Playing Well with Others: The Social Edition and Computational Collaboration

Constance Crompton & Cole Mash
University of British Columbia, Okanagan

Raymond Siemens
University of Victoria

INKE Research Group

Abstract

This article draws on the Social Edition of the Devonshire Manuscript's RDFa encoding practice as a case study of how to formalize statements about entities on the Web in a way that is machine-parsable. RDFa encoding allows machines to become collaborators with human readers in the discovery of new connections between entities (people, places, and events) even between websites. The edition's encoding is motivated by the INKE Modelling and Prototyping team's guiding research question about the implications and impact of real-time applications in relation to traditionally static knowledge objects. The authors argue for the value of bringing texts into communication with other texts, through RDFa, allowing virtual collaboration even when the scholars behind the projects do not know one another.

Keywords

Social knowledge creation; Social edition; Linked data; Collaboration

Constance Crompton is Assistant Professor in the Department of Critical Studies at the University of British Columbia, Okanagan, 1148 Research Road, Kelowna BC. Email constance .crompton@ubc.ca .

Cole Mash is a master's student in Creative Studies at the University of the Okanagan, 1148 Research Road, Kelowna BC. Email: coleallenmash@gmail.com .

Raymond Siemens is Professor in English at the University of Victoria, 3800 Finnerty Road, Victoria BC. Email: siemens@uvic.ca .

Implementing New Knowledge Environments (INKE) is a collaborative research intervention exploring electronic text, digital humanities, and scholarly communication. The international team involves over 42 researchers, 53 GRAs, 4 staff, 19 postdocs, and 30 partners. Website: inke.ca

Collaboration between scholars drives large research projects and can increase a project's reach by bringing together a wider group of engaged traditional and citizen scholars to collaborate. This article investigates the use of microdata formats to extend that reach, with the goal of bringing in a wider group of engaged researchers and editors. At the 2013 INKE Birds of a Feather gathering, William Bowen and Constance Crompton made a case for selecting a publication content management system for collaboration, based on the preferences of the community (in our case the preferences expressed by the Social Edition of the Devonshire Manuscript advisory group).[1] In response to community preferences, the Devonshire Manuscript Editorial Group (DMSEG)[2] turned to CommentPress, a WordPress plug-in by the Institute for the Future of the Book, which leaves the manuscript's poems static but lets community add commentary, in keeping with the advisory group's preferences. This new iteration of the Social Edition of the Devonshire Manuscript has been launched in Iter Community, a social space and publication platform for Early Modern and Renaissance scholarship, at http://dms.itercommunity.org .

The DMSEG, however, is not solely encoding for direct collaborators and editors interested in adding commentary. The project not only aims to engage human readers, but also provide readily parsable data about the content of the edition. In the interest of serving machine readers[3] (search algorithms, inferencing engines, etc.) and the human readers who employ them, this current phase incorporates Resource Description Framework in Attributes (RDFa) into the Social Edition of the Devonshire Manuscript not only to allow for structured data extraction and algorithmic inferencing about the relationships between the texts and contributors to the Devonshire Manuscript (BL MS Add. 17, 492), but also to build new knowledge from information in both the social edition and other digital scholarship about the sixteenth century. This article explores another facet of collaboration in a digital age, the adoption of standards that let machine readers disambiguate real-world entities referenced in the text from other entities (e.g., people or places) with the same names. Motivated by the INKE Modelling and Prototyping team's guiding research question about the implications and impact of real-time applications in relation to traditionally static knowledge objects, we argue that, far from stifling creativity, adopting linked data standards, like RDFa, even at the prototyping stage, creates the conditions to bring texts into communication with other texts, allowing virtual collaboration across projects, even when the scholars behind the projects do not know one another. Machine readers can extract connections between the content in disparate RDFa encoded projects, in short allowing one project's texts to "play well" with other encoded texts. Following an outline of the process that led to the selection of WordPress and CommentPress in the creation of the social edition within Iter Community, we reflect on the promise of RDFa, describe the process of using RDFa microdata to meet the needs of machine readers, and conclude by providing the results of our experiments in engaging with RDFa while attempting to address the advice of our advisory group and suggested directions for future research.

Community-informed platforms for collaboration

Devising appropriate protocols for collaboration between readers of all types requires integration and incremental development. The DMSEG has engaged in iterative development, creating the Social Edition of the Devonshire Manuscript in Wikibooks,

seeking direction from the advisory group to work out the best ways to meet the needs of the community of Early Modern and Renaissance scholars that the group represents (Siemens et al., 2012). The advisory group applauded both the content and uptake of the Social Edition of the Devonshire Manuscript in Wikibooks, but expressed reservations about both the mutability of the text and, for scholars outside of the Digital Humanities, the participation barrier created by having to learn to write wikicode in order to contribute to the edition.

