foundations, Development, Applications* (New York: George Braziller, 1968).

76 Peter M. Senge, *The Fifth Discipline: The Art and Practice of the Learning Organization* (New York: Doubleday Business, 1990).

77 Karin Knorr Cetina, "How Are Global Markets Global? The Architecture of the Flow World," in *The Sociology of Financial Markets*, ed. Karin Knorr Cetina and Alex Preda (Oxford: Oxford University Press, 2005), 39.

78 Knorr Cetina, "How Are Global Markets Global?," 40.

79 See, for example, Cook, "Mind Over Matter," 38–70.

of the three major time zones: in New York, London, Tokyo and Zurich, Frankfurt or Singapore. The moving market "rests" in these centers where it becomes articulated and revised.[80]

Neave discusses the boundaries of a financial system and provides a framework for making decisions about where a national financial system begins and where it ends.[81] He places emphasis on deciding a boundary in relation to juridical authority – the political and legal authorities to whom users, financial intermediaries, and markets primarily are accountable and that establish the rules of law that create contractual obligations in the exchange of instruments or the settling of claims. As the laws and regulations that shape the Canadian financial system become increasingly defined by global regulatory bodies, the boundaries between the Canadian financial system and the global financial system become blurred. Thus, it is impossible to speak of a Canadian financial system in any pure sense. Not only is it difficult to distinguish the boundaries of a financial system, but financial systems operate in an increasingly interdependent manner: from the financial crisis of 2007–2009, it is possible to observe how difficult it is to shelter the financial system of one country from imbalances in the operation of another country's financial system. In spite of these complications, we relied on Neave's guidance to establish a boundary for our domain ontology.

Building a Knowledge Base

Having determined the boundaries of our domain ontology in as theoretically and rigorous a way as we could in light of the difficulties of assigning boundaries to an entity as porous and fluid as a financial system, we then began to construct a knowledge base about the Canadian context of current financial electronic records, which we would be able to use to determine the abstractions for our domain (i.e., the classes of things in our domain and the relationships among them). To begin the process of building the knowledge base, the research team[82] established relationships with individuals who work in or with five of the top

80 Knorr Cetina, "How Are Global Markets Global?," 57.

81 Edwin H. Neave, *Financial Systems: Principles and Organization* (London, New York: Routledge, 1998).

82 Initially, our research team comprised two people: the principal investigator and a PhD student at the University of British Columbia's School of Library, Archival and Information Studies (Sherry Xie). The team later expanded to include two more PhD students from the University of British Columbia's Sauder School of Business, Management Information Systems Department (Lior Limonad and Kafui Monu), another PhD student from the School of Library, Archival and Information Studies (Elaine Goh), a master's student in Computer Science (Thomas Dang), and a Library and Information Science master's student (Jack Hallin). I offer my thanks to all of these students for their valuable contribution to this project.

Canadian financial institutions[83] and who have expert knowledge of the context of records creation and recordkeeping. The team also reached out to individuals from a variety of disciplines from each of the target organizations – records managers, archivists, IT specialists, compliance officers, privacy officers, regulators, operational risk specialists, and data management experts. Once the key relationships were established, the team organized three workshops in Toronto and two in Vancouver as a first data-gathering exercise aimed at gaining a deeper understanding of the Canadian business context in which financial electronic records are created and maintained; the different types of records within Canadian financial institutions; and the laws and regulations, processes, people, and technologies that first give rise to financial electronic records creation and which influence how they are managed over time. In Vancouver, the workshops were led by experienced regulators who provided high-level overviews of the institutions, processes, records and information, and regulations governing the Canadian financial system. The subsequent Toronto workshops focused on key issues and areas of concern for those responsible for managing financial records and information. Research data collected during the workshops was supplemented, as necessary, by follow-up interviews with workshop participants or other relevant respondents and through an analysis of relevant texts, such as laws and regulations related to the creation and keeping of financial electronic records. To guide data collection in the supplementary interviews, the research team used a set of semi-structured questions derived from the preliminary findings of the workshops. Participants were asked to articulate the context of record creation and recordkeeping in Canadian financial institutions by first identifying key business processes and product lines. They were then asked to identify the legal, regulatory, or other internal policy constraints on those processes or product lines; the records arising from those processes and products; and the technologies involved in the creation and keeping of these records. The results of this exploratory research, which included notes and documentation from the workshops, transcripts of supplementary interviews, texts of internal policy documents, and texts of financial regulation and industry guidance, were captured as a corpus from which the conceptual elements of the ontology could be extracted. It should be noted that our approach of building a knowledge base, or corpus of relevant texts, as a basis for our ontology differs from the approaches used in archival applications of ontologies (e.g., The National Archives' initiative noted above) whereby the archives' holdings of records are used as a corpus of texts from which to extract the constructs and relationships for the ontology. Which method is preferable and when it may be more appropriate or effective to use

83 In no particular order, the financial institutions were TD Canada Trust, Scotiabank, CIBC, RBC Royal Bank, and BMO Financial Group.

one approach or another, or even a combination, remains an open question deserving of further research.[84]

Defining Abstractions: The Classes and Their Relationships

Uschold and Gruninger explain that an ontology will include a vocabulary of terms and some specification of their meaning to varying degrees of formality.[85] An explicit ontology will include a formal description of the concepts in the domain of interest. These are also called classes. Noy and McGuinness explain the notion of classes:

> For example, a class of wines represents all wines. Specific wines are instances of this class. The Bordeaux wine in the glass in front of you while you read this document is an instance of the class of Bordeaux wines. A class can have **subclasses** that represent concepts that are more specific than the superclass. For example, we can divide the class of all wines into red, white, and rosé wines. Alternatively, we can divide a class of all wines into sparkling and non-sparkling wines.[86]

Each class will have properties describing various features or attributes of the concept as well as restrictions on concepts. These are sometimes also referred to as roles or properties or, following Noy and McGuinness, slots.[87] Slots describe properties of classes and, by extension, instances that fall within classes. For example:

> Château Lafite Rothschild Pauillac wine has a full body; it is produced by the Château Lafite Rothschild winery. We have two slots describing the wine in this example: the slot body with the value full and the slot maker with the value Château Lafite Rothschild winery. At the class level, we can say that instances of the class Wine will have slots describing their flavor, body, sugar level, the maker of the wine and so on.[88]

Thus, we might arrive at an ontology wherein the instance or particular wine Château Lafite Rothschild is defined by a number of classes (italicized in the following) and relationships (in bold), such as: 1) **is a type** of *Pauillac*; 2) **made**

84 Uschold and Gruninger, "Ontologies," 20–21, suggest a "middle-out" approach, noting that a bottom-up approach can result in too much detail without an understanding of the commonality between terms, while a top-down approach results in better control of detail but imposes arbitrary high-level categories.

85 Uschold and Gruninger, "Ontologies," 6. We tend to agree, given the representational problems associated with records, which means that records may not fully represent the context of their creation and thus produce an incomplete domain ontology. On the other hand, the resulting ontology may better represent, even if incomplete, the world view of the records' creators/preservers.

