CHAPTER 11

Publishing Sumerian Literature on the Semantic Web

Terhi Nurmikko-Fuller

Introduction

In the three decades since the invention of the World Wide Web (henceforth, “the web”), archaeologists, museums, and experts in the study of ancient languages have published their knowledge regarding the archaeological, curatorial, and philological material of the ancient world online. For the most part, it has been in the form of text and images intended for humans, both experts and interested laypersons, to read. Technological developments, however, now enable us to publish data online in machine-readable formats, where it is as accessible to software (a program or application which runs on a computer, as opposed to the hardware, which is the physical machine itself) as it is to a human being. It is in those formats that a family of technologies around the concept of the Semantic Web (SW) comes into play. These tools can be used to make information accessible in ways that allows machines to “understand” the meaning of the content they encounter.

The pragmatic implications of publishing datasets as Linked Data (LD) involve the social and academic considerations of trust, authority, data ownership, and copyright. The technical practicalities of data storage, server management, and user-interface generation represent another set of issues. From the perspective of capturing and representing expert knowledge, the development of the appropriate structural frameworks (known as “ontologies”)—through both the evaluation of existing models and the creation of new ones—plays a crucial role. In this chapter, three existing ontologies—the CIDOC CRM, FRBRoO, and OntoMedia (OM)—are evaluated for their suitability for adequately representing and capturing data made publicly available by an existing web resource, the Electronic Text Corpus of Sumerian Literature.

1 Crofts 2008, <http://www.cidoc-crm.org/get-last-official-release> (accessed May 9, 2017).
2 Bekiari et al. 2015, <https://www.ifla.org/files/assets/cataloguing/frbr/frbroo_v2.2.pdf> (accessed May 9, 2017).
3 Jewell et al. 2005, <https://eprints.soton.ac.uk/261024/> (accessed May 7, 2017).

© Terhi Nurmikko-Fuller, 2018 | DOI 10.1163/9789004375086_013
This is an open access chapter distributed under the terms of the prevailing CC-BY-NC License.
The applied perspective is that of critical evaluation in the context of the inter- and multidisciplinary digital humanities, with equal consideration given to the opportunities and challenges of the data and the technical implementations and solutions.

On Data, Information, and Knowledge

In order to understand the motivations behind the adoption of SW technologies and their unique advantages, we need to differentiate between three key concepts: data, information, and knowledge. The first, sometimes described in its un-analyzed state as being “raw,” consists of values and both quantitative and qualitative variables. We can easily recognize a spreadsheet as data: there is a column of numbers and, adjacent to it, a column with corresponding unit descriptions. The ancient Near East provides examples of clay tablets, which closely resemble modern spreadsheets, created almost five millennia before the invention of Microsoft Excel. We can use these to illustrate the differences between data, information, and knowledge: data is a list that stipulates that there are things that are sheep, and that there are, for example, five of them. The more data there is, the more information we can gain from it: if the list also contains goats, we can see that the owner has a mixed herd of animals of the subfamily of Caprinae. Knowledge is often generated as a result of much available, diverse information. The knowledge generated here would be that such mixed herds were and continue to be a popular approach to keeping livestock, due to their similar needs and non-competitive grazing habits (low and high, respectively). There are some clear caveats, too: the same data can be interpreted in many different ways; information is affected by bias and the limitations of the data; knowledge is often tacit and thus difficult to precisely pinpoint or define. Without further data about the animals (their sale price, for example) and additional information about the owner, our knowledge of the wider context remains incomplete. Was this a poor man who lived in a society in which sheep were more expensive than goats, and who could afford more animals if he had a mixed herd? Or was this a well-off individual wanting to display his wealth in a society that valued goats?
From Lovelace to Berners-Lee

Traditionally, computers—both the electronic machines we know today and their mechanical predecessors—have been well suited to creating and carrying out calculations on data. Several technological revolutions with far-reaching social ramifications have occurred since Ada Lovelace, the first programmer, wrote an algorithm for Charles Babbage’s Analytical Engine in 1842. Not the least of these are the realization and continuing growth of the internet (a global system of networks of interconnected computers) since the 1960s and the invention of the web (a hypermedia system of pages, sites, and other types of resources, accessed and connected through the use of hyperlinks) by Sir Tim Berners-Lee at the end of the 1980s (and its exponential growth and worldwide adoption since the early 1990s). Although its current manifestation is one of interconnected pages, the three technologies that lie at the heart of the web—HTML, HTTP, and URIs, all of which are discussed in greater detail below—are also at the core of LD. This information publication paradigm is a practical solution for bringing forth a SW, where connections occur at the level of data entities, not sites or pages.

Acronyms of Web Architecture

In the course of our daily browsing, we encounter a series of platforms, applications, projects, resources, and content. Behind the scenes, a number of technologies are at play.

The web currently consists of over 4 billion pages, all connected internally (within the page and a site) and externally (to other pages on other sites) through hyperlinks. Often visualized as blue font and underlined text, hyperlinks allow users to move between websites. The technology enabling this interconnectedness is the HyperText Transfer Protocol (HTTP); the acronym is familiar to us from the address bar displayed at the top of our browsers, starting with “http://”. These clusters of letters and characters form unique identifiers for each and every page on the web. They are known as Universal Resource Identifiers (URIs), and although other types of identifiers are possible, the ones we see when browsing are HTTP URIs. So common are the “http://” and “www.” prefixes that, when we discuss these resources, both are taken to be givens and thus are completely omitted; we refer to “Google,” for example, rather than “http://www.google.com.” On a conceptual level, these URIs do not differ from any other unique identifier, such as an ISBN for a book, a national identity...
number for a person, or an object identifier in a museum, but they are unique on the scale of the entire web.

For the most part, HTTP URIs point to specific pages; in this case, they are largely URLs, or Universal Resource Locators. The majority of the content of these pages is text, encoded using the HyperText Markup Language (HTML) to display in desired sizes, colors, and fonts. HTML does not capture any meaning, and, as such, is not strictly speaking fundamental to SW technologies—unlike HTTP URIs, which are intrinsically necessary. On the Semantic Web, rather than pointing to web pages, HTTP URIs serve as identifiers for specific data entities and information published on a page, a site, or in a resource. When utilized consistently and documented appropriately, URIs can and are used by large numbers of data publishers (people and institutions who make data available online) to show that completely disparate and unconnected pages contain information about the same entity or resource. As they refer to the same “thing,” which can occur on many different pages, these URIs are identifiers for resources, not for locations. They are thus URIs and not URLs. They are fundamental in the move from a “Web of Documents” to a “Web of Data.”

HTML is used for formatting and the visual appearance of the content of a web page, but it does not capture meaning. Enter the Extensible Markup Language (XML) and the tags, which appear in the code as angled brackets that allow us to encode the content as machine-readable. Agreed-upon standards such as the eponymous one maintained by the Text Encoding Initiative (TEI-XML) ensure that the tags used in the encoding of texts are systematically adhered to, and that different projects and sites can be shown to share relevant information. Consider two or more texts from independent projects that have been tagged to contain instances of people. If, for one, the utilized tag is `<person>`, but, for the other, it is `<human>`, an XML-based application would not necessarily consider them as containing information about the same kind of entity, unless such a relationship is explicitly defined in a relevant schema.

