

Describing Scholarly Information Resources with a Unified Temporal Map

Robert B. Allen , Hanna Song, Bo Eun Lee, and Jiyoung Lee

Yonsei University, Seoul, South Korea
{rballen,hnsong,belee}@yonsei.ac.kr, jiyoung.lee@simmons.edu

Abstract. We consider the use of procedures for providing structured descriptions of information resources such as scholarly works and of their contents. This goes beyond the usual view of metadata as discrete elements. For instance, we consider mapping the structured and interdependent activities in the publication of Ulysses. We discuss some specific representations and discuss the development of structured scholarly guides. Finally, we consider how the activities associated with publication, along with other historical activities, can be positioned on a unified temporal map. Ultimately, there should be a unified framework for the description of individual information resources and collections of information resources across periods and technologies.

Keywords: Archives · Metadata · Scholarly records · Structured continuum · Structured scholarly guides · Threads

1 Procedures and Information Artifacts

Formal modeling of processes and procedures goes back at least to Herbert A. Simon's research on business decision making and human cognition. That work led to studies of office and organizational processes and eventually to the development of object-oriented programming languages. Petri Nets are a specific modeling technique that are the key to many formal approaches to modeling procedures. Petri Nets set conditions for the gating of transitions in a procedure and generally require that all conditions are met before a procedure can proceed. Thus, Petri Nets provide a way to specify and manage flow control. They are the foundation of tools such as UML Activity Diagrams and workflow specifications such as Taverna.

There would be many advantages to combining a powerful content ontology with computer-based procedure specification. [4] attempted to reconcile the notion of Procedures derived from the programming language community with the notion of Process from the Basic Formal Ontology (BFO) [6]. That work extended the BFO notion of Process, which is conceived of as a sequence of present participles (e.g., racing, running, walking), to include descriptions of how the transitions are connected. Procedures are viewed as Processes which include some sort of condition, contingency, or gating. This broader perspective considers interaction with other entities in the context of Scenarios

and attempts to specify the causes of the transitions within and across Processes and Procedures.

The production, publication, and use of information resources follow many procedures. Capturing these procedures, using them to organize the materials, and then using them to support access should have substantial benefits over current approaches. The procedures would weave together what might otherwise appear as disjoint events. When supported with rich historical knowledge sources and powerful inferencing tools, the procedures should provide a useful perspective for many different levels of scholarship.

The earliest formal work for procedures associated with information resources was by [18] who used Petri Nets to model the procedures for the “office of the future”. While there can be difficulties in taking office procedures too literally [15], such specifications are now fairly common. An emphasis on business processes also became established as a core principle in the archives community, which argued that archival records should be organized by their business or organizational purpose. However, the descriptive systems that have been developed for records management were based on static metadata models [12]. Some recent work is beginning to address the preservation of the temporal aspects of business processes. [1] gives an example using a business Role Activity Diagram (RAD) to structure and then support visualization of project-management records. The notable and extensive work of Rauber and colleagues (e.g., [10, 11]) describes preservation of templates for business processes and scientific workflows along with consideration of the reuse of those procedures.

In many cases, procedures have a simple path and straightforward results. In other cases (e.g., a student who juggles schoolwork, family, and a job), there may be many interacting procedures in which some of the procedures may be blocked and suspended and then restarted later. We can distinguish these as closed and open worlds from which the models are drawn. We see both situations in descriptions of scholarly activity. The prototype procedure for processing conference or journal papers assumes a closed world while post-hoc descriptions, which are typical of scholarly editions or other histories, are open worlds with many interacting factors. Just as Petri Nets can be used to coordinate the parts of an individual procedure, they may also be helpful in post-hoc descriptions of interacting procedures.

2 Scholarly Records and Modeling Scholarly Activity

Like other complex human activities, scholarly authoring, publishing, and commentary are threads across time (e.g., [13]). Systems for organizing individual scholarly works and collections of works should go beyond traditional metadata and capture these temporal threads or events. This approach should allow us to describe a wide variety of technologies and media and to provide richer data than traditional approaches that deliver decontextualized metadata.

We believe that procedure descriptions should be used to develop a new generation of scholarly description. The production of scholarly materials is, generally, highly

sequential and routine. It should be helpful to capture that routine. When there are failures in the routine or other exceptions, we can highlight them by contrasting them to the expected procedures.

Consider the possible variations in the route of a modern research article before it is “published” in a journal. Descriptions of the research may appear as a technical report, be posted on an open-access repository, and/or be presented at a conference. Even within the context of a journal, the article may appear in an online repository long before it appears in print and arrives on a subscriber’s desk or in a library. In short, the very meaning of publication and publication date are increasingly broadened. Yet, accurate records are, ultimately, essential for the credibility of scholarship.