86 Noy and McGuinness, "Ontology Development 101," 3 (emphasis in original).

87 Ibid., 3.

88 Ibid., 3.

by Château Lafite Rothschild; 3) **a type of** *winery*; and 4) one of the *best wineries* in the Bordeaux *region*. The totality of these structures – classes, roles, properties, restrictions, and relationships – can be referred to as the modelling grammar for a domain ontology.

For our domain ontology, our research team began the process of creating the modelling grammar using a top-down approach.[89] In order to identify initial classes, we started with a review of Howells and Bain, who describe the financial system as comprising markets, individuals, organizations, plus the supervisory bodies with oversight of the system.[90] They further refer to financial markets as an organizational framework within which financial instruments can be bought and sold. They segregate the types of markets into those trading in instruments with long maturities (e.g., capital markets) versus those trading in instruments with short maturities (e.g., money markets). Within each market, there are financial intermediaries of many different types, each with its own role in relation to the operation of the financial system as a whole. Examples of these types of institutions are deposit-taking and non-depository banks, insurance companies, and pension funds.

For our model of the Canadian financial system, we also drew upon the data in our knowledge base, and, in particular, Canadian financial legislation, to identify and categorize different types of financial intermediaries (e.g., Schedule I, II, and III banks).[91] We continued to model the structure of the Canadian financial system by identifying specific instances of the classes of the entities we had identified (e.g., specific banks, markets, regulators, etc.). We pursued an iterative procedure of exploring our knowledge base in a progressive pattern to extract instances, abstract them into classes, define roles and properties for each class, and define relationships between classes. For each iteration, the knowledge base played two roles: a source from which new modelling constructs were identified; and a target, according to which the conceptual models produced using the modelling grammar could be verified to validate the consistency and effectiveness of our abstractions. We then used "rich pictures" as a technique for transforming informal representations of the context of financial electronic records, iteratively extracted from our knowledge base, into more formal representations.[92]

89 Ibid., 6.

90 Peter Howells and Keith Bain, *Financial Markets and Institutions*, 3rd ed. (Harlow, Essex, UK: Financial Times Prentice Hall, 2000).

91 Government of Canada, Justice Laws Website, *Bank Act, Statutes of Canada* 1991, c.46, under "Laws: Consolidated Acts," http://laws-lois.justice.gc.ca/eng/acts/B-1.01/FullText.html.

92 Peter Checkland, *Systems Thinking, Systems Practice* (New York: Wiley, 1999); Peter Checkland and Jim Scholes, *Soft Systems Methodology in Action* (New York: Wiley, 1999). Soft systems methodology (SSM) attempts to make sense of complex problem situations that involve human activity. Rich Pictures was developed to support the SSM method by

Following this stage in the development of our ontology, we formalized it by mapping each grammatical construct to concepts in two formal upper-level ontologies that we identified as being most similar. The first ontology was Bunge's ontology, which is often used to assess information systems designs as representations of the real world.[93] A core concept in Bunge's ontology is a *thing*. According to the ontology, the world is made of things. A thing is a substantial individual endowed with all its (substantial) properties, or intrinsic properties. Some properties, called mutual properties, are properties shared by two or more things.[94] A role, for example, is a set of mutual properties, acquired by some instances of a class when they engage in interaction. Bunge's ontology, however, is designed to represent concrete objects in the world, despite some relativistic features. Thus, it did not provide us with a way to account for both the concrete (paper) and social (paper that has a value of $20) aspects of things (e.g., money) in our financial domain. We therefore drew upon Searle's social ontology[95] to give our modelling grammar greater expressiveness in relation to the social, or observer-relative, aspects of the Canadian context of financial electronic records. A key concept in Searle's ontology is a social institution. To explain more fully:

> At its core, Searle's ontology includes three primitives to describe the basic structure of social realities: *collective intentionality*, *the assignment of function* and *constitutive rules*. Collective intentionality is the intentionality that is shared by different people, where *intentionality* means the capacity of the mind to represent objects and states of affairs in the world other than itself. Individual (or subjective) intentionality allows a distinction between *objective entities* and *subjective entities*. The former is equivalent to the notion of "things" in Bunge's ontology. That is, objects in the universe whose mode of existence is [in Bunge's view] independent of any perceiver or any mental state (e.g., mountains). The latter are objects whose mode of existence depends on being perceived by individuals (e.g., pains). When this mode of existence is further acknowledged by a collective intentionality (i.e., by a certain social group), such objects are further promoted to a level of *social institutions* (e.g., the USA). Through the assignment of function, social institutions may be associated by people with certain *features*. Unlike substantial properties, these features are never intrinsic to the physics of any phenomenon but are assigned from outside by conscious observers (i.e., observer relative). The acknowledgement of such assignments is achieved within any given society through the representation ([e.g.,] in language) of constitutive rules in the form of "X counts as Y in Z." As an ontological entity, each social institution can

providing a mechanism for learning about complex or ill-defined problems by drawing detailed graphical representations of them.

93 Wand and Weber, "Mario Bunge's Ontology."

94 Victoria Lemieux and Lior Limonad, "What 'Good' Looks Like: Understanding Records Ontologically in the Context of the Global Financial Crisis," *Journal of Information Science* 37, no. 1 (February 2011): 33.

95 Searle, "Social Ontology."

be associated with an underlying substantial object (e.g., money and a piece of paper) for which constitutive rules can assign additional observer-relative features (e.g., a piece of paper (X) counts as $20 (Y) in the country Canada (Z)). However, collective intentionality also allows for the assignment of function to social institutions for which there is no underlying physical object (e.g., money realized with the use of debit cards). Furthermore, the establishment and acceptance of constitutive rules in a society creates a deontic power. That is, the mode of existence and functions assigned to a social institution continue to exist after its initial creation and even after all the participants involved have stopped thinking about its initial creation. In short, social realities introduce a new type of entity.…[96]

It should be noted that a major challenge for us in mapping Searle's ontological concepts to Bunge's and our own constructs is that they are not as tightly or logically defined as Bunge's.[97] Based on the relationships between the concepts in Bunge's and Searle's ontologies and the constructs in our modelling grammar, we further refined the meaning (i.e., ontological semantics) given to each construct. Thus, we arrived at a definition of a financial instrument, for example, as *a type of social entity assigned a certain role within a financial system*. Finally, we associated each construct and relationship in the grammar with a corresponding graphical notation (i.e., what we labelled "concrete syntax"). This resulted in a specification for a domain-specific language intended to describe the static aspects of the financial domain in general, and the Canadian financial domain in particular.

From Societal Context to an Ontological Understanding of the Record

We conceptualized our domain ontology of the Canadian context of financial electronic records as equating to the notion of societal provenance, but what of the records themselves? Our next challenge was to determine how to conceptualize them in relation to our domain ontology. In traditional approaches to theorizing about records, as noted above, records are viewed as objects or artifacts that reside *within* a context. That is, they form part of the context in which they are found. Taking this view, records may be seen, in the traditional view, as the result of, the by-product of, or the execution mechanism of some transaction (e.g., a financial transaction) that takes place in a context (e.g., the Canadian financial system). However, by once again marrying archival theory to information systems theory, we can draw upon representation theory to merge traditions of

96 Lemieux and Limonad, "What 'Good' Looks Like," 34–35.

97 Searle, in "Social Ontology," does not provide clear and exact definitions in some cases; in some places, concepts seem to overlap or shift meaning, making it challenging to extrapolate meaning from his definitions. Thus, our use of his ontology is only an interpretation of his meaning.

viewing records as representations from archival theory with similar theories on the representational qualities of information systems.