Machine comprehension can be achieved through RDF-XML, a syntax that can be used to express Resource Description Framework (RDF) in an XML document. RDF is an abstract data model with roots in metadata modelling, and it is used to represent information, knowledge, and data entities in a graph structure. This approach captures the relationships between entities by representing these entities and the connections between them using HTTP URIs. A more in-depth description of this technology follows later in this chapter, but it is RDF that enables the SW by allowing us to represent not just data, but, also, information and knowledge, in machine-understandable, as well as...
machine-readable, ways. The use of RDF and HTTP URIs to publish information online is the fundamental approach at the heart of LD.6

RDF can be represented in different but equally valid serialization formats. Of these, the most commonly occurring are TURTLE (.ttl), JSON-LD, and RDF/XML, each of which has its own proponents based on the needs, skills, and personal preferences of the people using them.

Linked Ancient World Data Online

The success of online projects as diverse as Nomisma.org, which has a focus on numismatics, a term applied to the study of coins and currency, and the Pleiades.stoa.org gazetteer, an index or dictionary of historic places and spaces, illustrates the suitability of ancient-world data as the focal point for SW technologies.

LD projects such as these utilize W3C-endorsed technologies such as RDF to represent and capture information and SPARQL (SPARQL Protocol and RDF Query Language)7 to access, retrieve, and manage it. The rich material from the ancient Near East is no different and can contribute to the development of LD through the promotion of shared vocabularies or schemas and exchange protocols.8 Assyriological data can be used to evaluate and test the robustness and flexibility of existing models and structural frameworks (ontologies). As a discipline, Assyriology would benefit from improved interconnectedness of information, from enriching external sources, improved discoverability, and, ultimately, the inference of new, implicit knowledge from explicitly declared facts.

Data

A multitude of domains, ranging from archaeology to curatorial and philological specializations regarding different aspects of the cultures of the ancient Near East, are brought together under the umbrella of Assyriological scholarship. Conservatively, the scope of the discipline is limited to the presence of objects carrying texts written in the distinctive cuneiform script. Even then,

---
6 For an example of the concrete use of RDF, see in this volume, Matskevich and Sharon, 46–47.
7 W3C is an acronym referring to the World Wide Web Consortium, the international standards organization for the World Wide Web, founded and chaired by Sir Tim Berners-Lee.
8 Protocol: a set of rules for the exchange or transmission of data between machines.
exemplars spanning the three millennia BCE can easily be referenced, provenanced to sites separated by time and space, and containing inscriptions in a number of unrelated ancient languages. The Semitic Akkadian, with two distinct dialects, Babylonian and Assyrian, and the linguistic isolate Sumerian, appear as a general rule; there are also the Indo-European Hittite, Hurrian, and Elamite, as well as other smaller language groups. Major professional entities, such as the International Association of Assyriologists,9 are similarly polyglot, accepting submissions in English, French, and German. At the digital periphery of the discipline, programming languages and diverse technologies are added to the mix. Geographically, the Mesopotamian plateau constitutes modern-day Iraq as well as much of the surrounding area. Specialties by modern scholars and major research threads within the community are frequently focused on specific temporal, spatial, and socio-cultural subsets or niches within this wider context.

The biographies of Assyriological objects are complex. Historically, collaborative excavations have led to the division of material culture from a single archaeological site into various collections. Charles Leonard Woolley’s excavations at the Royal Cemetery of Ur,10 for example, saw material allocated and distributed between funding institutions, and the material is currently shared between the University of Pennsylvania Museum of Archaeology and Anthropology in Philadelphia, the British Museum in London, the National Museum of Iraq, and a number of other sites. Limited access to data that is physically distant or stored in closed-off information siloes or entirely off-line collections management systems complicates the types of philological scholarship that can be based on the analysis of ancient texts. The online publication of object metadata and high-resolution images accompanied by the translations and transliterations of the texts can help bridge collections and contribute to solutions to overcome the challenges of access and distance.11 SW technologies have the potential to link heterogeneous datasets and to open up Assyriology as a discipline to enriching external data streams. Much work remains to be done until the full potential of the “Web of Data” can be harnessed and used to answer new and diverse scholarly questions, but the suitability of existing technologies for supporting this research is clear.

---

9 <https://iaassyriology.com/> (accessed May 10, 2017).
10 Woolley 1934.
11 Transliterations are conversions of the cuneiform script into the Latin alphabet without the translation of the ancient language to a modern one.
On Composite Texts

This chapter examines the suitability of existing technologies for the representation of information and narrative content of ancient pieces of narrative fiction, published as composite texts by the ETCSL. Consisting of parts of the text from different surviving physical tablets, these manifestations of the inscriptions represent the result of “an attempt, based on certain theoretical presuppositions, to reconstruct a hypothetical original version of a composition, when none of the original sources of the text have survived.” Composites may represent the content of a single instance of a tablet of any degree of preservation or completeness, or of any number of witness tablets, numbering in some cases in the hundreds. The process of creating these digitally published amalgamations was described by Jeremy Black and Gábor Zólyomi as working with ‘jigsaw pieces, which may not overlap at all.’ Each of the 400 or so inscriptions is a new entity and the product of a multi-stage process of interpretation. Where corrections or changes to the text have happened in the hard-coded HTML alone, the translation or transliteration in question can be described as being “born-digital”: this text, in this specific form, has no analogue or tangible version in existence. For a humanities scholar, the distinction is important, as it affects the critical framework in which the text is to be understood. From the perspective of a digital humanist, or a working ontologist, it is essential to identify and separate the composite text at a technical level from other manifestations. At the same time, it is crucial to capture and represent corresponding overlaps, original sources, and other inherent relationships between the witness tables—and their digital amalgam—in a way that makes those differences explicit to a specialist piece of software known as a “reasoner.”

Neutrality of Technology

Three decades prior to the publication of this book, Melvin Kranzberg published his article outlining the relationships between the social and the technological. The first of Kranzberg’s “laws,” where a law is a collection of truisms, describes technology as “neither good, nor bad; nor is it neutral.” Socio-politi-
Publishing Sumerian Literature On The Semantic Web

343

cal, economic, cultural, and intellectual aspects—all things human—play a
role within the design, development, implementation, evaluation, alteration,
and use of technology. For LD, and in particular in the evaluation and creation
of ontologies, which by their very function capture human understanding of
fact and human interpretation of the truth, the personal and idiosyncratic bi-
ases of those designing and using these structures must be acknowledged. It is
similarly so with the heterogeneous and incomplete datasets we have regard-
ing the ancient world, where gaps can only be bridged by the informed deduc-
tions of modern scholars. Investigative agendas utilizing SW technologies
should at all times maintain a hyper-awareness of the limitations imposed by
the data, appreciate the likelihood of differences between modern under-
standing and ancient thought and perspectives, and endeavor to explicitly cap-
ture tacit information. Extensive documentation of each aspect and stage of
the project workflow as well as the datasets, including version control, should
sit at the core of each research process in order to ensure accountability, trans-
parency, and reproducibility.

