Computational Theology and Information Visualization

Janelle Peters

Abstract Information visualization assists in interpreting theological texts by allowing for close and distant readings. For close reading, visualization allows the comparison of manuscripts and linguistic features of a theological text. Distant readings using information visualizations of theological texts follow the schema of Moretti's graphs, maps, and trees. Oxford's *Paul and Patristics* project, helmed by Jenn Strawbridge, and *Coptic Scriptorium,* created by Carrie Schroeder, use graphs to illustrate word counts. Maps can be made using Stanford's ORBIS, an interactive map of geographical, economic, and social data of the late antique Mediterranean. Stanford's Palladio, which was originally designed to analyze social networks of elite medieval families, can visualize the location of Dead Sea Scrolls among the various Qumran caves and the Cairo Genizah. Trees have followed in the formidable legacy of Busa's *Index Thomisticus.* Advantages of these readings include the ability to assemble data on word count, variants, emendations, morphology, syntax, and the cultural backgrounds of manuscripts. Disadvantages to information visualization range from distorting the certitude of the translation or network via aesthetics to the durability of the analytical programs and final produced media. Possible future applications of information visualization in theological studies could look to the Map of Science, visualizing the networks of various researchers and their projects, and recreations of locations described by texts and locations in which texts were performed, read, and reproduced through visualization projects using technology such as augmented and virtual reality.

Keywords History of Visualization, Graphs, Maps, Trees, Close Reading, Distant Reading, Theological Data Visualization

Different forms of information visualization have been created independently of innovations in publishing (e.g., scroll, codex, printing press). From the earliest cave art onward, maps have hinted at human interactions with their physical and social contexts. Often encapsulating economic and logistical data, charts have connected different categories of data with one another, particularly after the line, bar, and pie charts of William Playfair (1759–1823).

Vast amounts of data can form an assemblage that invites the specialist or casual reader to play with different interpretations. Relational graphics such as the scatterplot, for instance, ask for interpretation about the causal relationship between X and Y (Tukey 1983, 47). The boxplot, whether as imagined by John Tukey (1915–2000) or Edward Tufte (b. 1942), allows the variation of a measurement to be seen.

At the same time, many underlying features of information visualization have been present in theological studies since quite early. Concordances have visually

Janelle Peters: Computational Theology and Information Visualization, in: Christopher A. Nunn and Frederike van Oorschot (eds.): Compendium of Computational Theology 1. Introducing Digital Humanities to Theology, Heidelberg: heiBOOKS, 2024, S. 315–331. DOI: https://doi.org/10.11588/heibooks.1521.c21956

tabulated the intensity of word references in theological texts since the medieval period, allowing users to interactively look up instances for themselves. Hugh of Saint-Cher (1200–1263), a French Dominican friar who would be made the first Dominican cardinal, is credited with creating the first concordance of the Vulgate and thus the Bible. He was said to have been helped by as many as 500 Dominican associates to produce the concordance in 1230, with divisions by book and chapter. Hugh was also a theological commentator, glossing and adding longer commentary to established theologians such as Peter Lombard. It is not a stretch to speculate that were Hugh to be participating in the current digital turn of theology he would be enthused about the many of the current projects connecting texts such as those of Alexander of Hales, Aquinas, biblical authors, and ancient philosophers. Rabbi Isaac Nathan created a concordance for the Hebrew Bible between 1438 and 1448 that was printed in Venice in 1524 and appeared in a Latin edition of 1556. From the sixteenth to eighteenth centuries, concordances also were important for studies in the Greek Septuagint, the Greek New Testament, and the Syriac.

Given the importance of concordances for theological studies, it is perhaps no surprise that the birth of the Digital Humanities is widely attributed to Italian Jesuit priest Roberto Busa (1913–2011), who collaborated with IBM to produce a digital concordance of the *opera omnia* of Thomas Aquinas. Busa announced in the July 1950 *Speculum* his intention to use "mechanical devices" to produce an *Index Thomisticus* (Busa 1950, 424f.).[1] The very next year, 1951, James Collins wrote a *Speculum* review for a work by Busa on Aquinas by noting the general problem of situating Aquinas' words in an individual work within his larger life and its influences: "A comparative linguistic approach to the authors consulted by Aquinas cannot be carried out at present on any large scale, and the method cannot be applied even to the entire Thomistic corpus until such a project as Busa's Index is completed" (Collins 1951, 495f.). Within the decade, Roy Wisbey had created a High German word index (Jones & Yeandle 2021). Within three decades, Busa, IBM, and an Italian committee of scholars had produced 49 volumes of the *Index Thomisticus,* which analyzed 179 texts (118 of the Thomistic corpus and 61 of other authors from the ninth to sixteenth centuries). In addition, there are titles, quotations, paraphrases of other authors, references, and transcriptions of the author's spoken words (Burton 1984). The legacy of this work is available on paper, on CD-ROM, and online.[2] Other projects have developed based on the work of Busa's team. One of these is the *Index Thomisticus Treebank* at the CIRCSE research center of the Università Cattolica del Sacro Cuore in Milan, Italy,[3] which analyzes Aquinas' morphology, form, and meaning (Passarotti 2019). The collaborative

1 Busa had been inspired by his experience as a doctoral student of having to produce 10,000 cards to discuss *praesens* and *praesentia* in Aquinas and how they occur with the word "in". Cf. Busa 1980, 83.