To elaborate, in information systems theory, an information system is defined as an artifact that pertains to three specific models: representational, state-tracking, and good-decomposition.[98] All three models rely on the premise that an information system is an iconic representation of concrete or conceived (i.e., in someone's mind) real-world systems (namely, the domain-of-discourse), as perceived by someone's or some group's view. This theoretical perspective is consistent with Geoffrey Yeo's work on concepts of records as being "*persistent representations of activities or other occurrents, created by participants or observers of those occurrents or by their proxies; or sets of such representations representing particular occurrents*."[99] In this work, we accept the general representation premise. Records may, however, be more than just representations; they may, in fact, *constitute* transactions or be the means by which transactions take place. Nevertheless, representation is necessary in order for a record to exist. Yeo argues that, in representing, records may offer a variety of affordances (here drawing upon the theory of affordances) such as information, evidence, and even the constitutive powers that provide the means by which transactions may occur. Once this leap of understanding about the record as a type of information system is made, it is then possible to formally link records theory to information systems theory and to suggest that the record is a type of information system. Thus, we propose the first amendment to Yeo's definition of the record as follows:

> *(1) A record is a type of information system that serves as a persistent representation of activities or other occurrents, created by participants or observers of those occurrents, or by their proxies; or sets of such representations representing particular occurrents.*

Though we accept the idea that the record is a type of information system, we note that, unlike other information systems, it is not state-tracking in that it does not track the state changes in the objects it represents across time and space. Instead, records aim to arrest state-tracking in order to create a *persistent* representation of a state as it is (or was) in space-time.

We also note that Yeo's definition is silent on the fact that activities and occurrents take place within a certain domain of discourse. The idea is implicit in Yeo's concept of a record, but not as explicit as we think it needs to be. As the economist Hernando de Soto says in relation to financial assets, "Remember, it

98 Yair Wand and Ron Weber, "On the Deep Structure of Information Systems," *Information Systems Journal* 5, no. 3 (July 1995): 203–23.

99 Yeo, "Concepts of Record (2)," 136 (italics in original).

is not your own mind that gives you certain exclusive rights over a specific asset, but other minds thinking about your rights in the same way as you do."[100] This view is also consistent with archival theorists who suggest that records are socio-technical constructs[101] and that records do not so much reflect the bodies that create them as they do the way that a particular body organizes its own memory.[102] We therefore suggest that the concept of belief by a social group about the conceptualization of a record is important enough to make more explicit in the definition of a record and suggest a revision of Yeo's definition as follows:

> *(2) A record is a type of information system that serves as a persistent representation of a set of beliefs about activities or other occurrents, taking place in a certain domain of discourse, created by participants or observers of those occurrents, or by their proxies; or sets of such representations representing particular occurrents.*[103]

We further note that if records are viewed as representations, instead of as objects, by-products, or execution mechanisms, then financial records represent the abstractions we have identified as comprising our domain ontology. Specifically, a financial record will represent a financial instrument (e.g., a mortgage) which is, according to our ontological definition and following Searle, a type of social entity assigned a certain role in a financial system and, following Bunge, "a set of mutual properties, acquired by some instances of a class, when they engage in interaction"[104] (or what Geoffrey Yeo might call an "occurrent"). In other words, by engaging in financial transactions, two actors create a relationship and interact.[105] Once we establish a theoretical basis for describing records as representations of occurrents and if occurrents are actually the mutual properties of two counterparties (i.e., entities in relationship), we may understand records as representing interactions. Thus, we propose a further revision to Yeo's definition of the records as follows:

100 Hernando de Soto, *The Mystery of Capital: Why Capitalism Triumphs in the West and Fails Everywhere Else* (New York: Basic Books, 2000), 177.

101 See Tom Nesmith, "Still Fuzzy, But More Accurate: Some Thoughts on the 'Ghosts' of Archival Theory," *Archivaria* 47 (Spring 1999): 136–50; and Victoria L. Lemieux, "Let the Ghosts Speak: An Empirical Exploration of the 'Nature' of the Record," *Archivaria* 51 (Spring 2001): 81–111.

102 Douglas discusses this point in "Origins" (p. 32) in reference to works by contemporary Italian archival thinkers, such as Claudio Pavone.

103 Note that both the activities and occurrents *and* the representation itself take place in a domain of discourse. It is the latter point that I wish to emphasize, since this shapes the record as a socio-technical construct.

104 Lemieux and Limonad, "What 'Good' Looks Like," 37.

105 I do not say that financial records always do fully represent financial transactions because records often fail to represent or are poor representations.

(3) A record is a type of information system that serves as a persistent representation of a set of beliefs about the assignment of status to actors within a certain domain of discourse when they engage in interaction, which is created by actors or observers of those interactions, or by their proxies; or sets of such representations representing particular occurrents.

Theoretically, this merging of archival with information systems theory, which allows us to theorize about records as representations of interactions, also gives us the opportunity to analyze records as representations of networks using, for example, graph analysis (i.e., computations of centrality and distance in the network). This opens up some potentially intriguing areas of future research on the theory of the record. It also further supports analysis of records as artifacts which represent and are constructed by social relationships, thereby giving us a theoretical basis to look at how social relations shape the record as a representational form (e.g., how incentive structures in social networks might impact upon the representational quality of records), a theme that this author has explored in a previous work.[106]

Not only can we theorize, drawing upon archival and information systems theory, that records are (or, at least, may be) representations of occurrents, and therefore also of relationships between counterparties as they engage in interactions, but also that their representational scope may go beyond even the occurrent. Here, we may further merge archival and information systems theory, with the particular economic theories of Hernando de Soto. De Soto writes:

> Throughout history human beings have invented representational systems – writing, musical notation, double-entry bookkeeping – to grasp with the mind what human hands could never touch. In the same way, the great practitioners of capitalism … were able to reveal and extract capital where others saw only junk by devising new ways to represent the invisible potential that is locked up in the assets we accumulate…. Westerners take this mechanism so completely for granted that they have lost all awareness of its existence…. **It is hidden in thousands of pieces of legislation, statutes,** regulations, and institutions that govern the system. Anyone trapped in such a legal morass would be hard pressed to see how the process actually works. The only way to see it is from outside the system…. [107]

Thus, following de Soto's ideas, a domain ontology gives us the power to view from outside the system the morass of laws, regulations, institutions, and other elements that govern it. All of these elements are what records as representational information systems can represent. As such, to discuss records as just representing

106 Lemieux, "Let the Ghosts Speak."
107 de Soto, *The Mystery of Capital*, 7–8, 48.

occurrents is too limited a view. Rather, their representational nature comprises microcosms of entire domains, though perhaps only fragmentary elements of these domains owing to representational limitations, or fractals of networked interactions defined by social institutions, such as (in the case of the financial domain) money and markets, and the physical or technological forms of their embodiment as artifacts. Thus, we should consider a further revision of the definition of a record as follows:

> *(4) A record is a type of information system that serves as a persistent representation of a set of beliefs about the assignment of status to actors within a certain domain of discourse when they engage in interaction, and of the other classes, roles, properties, restrictions and relationships that relate to that interaction, which is created by actors or observers of that interaction, or by their proxies; or sets of such representations.*

This brings the discussion to a further observation: though we may theorize that records represent occurrents or interactions, we note that, in many cases, they fail to do so. We need only point to the failings of mortgage documentation during the financial crisis of 2007–2009 as evidence of this fact.[108] Yeo also acknowledges that, as representations, records may not be accurate reflections of reality. Thus, it is necessary to acknowledge that, to the extent that individuals actually intend to create persistent representations, records are often only partial representations of perceived reality (i.e., transactional relationships between counterparties). Moreover, though individuals may not intend to create records, non-record types of information systems (i.e., those not constructed to stop state change in order to create a persistent representation) may still serve as persistent representations and, thus, as records. From this perspective, it is theoretically more accurate to think of records as *emergent properties* of things (i.e., information systems) that may arise from the complex interaction of entities (e.g., actors, things) in a domain. Emergent properties may be characterized as properties that "'arise' out of more fundamental entities and yet are 'novel' or 'irreducible' with respect to them."[109] Thus, it is more appropriate to refer to

108 See, for example, James R. Barth, Tong Li, Triphon Phumiwasana, and Glenn Yago, "Perspectives on the Subprime Mortgage Market" (working paper, Social Science Research Network, 2008), http://ssrn.com/abstract=1070404); Kenneth E. Scott and John B. Taylor, "Why Toxic Assets Are So Hard to Clean Up," *The Wall Street Journal*, accessed 21 July 2009, http://online.wsj.com/article/SB124804469056163533.html; Kathleen C. Engel and Patricia A. McCoy, *The Subprime Virus: Reckless Credit, Regulatory Failure, and Next Steps* (New York: Oxford University Press, 2011).

109 *The Stanford Encyclopedia of Philosophy*, s.v. "Emergent Properties," by Timothy O'Connor and Hong Yu Wong, http://plato.stanford.edu/archives/spr2012/entries/properties-emergent/ accessed 25 October 2012.

"recordness" than records per se. These ideas echo earlier reflections on the nature of records by Sue McKemmish, who noted that records are always in a "state of becoming."[110] Our definition should now therefore read as follows:

> *(5) A record is an emergent property ("recordness") of an information system that allows such an information system to serve as a persistent representation of a set of beliefs about the assignment of status to actors within a certain domain of discourse when they engage in interaction, and of the other classes, roles, properties, restrictions, and relationships that relate to that interaction, which is formed by actors or observers of those interactions, or by their proxies; or sets of such representations.*

A discussion of the reasons for which a record may fail to represent is the subject of further study beyond the scope of this paper.

We may sum up by stating that an ontological understanding of records accepts the representational premise and therefore accepts records as types of information systems that serve as persistent representations by virtue of the fact that they, unlike other types of information systems, do not attempt to track state changes. Furthermore, it views records as (potentially) representing occurrents, as well as all the ontological elements even beyond occurrents of a particular domain of interest as understood and constructed by a specific social group. Given an ontological understanding of records, we observe that occurrents are mutual properties that things take on as they interact and therefore they equate to relationships; thus, records represent networks of relationships. These relationships have properties of roles or functions, but unlike traditional approaches to archival abstraction, an ontological understanding of records does not give first-order representational primacy to the construct of a function. At the same time, while there may be the intention to represent, certain representational problems may prevent records from actually representing what they are intended to represent. In addition, even types of information systems that are not intended to represent interactions persistently may still represent them. Thus, we conclude that it is more ontologically accurate to refer to records as emergent properties of information systems (i.e., to speak of "recordness" rather than of records).

110 Sue McKemmish, "Are Records Ever Actual?," first published in *The Records Continuum: Ian Maclean and Australian Archives First Fifty Years*, ed. Sue McKemmish and Michael Piggott (Melbourne: Ancora Press in association with Australian Archives, 1994), 187–203; available at Monash University, Information Technology, under "Records Continuum Research Group Publications," accessed 20 May 2013, http://www.infotech.monash.edu.au/research/groups/rcrg/publications/smcktrc.html.

Part III: Representation

Visually Representing the Canadian Context of Financial Electronic Records

Once we had articulated the constructs comprising our domain ontology and had enough instances of these constructs to work with, we began to consider how best to represent and provide access to our knowledge base. As noted above, we sought a representational form that would be both parsimonious and expressive. For this reason, we ruled out text-based descriptions, which we saw as being less than parsimonious if fully expressive or reductionist if not fully expressive. Instead, we chose to represent our domain visually, drawing upon theories of visual perception and cognition.[111] Visualization, as a representational system, has emerged as a growing theme in archival discourse. One of the most well-known examples is the work of Mitchell Whitelaw on the visible archive and generous interfaces, but many others too numerous to discuss in the scope of this paper have been conducting research and actively engaging in applying visualization to aspects of archival work.[112] Visualization is said to have advantages over other modes of representation and communication. In visual design, the image acts as a repository of data that relieves our "working memory" from having to remember the features of an entire data set. This visual aid lets the user remember and compare more data faster. Certain graphical features can be observed with "pre-attentive processing" so that they are understandable at a glance.[113] These pre-attentive features are understood much more rapidly than others because they are perceived prior to conscious attention. As a result,

111 See, for example, Jacques Bertin, *Sémiologie graphique. Les diagrammes, les réseaux, les cartes* (Paris: Gauthier-Villars, 1967); Edward R. Tufte, *Envisioning Information* (Cheshire, CT: Graphics Press, 1990); Edward R. Tufte, *Images and Quantities, Evidence and Narrative* (Cheshire, CT: Graphics Press, 1997); Edward R. Tufte, *The Cognitive Style of PowerPoint* (Cheshire, CT: Graphics Press, 2003); Edward R. Tufte, *Beautiful Evidence* (Cheshire, CT: Graphics Press, 2006).

112 Mitchell Whitelaw, "Visualising Archival Collections: The Visible Archive Project," *Archives and Manuscripts* 37, no. 2 (October 2009): 22–40; and Mitchell Whitelaw, "Towards Generous Interfaces for Archival Collections," paper presented at the International Council on Archives Congress, Brisbane, Australia, 20–24 August 2012. For a fuller discussion of the application of information visualization and visual analytics in the archival domain, see Victoria L. Lemieux, "Visual Analytics, Cognition and Archival Arrangement and Description: Studying Archivists' Cognitive Tasks to Leverage Visual Thinking for a Sustainable Archival Future," *Archival Science* (September 2013), http://link.springer.com/article/10.1007%2Fs10502-013-9212-y; and Victoria L. Lemieux, "Using Information Visualization and Visual Analytics to Achieve a More Sustainable Future for Archives: A Survey and Critical Analysis of Some Developments," *Comma* 2012, no. 2 (2012): 55–70.

113 Matthew O. Ward, Georges Grinstein, and Daniel Keim, *Interactive Data Visualization: Foundations, Techniques, and Applications* (Boca Raton, FL: A.K. Peters Ltd., 2010).

well-designed pre-attentive graphical elements pop out in one's field of vision.[114] Another advantage of graphical representation is the enormous pattern-finding capacity of human cognition. A point of meaning can be clearly supported by a graphic representation of underlying data, lined up to display the intended pattern.[115] Finally, a well-crafted representation makes the mental model of its creator explicit and can be used to test alternative hypotheses or understandings of the data visually. For these reasons, we concluded that a visual representation of our domain ontology could provide us with an effective representational form.