On the Importance of Words

Disambiguation and clarity of communication are essential to the interdisci-
plinary research of the digital humanities. Even the simplest of terminologies
can be misinterpreted. Consider the term “tablet”: it has clear and unambigu-
ous meaning to an Assyriologist, who envisions a clay object, probably one that
fits comfortably in the hand and is inscribed with the distinctive wedge-like
signs of the cuneiform script. The term has an equally clear meaning to those
from the broader field of historical written media, but as a different type of
writing implement, perhaps one based on wax (e.g., the Vindolanda tablets).
The same word is just as unambiguously applied by all of us when referring to
iPads or other similar mobile technologies. Context, then, is key. Other such
examples can be readily identified in the data of the GLAM (Galleries, Libraries,
Archives, and Museums) sector and the wider digital humanities: “record,”
which could be understood as object metadata, digital metadata, or an object
carrying a piece of audio, and “ontology” which could be understood either
within the remit of the analysis of existence (philosophy), or within a struc-
tural framework for capturing information about entities and their relation-
ships (computer science and digital information technologies, the SW). Other
terms, such as “subject,” “predicate,” and “object,” have clear meanings to
linguists, but those meanings have little in common with the semantics an
ontologist would assign to them. Even the term “semantic” is understood by
linguists as relating to the human process of assigning meaning, but, for those working in Knowledge Representation (KR), the term refers specifically to machine-readability and beyond, into the domain of machine-understanding, accomplished not necessarily though the complexities of Artificial Intelligence, but, rather, through the assignment of HTTP URIs.

The SW and LD as terms are frequently used erroneously as if they were interchangeable nouns referring to the same thing, much like the internet and the web. In fact, the SW refers to a thing, while LD is a process, namely one of data publication online in accordance with determined and agreed-upon criteria. Within this terminology there are clear, unambiguously separated terms which ought to be used precisely, not synonymously: Linked Data, Linked Open Data, and Open Data: the first refers to information that has been captured and published using the W3C standards of RDF and HTTP URIs. The term does not, however, signify that the data in question has been made openly available. RDF contained on a computer that is not connected to the internet is still Linked Data; RDF published online but only accessible to certain individuals is similarly not Open Data (OD). OD is data that has been made publicly available without access restrictions resulting from copyrights or paywalls. It often manifests as raw, or largely unanalyzed, datasets, made downloadable in formats such as CSV files. Linked Open Data (LOD) combines both these types: it is often based on OD that has been converted to RDF, and it is similarly available for consumption and reuse without restrictions or limitations. Since the term is more inclusive, this chapter uses Linked Data (LD) throughout.

A clear and distinct line needs to be drawn between semantic technologies and SW technologies: the former are understood to constitute a heterogeneous collection of tools and algorithms that can be used to bring structure to information. These include Natural Language Processing (NLP), data mining, Artificial Intelligence (AI), and a number of other approaches and methodologies. SW technologies refers to a family of technologies and standards provided by the W3C for the purposes of describing and connecting information online. A by-no-means-exhaustive list of these technologies includes RDF, RDFS, OWL (RDF Schema and the Web Ontology Language; for ontologies as will be dis-

---

17 For further information, see in this volume, Ramazzotti, 62.
18 “Linked Data,” W3C, <https://www.w3.org/DesignIssues/LinkedData.html> (accessed May 7, 2017).
19 For general but concrete examples related to the use of Linked Data, see in this volume, Matskevich and Sharon, 47.
20 For further information, see in this volume, Ramazzotti, 62n9.
Publishing Sumerian Literature On The Semantic Web

cussed in detail momentarily), and SPARQL. It is on these W3C standards that this chapter will focus.

The Triple

The grand promise of LD has been one of harnessing all the intellectual, factual, and raw data anywhere online. Information amalgamated from all corners of the web, across disciplinary barriers, data siloes, and automated inference, would, at the stroke of a key, bring to light the types of implicit connections that would take human scholars generations to uncover. As a flexible data structure, RDF is both powerful and capable of capturing nuanced and domain-specific relationships between data entities. Information stored in this way can also be restructured, manipulated, and edited with comparative ease without the extensive restructuring that a relational database would require. RDF as a technology imposes few limitations on the type of information that can be captured and represented, whether at the level of entities (things, people, places, concepts, notions) or at the level of the relationships between them. All RDF data structures are based around a data cluster of three components, known as a “triple.”

A triple consists of three parts: a SUBJECT, a PREDICATE, and an OBJECT. These are used to define entity types and directional relationships between them, both at the general (or schema) and specific (instance) level.

Let us consider the instance level first, since the schema level is captured using ontologies, which are described in greater detail in the section below. At an instance level, we would consider a specific cuneiform tablet. It will have an author, perhaps an ancient scribe named in the colophon. A natural language utterance capturing that information might take the form of the sentence “Tablets are written by scribes.” Since not all scribes wrote all tablets, differentiation at the level of specific instances is necessary, and it is achieved through the use of personal names for scribes (e.g., Lukalla) and museum

21 “OWL 2 Web Ontology Language Document Overview (Second Edition),” W3C, <http://www.w3.org/TR/ow2-overview/> (accessed May 7, 2017).
22 See in this volume, a) applied to graph database, Matskevich and Sharon, 46–47, and b) for a different meaning, Pagé-Perron, 202–203, who uses triples for network graph.
23 A colophon is a brief statement on a text, such as a cuneiform metadata, that captures the equivalent of what could now be considered bibliographical metadata and can contain information such as the name of the authoring individual.
numbers for the tablets they wrote (e.g., So02932). The same utterance now becomes more specific: “the specific tablet to which we now refer by its modern identifier So02932 was written by the named scribe Lukalla.” Both granularities of information can be captured using RDF triples. In the case of the specific (instance-level) relationship between Lukalla and So02932, the subject could be the tablet, the predicate would be the notion of creation, and the object could be the scribe. This would result in a triple of

\[
\text{tablet So02932 (subject) was written by (predicate) Lukalla (object).}
\]

If we define the relationship between the tablet and the scribe as “wrote,” we can capture this relationship from the other direction, resulting in an equally true and accurate but slightly different triple:

\[
\text{Lukalla (subject) wrote (predicate) tablet So02932 (object).}
\]

In this example, the personal name and the museum number are expressed as strings of characters (letters and numbers), rather than as semantically meaningful identifiers. When this information is captured as RDF, each of the three parts is assigned a HTTP URI:

\[
<\text{http://example.org/abcd}> <\text{http://example.org/12345}> <\text{http://example.org/xyzz}>
\]

HTTP URIs can be simplified in .ttl, making them easier for humans to read. This is accomplished using a prefix or a disclosed shorthand: if we define “http://example.org/” as prefix:example, the same triple can be expressed more succinctly (without any loss of meaning) as:

\[
\text{prefix:example:tablet So02932 (subject) was written by (predicate) prefix:example:Lukalla (object).}
\]

---

24 Tablet So02932 is also known as P249185 in CDLI (<http://cdli.ucla.edu/search/search_results.php?SearchMode=Text&ObjectId=P249089> [accessed May 10, 2017]).

25 This example may have been clearer if HTTP URIs with embedded semantics had been used, e.g., Lukalla had been identified as http://example.org/scribe/Lukalla and the tablet as http://example.org/tablet/So02932. Although these are easy for humans to read and understand, the current best practice in the LD and SW communities state that it is better to use URIs that do not have human-readable meanings. In this example, http://example.org/abcde refers to the scribe, http://example.org/12345 to the relationship between scribe and tablet, and http://example.org/xyzz to the tablet.
It is here that a knowledge graph begins to emerge: the object in one triple, when represented using an HTTP URI, can be the subject in another. This enables us to produce an interconnected graph of nodes (the entities) and arcs (the relationships). Unlike hierarchical data structures such as XML, RDF graphs can extend to direction at any point. The end of the sprawling network only comes at points in which the object of the triple is a string (or indeed anything but a HTTP URI), as no new triples can be added.