2 S. http://www.corpusthomisticum.org, accessed 24 June 2024.

3 S. http://centridiricerca.unicatt.it/circse, accessed 24 June 2024.

and yet still prestigious ethos of Aquinas information visualization is a real strength and, hopefully, a model for future scholarship.

Information visualization has advantages in close, mid-range, and distant reading of theological texts. For close reading, morphology, textual structures, and semantic and cultural meanings can allow readers to annotate the text with richer interpretations. Variant readings can be inserted into the text, and intralinear comparison to major intertext of the text might be made (cf. J. Nantke's chapter in this volume). For distant reading, it allows viewers to interrogate correspondences among data and to intellectually apprehend large quantities of data within a data set, which can bring together traditionally disparate items. Named entities of a text can be made apparent, even so-called "unnamed" figures such as the biblical "Lot's wife," "Jepthah's daughter," "Chloe's people," "Pilate's wife," or "Rufus' mother" (cf. E. Gius's chapter in this volume). Timelines have the capacity to illumine when different named entries appear in the text, allowing the reader with knowledge of the plot or document outline to better analyze when references occur.

Disadvantages of information visualization include: (1) the impression that the data and its metadata underlying the visualization are more securely labeled than they are in reality (e.g., literary allusions that could be alternatively religious founder sayings or later follower expansions are pinned down historically leading to a falsely strong picture of either the expressions of the founder or the followers in the allusion aggregation); (2) the impression that networks are more cohesive or relational than they are (e.g., a prosopography of indifferent acquaintances); (3) the impression that the categories of the data visualization are static over time and cultures (e.g., a world map of the phoenix); (4) the impression that the data visualization's categories constitute a particularly salient definition of the phenomenon (e.g., portable altars); (5) the impression that the data visualization's neat and even "beautiful" ordering makes just-so sense of tragedies and atrocities; (6) the impression that statistical frequency is more important than momentous silences or rare occurrences.

In this article, I will analyze several aspects of data visualization in theological and religious studies. First, I will describe current data visualization projects on different aspects of theology, focusing on the use of *Close Reading, Distant Reading,* and *Mapping* to make sense of large quantities of data for both scholars and the general public. Then, I will analyze some of the problems with data visualization as they occur in theological data visualization. Lastly, I will compare the work on data visualization being performed in theological and religious studies with the work on data visualization being performed in other fields (e.g., various sciences) to suggest possible avenues for development of theological data visualization such as virtual and augmented reality worship setting visualizations.

1. Information visualization with close and distant reading

1.1 Close reading visualization

Close readings exist at the story-level of the text, balancing the literal and the allegorical meanings of the text (Szondi 1978). The word's definition, its meaning in that text, and its appearances elsewhere in contemporaneous other literature (e.g., other instances of *pneuma* in Greek sources contemporary to the Gospel of Mark) – rather than contemporaneous manuscripts of the same document (e.g., 11th-century copies of the Gospel of Mark) – all matter. If objects described in the text have a somewhat standardized appearance in material culture, it is helpful to include them, though some might have wide variation (e.g., the plethora of animals on rhytons) and some might only be attested by much later artistic representations (e.g., the Ark of the Covenant).

Digital visualizations have allowed readers to click on words of the text for expanded information about meaning, grammar, statistical frequency, and other relevant information. In this capacity, data visualizations serve not only to further the goals of the exegesis of literary texts but to illuminate for individual interpreters without formal theological training what comprises critical theological interpretation. Of note is the fact that the components of theological interpretive practice are most closely related to the close reading practices of literary studies with linguistic and historical additions. When a viewer encounters a data visualization that gives a manuscript's text along with information about each lexical unit on the page, practices like sifting through the *peshat* interpretations of others along with the *derash* interpretations are thus not included.[4] These must be left for other visualizations, such as connecting the thought of contemporary authors through their shared themes and literary allusions in close, distant, or even midrange readings.[5]

Linguistic analysis is essential to close reading, and many projects allow users to explore the linguistic features of words. A particularly elegant reader is available from the Quranic Arabic Corpus[6], which is housed at the University of Leeds. The grammar, syntax, and morphology for each Quranic verse are displayed. The Quranic Arabic Corpus features the Arabic text along with the transliteration and definition, meaning that a user with no Arabic knowledge may click on the transliterated word and be provided with a list of other occurrences of that defined word. A variety of colors help to introduce relevant lexical units of information for non-specialists. As it

4 For an explanation of how the work of finding the peshat interpretations must be ongoing at the same time that previous peshat interpretations may be appreciated, see Simon & Greenstein 1988.

5 For a description of how human-edited outlines of a text might be analyzed for digital visualization, see Booth 2017.

6 See https://corpus.quran.com (Accessed: 24 June 2024).

does with classical Latin and Greek texts, *Perseus*[7] provides the standard Greek text of New Testament books and allows the user to click on each word for morphological analysis. This feature is available with paid applications for electronic readers such as the Kindle that allow the user to look up the form of any ancient Greek word in any ancient Greek book loaded onto the device, as happened on the first day of my *1 Clement* reading group. It is also available with paid products with carefully curated Greek texts such as *Logos*, which has supplanted the now-defunct *BibleWorks*. These projects are oriented toward allowing both specialists and those with little ancient language knowledge to use the induction method of engaging with the ancient language of the text. They greatly improve the speed at which someone may read any Greek text, though the semantic range of the provided definitions and the relevance of the comparanda of each word might still be improved with advanced in visual representation capacities. Advanced textual tools are available to assist specialists with analyzing user-provided segments of texts. These include those of the *Coptic Scriptorium*[8]. All of these tools allow readers to engage texts with more fluidity and knowledge.