An "Institutional" View

As already discussed, archival fonds have been represented using a simple hierarchical diagram. We rejected the hierarchical metaphor in the first instance as we agreed with others that it would constrain our ability to represent the complex interconnections in our domain ontology (a.k.a. societal provenance). The individual entities comprising the financial system (e.g., banks, regulatory agencies, stock exchanges, etc.), though interacting with one another, were not related to one another in the same manner as entities in a hierarchical scheme (i.e., stock exchanges are not types of banks, nor are banks types of stock exchanges; rather, they transact with one another in financial markets). Thus, another representational form would have to be found. Instead, we drew inspiration from the notion of financial systems as networks of relationships in adopting the network graph as the best visual metaphor for representing the Canadian context of financial electronic records. Moreover, ontologies are typically visualized as network graphs (e.g., Resource Description Framework (RDF) triples when visualized as nodes and links[116]). It must be said that the network graph representation is, like hierarchical representations of fonds, only one way of understanding and representing financial markets and the institutions that comprise them. As previously mentioned, Knorr Cetina sets out a view of financial markets as a flow world in which the architecture is not networked, but "scopic" in nature.[117] Although we

114 Pasha Roberts, "Information Visualization for Stock Market Ticks: Toward a New Trading Interface" (master's thesis, Massachusetts Institute of Technology, 2003), http://dspace.mit.edu/bitstream/handle/1721.1/16668/56675083.pdf.

115 Ibid.

116 For more information about RDF triples, see W3C Semantic Web, "Resource Description Framework (RDF)," W3C RDF Working Group, 3 February 2012, http://www.w3.org/RDF/.

117 Knorr Cetina, "How Are Global Markets Global?," 40. Knorr Cetina explains that, in contrast to the network view of such markets that sees the world as a physical place or a totality of objects wherein we live or in which symbolic processes take place, a scopic view conceives of the markets as "processual": scopic markets emerge and project market reality, carrying it forward and allowing it to flow. Knorr Cetina stresses that scopic architectures are in a state of constant flux and operate in accordance with reflexive mechanisms of observation and projection (i.e., through what is projected to a trader on the computer screen and how that trader responds to what is projected). In this scopic view of financial markets, both

recognize that there are alternate conceptualizations of our domain space, and, consequently, alternate representational forms and visual metaphors that could be used to represent the domain – i.e., "knowledge doesn't have *a* shape" – we ultimately chose to keep to the network conceptualization and representational form for our first effort at developing a high-level visual reference model of the Canadian context of financial electronic records.

Identifying a suitable visualization method and tool to create an interactive network visualization of our domain proved to be another interesting challenge. Katifori et al. note that there exist several ontology visualization methods: the indented list; the node-link tree; the zoomable display; the space-filling visualization; the focus+context or distortion visualization; and 3D information landscapes.[118] They also note that there are a number of techniques from other contexts that could be used for ontology visualization. Ontology visualization researchers da Silva and Freitas conclude that the ideal ontology visualization tool will possess the following features:

- Provide overview of the ontology hierarchy, with the possibility of detailing some parts.
- Avoid presenting the different aspects of a specific ontology (classes, description, relationships, instances) together in a unique visualization.
- Optimize the results from ontology validation generated by inference processes.
- Explore the use of visual attributes such as colour, transparency, and shapes.
- Provide display filters based on different techniques of focus+context and/or overview+detail, zoom, pan, and rotation of the image.
- Allow rapid and simple inclusion of visual elements in the visualization, as well as their removal.[119]

A survey of existing tools revealed that they generally fall into one of two categories: 1) ontology-building and editing tools with some visualization functionality; or 2) visualization tools with some capability to represent networks.[120]

markets and ontologies are liquid.

118 A. Katifori, C. Halatsis, G. Lepouras, C. Vassilakis, E. Gannopoulos, "Ontology Visualization Methods – A Survey," *ACM Computing Surveys* 3, no. 4, article 10 (October 2007): 1–43.

119 Isabel Cristina Siqueira da Silva and Carla Maria Dal Sasso Freitas, "Using Multiple Views for Visual Exploration of Ontologies," in *Proceedings of Joint IV Seminar on Ontology Research in Brazil and VI International Workshop on Metamodels, Ontologies and Semantic Technologies*, ed. Renata Vieira, Giancarlo Guizzardi, and Sandro Rama Fiorini (CEUR Workshop Proceedings, 2011), 28, accessed 20 May 2013, http://ceur-ws.org/Vol-776/ontobras-mos 2011_completeProc.pdf.

120 J.R.G. Pulido, R. Herrera, M. Aréchiga, A. Block, R. Acosta, and S. Legrand, "Identifying Ontology Components from Digital Archives for the Semantic Web," in *Proceedings of the 2nd IASTED International Conference on Advances in Computer Science and Technology*, ed. S. Sahni (Anaheim, CA: ACTA Press, 2006), 7–12.

None of the tools perfectly matched da Silva and Freitas's requirements[121] or our own representational requirements. Protégé, a free, open-source tool used to develop and represent domain ontologies, exemplifies the first category.[122] An extension of the base platform, OWLViz, allows for visualization of OWL ontologies built using Protégé-OWL. This feature allows for visualization of the generic relationship (e.g., *x* is a type of *y*) in the domain ontology. Though this tool was capable of representing our modelling grammar and instances of entities in our domain ontology, it lacked the ability to project multiple views of our ontology that would support users in exploring the model and its underlying data – features we identified as being necessary in a third-order archival system. We also looked at several tools in the second category, i.e., those designed to represent and visualize networks, but many of these lacked interactivity and the multi-view functionality we desired as well. In the end, we settled on Quantum4D, an interactive visual interface that supported visual analysis by allowing users to view large volumes of data in dynamic two- and three-dimensional models.[123] Quantum4D was still not perfectly suited to our objectives, however, owing to the fact that it did not support features found in an ontology builder and editor, such as Protégé, including the use of a modelling grammar to reveal inconsistencies in and relationships among classes. The absence of this functionality ultimately proved problematic: it was possible to introduce new nodes representing classes that were inconsistent with our domain ontology, resulting in logical inconsistencies in our model. In spite of this, our experience with using Quantum4D revealed that it had more of the interactive functionality we thought useful in experimenting with building our reference model than other tools considered at the time. In our view, the ideal tool will possess features of both an interactive visual analytics solution, with the capacity to represent

121 da Silva and Freitas, "Using Multiple Views," 34.

122 See The Protégé Ontology Editor and Knowledge Acquisition System, "Home," Stanford Center for Biomedical Informatics Research, accessed 25 October 2012, http://protege.stanford.edu.