Navigating and querying the resulting knowledge graph necessitates understanding the entity and relationship types that have been used in capturing this information both on the instance and schema levels. We use information structures expressed in .ttl to do this.

**Ontologies**

The SW is not a new concept. KR, one of its cornerstones, has been discussed in the context of the Artificial Intelligence community since at least the early 1980s. Berners-Lee’s original vision for the web was one of a hypertext system not only linking human-readable documents but also including machine-readability from the onset. Yorick Wilks and Christopher Brewster described the SW as a “more powerful, more functional and more capable version of [the] document and language-centric Web.” In recent years, the SW has increasingly manifested through the adoption of the LD publication paradigm, with the more complex processes of KR relying on complex information structures (ontologies).

Ontologies are formalized structures that form parts of expert systems. They constitute the underlying data infrastructure for projects, tools, and datasets published as LD. Described as “crucial” to the realization of the SW, they are frequently defined through Thomas R. Gruber’s description as “an explicit specification of a conceptualization.” Ontologies are used to define the types of entities that occur in the dataset (classes) and the possible relationships

26 Levesque 1984.
27 Berners-Lee 2000.
28 Wilks and Brewster 2006.
29 Wilks and Brewster, 2006.
30 Gruber 1993, 199.
between those entities (properties). The classes of an ontology define the type of entity a subject or object is, and the properties describe predicates, including the directionality of the relationship it captures. The class from which the property runs is the domain; the class to which it runs is the range.\(^{31}\) While many of the relationships between entities are explicitly declared, one of the main advantages of SW technologies, and the publication of data in machine-readable formats, is automated reasoning (the finding of implicit connections between explicitly declared facts by software agents).

There are two main schools of thought regarding ontology design:\(^{32}\) the Newtonian reductionist model and the Leibnizian model, which concerned with the capture of the nuanced complexities of experience. The former allows for a greater degree of control and administrative clarity, while the fuzziness of the latter can make it easier to apply. Newtonian models depend on a degree of omniscience over a domain; Leibnizian models assert that true and absolute objectivity is in all pragmatic terms unachievable. Classification of ontologies is frequently limited to two broad types: domain-specific or upper-level. The former focus on a given, defined, and specific area of information. The latter are more generic and have applicability across a large number of disciplines and areas of expertise.

One of the fundamental cornerstones of the successful application of LD is the ability to share information and knowledge across disparate datasets. This is accomplished through several methods: adherence to W3C recommendations (RDF, HTTP URIs) and the publication of data according to the Five Star standard are examples of the adoption and use of agreed-upon standards to facilitate this information exchange.\(^{33}\) The reuse of existing ontologies to point to shared entity types is another significant approach. While it is possible to define a new ontological structure for each new project or resource, reusing existing ones is a simple, relatively straightforward method that is highly likely to effectively and efficiently facilitate links between datasets. Reuse is dependent on awareness of an existing model, and the latter’s adequate, often necessarily extensive, documentation enables the data publisher to be sure that they are using the ontology appropriately. Evaluations of existing ontologies are

\(^{31}\) For examples of concrete applications of ontologies in the philological field, see in this volume, Bigot Juloux, in particular 165–181; Prosser, 320. For ontology in data sharing, see also in this volume, Matskevich and Sharon, 47.

\(^{32}\) Brewster and O’Hara 2004.

\(^{33}\) "Linked Data Glossary," W3C, <https://www.w3.org/TR/ld-glossary/> (accessed May 7, 2017).
heavily dependent on the availability and completeness of relevant documentation and thus can be time-consuming. Yet they are a significant part of the interlinking process for LD.

Assertion of equivalence between ontological class and dataset entity may seem relatively straightforward when considering things that have a degree of tangibility or generic familiarity, such as person, place, or event. A closer examination highlights a number of complicating uncertainties, such as our frequent inability to point to very specific start- or endpoints for events (those moments when something begins, or when an event draws to a conclusion), or the challenges created by spatio-temporal changes to the geographical areas of given locations, changed as territory is won and lost. Disambiguation between individuals may also be complicated by naming conventions and inherited professions, resulting in a number of persons from the same location and time's sharing both name and societal role. Incomplete data exacerbates the problem. Consideration and evaluation of existing ontologies ought to form a significant part of any research project examining and using LD.

**Ontologies for Sumerian Data**

The Assyriological community has yet to extensively embrace SW technologies. Although other existing models will be shown to be relevant and appropriate for the representation of our data, at the time of this printing, only two ontologies specific to Sumerological data are known (with a third included in a PhD dissertation).  

Wojciech Jaworski describes an application-oriented system focused on economic tables of the Ur III period. It includes an ontology representing a selected branch of economic activities, with translations of texts into a meaning representation language through a semantic grammar. This approach allows for the representation of ambiguities caused by the limitations of the document simplicity, any incompleteness of current understanding of the Sumerian language, and omissions in the texts caused by damage to the original primary sources.

---

34 Nurmikko-Fuller 2015.

35 Jaworski 2008.
Sumerian was used by Sergio Alivernini as an example for the representation of natural language grammars using ontologies.\(^{36}\) Described as “an experiment,”\(^ {37}\) it consists of two parts: the **T-box** (terminological) and **A-box** (assertions), where the former is Sumerian grammar, and the latter is the content of ancient inscriptions. Of these, the latter utilizes foundation-brick inscriptions from the reign of Ur-Nammu (c. 2000 BCE), but the source material for the grammatical features is not declared.

Although focused on data that is socio-culturally, linguistically, and temporally complementary, these models are insufficient or unsuitable for the representation of the content of the literary inscriptions as published on the ETCSL. Before we evaluate the suitability of ontologies created from the perspective of other domains, the niche of Sumerian literature deserves definition and description.

### Literary Sumerology

Literary Sumerology (the study of literature written in Sumerian) is one of the most recent areas of expertise within the sphere of Assyriology, having emerged only in the 1950s.\(^ {38}\) The term “Sumerian” is here to be understood as an exclusively linguistic label, referring to the language of the compositions themselves, rather than being indicative of the ethnic or socio-cultural identity of the authoring scribe or the patron commissioning the piece. Data is not limited to sources with a proven provenance in Sumerian periods for two reasons: First, literary pieces unearthed from third-millennium BCE contexts at geographically diverse sites (Ebla and Mari in the north and Girsu, Abu Salabih, Nippur, and Adab in the south) suggest these compositions were conceived of even earlier;\(^ {39}\) second, while primary sources for Sumerian literature originate from a range of temporal, geographical, and cultural settings,\(^ {40}\) many are dated to the Old Babylonian (OB) period (c. 20th-16th centuries BCE). Sumerian is absent from administrative and legal documents, as well as letters, from around 1730 BCE onwards,\(^ {41}\) supporting the theory that at this time it no longer sur-

---

\(^{36}\) Alivernini, D’Agostino, and Romano 2006, <http://www.epistematica.com/docs/Ur_Namma.pdf> (accessed May 12, 2017).