Close reading data visualizations have the capacity to display one to two related texts by intralinear comparison. Readers of the New Testament have long performed such comparison through printed apparatuses putting related portions of the Synoptic Gospels together, and a similar data visualization is possible digitally. As the *digital Dante* project at Columbia University[9] shows, it is possible to compare Dante with Ovid in the original languages through the use of highlighting in yellow the lines of Ovid that inspired Dante and then including a caption that explains in English how the connection is relevant. Many more such projects could exist, particularly for texts such as biblical texts that have myriad texts from which they drew (e.g., Babylonian literature, Egyptian literature, Greek literature), myriad texts with which they shared a cultural milieu (e.g., early rabbinic texts, Greek and Roman texts), and myriad texts they inspired.

Variant readings demonstrate another area where visualization helps in the interpretation of the text. A visualization tool called *CollateX* in the VMR *(Virtual Manuscript Room)* brings together alternate attestations of the words of a text. In the *Codex Sinaiticus* Project, the manuscript and its transcription are displayed next to each other and suggested textual emendations may be found by hovering the cursor over highlighted text (Batovici & Verheyden 2021). As most scholars accept that there was a "Gospel of Mark" or a historical Jesus who uttered sayings, these attested words and suggested emendations are important because it is possible that they represent earlier and more authentic versions of the passage. For modern authors, variant readings are important, because there are authors such as Emily Dickinson who had variations

7 See http://www.perseus.tufts.edu (Accessed: 24 June 2024).
8 See https://copticscriptorium.org (Accessed: 24 June 2024).
9 See http://digitaldante.columbia.edu (Accessed: 24 June 2024).

within their own poems. In any event, variant readings preserve ancient and alternative interpretations of how a phrase could be imagined, and being able to see these even momentarily as part of the text instead of a footnote would be the ultimate goal of a reader. With practical limitations being what they are, variant readings in the collations provide a glimpse into a multiverse of interpretative potential.

1.2 Distant reading visualization

Distant readings of a text or multiple texts investigate detail patterns that might be missed by close reading. Writing of general literary studies and not in theological literary studies, Moretti posited that distant reading involved three separate domains: graphs for the linguistic details, maps for geographical associations, and trees for genres (Moretti 2005). I will apply this general tripartite structure in modified form to theological texts. In responding to comparisons of close and distant reading more than a decade after his original work, Moretti clarified that "attention to linguistic detail is common to both choices; all the rest is different" (Id. 2017, 688). Jänicke et al. frame the difference between close reading and distant reading as one of structure: "While close reading retains the ability to read the source text without dissolving its structure, distant reading does the exact opposite. It aims to generate an abstract view by shifting from observing textual content to visualizing global features of a single or of multiple text(s)" (Jänicke et al. 2017). Information visualizations have the advantage of providing statistical rigor to one's general impression that certain features of a text are recurring often, and they also have the ability to identify significant features of a large corpus that multiple readings and readers might have missed, whether a repetition of the word "one" or a greater preponderance of unnamed entities in some texts as opposed to other texts (Esposito 2022).

a. Graphs

Many projects within the realm of early Christianity have sought to provide what Digital Humanists have called distant readings of the corpora of early Christian texts in order to map common convergences.[10] One example of such a project is the *Paul and Patristics* project[11] of Jenn Strawbridge and Martin Hadley at the University of Oxford. Hadley organized Strawbridge's research into interactive tables, bar charts, and bubble matrix plots, all common forms of data visualization. The interactive table allows visitors to ask the database to display patristic authors by scriptural

10 For a definition of distant reading as pulling features of books out of their context to read against one another, see Hayles 2013.

11 See http://idn.it.ox.ac.uk/article/paul-and-patristics (Accessed: 24 June 2024).

references, region (Italy, Greece, Asia Minor, Egypt, North Africa, Palestine, Gaul, and unknown), and more. The bar chart is supposed to display references to authentic and disputed Pauline letters by century, early Christian author, and region; however, only the century option is consistently successful, and it highlights the predominance of Romans and 1 Corinthians diachronically. With the early Christian writer bubble matrix plot, one receives information by author about the author's Pauline references. Each graph stands on its own, and the information, though related, does not color one's view of the other graphs.

