

Sharing Knowledge about Interreligious Connections

A Linked Data Approach to Medieval Studies

By Postdoctoral Researcher Jacob Langeloh



All research eventually leads to the question: How do we share what we know? In this article, I will explain why I have used technologies from the Semantic Web to model historical interreligious connections and how this procedure has helped me to shed new light on the origins and reception of a key work by Riccoldo da Monte di Croce (d. 1320), a Medieval Dominican and missionary among Muslims in the Arab world.

From ‘Orientalism’ to Data

During the past two years, I have worked at the Faculty of Theology as a Marie Skłodowska-Curie fellow on a project called OTRA. The initial question seemed simple: Was Edward Said correct in claiming that the Western picture of Islam was both distorted and unchanging over centuries? Or: Do Christian texts from a Christian perspective, despite being polemical and one-sided, reflect some development due to changing circumstances and increased scholarship in early-modern times?

Traditionally, this question would be approached by reading as many texts as possible, extracting tendencies, and then

producing a scholarly synthesis. But is this method really suitable for such a long history with an almost inexhaustible number of sources?

To simplify the task, I made one basic assumption. I posited that interreligious discourse always relies on previous texts, for two reasons. First, to prove one’s own religion right and the other wrong, authors must cite proof from relevant scriptures. Second, information about the religious ‘other’ is often difficult to obtain.

Two translations of the Qur’an into Latin existed in the Middle Ages: one by Robert of Ketton (c. 1140) and one by Mark of Toledo (c. 1210). However, these translations were not always accessible, leading authors to frequently rely on secondary texts without proper attribution (in modern parlance, ‘plagiarizing’).

My approach was to document which scriptures or secondary texts were cited and what arguments were based on them. By comparing the citations, we can see if the areas of interest changed over time or remained focused on the same verses of the Bible and Surahs of the Qur’an. Similarly, by analysing the arguments,

we may be able to register whether the interpretation of Quranic or biblical verses evolved over time.

Despite this conceptual simplification, the sheer number of sources remains vast. The final component, therefore, was to represent this knowledge – citation and argumentation – through a consistent, digital format named Linked (Open) Data. Doing so, I could share my complete findings, not just the synthesis, and enable others to add their insights. Even though *I* might not be able to completely answer the question I started out with, I would provide part of the data necessary to do so and provide a blueprint for others to do the same. In time, one would collectively be able to give a detailed and substantive answer to the challenge raised by Said.

My Contribution to the Story

For my contribution, I studied one particular author and his influence, namely, the Dominican Riccoldo da Monte di Croce (d. 1320). In the late 13th century, this Florentine friar first travelled to the Holy Land, and then went east to convert Muslims. He reached Baghdad, learned Arabic, and tried to have religious disputations with the learned men there. Around 1300, he returned to Italy, putting his experiences down in an acrimonious treatise titled “Against the Law of the Saracens” (*Contra legem Sarracenorum*). His work is full of allusions to and citations of the Qur’an and other related works such as Hadith (sayings of the Prophet Muhammad) and the writings of other Christians such as Peter the Venerable, Thomas Aquinas, and the *Liber denudationis*.

As an authoritative work that exuded first-hand knowledge, *Contra legem* became highly influential. For instance, it was translated into Greek by Demetrios Kydones (around 1370), this version was then re-translated into Latin by Bartholomaeus Picensis (1506), whose rendering was subsequently turned into German by none other than Martin Luther as his *Verlegung des Alcoran* (1542). What a curious chain of transmission! Aside from direct translations, portions of *Contra legem* were frequently reused to create other, more or less independent works.

Throughout my project, I digitally modelled the references and arguments that *Contra legem* contains, and how other texts from the following centuries use *Contra legem* to make their arguments.

What is Linked Data?

What do I mean by digitally modelling references, and arguments? There are many ways to do so, but I chose Linked Data, a concept proposed by the inventor of the internet, Tim Berners-Lee in 2006. The evolution of the web has taken us from a web of documents – pages filled mostly with text and images (like news websites) – to a web of data, where we can search and retrieve information (such as library catalogues). The next step envisioned by Tim Berners-Lee was to connect the data through expressive links. If we could link all individual sources of data with each other, an all-encompassing network of knowledge would emerge. This vision is known as the Semantic Web. How does this work in practice?