Scholarly records are used in many ways. The most common is to assign appropriate credit to the discoverers of new insights and ideas. In other cases, we want to better understand the nuances and creative pressures in works. This has often been done retrospectively through scholarly editions. Scholarly editions are an important resource for research. There are extensive recommendations for the development of scholarly editions¹ and a recent statement about the development of digital scholarly editions² but systematic approaches have not been proposed.

Beyond publication of scholarly texts, there is an increasing number of other scenarios with a temporal or causal component (e.g., performance or of new musical works) that are important for scholarly records, which may benefit from a systematic approach. Literature, for example, is an area in which complex publication histories are common. The novel *Ulysses* by James Joyce has a particularly complex history. There were many editions with many revisions. Moreover, its distribution was disrupted by court battles. This is an example of publication in an open-world context in which many, sometimes unexpected, procedures interact.

3 Next Steps

3.1 Standard Sets of Procedures

Scholarly description has separate traditions and approaches across archives, literary studies, intellectual history, librarianship, and even genomics [8] that all have common traits. For instance, both archives and literary theory place great emphasis on understanding the context (i.e., “original order”) in which works are created. We believe that especially with the many similarities across the fields both conceptual frameworks and specific procedures should be better coordinated.

Our emphasis on standardization comes from an information science perspective rather than a humanities perspective. However, just as there are clear benefits from

¹ Modern Language Association, Guidelines for Editors of Scholarly Editions. Last revised 29 June 2011, <https://www.mla.org/Resources/Research/Surveys-Reports-and-Other-Documents/Publishing-and-Scholarship/Reports-from-the-MLA-Committee-on-Scholarly-Editions/Guidelines-for-Editors-of-Scholarly-Editions>.

² MLA Statement on the Scholarly Edition in the Digital Age, MLA Committee on Scholarly Editions, May 2016, <https://www.mla.org/content/download/52050/1810116/rptCSE16.pdf>.

accepting the EAD and TEI standards, we believe there would be benefits from adoption of standard descriptions of scholarship based on procedures. Moreover, once standard frameworks are developed, they could be populated with data and would become valuable resources in their own right.

There are many procedures involved in the production and publication of a scholarly work. We would need to categorize then and develop a flexible framework for coordinating those variations. There is an even greater challenge in developing a broad set of procedures for an open-world approach. For the immediate future, guidelines could be proposed for developers to use initially while a standard set of procedures is developed to cover general cases. Ultimately, a flexible framework such as coordination theory [9] could be applied to less structured situations even including the process of authoring.

3.2 From Scholarly Editions to Scholarly Guides

There may be many scholarly editions of a single work. Many works are published in a traditional bound paper format and in multiple electronic formats. One web site with an ad hoc list of scholarly electronic editions has 399 entries³. As noted above, there is no standard structure even for the core elements of a scholarly edition. Some editions have many types of material beyond the annotations to the text itself. Moreover, this additional material is unstructured. The situation with scholarly editions seems analogous to archival finding aids before the EAD [7] was introduced. The EAD has proven extremely useful in the coordination of archival materials by providing an XML DTD that provides a common structure across finding aids. We believe there would be similar benefits in standardizing scholarly editions with consistent sections and data structures. We call these standardized versions of scholarly editions Structured Scholarly Guides. Once a general framework is established, it could be applied to support applications such as structuring digital articles and guides, such as Wikipedia, about books and scholarly works.

3.3 Unified Temporal Map and Interactive Services

An important aspect of our proposal is that procedures could be mapped into a single over-arching framework. Not only would the creation and production of scholarly works be mapped but post-publication events such as reviews and citations could all be included. This may be thought of as providing a Continuum approach [16, 17] rather than a finite lifecycle approach to records about scholarly publishing.

Beyond enabling viewing histories of individual works, a unified temporal map could allow coordinated views of several works by the same author, by the same publisher, or by authors working in a specific media or in specific geographical regions. Indeed, a broad range of other historical events and procedures could all be coordinated into one view (e.g., [3]). In addition, the temporal knowledgebase could be linked to other factors such as genres, technologies, and cultural and intellectual trends that may be only

³ A catalog of Digital Scholarly Editions, v 3.0, snapshot 2008ff, compiled by P. Sahle, last change 2016/05/19, <http://www.digitale-edition.de/>.

indirectly based on temporal factors. Views of individual works could be implemented as threads which are highly interconnected sub-sets of sub-views of the broader map.

The rich models proposed here should be able to support several types of structured interaction. For instance, scholarly procedures could be mapped onto interactive timelines (e.g., [2]). In addition, there has been considerable discussion about the possibility of interactive digital scholarly editions (e.g., [14]). Our approach suggests some services that should be useful for such editions. In addition, as we noted above, there are similarities between archival finding aids and structured scholarly guides. Interactive versions of those guides could be implemented which would be analogous to the interactive archival finding aids proposed by [5].