\(^{37}\) Alivernini, D’Agostino, and Romano, 2006, 3.

\(^{38}\) Robson 2013.

\(^{39}\) Taylor 2013.

\(^{40}\) Goodnick-Westenholz 2013.

\(^{41}\) Black and Zólyomi 2007.
vived as a spoken language in any of the urban centers of the Mesopotamian plateau. Its use in the vernacular registers of personal correspondences had ceased some two centuries earlier. Scribes who most likely spoke Akkadian as their mother tongue carried on the tradition of Sumerian literature via copies of earlier compositions (e.g., Instructions of Šuruppak) as well as new compositions and those created through extensive modification of known material such as *Lugalbanda*. Separating instances of deliberate variation (a reflection of a particular scribal tradition, for example) from those of scribal error can be difficult, but it is possible that the practice of creating Sumerian literary compositions outspan, in both directions, the third millennium BCE.

The Electronic Text Corpus of Sumerian Literature (ETCSL)

The ETCSL is an online resource that provides access to the transliterations (in both Unicode and ASCII) and translations of some 400 literary compositions, all dated to the late third or early second millennium BCE. The name encapsulates the form and function of this bi-part, web-based resource, which Jarle Ebeling describes as a “diachronic, transliterated, annotated corpus of Sumerian literature, and a Sumerian-English parallel corpus.” Internally, it has a seven-part thematic structure that mirrors the original non-digital and largely unpublished corpus created by Miguel Civil.

For the Sumerian, there are transliterations rendered using the Latin alphabet and modern philological conventions (with no accompanying cuneiform, either Unicode or otherwise) on a line-by-line basis. For the English translation, there are paragraphs of prose.

The ETCSL provides external links to other related and complementary projects. These include other online corpora, resources for museological and palaeographical data, and a dictionary. These projects include the Diachronic Corpus of Sumerian Literature, Database of Neo-Sumerian Texts, the Cuneiform Digital Palaeography Project, the Cuneiform Digital Library Initia-
tive (CDLI), and the Pennsylvania Sumerian Dictionary (EPSD). Many of the links run in one direction: the Diachronic Corpus links to the transliterations published on the etcsl, but cannot be accessed via it. The EPSD links externally, providing a list of texts and a short passage in which a specific lexical item occurs, but the hyperlink is not bidirectional. The etcsl record is a pop-up window and can be compared with the original search results from the EPSD.

Although not actively developed since 2006, the etcsl is a cited resource and is itself the focus of research. It was an innovative project in the realm of study of Sumerian literature in that it offered free and unrestricted online access to the data, with the aim of interlinking with external resources. The live site was created using HTML and JavaScript (a programming language for HTML and the web) only, with the lemmatisation hard-coded into the latter, but the content is exclusively output as customized TEI-XML P4, available from the Oxford Text Archive.

Research Questions for the etcsl

Ontological representation and the publication of data as RDF allows for a new, more flexible and complex type of research question. These enabled avenues of investigation can span beyond the expected searches facilitated by most search fields on online projects and resources. Traditional investigative agendas are frequently limited to the defining characteristics of geographical location, time, and an area of expertise. Capturing the narrative thread of these literary compositions will enable a new type of question: Which kings have claimed a divine or semi-divine character as their brother? What are the demographics and dynamics of giving and receiving advice? Are men more likely to advise or be advised by women? Is advice predominantly given by seniors to juniors? What are the shared characteristics of persons associated with a birth legend? Adherence to LD will not accomplish data collection tasks that would be beyond a human scholar intellectually, but LD can speed them up, make

50 <http://cdli.ucla.edu/> (accessed May 10, 2017). For further information, see in this volume, Pagé-Perron, 198–200, and Eraslan 285.
51 <http://psd.museum.upenn.edu/epsd/index.html> (accessed May 9, 2017).
52 Ebeling and Cunningham 2007; Delnero 2012; Crawford 2013.
53 <https://ota.ox.ac.uk/>, <http://etcsl.orinst.ox.ac.uk/> (both accessed May 7, 2017). Robson 2013; Nurmikko-Fuller 2014, <http://dlib.nyu.edu/awdl/isaw/isaw-papers/7/nurmikko-fuller/> (accessed May 9, 2017).
them more precise, and bring to light complementary compositions from the surrounding areas (such as the oral traditions of the Bedouin). The addition of data features published by the etcsl for compositions could add layers of complexity: Are there objects types that are frequently associated with tablets containing a specified literary motif? What types of material culture are curated by professionals responsible for tablets that carry a specific narrative feature? Do tablets carrying specific storylines tend to be displayed in a particular interpretative context within a museum setting? On what other areas do scholars who have published on a given literary composition, or one containing a particular feature, tend to focus? These research agendas are possible not only through the richness of the transcriptions contained in etcsl data, which open up the analysis to specific narrative motifs, but also through museological information regarding specific witness tablets and through bibliographical metadata regarding publications of translations.

Evaluating the Suitability of Existing owl Ontologies to Represent etcsl Data

The three existing owl ontologies with structures complementary to different aspects of etcsl data that are evaluated here are the cidoc crm, frbroo, and om.

**CIDOC CRM**

CIDOC CRM is a domain-specific, event-based ontology designed for the representation of cultural heritage data.\(^{54}\) It has been an official International Organization for Standardization (ISO)\(^{55}\) standard from 2006 and lists some 30 examples of extensions and compatible models that have been launched since. The purpose of the CIDOC CRM is to establish a common framework for the sharing of data between GLAM institutions, and to function as an example of best practices in the cultural heritage domain.

A number of other ontologies have already been incorporated into the CIDOC CRM. These include Dublin Core, SKOS, and FOAF, with additional available extensions and use cases further enabling the representation of archaeological

---

54 For a description of CIDOC CRM and its use in material heritage and field archaeology recording, see in this volume, Matskevich and Sharon, 45, 48–49.

55 International Organization for Standardization (ISO): the world’s largest developer and publisher of international standards. It is an independent, voluntary organization, with representatives from each of the 162 member countries.
data and processes, the provenance of digital artifacts, and bibliographies. All of these provide opportunities for linking to other data streams.

**CIDOC CRM** is a large and complex ontology consisting of 90 distinct entities and 149 property declarations. Extensively documented, it makes the identification of suitable classes and properties relatively easy. The focus of the structure is to map cultural heritage information or details regarding the biography of a given object, but neither is directly available through ETCSL. The **CIDOC CRM** alone is insufficient for the representation of the project data, but it is crucial in its role linking FRBRoo and OM, as well as in its enabling possible future data exchange with other digital heritage projects.

**Mapping ETCSL data onto the CIDOC CRM**
One future potential is clear: the **CIDOC CRM** allows for the differentiation between the text content (E33 Linguistic Object) and the physical item that carries a composition (E84 Information Carrier). They are linked through E73 Information Object (superclass of E33). This is particularly useful when representing a composite text, an intangible construct, and any instance of E33, manifesting as a transliteration on the ETCSL site (itself a digital instance of E73). The transliteration and translation can each be mapped to have a language (Sumerian and English, respectively). Similarly, each witness tablet is an instance of E73, and each text carried on each physical tablet is a separate instance of E33. The use of E90 Symbolic Object (the superclass of E73) makes it possible to represent the notion of the composite as a separate entity. The P106 is composed of—property maps the composite (an instance of E73) as an amalgamation of several other texts (also all instances of E73). This nesting of super- and subclasses is significant because subclasses inherit the properties of the superclasses.