As mentioned above, in response to the advisory group's reservations, the Devonshire Manuscript Editorial Group turned to CommentPress, a WordPress plug-in by Bob Stein's Institute for the Future of the Book that leaves the manuscript's poems static but lets community add commentary, in keeping with the advisory group's preferences. CommentPress has emerged as a standard plug-in to facilitate commenting, and has been used for peer review and commentary by the authors and teams behind publications as diverse as the *Shakespeare Quarterly*, *Planned Obsolescence*, and *Off the Tracks* (see Clement & Reside, 2011; Fitzpatrick, 2011; Rowe, 2010). Its technical specifications aside, broad uptake of CommentPress has helped the DMSEG address one of the concerns raised by the advisory group: since scholars are likely to have come across CommentPress-enabled sites before, they will also likely know that comments are welcome. The CommentPress interface itself encourages commenting. By comparison, the process for adding comments in Wikibooks is rather opaque, and it requires that users leave the editing environment and read documentation in Wikibooks and *Wikipedia*'s special Help namespace pages. CommentPress is self-documenting, telling contributors how to comment without leaving the edition to visit external documentation pages. Furthermore, the plug-in uses a WYSIWYG editor, with features familiar to anyone who has ever used Microsoft Word or other GUI text editors. CommentPress also helps the DMSEG meet the advisory group's final concern: within WordPress the text of the Devonshire Manuscript poems and the critical apparatus remain static, but through CommentPress readers can add commentary on the text, facilitating what the advisory group expects will be a greater trust in the text among readers and contemporary contributors. This is in contrast to the lack of trust many contributors had in the Wikibooks edition as a result of validity questions raised by the open editing practices of Wikibooks, a platform where almost anyone is able to edit the text (even though Wikibooks meticulously records when edits were made and which edits were made by whom).

The promise of RDFa

The CommentPress version of the Social Edition of the Devonshire Manuscript is a born-digital edition, housed on servers at the University of Toronto Scarborough (Siemens et al., 2015). While the distribution of born-digital projects across servers, connected via hyperlinks, constituted a revolution in information distribution and access in the 1990s, the early Web was (and is) more concerned with the appearance of HTML documents to end users than with the consistency and interoperability of Web data. As a result, as Dean Allemang and James Hendler note, "the web often feels like it is 'a mile wide but an inch deep,'" which motivates them to ask, "[H]ow can we build a more integrated, consistent, deep web experience" for web application users?" (2011, p. 2). Encoding for the semantic Web, using systems such as RDFa, is one way to make

sure that machine readers can aggregate information about real-world entities from various locations on the Web (people, places, and events, in the case of the Social Edition, the sixteenth-century contributors to the Devonshire Manuscript) (Berners-Lee, Hendler, & Lassila, 2001). RDFa encoding is conceptually well suited to digital academic projects, since RDF's purpose is reminiscent of traditional scholarship. As Allemang and Hendler remind their readers,

> when two (or more!) viewpoints come together in a web of knowledge, there will typically be overlap, disagreement, and confusion before there is synergy, cooperation, and collaboration. If the infrastructure of the Web is to help us to find our way through the wild stage of information sharing, an informal notion of how things fit together, or should fit together, will not suffice. (2011, p. 22)

Allemang and Hendler recommend the use of RDF to formalize statements about entities on the Web in a way that machine readers can parse, allowing them to become collaborators with human readers in the discovery of new connections between entities, even when the RDFa encoded information about those entities is on different websites. Following Allemang and Hendler's lead, the UBC Okanagan team has taken up the RDFa standard, with its formal modelling principles and ready crosswalks between ontologies, as a prototype for digital publishing that continues to serve computer and machine readers in pursuit of what scholars do best: discovering, comparing, evaluating, annotating, and, in a digital publishing context, collaborating (Unsworth, 2000).

Encoding for the semantic Web is the means of building that integrated, consistent, deep Internet through the use of Web standard languages. The user experience of the semantic Web may already be familiar to those who have seen the information boxes in the upper-right-hand corner of Google search results, or who have used research tools and portals such as Europeana, Out of the Trenches, or Linked Jazz. However, as John Simpson noted at the 2014 INKE Birds of a Feather gathering, semantic data makes up only 1% of the Web (Simpson & Brown, 2014). And yet, as the Web grows, the need for semantic data grows. In order to address this demand, "search engines have started to provide richer search results by extracting fine-grained structured details from the Web pages they crawl" and "publishers are producing increasing amounts of structured data within their Web content to improve their standing with search engines" (Herman, 2010). One of the "key enabling technologies" in this rich result production domain is RDFa (Resource Description Framework in Attributes), markup that adds structured data directly to HTML pages in the form of ontology declarations and specific HTML attributes (Herman, 2010).