Another project that analyzes early Christian literature through charts is *Coptic Scriptorium.* The provided data visualizations focus on named/unnamed entities in a variety of corpora in world literature, ranging from Ruth to Cyrus to "English literature" to Shenoute to the New Testament. The data visualizations show that Coptic texts vary in named/unnamed entities but fall on the lower end of the spectrum of all world literature, with a traditional bar chart showing Shenoute below Mark and Besa below 1 Corinthians. This data visualization is particularly interesting, because it combines the computing interest in *named entity recognition* (cf. the chapter by E. Gius in this volume) with the theological interest in why some characters in texts appear unnamed and whether the lack of specificity indicates a lower status of the unnamed. With this data visualization, it is possible to see that the named and unnamed elements of the text might have meanings other than those attributed to them by textual scholars only reading for Jephthah's daughter, Rufus' mother, or Valens' wife. Such a possibility has been actualized in Schauf & Escobar Varela's (2018) network analysis that found female characters in the Javanese *wayang kulit* of the *Mahabharata* epic appear only sporadically and yet demonstrate high frequency of establishing bridges or betweenness in the story.

b. Maps

Maps are important supplements to textual analysis, because many of our religious texts mention multiple locales – often through travel – for the purposes of conveying the image of spreading the word extensively.[12] A major advantage of the digital turn of scholarship is the expanded capacities for the widespread dissemination of maps that it yields. As a visualization that typically relies heavily on color to convey meaning, a map is usually an expensive figure to include in a printed book, and its color possibilities are somewhat limited. With the progression of graphic user interface technology, it is now possible for maps to have a brilliant array of colors and to even allow for close-ups of the terrain on a platform like *Google Maps.* Not all time periods

12 As John Elsner says in analyzing the far-flung travels of Apollonius of Tyana, "[b]oth travels and miracles are instances of *thaumata,* the kinds of wonders which one would normally associate with a late antique holy man." (Elsner 1997, 23).

have equal access to the mapping possibilities, even for platforms such as Stanford's ORBIS that are built for Mediterranean antiquity.[13] Map interpretations still customarily limit one to a certain configuration of political geography that was extant for only a few years or decades. Coastlines have similar limitations, meaning that magnifying the terrain on *Google Earth* for every time period will simply not produce an accurate representation consistently as many ancient coastal cities have moved closer or farther from the sea due to silting and seismic activity (cf. the chapter by M. Wilkens in this volume).

Some of these problems may be viewed on non-scholarly interactive maps of Paul's journeys, which are the only maps the major search engine Google displays on the first few pages of results on a search for Pauline travels. *Google Maps* has a map of the journeys of Paul that has been created by simply pinning potential stops of Paul on a modern map. *Viz.Bible* is a project created by a sole Christian layperson, Robert Rouse. Unlike the *Google Maps* visualization, *Viz.Bible* has both an ancient map and a modern map. Four colors – green, purple, orange, and blue – indicate the potential routes of Paul. Yet, the main *interactive* component of this digitalization comes from the city links that lead to what should be information pages on a Classics site. Unfortunately, the information of the linked cities is not accessible, rendering the slider that changes the map from ancient to modern the only functioning interactive aspect of the visualization. Though *Viz.Bible* uses *OpenStreetMap,* which has some university affiliation through UCL *(University College London),* the two physical geography maps do not have a clear connection with one another. The lack of interactive mapping for figures from the New Testament, such as Paul, who have clearly delineated cities that they visit must be attributable to the scarcity of digital resources for biblical studies and the fear that technology is advancing too rapidly to make an interactive visualization viable for a sufficient amount of time to justify committing financial and temporal resources to such a project.

One workaround is to modify one's research question to fit the resources available by a broadly available humanities tool. For my undergraduate classes on early Judaism and Christianity, I had the students analyze the languages, locations, and purposes of Dead Sea Scrolls using Stanford's *Palladio* tool[14], which was developed for later historical periods. By interacting with the data themselves, students were able to understand the diversity of scrolls found at Qumran and in the Judean Desert, noticing which caves held which documents and whether there might be patterns. In addition, they also created word heat maps with readily available online tools and *R* from the so-called "sectarian scrolls," revealing a concern for concepts like the *Temple,* a concept that exist in physical form in Jerusalem.

13 See https://orbis.stanford.edu (Accessed: 24 June 2024).

14 See https://hdlab.stanford.edu/palladio (Accessed: 24 June 2024).

c. Trees

Moretti drew his distant-reading trees from evolutionary theory (Moretti 2003, 67), but trees as a data visualization tool for theology might also be seen in the linguistic treebanks associated with corpora, possibly influenced by Busa's use of the Porphyrian tree (Testori 2017). Many of these treebanks use *Universal Dependencies*[15], including the Latin treebanks: *Index Thomisticus Treebank* (ITTB); *Late Latin Charter Treebank* (LLCT); *Perseus; PROIEL; UDante.* Because of the diverse and diachronic usage of Latin, no single treebank can represent all theological Latin, stretching from earlier than the Vulgate to later than Vatican II (Cecchini et al. 2020). The treebanks of theological corpora, then, do not follow an evolutionary framework, but they point in the aggregate to the development of Latin as a language with a strong theological tendency. In exploring the treebanks, one might approximate the passage of theological time without having to slog through each minute itself as in close reading.

1.3 Summary

Data visualizations performed by modern scholars are able to bring together hundreds of manuscripts that were written over many centuries in a way never before possible. Access to manuscripts – provided they have been digitized and authorized – is radically democratized. In such a way, it is easier to compare manuscripts to one another and to see the possibilities for diachronic and synchronic comparison. Close reading is helped by visualizations of other occurrences of a word, morphology, syntax, definition(s), variants, emendations, cultural background, and even geography. Distant readings help create graphs of character and textual relations, maps arising from manuscript sources, and trees of indices. All in all, visualization of theological texts serves to enhance the exegetical reading of texts and corresponding understandings of how texts fit into the broader literary, theological, social, and political landscape.