123 See "Quantum4D," accessed 3 February 2014, http://www.youtube.com/watch?v=cohxdqDqO98. At the time that CIFER and Quantum4D agreed to collaborate on this project, Quantum4D was a rising start-up company that was working with SWIFT and several financial institutions, and which subsequently received several rounds of start-up venture funding. However, as of February 2014, the company had ceased operation and its assets were placed on hold; our model is no longer accessible online. This, in itself, makes an interesting story about digital innovation, the commercialization of labware, digital preservation, and open source software. What this means, however, for continued use of Quantum4D experimentation with Third Order archival systems remains to be seen. Fortunately, the basic concepts and approaches implemented in Quantum4D transcend this specific software. Any other platform with similar functionality could be used to prototype further development of the ideas. Ideally, the archival community should be looking to create its own open source alternative to support Third Order systems such as that prototyped using Quantum4D.

instances in many ways, and an ontology builder/editor that structures and controls the underlying abstractions.

Quantum4D had three basic constructs: objects, relations, and inter-linking spaces. To create our interactive visualization of the Canadian context of financial electronic records, we first created a "space" in our model called the Canadian financial system. Within that space, we mapped each class of entity in our domain model to an object in Quantum4D. We then created a hierarchical relationship between each class of entity and the higher-order entity (as per the class hierarchy in our modelling grammar – see Figure 1). In our "institutional" view (Figure 2), instances of Canadian banks are banks, which, in our modelling grammar, are types of institutions with specific roles, i.e., the role of financial intermediary, or bank. We then uploaded a file containing all the names of Canadian banks into Quantum4D to generate a 3D ring-shaped graphical representation of all the named banks.

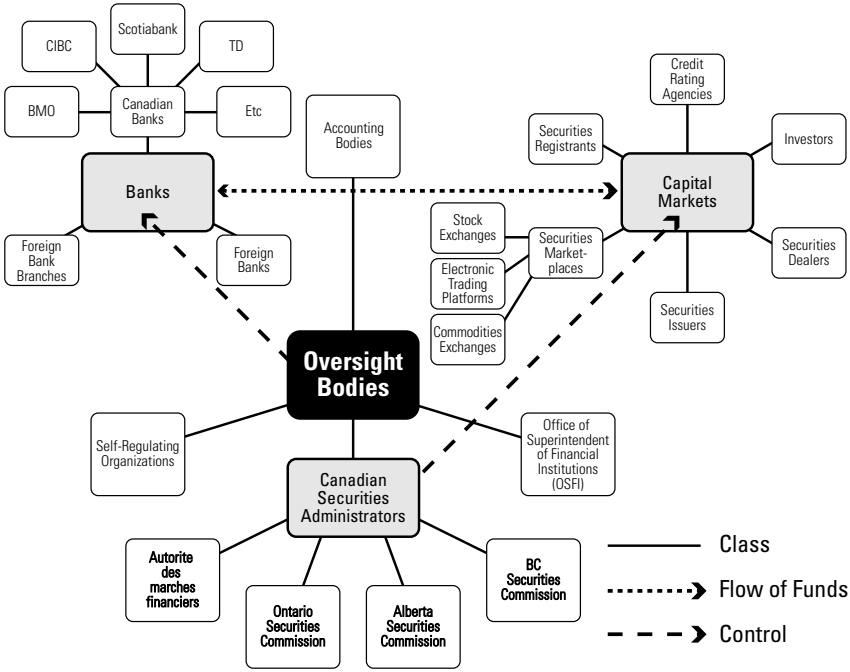

Figure 1. Canadian Context of Financial Electronic Records Model, showing a node-link representation of the class hierarchy.

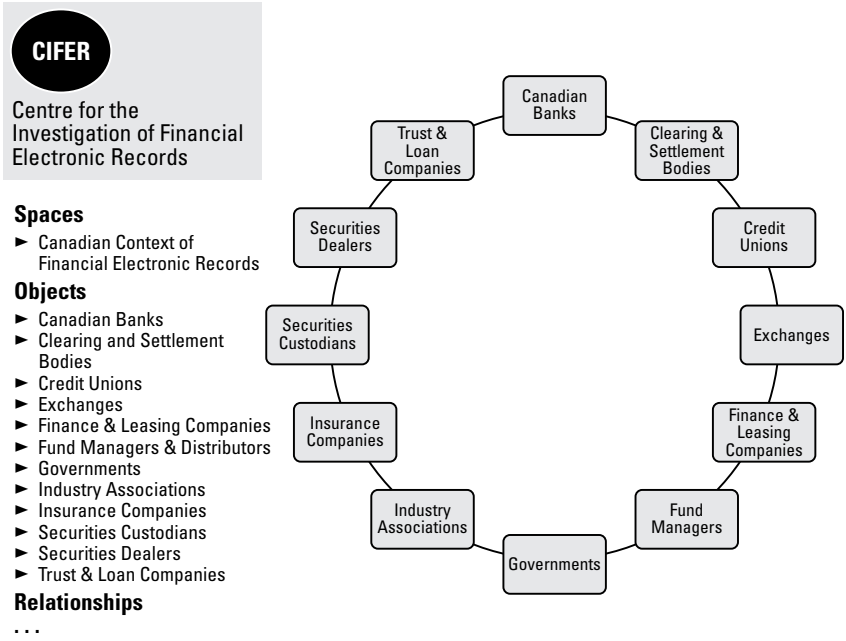

Figure 2. "Institutional" 3D ring view.

Each financial institution or other entity in the financial system is connected to other entities via a network of transactional relationships (or, in our modelling grammar, financial instruments). These we have not defined other than conceptually, owing to the fact that we do not yet have a full theoretical understanding of these relationships and do not have access to instance data to represent them (e.g., we had no data on financial transactions between financial institutions and other entities at the time of developing our model). Functions, as mutual properties of financial instruments rather than as first-order constructs, can be represented by colour-coding each instance to indicate the "role" or transaction type of the instance. Thus, equities transactions can be represented by one colour and fixed-income trades by another colour. In the resultant model, users are able to navigate through the spaces to achieve "Context+Focus" – for example, by zooming in and out of specific areas of the visualization. The higher-level spaces provide an overview of the Canadian financial system, while the lower-level spaces are intended to provide details about specific institutions and transactions. In addition, users are able to rotate the 3D visualizations for greater visibility of objects. Finally, Quantum4D allowed users to represent the objects and relationships within spaces using other visual metaphors, called lenses (e.g., bar charts, graphs, etc.).

A "Functional" View

We also created a "functional" view, shown in Figure 3, of the Canadian context of financial electronic records, which represents the various roles that different business units within financial institutions may have and, in contrast to our "institutional" view, gives functions greater primacy as a first-order construct in the representation. This view is very closely aligned with traditional abstractions and representations of archival records; that is, abstractions and representations that are broken down according to functions, activities, processes, and transactions. For the "functional" view, we followed a traditional archival functional analysis methodology to create a 3D spring-loaded network graph manually, where a node represents a function or sub-function, and each edge a hierarchical relationship between a higher class of functions and sub-classes of functions (e.g., a line in the graph links Investment Banking with its two sub-functions, Corporate Finance and Trading & Sales). Though we did not create this type of representation, Quantum4D theoretically allowed us to link aggregates or instances of transactional data to particular banks within the "institutional" view and across spaces to the "functional" view to specific functions (or roles) so as to be able to characterize the functional nature of each financial transaction in a manner that differs from the method that uses colour, suggested above.