Adding invisible ink for computational readers

The DMSEG committed to use the Institute for the Future of the Book's CommentPress plug-in to meet the stated needs of our advisory group, which anchors us to the WordPress platform, leaving us with the challenge of adding RDFa to software built primarily for human readers. Parsers and other machine readers, including search engines, may ignore CDATA content, which includes the microdata we are inserting into the HTML tags on each page of the edition. To solve this problem, we turned to

the RDFa Content Editor (RDFaCE), a WordPress plug-in that allows us to add RDFa to WordPress pages and direct parsers to attend to the RDFa within CDATA, rather than ignore it. RDFaCE was developed and is maintained by the AKSW research group at the University of Leipzig, by Ali Khalili and Sören Auer. The purpose of the plug-in is to support "different views for semantic content authoring and [using] existing semantic Web APIs to facilitate annotating and editing of RDFa contents" (RDFaCE, n.d.). Built on the TinyMCE rich text editor, it enables users to annotate blog posts with RDFa and microdata through a series of user-friendly GUI fields. In adopting RDFaCE, we hoped to offer editors and commenters an interface as easy to use as CommentPress, allowing content experts outside of the Digital Humanities to contribute their expertise to the semantic Web. Ideally, editors and commenters using RDFaCE can highlight the parts of the text they wish to mark up with RDFa, and, with the click of a button, open a list of optional RDF attributes to add to the text. This allows users to classify the selected text as a type of entity (e.g., person, place) and mark it up with metadata about that entity (e.g., parents, siblings, birth and death date). The RDFaCE plug-in then adds the annotations directly into the page's code, and most importantly, exposes the RDFa to machine readers, resolving the parsing problems that may be introduced by CDATA.

Although our final goal is to mark up all the poem commentary in the Social Edition of the Devonshire Manuscript in such a way that would let us trace contemporary citation networks as well as sixteenth-century authoring and annotating habits, we started with a test markup of the edition's biography page. With so many familial relations and overlapping names in Henry VIII's court (Thomas Howard, Devonshire Manuscript contributor, ought not, for example, be confused with his uncle Thomas Howard or his half-brother Thomas Howard), we had a modelling challenge suitable to semantic Web markup, designed, as it is, to disambiguate Web content and connect to existing ontologies and authorities.

On our first experimental pass we encoded person entities and their relations, disambiguated using schema.org's ontology and Ian Davis' relationship ontology in combination with the URIs provided by the Virtual International Authority File (VIAF), the *Oxford Dictionary of National Biography* (ODNB), and GeoNames. Schema.org's *Person*, for example, offered us the following attributes: @name; @uri, which we pointed to each person's Virtual International Authority File; @sameAs, which included their *Oxford Dictionary of National Biography* permalink; @affiliations; @birthDate; @deathDate; @children; @nationality, which we pointed to GeoNames; @parent; @sibling; and finally, @spouse, which we pointed to URIs and permalinks. Finally we used Ian Davis' relationship ontology, also pointing to ODNB and VIAF URIs and permalinks, to clarify the relationships between affiliated Devonshire Manuscript contributors. For example Anne Boleyn makes it clear to a machine reader that the letters A-n-n-e B-o-l-e-y-n refer to a person entity as defined by schema.org at https://schema.org/Person and that this particular person is Anne Boleyn as defined at http://viaf.org/viaf/29521340 and not some other woman of the same name.

As part of the team's test markup, DMSEG research assistant Cole Mash marked up a private, purpose-built sample page that had examples of all the entity types, fields, and