2. Problems with information visualization

Although visualizations of theological texts are inherently useful, their creation is fraught with potential problems. These range from creating the impression that numerical frequency is what constitutes religious feeling in a text to detracting from the beauty of the text with a distractingly aesthetically pleasing chart. Some of our data simply has not found its way into our data sets due to a lack of sufficient resources

15 See https://universaldependencies.org (Accessed: 24 June 2024).

to digitize all available liturgical and other theological texts, and visualizations of the current categories might reinforce the neglect of theological texts outside of liturgical scrolls and codices, such as *mezuzot*, door lintel inscriptions, and other texts placed on practical objects. While these problems do not seem of a magnitude to preclude visualizations of theological texts, it is important to be aware of opportunities to better envision the textual world.

Aesthetics are an important part of visualization charts, maps, and trees, and they have the potential to distract from the message of the information. Though most data visualizations are produced digitally, they are expected to resonate artistically and occasionally have hand-drawn aesthetics. This means that data visualizations can inspire feeling in a way that is potentially unconnected with the data the visualizations interpret (Brinch 2020, 259f.; Simpson 2020,164). As many scholars have noted, it is possible for exceptionally "beautiful visual renderings" to signal authority and accuracy where none exists (Laaksonen & Pääkkönen 2020, 104.). In instances where the data visualization is of a negative phenomenon, such as the spread of tuberculosis evidenced by parish records, there is the potential for the beautiful aesthetics of the data visualization to dull the depth of the anguish felt by individuals and their social circles without some complementary information such as personal narratives (Nygren et al. 2016). On the other hand, it would seem that some readers have become so accustomed to data visualizations that readers do not interact with them. A study of a *New York Times* website identified that only 10–15 percent of readers attempted to absorb information in its interactive data visualizations, potentially signaling the data visualization is accepted as an unquestioned authority for a story (Ridley & Birchall 2020, 132). Literary critics have also pointed out that the rationale for close reading is to find the beauty in the text, which seems to be an enterprise that is not at all present in information visualizations, which satisfy themselves with being aesthetically pleasing in and of themselves instead of pointing the way to innate beautiful texts, whether theologically or secularly literary ones (Aquilina 2017, 513).

In processing the texts, different subspecialties of theology might encounter different problems and certain theological authors might have a way of framing linguistic conventions that diverge from other contemporary linguistic practice. Chinese and Arabic, for instance, require more preprocessing for tokenization than languages such as English, requiring their own segmenter.[16] While the concept of *presence* in Latin can take on very mundane meanings, Busa began his Index specifically because he wanted to account for the connection in Aquinas' thought of *praesens* with the preposition *in*.[17] This would complicate visualizations, because not all word heat maps or word clouds with *praesens* would accurately convey the nuance found in Aquinas or another instance where tokenization becomes crucially important.

16 See https://nlp.stanford.edu/software/segmenter.html (Accessed: 24 June 2024).

17 According to Busa, "I soon learned that such words in Thomas Aquinas are peripheral: his doctrine of presence is linked with the preposition *in*" (Busa 1980, 83).

Another potential concern of information visualization is the perception that statistical frequency automatically yields greater textual and social significance. A class data visualization project on Islamic sources run by Jeff Blecher at George Washington University demonstrates several areas where number does not automatically equal the correct and singular response. If sources attested multiple spellings for the same city (e.g., Medina, Medinah, and Madinah), would a standard spelling be necessary and would the standard spelling have to be the most attested? If listed occupations range from judge to scholar, would the most frequent automatically be the most powerful? How would individuals with multiple occupations divide their importance (Blecher 2016, 238)? Scholarly studies into the Septuagint find a similar diversity in that the LXX treats Hebrew *hapax legomena* in a variety of ways: omission, transliteration, association with similar-looking words, and so forth (Verbeke 2008; 2011). Information visualizations relying on statistical frequency lose the richness of *hapax* words and their transference into various texts. Pilate's wife, for instance, might only feature in one of the four canonical gospels, but that does not make her action of unsuccessfully attempting to warn her husband just as Calpurnia was thought to have tried to warn Caesar any less portentous.

Likewise, information visualizations have to contend with very sporadic manuscript availability, with certain regions and languages heavily privileged over others due to historical and climactic circumstances. In order to accommodate manuscript collections and academic training patterns, projects form around texts, textual traditions, languages, regions, and manuscript dates. Choices have to be made: how should we organize the scant information we have? One answer is from the item generating system at the *Trismegistos* Project[18], which creates items based on the material manuscript. With *Codex Alexandrinus,* there is an example of how a rarely found text in manuscripts and reception history *(1 Clement)* can have an early instance of it being included with the canonical New Testament. In projects such as that of *Trismegistos,* the material fact of the text's inclusion may be accounted for without recourse to other instances of a New Testament sans *1 Clement.* Thus, TM 62318 is linked to British Library images of *Codex Alexandrinus* and described as having the Old Testament, the New Testament, Athanasius, and Clement of Rome despite missing some verses of *1 Clement.* Allowing manuscripts to retain the particular circumstances of their creation and re-finding gives a more historically accurate and less anachronistic version of history. Another answer is that texts might be virtually reunited, whether in the sense of rejoining a manuscript that has its fragments scattered across the globe or in the sense of bringing together a number of witnesses to a sketchily present text (Miyagawa 2021, 181; Wolf et al. 2011).