An example: Consider we want to link data about actors and politicians. How do

we know that the Arnold Schwarzenegger who played Conan the Barbarian is also the one who served as Governor of California? The first step is ensuring that we are referring to the same Arnold. To do so, we assign him a URI, a Uniform Resource Identifier. On Wikidata, a great repository of such URIs, Arnold is listed as <https://www.wikidata.org/wiki/Q2685>. By using this URI, we are guaranteed to be talking about the same individual. This URI doubles as a URL, allowing us to learn more about him directly from our browsers.

But identifying Arnold is just the beginning – he must also be linked to other relevant data! In Wikidata, he is linked to the movie ‘Conan the Barbarian’ under the property ‘notable work’ (Property:P800) and to the office ‘Governor of California’ through the property ‘position held’. We can read these connections as complete statements like ‘Arnold Schwarzenegger’ ‘has as notable work’ ‘Conan the Barbarian’: knowledge is expressed in simple sentences of subject, predicate, and object. This is exactly what the so-called Resource Description Framework (RDF) that underlies the Semantic Web mandates, and such combinations of subject, predicate, and object are called ‘triples’.

As a final component, we need general terms to specify what *kinds of things* we are linking to each other. Arnold Schwarzenegger is probably not a cybernetic organism but a ‘person’ and ‘Conan the Barbarian’ is a ‘film’. To maintain consistency, these terms are regulated through controlled vocabularies called ‘ontologies.’ Ontologies contain ‘classes’, which we use to categorize individ-

ual entities, and ‘predicates’, which are used to link them. Both classes and predicates can have subclasses and sub-predicates that allow us to further specify the type of entity or connection. For example, on Arnold’s page, we would not find that he is a ‘person’, but rather that he is a ‘human’. However, at least in Wikidata, ‘human’ is a subclass of ‘person’. This entails that every human is also a person and that everything that applies to a person also applies to a human. The inverse is of course not true: not every person is human. (On Wikidata ‘god’ is also a subclass of ‘person’!)

Connecting Text Passages

The elements just described – giving pieces of data unique identifiers, linking them, and specifying them through classes and predicates – are not only the foundations of the Semantic Web, but they were also the challenges I faced while implementing my project on interreligious relations. Thankfully, existing ontologies provided vocabularies that I could ‘steal’ from. In fact, this is a best practice of the Semantic Web: *go as far as you can with the vocabulary that already exists, and only then invent your own!*

I spent a great deal of time addressing the challenge of uniquely identifying text passages. Texts exist in different editions and are cited by different publications in different ways. As a first step, I created unique IDs for the paragraphs of my texts. Text passages are then identified by providing the IDs of the paragraphs where they start and end, and by specifying the exact word number within these paragraphs where they start or end.

Then, I had to find a way to express that two text passages are connected with each other. Instead of stating that two passages *are* related, I chose to model that scholars *claim* they are related. In my ontology, I therefore defined ‘Connective Claims’ as statements made by scholars that suggest ‘passage a’ and ‘passage b’ have a certain connection which can lead to a better understanding of ‘passage a’. For example, the historical writer might allude to a certain passage of the Qur’an, and the modern editor would supply the best match. These first two steps build on an ontology by a project called the *Hypermedia Dante Network*.

As the next step, I classified what kind of connection the scholar suggests. For this, I took inspiration from a project called *Sharing Ancient Wisdoms*. Using their vocabulary allowed me to describe precisely how two texts are related. If they are labelled as ‘verbatim’, all words need to be the same. At the other end of the spectrum is ‘loose rendering’, which means there is some connection, but the material has been heavily transformed.

Finally, I described the argumentative content of the passage in simple statements and also linked them to text passages. It would have been beneficial to distil these statements into a limited number, so that parallels can be found more easily, but I have not managed to do this so far – work for the future!

What Is It Good For?

The intention behind OTRA, which, as I may now reveal, stands for an “Ontology for the Transmission and Re-Use of Argumentative Patterns”, was to study patterns of interreligious argumentation. The corpus I created offers some insights. For instance, I can ask my data which parts of Riccoldo’s *Contra legem* were the most used in the two centuries after its writing and I can retrieve how his work was used to support different arguments.

I can also study the long influence of sources. As mentioned earlier, Riccoldo’s work drew from the Latin version of the *Liber denudationis*, which itself is based on an Arabic text from 11th century Egypt. Modern editions of texts that use *Contra legem* will generally only mark their dependence on Riccoldo. However, my data allows me to go one step (or several steps) further, and find out how later works indirectly depended on this original source.

My dataset is still too limited to study patterns of interreligious argumentation over long periods. However, it embodies the spirit of Linked Data: if everyone does something, eventually a more comprehensive picture will emerge that goes far beyond our capacity as individual scholars.