Crompton, Constance, Mash, Cole, & Siemens, Raymond. (2015). Playing Well with Others: The Social Edition and Conceptual Collaboration. *Scholarly and Research Communication, 6*(3): 0301111, 9 pp.
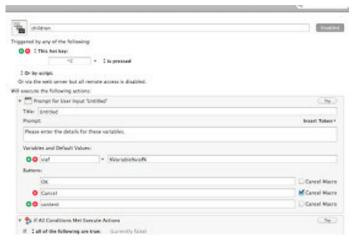
5

relationships in the manuscript. His goal was to test the RDFaCE plug-in, still in beta, to assess whether it would be suitable for long-term use by the social edition project. RDFaCE was supposed to mark up the sample page with all of the RDFa for Cole, saving him from having to enter each piece of code himself; however, the plug-in did not work as planned. He found that RDFaCE would not save all of the attribute values he entered. The only property that RDFaCE would preserve was the URI field. Furthermore, RDFaCE would also not let him cross-reference the entities already entered. Each time he came across an entity RDFaCE would treat it as a completely new one, failing to offer a list of entities he had already entered. Fortunately, of all the things that RDFaCE could have saved, @uri is the most important, connecting the entity to an authority record. Not being able to record anything other than a URI is suitable provided that the authority the @uri points to contains all information the encoder wants to reference. The DMSEG, however, wants to record affiliations that are central to the Devonshire Manuscript's production and circulation, but which are not captured by the VIAF or ODNB.

RDFaCE also allowed Cole to add a person's relationship to a spouse or to children. Though the plug-in failed to save the information he added about the spouse or children, it saved the fact that the person has relationships to others. In the final analysis Cole used RDFaCE to point from each Devonshire Manuscript contributor to their VIAF or ODNB ID, and then give them the ability to have a spouse, children, or relationships. This way, the RDFaCE recorded the fact of the person's relationship using schema.org attributes, leaving him to add the URIs by hand. RDFaCE is a plug-in that promises to put RDFa directly into a WordPress site without going into the raw code, but as a beta product, at the time of writing, does not offer that full functionality.

Figure 1: Macro that creates RDFa for people's names

Figure 2: Macro that creates an RDFa attribute connecting the named entity to a VIAF record
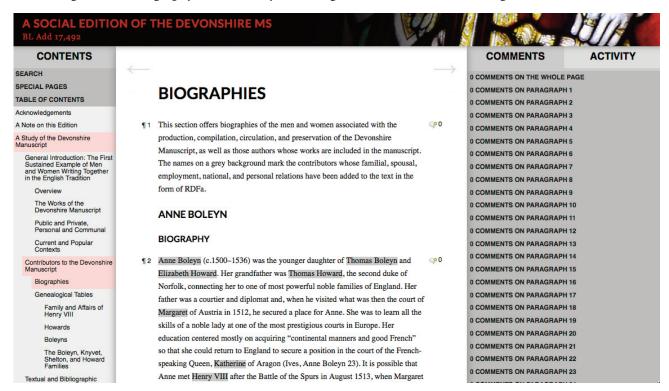

To complete the markup, Cole used a program called Keyboard Maestro to build a series of macros that offered him user-friendly fields to enter an entity's data and output RDFa (see Figures 1 and 2). Rather than having to enter each line by hand, paying particular attention to opening and closing quotation marks, missing carets, or misspelled attribute names, the macro automated the encoding that is common to each entity, leaving the human encoder with the sole responsibility of accurately entering the data that is specific to that entity. For example, when Cole entered in the VIAF for each person, he would click a hotkey that would open the macro, then there would be an empty field where he would copy and paste the VIAF URI, then press a "complete" GUI button. The macro would then wrap that VIAF URI in the necessary RDFa and enter it into the WordPress page's HTML.

He designed a macro for each attribute. Using macros and working from a spreadsheet containing all the known Devonshire Manuscript contributors sped up his encoding process considerably.

RDFa allowed the project to define the manuscript's sixteenth-century contributors as people and relate them to other people in the manuscript based on their connections as parents, children, siblings, and spouses. It also allowed us to include biographical information such as birthdates, death dates, and nationality. We could also affiliate contributors to other important people not directly involved in the manuscript based on connections such as employee/employer and enemy/ally relationships; however, formalizing these affiliations proved challenging. As with all regularization, the process of defining relationships was slightly lossy. Many of the contributors had complex relationships: friendships and rivalries changed over the course of contributors' lifetimes. In the end, we were able to indicate when someone was employed by someone else, or when someone was the enemy of someone else, or if someone knew someone else. At the level of code we marked these affiliations, letting the prose of the social edition's biographies page fill in any more specific information about contributors' relationships over time (see Figure 3).

Figure 3: The biography of Anne Boleyn, showing which stretches of text are augmented with RDFa

We invite readers to visit the WordPress-based edition at http://dms.itercommunity.org. We anticipate that the Social Edition of the Devonshire Manuscript, enhanced with RDFa to help machine readers aggregate data from all projects that point to VIAF and ODNB URIs, will help scholars develop new research questions and generate new knowledge about the culture and contexts of the Tudor court and a deeper understanding of the networked scholarship through which we come to engage it.