The carrier in this case is a cuneiform tablet purposely made for the task of storing data in the form of written text (E84 Information Carrier is a nested subclass of E24 Physical Man-Made Thing, connected to E90 via the P128 carried by (is carried by)—property). The **CIDOC CRM** enables the mapping of the physical features of the physical items (E57 Material, E58...
Measurement Unit), and although presently that data is not directly available through the ETCSL, it could be found by identifying the corresponding record from the housing institution’s online collections (Musée du Louvre, the British Museum). This is an ideal example of the type of information-symbiosis that could be used to enrich the ETCSL and museum records, providing access to the transliteration and translation of the text content and helping to place objects in a wider literary context. With the addition of high-resolution photography from projects such as CDLI, the collections and the knowledge within the discipline could be brought together efficiently and accessed from a single entry point by anyone with access to the web.

FRBROO
FRBROO is a formal ontology for the representation of bibliographic information. It has been designed to merge with the CIDOC CRM and to facilitate the integration of museum and library data. It is based on the Functional Requirements for Bibliographic Records (FRBR), which was originally designed as an entity-relationship model. FRBR was developed independently of CIDOC CRM, but, coincidentally, at approximately the same time (1991–1997). It was designed by a group appointed by the IFLA (International Federation of Library Associations and Institutions) and approved by the IFLA Cataloguing Section in 1997. In 2003, an international working group began to examine the potential of merging FRBR with the CIDOC CRM; the first draft of FRBROO was completed in 2006, with the official publication of version 1.0 at the end of 2009. It was approved and issued by January 2010. Unlike the CIDOC CRM, FRBR models products, not processes. It is smaller and less complex, with 52 classes and 64 properties, compared to 90 classes and 149 properties in the CIDOC CRM. The process of merging the two existing structures necessitated changes in both, resulting in changes to the CIDOC CRM ISO standard. The combination of FRBR and the CIDOC CRM brought about FRBROO—a new, object-oriented mapping. The process was one of extensive merging: FRBROO is referred to in 60 of the CIDOC CRM classes and in 55 of the properties.

The international working group for FRBROO cited the complementary and interlinked nature of bibliographic and cultural heritage data as the incentive for the project: “Libraries and museums are memory institutions—both strive to preserve cultural heritage objects, and information about such objects, and they often share the same users ... the boundary between them is often blurred.

60 This assumes that these heritage institutions’ required object catalogues and collections management data had been published in adherence to the Five Star LD criteria.
61 Bekiari et al. 2015.
62 Bekiari et al. 2015, 11.
... the cultural heritage objects preserved in both types of institutions were created in the same cultural context or period, sometimes by the same agents ... it seems therefore appropriate to build a common conceptualisation of the information gathered by the two types of organization."63 An example of the use of CIDOC CRM and FRBROO for the representation of data related to ancient material is the British Museum’s project for the digitization of Malcolm Mosh-er’s work on the ancient Egyptian Book of the Dead.64

Mapping etcsl data onto FRBROO

There are two types of bibliographies embedded into the etcsl. The first is the metadata of the transliteration, including the title and the revision history. The second is the text-only page listing the items cited in the transliterations, including the author name, article title, year, name of the publication type, and name of the publisher. These six categories of information can be mapped onto the classes of FRBROO and connected to the aforementioned class of E73 Information Object: FRBROO class F2 Expression is a subclass of E73. F2 has two subclasses, which allow for the mapping of the composite text: F22 Self-contained Expression is a class for the composites as they appear on ETCSL (as a cohesive text), while each separate segment is an instance of F23 Expression Fragment, since they are sections of the composite and sections of the original witness tablets. The two classes are connected as subclasses of E2, and their relationships to E2 additionally include R5 has Component (is component of) and R14 incorporates (is incorporated in) for E22, and R15b is fragment of (has Fragment) for E23. The relationship of text to the object is via P128 carries (is carried by). It connects F4 Manifestation Singleton to F2. Modern scholars named in the bibliography can be mapped as instances of F10 Person (an equivalent of E21), so these classes provide an example of one of the many points where the two ontologies merge. Another example relevant here is F4 (a subclass of E24). FRBROO allows for the representation of both physical and electronic publishing.

In terms of content, FRBROO includes F38 Character, ideally suited for the capture of protagonists, antagonists, and other characters in Sumerian literature. The scope notes specify this class as being for “fictional and iconographic individuals ... appearing in works in a way relevant as subjects. Characters may be purely fictitious or based on real persons.”65 Applicable to a wide spectrum

63 Bekiari et al. 2015, 11.
64 Oldman and Norton 2014, <https://www.youtube.com/watch?v=cK54YlY-xZs> (accessed May 12, 2017).
65 Bekiari et al. 2015.
of protagonists, it is equally valid for the unnamed characters (the slave-girl, the young scribe, the ox-driver) as it is for Gilgameš, the eponymous main character of the oldest known piece of epic literature,⁶⁶ a semi-divine being, and a powerful ruler.

For Gilgameš, there are three disjoint categories: the man, who may have been a genuine, historical person, and is treated in the context of the literature as such, and ought to be mapped as an instance of $F_{10}=E_{21}$; the later-deified, $d$Gilgameš,⁶⁷ a myth, a fabricated construct serving an educational, cultic or socio-political purpose; and the character in a literary composition ($F_{38}$). This character is separate from the divine manifestation, as he exists solely in the fabula of the story and could, from the perspective of those who do not believe in the existence of a divine entity, be considered wholly fictitious. The deified $d$Gilgameš is disjoint from the man, as the latter has to transform into the former (they cannot coexist). Neither CIDOC CRM nor FRBR00 has a class that truly captures this notion, although FRBR00 does allow for the capture of the third manifestation of Gilgameš: the literary representation. It is possible to map the relationship between the protagonist and his historical counterpart by $R_{57}$ is based on (with $E_{39} \text{Actor}$ as range).

**OntoMedia (OM)**

OM, like the CIDOC CRM, is event-based. It focuses on the representation of the narrative in multi-media and has been designed as linkable to the CIDOC CRM.⁶⁸ The aim of this ontology is to enable the human-like, vague questions we might use in conversation when trying to identify a given story.⁶⁹ OM, unlike CIDOC CRM or FRBR00, is not a widely utilized ontology, nor has it been awarded ISO (or equivalent) status; the first version was the product of a collaborative doctoral thesis, and, since then, the constituent .owl files have been made (somewhat inconsistently) accessible online. The widespread adoption of OM is likely to be hindered by the lack of extensive, clear, up-to-date, and systematic documentation.

OM represents the narrative content of heterogeneous media. It is based largely on two interlinked topics: the literary genres of fantasy and science fiction and the fan-fiction associated with them, both manifesting as a number of genre-specific classes (Void-Travel, Pegasii, Unicorn and Faerie).

---

⁶⁶ George 1999; 2010.
⁶⁷ This addition of the lowercase "d" in superscript before the personal name denotes that the entity in question is a deity.
⁶⁸ Jewell et al. 2005.
⁶⁹ Lawrence 2008.
Some classes are indicative not only of genre, but, also, the works of a specific author (Hobbit).