4 Conclusion

We have proposed the use of procedures for the description of information resources that goes far beyond scholarly editions and requires a broader integration of approaches. We have proposed initial steps toward a new structure for organizing descriptions of scholarly material. Rather than having bibliographic description as a set of disjoint and decontextualized metadata categories, we propose focusing on the continuity and the inter-relatedness of activities. Our solution is to provide process-based descriptions that allow richer structures. If needed, those richer descriptions could be reverted to basic category labels. Evidence and argumentation about evidence could be supported with “structured applied epistemology” [3]. Based on the structured description of processes, we foresee a unified framework for the description of individual information resources and collections of information resources that covers different periods and technologies.

References

1. Allen, R.B.: Using information visualization to support access of archival records. *J. Archival Organ.* **3**, 37–49 (2005). doi:[10.1300/J201v03n01_04](https://doi.org/10.1300/J201v03n01_04)
2. Allen, R.B.: Visualization, Causation, and History. In: *iConference* (2011). doi:[10.1145/1940761.1940835](https://doi.org/10.1145/1940761.1940835)
3. Allen, R.B.: Issues for the Direct Representation of History. In: Morishima, A., Rauber, A., Liew, C.L. (eds.) *ICADL 2016. LNCS*, vol. 10075, pp. 218–224. Springer, Heidelberg (2016)
4. Allen, R.B., Chu, Y.M.: Architectures for complex semantic models. In: *IEEE Conference on Big Data and Smart Computing*, pp. 254–261 (2015). doi:[10.1109/35021BIGCOMP.2015.7072809](https://doi.org/10.1109/35021BIGCOMP.2015.7072809)
5. Anderson, S., Allen, R.B.: Malleable Finding Aids. *TPDL*, pp. 402–407 (2012). doi:[10.1007%2F978-3-642-33290-6_43](https://doi.org/10.1007%2F978-3-642-33290-6_43)
6. Arp, R., Smith, B., Spear, A.D.: *Building Ontologies with Basic Format Ontology*. MIT Press, Cambridge (2015). <http://purl.obolibrary.org/obo/bfo/Reference>
7. DeRose, S.J.: Navigation, access, and control using structured information. *Am. Archivist* **60**, 298–309 (1997). doi:[10.17723/aarc.60.3.0777u1361u62tp6](https://doi.org/10.17723/aarc.60.3.0777u1361u62tp6)
8. Huntley, R.P., Sawford, T., Martin, M.J., O’Donovan, C.: Understanding how and why the Gene Ontology and its annotations evolve: the GO within UniProt, *GigaScience*, **3** (2014). doi:[10.1186/2047-217X-3-4](https://doi.org/10.1186/2047-217X-3-4)

9. Malone, T.W., Crowston, K.: The interdisciplinary Study of coordination. *Comput. Surv.* **26**, 87–119 (1994). doi:[10.1145/174666.174668](https://doi.org/10.1145/174666.174668)
10. Mayer, R., Antunes, G., Caetano, A., Bakhshandeh, M., Rauber, A., Borbinha, J.: Using ontologies to capture the semantics of a (business) process for digital preservation. *Int. J. Digital Librar.* **15**, 129–152 (2015). doi:[10.1007/s00799-015-0141-7](https://doi.org/10.1007/s00799-015-0141-7)
11. Mayer, R., Proell, S., Rauber, A.: On the Applicability of Workflow Management Systems for the Preservation of Business Processes, *iPres* (2012). http://www.ifs.tuwien.ac.at/~mayer/publications/pdf/may_ipres12-workflows.pdf
12. McKemmish, S., Acland, G., Ward, N., Reed, B.: Describing Records in Context in the Continuum: the Australian Recordkeeping Metadata Schema, *Archivaria*, (48) (1999)
13. Rico, F.: Scholarly editions and real readers. *Variants* **5**, 1–13 (2006)
14. Robinson, P.: Current Directions in the Making of Digital Editions: Towards Interactive Editions. *Ecdotica* **4**, 176–191 (2007)
15. Suchman, L.: *Plans and Situated Action*. Cambridge U. Press, Cambridge (1987)
16. Upward, F.M.: Structuring the Records Continuum, Part One: Postcustodial Principles and Properties. *Archives and Manuscripts* **25** (1997). <http://www.infotech.monash.edu.au/research/groups/rcrg/publications/recordscontinuum-fupp1.html>
17. Upward, F.M.: Structuring the Records Continuum, Part Two: Structuration Theory and Recordkeeping. *Archives and Manuscripts*, **25** (1997). <http://www.infotech.monash.edu.au/research/groups/rcrg/publications/recordscontinuum-fupp2.html>
18. Zisman, M.D.: *Representation, Specification and Automation of Office Procedures*. Ph.D. thesis, University of Pennsylvania Wharton School of Business, Pennsylvania (1977)