Notes

1. *Members of the advisory group*: Robert E. Bjork (Director, Arizona Center for Medieval and Renaissance Studies; Arizona State University); William R. Bowen, chair (Director, Iter; University of Toronto Scarborough); Michael Ullyot (University of Calgary); Diane Jakacki (Georgia Institute of Technology); Jessica Murphy (University of Texas at Dallas); Jason Boyd (Ryerson University); Elizabeth Heale (University of Reading); Steven W. May (Georgetown College); Arthur F. Marotti (Wayne State University); Jennifer Summit (Stanford University); Jonathan Gibson (Queen Mary, University of London); John Lavignino (King's College London); and Katherine Rowe (Bryn Mawr College).

2. *The DMSEG comprises* Raymond Siemens, Karin Armstrong, Constance Crompton, Barbara Bond, Terra Dickson, Johanne Paquette, Jonathan Podracky, Ingrid Weber, Cara Leitch, Melanie Chernyk, Brett D. Hirsch, Daniel Powell, Alyssa Anne McLeod, Alyssa Arbuckle, Jonathan Gibson, Maggie Shirley, Cole Mash, Chris Gaudet, Eric Haswell, Arianna Ciula, Daniel Starza-Smith, James Cummings with Martin Holmes, Greg Newton, Paul Remley, Erik Kwakkel, and Aimie Shirkie.

3. We use the term *machine readers* to refer to what others have productively called the algorithmic or computer readers/reading, ever mindful, however, of Ian Bogost's admonition not to use the terms *algorithm, machine,* and *computer* to obfuscate the human and material systems that produce the products of those terms (from financial markets to server farms, from security guards union representatives to Chinese factory labourers) (Bogost, 2015; Manovich, 2013; Ramsay, 2011).

References

Allemang, D., & Hendler, J. A. (2011). *Semantic web for the working ontologist: Modeling in RDF, RDFS and OWL* (2nd ed.). Boston, MA: Elsevier.

Bogost, I. (2015, January 15). The cathedral of computation. *The Atlantic.* URL: http://www.theatlantic .com/technology/archive/2015/01/the-cathedral-of-computation/384300 [April 8, 2015].

Clement, T., & Reside, D. (2011). *Off the tracks: Laying new lines for digital humanities scholars.* National Endowment for the Humanities white paper. College Park, MD: Maryland Institute for Technology in the Humanities / MediaCommons Press. URL: http://mcpress.media-commons .org/offthetracks [April 2, 2015].

Fitzpatrick, K. (2011). *Planned obsolescence: Publishing, technology, and the future of the academy*. New York, NY: New York University Press.

Herman, I. (2010). *RDFa primer.* URL: http://www.w3.org/TR/xhtml-rdfa-primer [April 8, 2015].

Manovich, L. (2013, December 16). The algorithms of our lives. *Chronicle of Higher Education.* URL: https://chronicle.com/article/The-Algorithms-of-Our-Lives-/143557 [April 8, 2015].

RDFaCE. (n.d.). *RDFaCE: RDFa Content Editor.* URL: http://aksw.org/Projects/RDFaCE.html [November 18, 2014].

Ramsay, S. (2011). *Reading machines: Toward an algorithmic criticism.* Chicago, IL: University of Illinois Press.

Rowe, K. (Ed.). (2010). *Shakespeare Quarterly*. MediaCommons Press. URL: http://mcpress.media-commons.org/ShakespeareQuarterly_NewMedia [June 2, 2015].

Siemens, R., et al. (2012). *The Social Edition of the Devonshire Manuscript.* URL: https://en.wikibooks
.org/wiki/The_Devonshire_Manuscript

Siemens, R., et al. (2015). *A social edition of the Devonshire Manuscript.* URL: http://dms
.itercommunity.org

Simpson, J., & Brown, S. (2014, February). *Inference and linking of the humanist's semantic Web.* Paper
presented at the INKE-hosted Partner gathering, "Implementing New Knowledge Environments,"
Whistler, BC.

Unsworth, J. (2000, May). *Scholarly primitives: What methods do humanities researchers have in
common, and how might our tools reflect this?* Paper presented at symposium "Humanities
computing: Formal methods, experimental practice," King's College London. URL: http://
people.brandeis.edu/~unsworth/Kings.5-00/primitives.html [January 2, 2015].

Crompton, Constance, Mash, Cole, & Siemens, Raymond. (2015). Playing Well with Others: The Social
Edition and Conceptual Collaboration. *Scholarly and Research Communication, 6*(3): 0301111, 9 pp.

9