Design decisions complicate the structure and result in the repetition of equivalent classes across sub-ontologies. Reliance on nested subclasses and the duplication of classes is illustrated by the Trait: it is simultaneously a parallel class and a superclass for over 50 other classes. Another example is Person≡Person≡Person in the Being sub-ontology, where one is foaf:Person,70 and the other two are nested classes of OM repeated within separate sub-ontologies.

Protagonists are mapped as instances of Character, although this class does not allow for any differentiation between people, deities, sentient supernatural beings, demigods, or anthropomorphized creatures. Incorporation of the FOAF ontology enables the mapping of interpersonal relationships and physical characteristics through a series of additional classes. Regarding relationships, it is possible to differentiate between four distinct subclasses of Family Bonds (Adopted, Blood, Foster, and Step), as well as to map alliances (including a separate subclass Friendship). There are classes for Deal, Enmity, Pledge, and Possession. The Agent class has subclasses for Group, Organisation, and Person.

Mapping ETCSL data onto OM

OM’s Character Trait class contains many characteristics of individuals, including Gender, Name, Stages Of Life, Stages Of Being, and State Of Consciousness, as well as Knowledge and Motivation. Various subclasses of the Detail class map human physical features (Eye Colour, Body Type, Hair Colour), and enable the mapping of the descriptions of protagonists: “his angry brow…his bison eyes…his lapis lazuli beard…his elegant fingers…” (lines 70–75 in Gilgamesh and Aga).71 The Attire class can also be used to represent different parts of the arc of the story, such as the undressing of the eponymous protagonist in Inanna’s descent to the Netherworld.72

Another tangible data entity within the content of the literary corpora is geographical location. The OM class Space has three distinct subclasses. Two of these are structurally simple, consisting of a small number of nested sub-

70 That is to say, instances of this class are people, as defined by the FOAF ontology.
71 “Gilgamesh and Aga,” Electronic Text Corpus of Sumerian Literature, <http://etcsl.orinst.ox.ac.uk/section1/tr1811.htm> (accessed May 10, 2017).
72 “Inanna’s Descent to the Netherworld,” Electronic Text Corpus of Sumerian Literature, <http://etcsl.orinst.ox.ac.uk/section1/tr141.htm> (accessed May 10, 2017).
classes. AKT Abstract Space (relevant nested subclasses include Surface Space and Biological Surface Space) and AKT Enclosed Space (including Vessel, Portal, and Container) allow for the representation of, for example, the coracle that held the infant Sargon as he was sent down the river according to his eponymous *Birth Legend*, or the much larger version built by *Atrahasis*.\(^3\) The extensive AKT Open Space contains almost 40 distinct classes and subclasses, most of which (with the exception, perhaps, of World, Galaxy, and Universe) are directly useful for the representation of place in these ancient literary compositions.

Although extensive, the Space sub-ontology does not include those architectural entities and features familiar to us from Sumerian narratives: the temple, the palace, the city wall, the canal, the dyke, and the reed bed. The closest class is Building, but since temples were often individually named, it does not seem infeasible to suggest that they were seen as a separate type of entity, disjoint from other, more mundane buildings, and that representing them as instances of this class would reflect a perspective radically differ from the perspectives of the ancient peoples.

The Core Expression class includes aspects with universal applicability. These include Abstract Item and Physical Item, as well as the classes of Timeline, Occurrence (with the subclass Event), Introduction, Gain, Loss, Transformation (with subclass Travel), Social, and Action. All encountered events in the Sumerian literary canon can be mapped using these classes, although the lack of granularity would result in many false positives when identifying stories that share some broad narrative motif. Query-based, automated differentiation between different narrative structures would necessarily be quite limited.

Although OM Profession class includes Military, challenges remain as to whether the differences between ancient (as described and recorded in the texts), modern (as per our own cultural perspectives), and hypothetical future (as expressed in the narratives of science fiction) warfare are sufficiently similar to justify the use of this class. The Bestiary and Zoology sub-ontologies allow for the specification of many animals, but they are not based on Linnaean taxonomy and rarely match those identified in the ETCSL.

Thirteen classes allow for the mapping of human and animal actions, movements of celestial bodies, and natural phenomena: Action, Being, Celestial, Environment, Eventprop, Gain, Group, Introduction, Loss, Social, Space, Transformation, and Travel. Action contains the parallel subclass system discussed in the context of the Trait class: here, Celestial

\(^3\) Finkel 2014.
and Environmental are subclasses of Action, as well as parallel to it. When parallel, they contain further subclasses. Environmental Event subclasses enable the mapping of natural phenomena that appear in the context of the Sumerian narrative, such as rising waters (Enki and Ninhursaga)\textsuperscript{74} or growing of crops (Enki and the World Order).\textsuperscript{75} Celestial is more complex due to the Mesopotamian perspective on the cosmos. At the most reductionist of levels, the Mesopotamian perspective could be described as viewing gods and goddesses as anthropomorphized celestial objects. Utu rises and sets; the sun is the god.

Several of the subclasses of the Action class can be used directly to represent the elements of a Sumerian story. The Victory of Utu-hegal has instances of Battle and War; the beating received by Birḫar-tura in Gilgameš and Aga populates Corporal Punishment. Other classes of Events can be shown to be equally relevant, and others as unnecessary (Space Travel).

For all the complexity of OM, its classes fall short of representing the potential richness of ETCSL data when they attempt to capture verbal exchanges. Various sorts of oration are a frequently utilized rhetoric device and contain additional explanations and information. Extensions to Social, including the subclasses of Monologue, Dialogue, Polylogue, Advice, Warning, Lamentation, Curse, Prayer, and Rejoicing would be necessary to adequately capture the richness of the rhetoric in these compositions.

### Conclusion

SW technologies facilitate the sharing of knowledge regarding identified entities, a concept familiar to Assyriologists, who already use non-LD online resources such as CDLI to access data collated from a large number of different databases into one cohesive unit and to query them via a single point of entry. LD differs from these existing models in that it is not limited to the content and structure of the selected databases. Relevant information, could potentially be identified anywhere online and incorporated into the processes of automated inference. Data published as RDF can be queried starting from any desired or arbitrary point in the knowledge graph without the need to adhere to a strict hierarchical structure. Both RDF and ontologies are as inherently flexible as they are extensible: it is possible to merge ontologies and to add RDF triples as

\textsuperscript{74} “Enki and Ninhursaga,” Electronic Text Corpus of Sumerian Literature, <http://etcsl.orinst.ox.ac.uk/cgi-bin/etcsl.cgi?text=t.1.1.1#> (accessed May 10, 2017).

\textsuperscript{75} “Enki and the World Order,” Electronic Text Corpus of Sumerian Literature, <http://etcsl.orinst.ox.ac.uk/section1/tr113.htm> (accessed May 10, 2017).
knowledge increases and new discoveries are made. Crucially, the addition of semantics via the ontology facilitates a degree of understanding of the content of a text by software agents as well as human readers. Data stored as RDF can be queried using another W3C standard, SPARQL, effectively creating searches that extend beyond keywords and Boolean operators.