Durability represents a further consideration of information visualization. Planned obsolescence is necessary for Digital Humanities projects. Projects require ongoing physical resources (e.g., servers, storage, electricity), and they may encounter

18 S. https://www.trismegistos.org (Accessed: 24 June 2024).

problems with links to external sites that change or with applications that are no longer supported. While it may be beneficial to students in a particular course to use a contemporary business trend such as *R*, *Python*, or *SQL* to create data visualizations, durability of a project will be assured by translating true humanities projects – such as the *Index Thomisticus* – to the digital platform and letting those inform data science rather than letting data science set the parameters of what constitutes a *good* project in data visualization. Just as IBM and punch cards are no longer as cutting edge as they once were, information visualization that is too *of the moment* will be restricted to that moment as technological advancement continues along the lines of Moore's Law or whatever law replaces it. Transferring digital information visualization back to the analog sphere via paper publishing might not only help to preserve the project longer, but it would also save electricity and carbon credits. Recyclable transcriptions and providing a "what you see is what you get" environment in which to produce major editions rather than forcing everyone into XML or its analogue should be viewed as a best practice (Houghton 2013).

3. Comparing data visualization in other fields to theology

Caveats aside, data visualization held a place in academic circles prior to the *Index Thomisticus,* and it represents a useful way to engage readers in the interpretation of theological texts. Information visualization will only continue to become more sophisticated in terms of metadata analysis and visual display. Some practices in academic data visualization would benefit data visualization in theological and religious studies. Among these are research maps and augmented and virtual reality projects.

In science, the idea of a *Map of Science* as initiated by UCSD and continued by Georgetown helps to analyze researchers, projects, and funders of various areas of science.[19] For UCSD, the *Map of Science* needed to bring together two sets of publishing data (Börner et al. 2012). The analyses performed with these maps are perhaps imprecise. Various scientific papers have to be incorporated in the dataset by title, keywords, and abstracts regardless of the academic departments of the researchers, leading to the possibility that some articles might be missing and some articles might be duplicated. Research clusters of 50 to 15,000 form when researchers cite one another in papers, leading to the possibility that new fields and subfields might be created. When gender, institutional affiliation, country, ethnicity, and other categories are known, analyses might be run about the strengths of various categories in various traditional and emerging fields of sciences. Researchers might find collaborators working on the same topics by consulting the latest iteration of the *Map of Science.*

19 See https://sciencemap.eto.tech (Accessed: 24 June 2024).

A comparable capability to visualize the scholarly intersections of the field could benefit theological studies. It would be possible for researchers on different biblical themes in different disciplines to find each other and collaborate. Areas of theology neglected regionally could be addressed. Potential gender gaps in areas of research could be found. Trends in established subdisciplines could be identified, and emerging subdisciplines could be discovered. As a Catholic biblical scholar, I have mostly been trained in biblical studies by Jewish, Protestant, Catholic, and atheist biblical scholars, but I have also used my Catholic education to publish on biblical reception and theological topics in the history of Catholicism. One wonders at how subdisciplines, religious traditions, theological training, and theological affiliation would intersect and diverge in a visualization of theological scholars and their research.

In addition to a map of theological studies that lays out subdisciplines and affiliations, another idea in the broader realm of digital data visualizations that would be helpful for theological studies is that of augmented reality and virtual reality. In the current digital landscape, there are scholarly virtual recreations of salient moments in religious history. One such project is the *Virtual Paul's Cross* Project, which takes its audience into a reconstruction of the church setting as John Donne gives a sermon to support James in his overcoming of the Gunpowder Plot. By developing the sounds of the audience and other features of the text, Wall has come to see the sermon less as a "text" than as a "trace" (Wall 2014). Our theological texts, however scriptural they may be, often existed originally primarily as speech in larger theological events. By being able to click from a description of a meeting between Peter and Paul in Jerusalem to a possible setting of such an encounter with its accoutrements based on archaeology of first-century Jerusalem, a layer of interpretive potential would be added to close reading as powerful as having other instances of a particular lexical choice attested by a majority of textual witnesses. Alternatively, were we situating our reading in a particular tome such as *Codex Alexandrinus,* we could have a visualization of a copying room in the fifth century C.E. Virtual and augmented reality projects are essentially multidimensional maps at close range, and they aid one in immersing oneself in circumstances that are no longer present.

4. Conclusion

Information visualization is an important tool for interpreting theological textual sources. Visualizing words in theological texts can improve their close reading through linguistic aids and literary-historical contextualization and enable their distant reading, whether in graph, map, or tree formats. In terms of the Digital Humanities, the entire field started auspiciously with the indexing of the complete works of Thomas Aquinas by Father Roberto Busa's scholarly team and the donated computing resources of IBM. Thus, in the sense that an index is a table visualizing word

frequency in corpora, digital data visualization too might be traced back to the late 1940s and 1950s. To claim that information visualization is particularly innovative, unknown, or threatening to existing ways of examining texts humanistically is hyperbolic. Data visualization enriches our understanding of concordances, a tool of theological study since at least the Middle Ages, and how they help us read our texts. Visualization allows us to see when authors are invoking a particular theme, text, or author repeatedly and diachronically. Language patterns across a collection of texts in a library can be better discerned through it.

