



The Index of Digital Humanities Conferences

DATA PAPER

MATTHEW D. LINCOLN 

SCOTT B. WEINGART

NICKOAL EICHMANN-KALWARA 

**Author affiliations can be found in the back matter of this article*

]u[ubiquity press

ABSTRACT

The Index of Digital Humanities Conferences is a database of conference metadata relevant to the digital humanities community dating back to the 1960s. It includes details about hundreds of conferences, and for a subset of them, details about work presented at those events. Weingart, Eichmann-Kalwara, Lincoln, and research assistants collected information by hand or with machine assistance from paper programs, listserv announcements, published proceedings, PDF programs, and conference websites. Data are in CSV tables with a data dictionary describing how they can be linked together, for use by digital humanists and bibliometricians to study the history of the digital humanities community.

CORRESPONDING AUTHOR:

Scott B. Weingart

University Libraries, Carnegie Mellon University, Pittsburgh, United States

weingart.scott@nd.edu

KEYWORDS:

digital humanities; humanities computing; bibliometrics; conference metadata

TO CITE THIS ARTICLE:

Lincoln, M. D., Weingart, S. B., & Eichmann-Kalwara, N. (2021). The Index of Digital Humanities Conferences. *Journal of Open Humanities Data*, 7: 12, pp. 1–5. DOI: <https://doi.org/10.5334/johd.26>

OVERVIEW

REPOSITORY LOCATION

KiltHub (<http://doi.org/10.1184/R1/12987959.v1>)

Live Website (<https://doi.org/10.34666/k1de-j489>)

CONTEXT

These data were created for *The Index of Digital Humanities Conferences*. It is a database of conference metadata relevant to the digital humanities community dating back to the 1960s, published online with a searchable interface. The website includes details about hundreds of conferences, and for a subset of them, details about work presented at those events, including authors, keywords, and so on. Weingart, Eichmann-Kalwara, Lincoln, and research assistants collected information by hand or with machine assistance from paper programs, listserv announcements, published proceedings, PDF programs, and conference websites.

Although the *Index* offers a historical record, it is an ahistorical initiative. It began with those individuals self-identifying with the phrase “digital humanities” in the year 2020 and looked backward towards those communities’ antecedents. The project’s intended scope, then, is to trace the communities which current digital humanities practitioners consider their antecedents. From that wide pool, the *Index* focuses specifically on conferences, symposia, and other scholarly events hosted by the subsets of those communities where technology or computation and the humanities intersect.

Conferences are sites of identity-building in scholarly communities. As opposed to other such sites, like academic departments or journals, few institutions exist to ensure their records survive and are centrally accessible. This project aims to fill that gap for the digital humanities.

METHOD

The *Index* is added to using a three-step process: first collect and enter metadata on relevant digital humanities and proto-digital humanities conferences, then find and enter metadata from the programs of those conferences, and finally reconcile entities such as people, institutions, and topics. Regrettably, it is impossible to reconstruct a list of all sources used to build the dataset, however a description of the history and process appears here. When possible, primary sources are recorded in the *references* field of the *conferences.csv* file.

STEPS

Weingart began collecting Alliance of Digital Humanities Organizations (ADHO) conference metadata in an Excel spreadsheet in 2012 and was joined by Eichmann-Kalwara in 2013. In 2018, Lincoln began developing a data model, and a Django-powered web application to enter and reconcile conference data, and eventually make it publicly browsable and searchable online.

In late 2019, a *crowdsourced Google Sheet* was used to generate a list of additional digital humanities and proto-digital humanities conferences dating back to 1960 from which to draw data. Using that sheet as a base, Weingart and Eichmann-Kalwara researched additional conferences to include in the *Index* and enter into the Django interface. As part of that process, Weingart and Eichmann-Kalwara collected as many paper and digital records for each conference as possible, to use as a source for conference work metadata.

After each conference record was entered, Weingart, Eichmann-Kalwara, and research assistants entered data on an incomplete subset of those conferences for which program information was available. Most records prior to 2000 were hand-entered from printed conference programs or text-formatted listserv announcements. For most conferences after 2000, data were pulled by hand from PDFs, websites, and other forms of digital conference programs. In the case of ADHO conferences specifically, for years where work-level TEI/XML data were available, work metadata was automatically extracted and imported into the database.

Finally, using both hand-checking and string similarity matching in OpenRefine, Weingart reconciled unique entities in the database. For example, in situations where one author was

listed with multiple names, he identified them as referring to the same person. In rarer situations when two authors with identical names were mistakenly linked to the same database entry, he attempted to correctly split them apart. The data model preserves the exact names for authors as given in the original source materials. However, spelling normalization was performed as part of the reconciliation process for locations, institutions, topics/keywords, and languages.

DATASET DESCRIPTION

OBJECT NAMES

dh_conferences_tables.zip, dh_conferences_works.zip, readme.txt

FORMAT NAMES AND VERSIONS

Within zip files are UTF-8 formatted CSV files. Readme.txt offers a detailed description of each file.

CREATION DATES

2012-01-01 to 2020-09-22

DATASET CREATORS

Scott B. Weingart (University of Notre Dame) – Project lead, data entry.

Nickoal Eichmann-Kalwara (University of Colorado, Boulder) – Project lead, data entry.

Matthew D. Lincoln (Carnegie Mellon University) – Technical lead, data entry.

Camille Chidsey (Carnegie Mellon University) – Data entry.

Aastha Jhunjhunwala (Carnegie Mellon University) – Data entry.

Gloria Kwakye (Carnegie Mellon University) – Data entry.

Harrison Lee (Carnegie Mellon University) – Data entry.