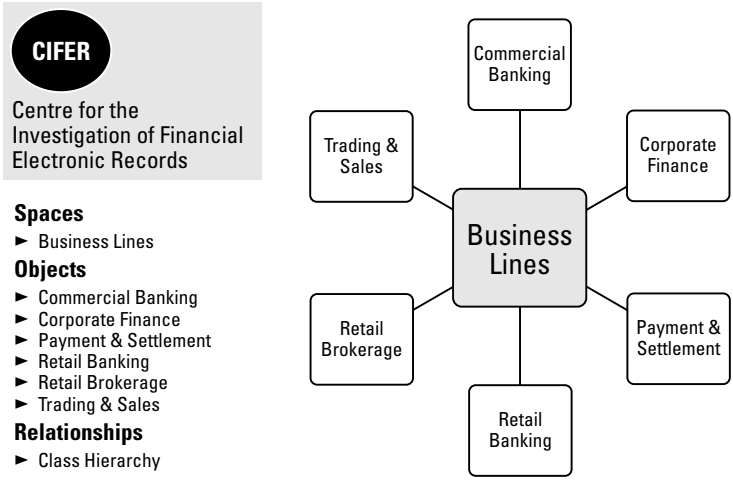

Figure 3. "Functional" view. Each one of the boxes represents a separate space that can be entered to view the sub-functions relating to the higher-level function.

Conclusion: From Experimentation to Interactive Third-Order System

The work on creating our prototype of a high-level interactive visual reference model of the Canadian context of financial electronic records has contributed a number of theoretical and practical insights into questions of abstraction and representation of records and archives. Among the theoretical contributions of this research is the use of ontology theory as a theoretical basis for conceptualizing of societal provenance. Ontologies provide a rich semantic framework for expressing the complex relationships that comprise provenance. In relying on ontology theory to articulate social provenance, we argue for, and demonstrate, how both a substantial upper-level ontology (i.e., Bunge's ontology) and a social ontology (i.e., Searle's ontology) are needed to express the complex reality of financial systems semantically as context for financial records. Use of substantive versus observer-relative upper-level ontologies remains a controversial area of research in information systems theory, and one deserving of continuing research.[124] In particular, we would like to conduct further research to evaluate whether our ontology does indeed offer expressiveness that other financial domain ontologies lack, as a means of making the case for combined use of substantive and social ontologies. We also demonstrate how systems and network theory can provide a theoretical framework for determining the boundaries of abstractions and representations of societal provenance. Finally, we use ontology theory to extend conceptualizations of records as representations, particularly Geoffrey Yeo's conceptualization of records as representations of occurrents, to arrive at new theoretical insights about records as emergent properties of information systems.

On a practical level, we have built a prototype of a third-order system in the form of our high-level interactive reference model of the Canadian context of financial electronic records. In creating our prototype reference model, we have demonstrated a manual "top-down" methodology for building a domain ontology to represent societal provenance, using financial records as an example. As our process was a time-consuming and manual one, we would also like to explore automated techniques for ontology development and evaluation of different approaches to ontology building in our future work.[125]

Much work remains to be done to complete our reference model – in par-

124 See, for example, Gove N. Allen and Salvatore T. March, "A Critical Assessment of the Bunge-Wand-Weber Ontology for Conceptual Modeling," *16th Annual Workshop on Information Technologies & Systems (WITS) Paper*, accessed 20 May 2013, http://dx.doi.org/10.2139/ssrn.951803.

125 There is pre-existing work in the area of automated ontology development that can be drawn upon. See, for example, Gómez-Pérez and Manzano-Macho, "An Overview of Methods and Tools for Ontology Learning from Texts"; and Pulido et al., "Identifying Ontology Components from Digital Archives."

ticular rebuilding our model using an open source ontology visualization tool. Other work includes the following: continuing to articulate the ontologically fluid classes of entities that comprise the Canadian financial system; linking these classes of entities to visual representations of time in our model so as to be able to show dynamic changes in the structure of the system; identifying specific instances of classes of entities and importing data about these instances into the visual representation; developing a sub-space to represent different types of financial instruments to provide an instrument-focused view of the Canadian financial system; and, finally, what may turn out to be one of the most challenging aspects, defining the different types of relationships between all of the objects in our model and gathering data related to instances of these. We also acknowledge that in conceptualizing our ontology as representing societal provenance, we have not yet explored other aspects of the context of records associated with expanded notions of the principle of provenance, such as Laura Millar's creator history, records history, and custodial history.[126]

In spite of these limitations, we see potential for our high-level reference model, or at least the future vision that we have for it, as a framework for the kind of third-order system referred to in Geoffrey Yeo's recent article.[127] In our vision of how this could work, we see interactive visual interfaces that represent domain ontologies as providing rich contextual information about the domain space that permits the user to view that domain space from a traditional hierarchical perspective, a networked perspective, or possibly to reconfigure the records into a multiverse of representational forms or visual metaphors in order to support visual exploration. Archives could link the records in their custody to specific domain elements (i.e., a particular node or an edge) within the visual model so that end users could see where the records fit into the domain, resulting in a better understanding of the context of the records. At the same time, these domain models would serve to represent elements of the internal structure of the records. Interactive features within the interface could provide users with the functionality to zoom into specific aggregations of records and to conduct exploratory visual analysis on demand. Our ultimate aim is to build an interactive visual platform for combining large, heterogeneous financial data sets into a holistic, descriptive system and end-user interface framework with which to derive faster and better insights about financial records. This should ultimately be possible, though we are a long way from having the full conceptual and technical capability to realize our vision. As Geoffrey Yeo suggests:

> we can respect provenance in ways that do not depend on the stability of particular orderings of the world. Far from feeling threatened by the fluidity of the "third order," archivists and records managers should be able to take advantage of the new capabili-

126 See Millar, "The Death of the Fonds and the Resurrection of Provenance."
127 Yeo, "Bringing Things Together."

ties to help overcome some of the contextual limitations of hierarchical classification schemes and paper-world methods.[128]

One approach to continuing to explore domain ontology-based interactive visual interfaces as the foundation of archival third-order systems would be to convene an archival working group along the lines of the former Canadian Working Group on Descriptive Standards, which led to the establishment of the *Rules for Archival Description (RAD)*. This group could work together to plan for and guide the elicitation of domain ontologies for the representation of archives using a combination of automated and manual techniques. This need not be an exercise in starting from scratch, as many domain ontologies already exist that could provide relevant points of departure for archivists. For example, one domain ontology that our research team has had some involvement with is the Object Management Group (OMG)/Enterprise Data Management Council Financial Industry Business Ontology (FIBO).[129] The objective of the work of the group developing FIBO is to make financial data more open and accessible in support of better financial risk analytics. We see potential for adaptation of this ontology for archival purposes, and for collaboration with financial domain experts to provide archivists with an opportunity to leverage the descriptive work of records creators and managers in a manner consistent with continuum approaches. With the rise of many "open data" initiatives, this approach provides a framework to support the linking of archival data, as, for example, in a manner implemented by the Europeana Linked Open Data Pilot.[130]

Other systems or networks that might be amenable to the work begun by our research team include health systems, governance systems, and educational systems. As these are not mutually exclusive but rather overlapping areas, in many respects these ontologies could be formalized using upper-level ontologies, such as those described by Bunge and Searle, in order to establish consistent semantics that would support semantic "crosswalks" between ontologies, linking them in a coherent, neural network-like framework of interconnecting ontologies. We further see this approach as being supportive of a web of functional authorities that provide contextual information about different documentary forms or types of records.[131]

128 Yeo, "Bringing Things Together," 61.

129 Enterprise Data Management Council, "Financial Industry Business Ontology," under "Projects," accessed 25 October 2012, http://www.edmcouncil.org/financialbusiness.