There are several potential problems with information visualization in theological studies, which are typically found elsewhere. Visualizations might become in an end unto themselves, losing sight of the delight they were trying to convey in the first place. Durability becomes an issue when technology advances or external sites reorganize the content to which visualizations were linked. In certain cases, an emphasis on statistical frequency diminishes the emotional impact of intense events upon individual lives and their texts. Tokenization, lemmatization, and other preprocessing issues might distort visualizations in terms of word frequency or definition. The categories of the visualization might preclude other valid interpretations, such as the choice between reading a manuscript as a localized document in its own right or as an ancient and somewhat universal collation of earlier manuscripts.

All in all, though, information visualization has much to offer with these considerations taken into account. Theological scholars broadly should have the opportunity with democratized access to conduct visualization experiments to see what patterns emerge. Funding sources should follow the example of IBM in letting scholars of theology and the humanities to lead the way and determine the course and goals of the project, so that theology and the humanities may humanize the development of data visualization happening in other sectors of knowledge creation. To this end, should theological scholars look farther afield to see what aspect of data visualization they might bring into their discipline, the *Map of Science* – which brings together researchers in different subfields by charting connections between researchers – would be a helpful model. Theological data visualization projects might continue to inspire each other, whether in making treebanks of an already extant index or in creating new virtual and augmented reality projects that allow scholars to visualize the settings in which the text was first created and rehearsed along with the word frequency, authorial allusions, and so forth of the texts within given corpora. Although the process of information visualization might seem a little contrived and susceptible to oversimplifying data, one must remember and take inspiration from the verbal data visualization of the creation of world at the beginning of the Book of Genesis – and, indeed, the Bible – in its organization of creation into seven simple categories of light, sky, land, plants, animals, people, and rest and in the canonical need to provide yet another interpretation of the same event with a different data set.

References

Aquilina, M. (2017). The Work of the Literary Critic in the Age of Big Data, *Interdisciplinary Literary Studies,* 19(4), 493–516.

Batovici, D., & Verheyden, J. (2021). Digitizing the Ancient Versions of the Apostolic Fathers. Preliminary Considerations. In T. Hutchings & C. Clivaz (Eds.), *Digital Humanities and Christianity. An Introduction* (pp. 103–124). Berlin/Boston: De Gruyter.

Blecher, J. (2016). Pedagogy and the Digital Humanities. Undergraduate Exploration into the Transmitters of Early Islamic Law. In E. Muhanna (Ed.), *The Digital Humanities and Islamic and Middle East Studies* (pp. 233–250). Berlin/Boston: De Gruyter.

Booth, A. (2017). Mid-Range Reading. Not a Manifesto, *Publications of the Modern Language Association of America,* 132(3), 620–627. URL: https://www.jstor.org/stable/27037375 (Accessed: 24 June 2024).

Börner, K., Klavans, R., Patek, M., Zoss, A.M., Biberstine, J.R., Light, R.P., Larivière, V., & Boyack, K.W. (2012). Design and Update of a Classification System. The UCSD Map of Science, *PLoS ONE,* 7. DOI: https://doi.org/10.1371/journal.pone.0039464 (Accessed: 24 June 2024).

Brinch, S. (2020). What we talk about when we talk about beautiful data visualizations. In M. Engebretsen & H. Kennedy (Eds.), *Data Visualization in Society* (pp. 259–276). Amsterdam: Amsterdam University Press, 2020).

Burton, D.M. (1984). Review of *Index Thomisticus. Sancti Thomae Aquinatis operum indices et concordantiae* by Roberto Busa. *Sancti Thomae Aquinatis opera omnia* by Thomas Aquinas and Roberto Busa, *Speculum,* 59(4), 891–894.

Busa, R. (1949). Announcements, *Speculum* 25, 424f.

Id. (1980) The Annals of Humanities Computing. The Index Thomisticus, *Computers and the Humanities,* 14, 83–90. URL: https://www.jstor.org/stable/30207304 (Accessed: 24 June 2024).

Cecchini, F.M., Korkiakangas, T., & Passarotti, M. (2020). A New Latin Treebank for Universal Dependencies. Charters between Ancient Latin and Romance Languages. In *Proceedings of the Twelfth Language Resources and Evaluation Conference* (pp. 933–942). Marseille: ELRA. URL: https://aclanthology.org/2020.lrec-1.117 (Accessed: 24 June 2024).

Collins, J. (1951). Review of *La terminologia tomistica dell'interiorità* von R. Busa, *Speculum,* 26, 495–497.

Elsner, J. (1997). Hagiographic geography. Travel and Allegory in the Life of Apollonius of Tyana, *Journal of Hellenic Studies,* 117, 22–37. DOI: https://doi.org/10.2307/632548 (Accessed: 24 June 2024).

Esposito, E. (2022). Reading Images. Visualization and Interpretation in Digital Text Analysis, *Artificial Communication. How Algorithms produce Social Intelligence*

[Blog]. Cambridge, Mass.: MIT Press. URL: https://artificialcommunication.mitpress.mit.edu/pub/jtkikw5z (Accessed: 24 June 2024).

Hayles, N.K. (2013). Combining Close Reading and Distant Reading. Jonathan Safran Foer's 'Tree of Codes' and the Aesthetic of Bookishness, *Publications of the Modern Language Association of America,* 128(1), 226–231.

