

Priani Saiso, Ernesto. "Digging into Early Colonial Mexico: A Large-Scale Computational Analysis of Historical Sources from the Sixteenth Century."

Citation details

Early Modern Digital Review, vol. 4, no. 3, 2021, <https://doi.org/10.33137/rr.v44i3.38001>.

Renaissance and Reformation / Renaissance et Réforme, vol. 44, no. 3, 2021, pp. 228-231, <https://doi.org/10.33137/rr.v44i3.38001>.

Peer review

This is a peer-reviewed article in *Early Modern Digital Review*, distributed in print by *Renaissance and Reformation / Renaissance et Réforme*.

Copyright

Early Modern Digital Review materials are published under a Creative Commons 4.0 license (CC BY 4.0) that permits the right to share (copy and redistribute the material in any medium or format) and adapt (remix, transform, and build upon the material for any purpose, even commercially) the material, provided that the author and source are credited. The full description of CC BY 4.0 can be found on creativecommons.org/licenses/by/4.0/.

Early Modern Digital Review

Early Modern Digital Review is an online, open-access, and refereed journal publishing high-quality reviews of digital projects related to early modern society and culture. It is committed to productive evaluation of both established digital resources and recent tools and projects. Its publications are distributed online by the journal and its partners, and in print by *Renaissance and Reformation / Renaissance et Réforme*.

Murrieta-Flores, Patricia, principal investigator.

Digging into Early Colonial Mexico: A Large-Scale Computational Analysis of Historical Sources from the Sixteenth Century. Other.

Lancaster: University of Lancaster, 2018. Accessed 4 May 2021.

lancaster.ac.uk/digging-ecm/es/inicio.

One of the debates within the digital humanities revolves around how the field produces knowledge. Among the most frequent criticisms is, as pointed out by Tanya Clement, “a decoupling of method from the theoretical perspectives that would ordinarily help situate the kind of intellectual effort being engaged.”¹ In particular, the critique that digital methods do not conform to the literary theories from which their results are later interpreted has been levelled at distant reading. However, projects like Digging into Early Colonial Mexico (DECM) demonstrate that collaborative work in digital humanities often begins with the development of a digital methodology, in which it seeks to couple or, perhaps more precisely, codify the theoretical framework of research, before even proceeding with the interpretation of the results.

DECM is a multidisciplinary project. It involves the participation of archaeologists, historians, geographers, and computer scientists based at Lancaster University (UK), INESC-ID-University of Lisbon (Portugal), and the National Institute of Anthropology and History (Mexico) who focus on the development of more efficient digital technologies and methods for historical corpus analysis. The project is funded by the Trans-Atlantic Platform for Social Sciences and Humanities (T-AP) 2016 Digging into Data Challenge. By its nature, the project may be of interest to scholars exploring the application of digital methods in historical documents in general, and particularly for those studying the history and geography of the Spanish colonies in America, with a specific focus on *Las relaciones geográficas de la Nueva España* (Geographical reports of New Spain, 1577–85). These are a collection of 168 reports (about 2.8 million words) containing the responses of local authorities and inhabitants in the four Spanish territories in America—New Spain, New Granada, Peru, and Rio de la Plata—to surveys sent by the Spanish Crown, which are the object of study of the project. The reports gather a wealth of information about

1. Tanya E. Clement, “Where Is Methodology in Digital Humanities?,” in *Debates in Digital Humanities 2016*, ed. Matthew K. Gold and Lauren F. Klein (Minneapolis: University of Minnesota Press, 2016), 153–75, 158, [dx.doi.org/10.5749/j.ctt1cn6thb.17](https://doi.org/10.5749/j.ctt1cn6thb.17).

nature, moral and political life, military organization, and religious beliefs of the different villages in those regions using a mixture of ancient Spanish and Indigenous languages, mainly Nahuatl, as well as seventy-eight images, mainly maps of municipalities, in which it is possible to identify names of towns, authorities, and other particularities of the places. Given its breadth and characteristics, *Relaciones geográficas* is now considered one of the most important sources for geographical information and daily life in New Spain in the last quarter of the sixteenth century.

So far, DECM has focused on the integration and application of a methodology called geographical text analysis that integrates various techniques such as natural language processing (NLP), corpus linguistics (CL), machine learning (ML), and geographic information systems (GIS) to identify and search for georeferenced data from *Las relaciones geográficas de la Nueva España* and a group of secondary texts on geographic information about New Spain. As explained below, geographical text analysis (GTA) is built through the integration of theories, methods, and techniques that incorporate, on the one hand, prior knowledge of the objects to which it will be applied—in this case historical documents with geographical information. It also incorporates the relevant secondary sources on *Relaciones geográfica* and, more generally, on the geography of New Spain—these were mainly used for the disambiguation of proper names; on the other hand, GTA relies on knowledge of the scope, limitations, and biases of the tools used, so that the proposed methodology for developing a remote reading of documents is consistent with the linguistic, historical, and geographical theories used to interpret the results.