130 Bernhard Haslhofer and Antoine Isaac, "data.europeana.eu: The Europeana Linked Open Data Pilot," *Proceedings of the International Conference on Dublin Core and Metadata Applications, North America, August 2011*, 94–104, accessed 20 May 2013, http://dcevents.dublincore.org/IntConf/dc-2011/paper/view/55/14.

131 See Madeleine McLuhan-Myers, "The Archival Web: Contextual Authority Files and the Representation of Institutional Textual Documents in Online Description" (master's thesis, University of Manitoba, 2012), http://hdl.handle.net/1993/8464.

This approach, we believe, also has the power to integrate representations of organizational and personal records, as the records of individuals could be analyzed in the same iterative fashion as we analyzed the Canadian context of financial electronic records. This analysis would result in the creation of ontological representations of an individual's domain, connecting the classes of entities in the individual's domain to those in other ontologies via the aforementioned crosswalks. Instead of drawing on theories of systems or networks, as we have done for financial records, personal records ontology development might use alternate approaches, drawing on the work of DiSalvo, which demonstrates how information and communication technologies, especially those meant to engage users in participation, can be effectively informed by recognizing and constituting "publics."[132] DiSalvo, in turn, draws on the work of John Dewey concerning how people organize around collective action. For Dewey, a public is brought into existence by action around a shared social condition, through mobilizing either to mitigate or promote its consequences.[133] This approach shares a focus on the notion of community found in the writings of Jeannette Bastian, who calls for archivists to embrace more expansive notions of context as community and community as context.[134]

One issue with the production of domain ontologies is that they are generally developed by an individual or group analyzing real-space (or a corpus of texts that stand in as proxies of real-space) to extract the ontologically representative classes and class hierarchies. Whenever people come together to build a consensus view, it is a time-consuming process and there is room for disagreement, a problem that is only partially resolved with automated techniques for ontology development.[135] Moreover, what may be ontologically correct in one geopolitical

132 Carl DiSalvo, "Design and the Construction of Publics," *Design Issues* 25, no. 1 (2009): 48–63. For an example of a project putting these ideas into practice in the design of information systems, see Christopher A. Le Dantec, Jim E. Christensen, Mark Bailey, Robert G. Farrell, Jason B. Ellis, Catalina M. Danis, Wendy A. Kellogg, and W. Keith Edwards, "A Tale of Two Publics: Democratizing Design at the Margins," in *Proceedings of the 8th ACM Conference on Designing Interactive Systems* (New York: ACM, 2010), 11–20.

133 John Dewey, *The Public and Its Problems* (Ohio: Swallow Press, 1954).

134 Jeannette Allis Bastian, "Reading Colonial Records Through an Archival Lens: The Provenance of Place, Space and Creation," *Archival Science* 6, no. 3–4 (December 2006): 267–84.

135 Shilton and Srnivasan, "Participatory Appraisal and Arrangement." The authors warn of these risks: "Of course, we cannot proclaim that there are no risks associated with the participatory model which we propose. Participatory approaches to archiving are likely to be time-consuming, requiring patience and an extended commitment by archival staff and community representatives alike. Community members do not always agree, and decisions on such difficult matters as authorship and relationship between narratives are not always easy to reach. And at a time when archival backlogs have risen to problematic levels, expenditures of greater amounts of time for appraisal and processing warrant serious consideration" (p. 100). On the limitations of automated techniques of ontology development, see Gómez-Pérez and Manzano-Macho, "An Overview of Methods and Tools for Ontology Learning from Texts."

or legal jurisdiction may not be true in another (e.g., a class of instances that are "legal persons" in one jurisdiction when the concept of legal persons does not exist in another jurisdiction). Thus, there will be a need to build flexibility into the abstractions we use in order to allow for alternative perspectives of different users, by use of extensions, for example. This approach moves archival representation in the direction of what many archivists have been calling for and, indeed, have experimented with, in relation to archival knowledge representation; namely, to open it up to the users of archives. Shilton and Srinivasan articulate the vision well when they write:

> Once a map of community ontologies is created, an online organizational system for our archive can be built. Future online participants in the South Asian Web can use the community ontologies as a jumping-off point to arrange and describe their uploaded representations and records. And through an ongoing, iterative, and emergent ontology-building process, wherein participants are asked to identify connections between their records and the records of others, future participants can continue to revise the community ontologies as the community itself changes over time.[136]

Ontologies, as a technology to support interoperability and translation of the perspectives of different users, would support a more open approach to archival abstraction and representation than has been the case in the past.

As ontologies tend to be large and complex, visually representing them, as opposed to listing or textually describing constructs, relationships, and instances, allows for leveraging of the unique aspects of human visual cognition and perception that underpin visual analytics to detect patterns and relationships and, ultimately, to extract meaning from large knowledge bases. As da Silva and Freitas observe, however, static graphs, commonly used for ontology representation, are not the best choice for such visualizations.[137] Thus, as our experimentation has uncovered, visualization tools require the functionality not only of ontology builders and editors, but also the functionality that allows for multiple views or metaphoric representations (i.e., hierarchical node-link graphs, network graphs, 3D rings, or flows), which permit the user to explore different visual projections of a knowledge base. Though Quantum4D, the tool with which we conducted our experiments in the creation of our high-level interactive reference model, proved to have a good deal of the functionality we were looking for, we did find it lacking in certain respects; for example, it lacked expressiveness in the variety of relationship types that can be created and how these can be combined. The "killer app" in ontology viewing has yet to be developed, so this remains an area of ongoing research for our team.

136 Shilton and Srinivasan, "Participatory Appraisal and Arrangement," 99.
137 da Silva and Freitas, "Using Multiple Views," 34.

Our final analysis concludes that the complex world of finance – comprising considerations of societal provenance and records – can be viewed through the lens of the hierarchy, the system, the network, or the flow; these are all just metaphors, or representations, that must not be confused with reality. Each approach captures aspects of that reality and allows us to see facets of a domain of interest that would remain unobserved through some other lens. Hence, the best representations or models will be dynamic, offering the ability to view societal provenance and records from many dynamic and emergent perspectives. This is a very difficult modelling challenge, particularly when the objects in the domain of analysis are ontologically fluid. Our high-level reference model of the Canadian financial system offers only an initial, exploratory step in this direction.

Victoria L. Lemieux *is an associate professor in the School of Library, Archival and Information Studies and former director of the Media and Graphics Interdisciplinary Centre, both at the University of British Columbia. She is currently on leave at the World Bank. Lemieux is also director of the Centre for the Investigation of Financial Electronic Records, which she established in 2008. Her interest in financial records and their relationship to risk stems from her 1999–2001 doctoral research on the information-related causes of the Jamaican banking crisis (University College London, 2002).*