The rich, heterogeneous, and at times incomplete information contained within the etcsl and the wider corpora of Sumerian literature offers many challenges to these technologies. Three ontologies were evaluated in the context of this data: the CIDOC CRM, FRBROO, and OM. The first captures cultural heritage data, the second represents bibliographic data, and the third maps narrative content. Each of the three was assessed in turn, and they were all found to be in many ways adequate for the representation of this information but also to contain numerous superfluous classes. They are all predominantly designed to capture data-types that do not exist in the etcsl.

The need for universality and the benefits of reuse are compelling arguments in favor of the use of these ontologies in further Assyriological investigations and of the large-scale publication of Sumerological data as RDF. The ontologies also facilitate linking between the considered data and myriad other contexts, both historical and contemporary.

Three conclusions are clear. First, the ontological representation of Sumerian literary narratives remains an interesting, rich, and engaging topic, with great potential for future development and investigation. Second, existing research has done little more than scratch the surface, and future scholarship is likely to bring to light further parallels and patterns that help us better understand the literature of ancient Sumer. Third, SW technologies are in a position to help support existing investigative paradigms across an array of topics, enabling new types of scholarship which can enrich—and be enriched by—the philological analysis of ancient Sumerian literary compositions.

References

Alivernini, Sergio, Franco D’Agostino, and Marco Romano. 2006. “Ur_Namma: An OWL Ontology of a Sumerian Grammar.” Epistematica. <http://www.epistematica.com/docs/Ur_Namma.pdf>.

Allemang, Dean, and Jim Hendler. 2011. Semantic Web for the Working Ontologist: Effective Modelling in RDFS and OWL. Amsterdam: Elsevier.

Bekiari, Chryssoula, Martin Doerr, Patrick Le Boeuf, and Pat Riva. 2015. “FRBR Object-orientated Definition and Mapping from FRBRER, FRAD and FRSD (version 2.2).”
International Working Group on FRBR and CIDOC CRM Harmonisation. <https://www.ifla.org/files/assets/cataloguing/frbr/frbroo_v2.2.pdf>.

Berners-Lee, Tim. 2000. Weaving the Web: The Past, Present and Future of the World Wide Web by its Inventor. London: Butler & Tanner.

Black, Jeremy, and Gábor Zólyomi. 2007. ‘Introduction to the Study of Sumerian.’ In Analysing Literary Sumerian Corpus-based Approaches, edited by Jarle Ebeling and Graham Cunningham, 1–32. London: Equinox.

Brewster, Christopher, and Kieron O’Hara. 2004. “Knowledge Representation with Ontologies: The Present and Future.” IEEE Intelligent Systems 19 (1): 72–73.

Crawford, Harriet. 2013. The Sumerian World. London: Routledge.

Crofts, Nick, Martin Doerr, Tony Gill, Stephen Stead, and Matthew Stiff. 2008. Definition of the CIDOC Conceptual Reference Model. ICOM/CIDOC Documentation Standards Group. CIDOC CRM Special Interest Group 5. <http://www.cidoc-crm.org/get-last-official-release>.

Delnero, Paul. 2012. The Textual Criticism of Sumerian Literature. JCSSS 3. Boston: ASOR.

Ebeling, Jarle. 2007. “Corpora, Corpus Linguistics and the Electronic Text Corpus of Sumerian Literature.” In Analysing Literary Sumerian Corpus-based Approaches, edited by Jarle Ebeling and Graham Cunningham, 33–50. London: Equinox.

Ebeling, Jarle, and Graham Cunningham, eds. 2007. Analysing Literary Sumerian Corpus-based Approaches. London: Equinox.

Finkel, Irving L. The Ark Before Noah: Decoding the Story of the Flood. London: Hodder & Stoughton, 2014.

George, Andrew. 1999. The Epic of Gilgamesh: The Babylonian Epic Poem and Other Texts in Akkadian and Sumerian. London: Penguin.

George, Andrew. 2010. The Epic of Gilgamesh: The Babylonian Epic Poem and Other Texts in Akkadian and Sumerian. London: The Folio Society.

Goodnick Westenholz, Joan. 2013. “In the Service of the Gods: The Ministering Clergy.” In The Sumerian World, edited by Harriet Crawford, 246–276. London: Routledge.

Gruber, Thomas R. 1993. “Translation Approach to Portable Ontology Specifications.” Knowledge Acquisition 5 (2): 199–220.

Jaworski, Wojciech. 2008. “Contents Modelling of Neo-Sumerian Ur III Economic Text Corpus.” In Coling 2008: Proceedings of the 22nd International Conference on Computational Linguistics, 24 August 2008, Manchester, UK, Vol. 1, 369–376. Stroudsburg, PA: Association for Computational Linguistics.

Jewell, Michael O., Faith K. Lawrence, Mischa M. Tuffield, Adam Prugel-Bennett, David E. Millard, Mark S. Nixon, m.c. schraefel, and Nigel R. Shadbolt. 2005. “OntoMedia: An Ontology for the Representation of Heterogeneous Media.” <https://eprints.soton.ac.uk/261024/>.

Kranzberg, Melvin. 1986. “Technology and History: ‘Kranzberg’s Laws.’” Technology and Culture 27 (3): 544–560.
Lawrence, Faith K. 2008. “The Web of Community Trust Amateur Fiction Online: A Case Study in Community Focused Design for the Semantic Web.” PhD diss., University of Southampton.

Levesque, Hector J. 1984. “Foundations of a Functional Approach to Knowledge Representation.” *Artificial Intelligence* 23: 155–212.

Nurmikko-Fuller, Terhi. 2014. “Assessing the Suitability of Existing owl Ontologies for the Representation of Narrative Structures in Sumerian Literature.” *Current Practice in Linked Open Data for the Ancient World, ISAW Papers, 7*, edited by Thomas Elliott, Sebastian Heath, and John Muccigrosso. <http://dlib.nyu.edu/awdl/isaw/isaw-papers/7/nurmikko-fuller/>.

Nurmikko-Fuller, Terhi. 2015. “Telling Ancient Tales to Modern Machines: The Ontological Representation of Sumerian Literary Narratives.” PhD diss., University of Southampton.

Oldman, Dominic, and Barry Norton. 2014. “A New Approach to Digital Editions of Ancient Manuscripts using CIDOC CRM, FRBROO and RDFa.” *The Digital Classicist*. <https://www.youtube.com/watch?v=cK54YlY-xZs>.

Robson, Eleanor. 2013. “Lone Heroes or Collaborative Communities? On Sumerian Literature and its Practitioners.” In *Ancient Egyptian Literature: Theory and Practice*, edited by Roland Enmarch, and Verena M. Lepper, 45–61. Oxford: Oxford University Press.

Taylor, Jon. 2013. “Administrators and Scholars: The First Scribes.” In *The Sumerian World*, edited by Harriet Crawford, 290–305. London: Routledge.

Wilks, Yorick, and Christopher Brewster. 2006. “Natural Language Processing as a Foundation of the Semantic Web.” *Foundation and Trends® in Web Science* 1 (3–4): 199–327.

Woolley, Charles Leonard. 1934. *Ur Excavations 2. The Royal Cemetery: A Report on the Predynastic and Sargonid Graves Excavated between 1926 and 1931*. London: The Joint Expedition of The British Museum and of the Museum of the University of Pennsylvania to Mesopotamia.