Houghton, H.A.G. (2014). The Electronic Scriptorium. Markup for New Testament Manuscripts. In C. Clivaz, A. Gregory & D. Hamidovic (Eds.), *Digital Humanities in Biblical, Early Jewish and Early Christian Studies* (pp. 31–60). Leiden: Brill [= *Scholarly Communication,* 2].

Jänicke, S., Franzini, G., Cheema, M.F., & Scheuermann, G. (2017). Visual Text Analysis in Digital Humanities, *Computer Graphics Forum,* 36, 226–250.

Jones, M., & Yeandle, D. (2021). Obiturary. Professor Roy Albert Wisbey (1929–2020), *German Life and Letters,* 74, 161–165.

Laaksonen, S.-M., & Pääkkönen, J. (2020). Between Automation and Interpretation. Using Data Visualization in Social Media Analytics Companies. In M. Engebretsen & H. Kennedy (Eds.), *Data Visualization in Society* (pp. 95–110). Amsterdam: Amsterdam University Press.

Miyagawa, S. (2022). *Shenoute, Besa and the Bible. Digital Text Reuse Analysis of Selected Monastic Writings from Egypt* [Diss.]. Philosophische Fakultät: Georg-August-Universität Göttingen. URL: https://ediss.uni-goettingen.de/handle/11858/13894 (Accessed: 24 June 2024).

Moretti, F. (2003). Graphs, Maps, Trees. 1, *New Left Review,* 24, no pag. URL: https://newleftreview.org/issues/ii24/articles/franco-moretti-graphs-maps-trees-1 (Accessed: 24 June 2024).

Id. (2005). *Graphs, Maps, Trees. Abstract Models for Literary History.* London: Verso.

Id. (2017). A Response, *Publications of the Modern Language Association of America,* 132, 686–689.

Nygren, Th., Frank, Z., Bauch, N., & Steiner, E. (2016). Connecting the Past. Opportunities and Challenges in Digital History. In M. Hayler & G. Griffin (Eds.), *Research Methods for Creating and Curating Data in the Digital Humanities* (pp. 62–86). Edinburgh: Edinburgh University Press.

Passarotti, M. (2019). The Project of the Index Thomisticus Treebank. In M. Berti (Ed.), *Digital Classical Philology. Ancient Greek and Latin in the Digital Revolution* (pp. 299–320). Berlin/Boston: De Gruyter Saur [= *Age of Access? Grundfragen der Informationsgesellschaft,* 10]. DOI: https://doi.org/10.1515/9783110599572-017 (Accessed: 24 June 2024).

Ridley, A.L., & Birchall, Ch. (2020). Evaluating Data Visualization: Broadening the Measurements of Success. In M. Engebretsen & H. Kennedy (Eds.), *Data Visualization in Society* (pp. 127–140). Amsterdam: Amsterdam University Press.

Schauf, A., & Escobar Varela, M. (2018). Searching for Hidden Bridges in Co-Occurrence Networks from Javanese *Wayang Kulit, Journal of Historical Network Research,* 2, 26–52.

Simpson, J. (2020). Visualizing data. A lived experience. In M. Engebretsen & H. Kennedy (Eds.), *Data Visualization in Society* (pp. 157–168). Amsterdam: Amsterdam University Press.

Szondi, P. (1978). Introduction to Literary Hermeneutics, *New Literary History,* 10(1), 17–29. DOI: https://doi.org/10.2307/468303 (Accessed: 24 June 2024).

Testori, M. (2017). Methods of Quality, Quality of Methods. What does Roberto Busa have to Communicate to Digital Humanists in the 21st Century? From Hermeneutics to Performativity, *Digital Humanities Quarterly,* 11(3), 1–9. URL: http://www.digitalhumanities.org/dhq/vol/11/3/000329/000329.html (Accessed: 24 June 2024).

Tufte, E.R. (1983). *The Visual Display of Quantitative Information.* Chesire, CT: Graphics Press.

Uriel, S., & Greenstein, E.L. (1988). The Religious Significance of the Peshat, *Tradition. A Journal of Orthodox Jewish Thought,* 23, 41–63.

Verbeke, E. (2008). The use of Hebrew hapax legomena in Septuagint studies. Preliminary remarks on methodology. In H. Ausloos, B. Lemmelijn & M. Vervenne (Eds.), *Florilegium Lovaniense. Studies in Septuagint and trextual criticism in honour of Florentino García Martínez* (pp. 507–521). Leuven: Peeters.

Ead. (2011). *Hebrew hapax legomena and their Greek rendering in LXX Job* [Diss.]. Faculteit Theologie en Religiewetenschappen: KU Leuven. URL: https://lirias.kuleuven.be/1875994 &lang=en (Accessed: 24 June 2024).

Wall, J.N. (2014). Transforming the Object of our Study. The Early Modern Sermon and the Virtual Paul's Cross Project, *Journal of Digital Humanities,* 3(1). URL: https://journalofdigitalhumanities.org/3-1/transforming-the-object-of-our-study-by-john-n-wall (Accessed: 24 June 2024).

Wolf, L., Littman, R., Mayer, N., German, T., Derschowitz, N., Shweka, R., & Choueka, Y. (2011). Identifying Join Candidates in the Cairo Genizah, *International Journal of Computer Vision,* 94, 118–135. DOI: https://doi.org/10.1007/s11263-010-0389-8 (Accessed: 24 June 2024).