The GTA method prepared various corpora of *Relaciones geográficas*: a machine ready corpus for linguistic analyses and visualizations, and a gold standard corpus annotated by project researchers using the Tagtog tool. For the latter, experiments with NLP and ML techniques were developed to identify and annotate the names of an ontological model, created by the research team, that functions as a dictionary that provides the exact definition of twenty-one entities and labels appropriate for the analysis of complex concepts in the corpus.

The next step was to relate the names of places obtained from the corpus to their geographical coordinates, for which the team developed, as one of its most important tasks, a digital sixteenth-century gazetteer of New Spain. This is a particularly valuable tool for anyone interested in the study of colonial America because it contains information about places and their physical

locations. Another step of the process was the collocation analysis. Taken from corpus linguistics theory, collocation consists of identifying words that appear together or in proximity to each other more often than they appear alone. The novelty is that geographical text analysis extends this notion to a geographic collocation analysis so that, using the developed ontology, any word in the corpus can be geolocated and put into a spatial context.

The last step was mapping the geolocation and geographic collocation results using GIS to visualize the data. It should be clarified that, so far, the project has worked with a corpus formed by modern editions of *Relaciones geográficas* because the current ML and NLP methods are not applicable to old Spanish and Nahuatl, so techniques for working with these linguistic variants will be confronted later. For those interested in the study of the history of New Spain, it is easy to identify the most significant contributions of the project so far, starting with the Digital Sixteenth Century Gazetteer of New Spain whose development is documented in greater detail on this site (storymaps.arcgis.com/stories/9c6efb33ef2b4afdab3c9c6865dbb4cc) and which can be accessed on GitHub by following this link: github.com/patymurrieta/Digging-into-Early-Colonial-Mexico. The same goes for the ontological model from which it was built, and which can be found in the same place. Both tools are of great value for upcoming projects because their availability will allow other researchers to use them and improve other tools and methods. All project production is open access with Creative Commons Attribution-ShareAlike licenses.

In addition, in “The Corpus and the Data” section of the website, DECM has integrated a map (lancaster.ac.uk/digging-ecm/es/el-corpus) with some of the geographical sites already identified by the project. The project has also prepared a “Pathways to Understanding Sixteenth-Century Mesoamerica” site (lancaster.ac.uk/digging-ecm/portfolio-items/pathways-to-understanding-16th-century-mesoamerica) with three stories: “Tracing Toponymy,” “A History of Mexico,” and “Depicting Geographies” where they present different approaches to the problem of the origin of places names in ancient Mexico, the description of the trip for the founding of Tenochtitlan, and how to represent geographical spaces in the pre-Hispanic documents, which have a wide general informational value.

The project, now in its second phase, consists of two case studies in which the developed methodology will be used. The initial case study seeks to establish, using the results of the project’s first phase, the meaning of certain geographical

terms such as *cabeceras* (county seats), *sujetos* (subjects), and *villas* (towns) in the documents, and to shed light on changes in settlements between the late postclassic era and the end of the sixteenth century. The second seeks to clarify the representation and political structures between 1577 and 1585.

DECM offers and will continue to deliver results of great value for the digital humanities, history, and archaeology, especially those areas focused on the study of the territories of New Spain. The GTA and other tools are not openly available, but it is expected that they will be released soon as the project is committed to open access. This project, along with many others, shows that digital humanities involves not only the application of digital tools to a given case, but the design and adaptation of digital tools and methodologies to the chosen theoretical framework.

ERNESTO PRIANI SAISO

Universidad Nacional Autónoma de México

<https://doi.org/10.33137/rr.v44i3.38001>

Jenstad, Janelle, project dir.

Map of Early Modern London. Other.

Victoria, BC: University of Victoria, 2021. Version 6.6. Accessed 15 August 2021.

mapoflondon.uvic.ca.

The Map of Early Modern London (MoEML) is a major, web-based project exploring information about the city of London from 1550 to 1650. The version of MoEML reviewed here (6.6, 2021) represents more than twenty years of scholarship, and the contributions of more than five hundred people. In reality, MoEML is not one project but rather seven interoperable projects, supported by six different databases. Put very simply, MoEML takes complex, historical urban data about places, people, and organizations mentioned in early modern primary source material and links that data together so that users can explore their interrelationships by moving from page to page on the project site. For studying London's built environment and human geography as documented in the early modern English corpus, MoEML has become the definitive scholarly starting point, and will remain so for some time to come.