Steffi Nazareth (Carnegie Mellon University) – Data entry.

LANGUAGE

English

LICENSE

CC-BY-4.0

REPOSITORY NAME

KiltHub

PUBLICATION DATE

2020-09-22

REUSE POTENTIAL

The Index of Digital Humanities Conferences dataset will likely be reused for further analysis, reference, and pedagogy.

Very few open, centralized sources exist for the history of digital humanities, and so far as the authors are aware, none beyond the *Index of Digital Humanities Conferences* are available as structured data. This resource, then, can be of profound use to those quantitatively or qualitatively analyzing the history of the community. Because so much care was put into reconciling entities, diachronic or cross-geographic trends might be seen in authorship, geography, institution, or topic. Additionally, work titles may prove useful in text analysis.

Beyond its use as an analyzable dataset, the information within the *Index* will prove useful as a bibliographic reference for those seeking more information on authors, works, conferences, or topics. These records will be of particular interest to those studying digital humanities-adjacent scholarly organizations, as the *Index* contains lists of affiliated conferences and references to published proceedings or program URLs.

The *Index* has additionally found use in digital humanities classrooms. Instructors have been known to send their students to the *Index* to learn more about the sort of topics that appear at digital humanities events, both recently and in its earliest years.

While the *Index* offers a valuable resource for the history of digital humanities, conferences themselves offer insight into only a small corner of a scholarly community. This project, and the database structure it uses, enforces a narrow view of digital humanities. Like any archive, the structure of the *Index* preferences certain types of scholars and work, often reinforcing oppressive or hegemonic perspectives. Additionally, the data on both conferences as well as individual works are necessarily biased in scope and incomplete in execution. The ingest of new conference metadata as well as metadata on individual abstracts may continue to be released in periodic future versions hosted in the same repository. The *Index*'s English-speaking creators drew primarily from anglophone conference programs, and its initial focus on ADHO tends to favor a particular subset of digital humanists that ought not be taken as representative of the whole community. The indexing of individual abstracts for conferences is also biased, both by the availability of original sources as well as the priorities of its creators: conferences hosted by ADHO, the Association of Computing in the Humanities, and the Association for Literary and Linguistic Computing, were prioritized in particular for entry of individual abstracts. This database, then, must be used as one narrow, incomplete avenue of historical exploration.

ACKNOWLEDGEMENTS

Thanks also must extend to the countless program committee members who organized digital humanities and proto-digital humanities conferences throughout the years and published their proceedings in any form.

COMPETING INTERESTS

The authors have no competing interests to declare.

AUTHOR CONTRIBUTIONS

The *Index* relies on the work of many contributors.

John Walsh (Indiana University) and John Unsworth (University of Virginia) began creating an infrastructure to collect ADHO conference abstracts as XML around 2010. Sara Schmidt (University of Illinois) gathered and encoded these abstracts, with language assistance from Thomas Dousa, Maria Esteva, Kevin S. Hawkins, and Maki Miyake. Their efforts were funded by the Alliance of Digital Humanities Organizations, the Association for Computers and the Humanities, and the University of Illinois. Later work on this project was undertaken by Richard Higgins (Indiana University). The site code and XML data is available in an ADHO GitHub repository (<https://github.com/ADHO/dh-abstracts>).

Additional efforts to assemble conference metadata and abstracts were undertaken several years later by ADHO infrastructure committee chair Christoph Schöch. Schöch assembled DH2013 and DH2015-2018 on ADHO's GitHub page. Several of the team's paper copies of ADHO (and precursor) conference programs were made available from the extensive personal collections of Joseph Rudman (Carnegie Mellon University).

The vast array of conferences featured on this site were collected on a collaborative spreadsheet, with contributors including Alex Gil, Alexander Huber, Amanda Visconti, Anna-Sophia Zingarelli-Sweet, Anne Ladyem McDivitt, Brian Rosenblum, Carlos A. Martinez, Celeste Sharpe, Christoph Kudella, Colette Leung, Daniel Evans, Edward Vanhoutte, Geert Kessels, Heidi Tebbe, Hoyeol Kim, Jason B. Jones, Joris van Zundert, Kristen Mapes, Laura Braunstein, Laurel Stvan, Matt Lincoln, May Ning, Michael Hendry, Michael Piotrowski, Miguel Escobar Varela, Nadezhda Povroznik, Pim

AUTHOR INFORMATION

For determining author roles, please use following taxonomy: <https://casrai.org/credit/>.

Scott B. Weingart: Conceptualization, data curation, investigation, methodology, project administration, supervision, writing – original draft.

Nickoal Eichmann-Kalwara: Conceptualization, data curation, investigation, methodology, project administration, writing – review & editing.

Matthew D. Lincoln: Conceptualization, data curation, methodology, software, writing – review & editing.

AUTHOR AFFILIATIONS

Matthew D. Lincoln  orcid.org/0000-0002-4387-3384
University Libraries, Carnegie Mellon University, Pittsburgh, United States

Scott B. Weingart
Hesburgh Library, University of Notre Dame, Notre Dame, United States

Nickoal Eichmann-Kalwara  orcid.org/0000-0002-8851-852X
University Libraries, University of Colorado, Boulder, Boulder, United States

TO CITE THIS ARTICLE:

Lincoln, M. D., Weingart, S. B., & Eichmann-Kalwara, N. (2021). The Index of Digital Humanities Conferences. *Journal of Open Humanities Data*, 7: 12, pp. 1–5. DOI: <https://doi.org/10.5334/johd.26>

Published: 09 July 2021

COPYRIGHT:

© 2021 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

Journal of Open Humanities Data is a peer-reviewed open access journal published by Ubiquity Press.