INTEGRATIVE, INTERDISCIPLINARY DATABASE DESIGN FOR THE SPATIAL HUMANITIES: THE CASE OF THE HOLOCAUST GHETTOS PROJECT

ANNE KELLY KNOWLES AND JUSTUS HILLEBRAND WITH PAUL B. JASKOT AND ANIKA WALKE

Abstract *Databases are central to the digital, spatial, and geohumanities. There is surprisingly little scholarly literature, however, on the process of database construction in humanities projects. This article describes a process of interdisciplinary database design that emerged in the course of building the core sections of an historical GIS of Holocaust ghettos. The process foregrounds collaborative design, testing that purposely flushes out paradigmatic differences and ontological problems, and revision to incorporate group decisions and agreed-upon meanings into data structures, field definitions, and instructions for data entry. The result is a deeply integrative form of mixed methodology that incorporates ethical standards along with data entry instructions and team training.*

Keywords: Integrative database design, HGIS, Holocaust ghettos

Databases are central to the digital, spatial, and geohumanities. Many scholars working in these fields have acknowledged the importance of database design because it directly influences research results.[1] Databases and the modes of inquiry and representation that they enable also carry philosophical and ethical implications that pose problems for humanistic scholars.[2] Historian David J. Bodenhamer, a leading spokesman for the spatial humanities and historical GIS (geographic information systems), wrote recently that 'the humanities pose epistemological and ontological issues that challenge [GIS] technology in a

International Journal of Humanities and Arts Computing 14.1-2 (2020): 64–80
DOI: 10.3366/ijhac.2020.0245

www.euppublishing.com/ijhac

number of ways, from the imprecision and uncertainty of humanities data to humanists' reliance on time (and time linked to space) as an organizing principle'.[3] Historians in particular traditionally use nuanced narrative to weave together the strands of surviving evidence about a particular event, period, or group of historical characters. Computational methods, in contrast, are predicated on expectations of precision, and they usually require a degree of completeness or comparability that historical sources often lack.[4] In addition, data structures impose "mutually exclusive categories" that simplify and break down sources whose complexity historians ordinarily seek to preserve.[5]

Various mixed-methods approaches have been developed to mitigate these problems. In qualitative GIS, for example, researchers augment conventional spatial databases with embedded images, video and audio clips, and other media. This can enrich and humanize mapped places that otherwise might be visualized only as points, lines, and areas.[6] Practitioners of deep mapping layer many kinds of data in a shared digital space to enable viewers to read, or digitally select, the particular information, perspective, or narrative stream that interests them.[7] While both of these approaches incorporate more qualitative data and facilitate modes of representation beyond two-dimensional cartography, they do not directly address the epistemological and ontological differences that continue to divide history and geography as academic disciplines.[8]

Finding ways to resolve those differences has been central to the developmental stage of our research on Holocaust ghettos in German-occupied Eastern Europe. The Holocaust Ghettos Project has three parts. The first is an HGIS of ghettos based on more than 1,300 places documented in the multi-volume *Encyclopedia of Camps and Ghettos* in progress at the United States Holocaust Memorial Museum (USHMM).[9] The second part will be an analysis of the spatial content of 1,800 transcripts of post-war interviews with Holocaust survivors. The third will be a multilingual gazetteer of ghettos, which will link evidence from the HGIS and interview transcripts through the place names they share. The project's over-arching goal is to map the events and conditions that shaped ghettos in relation to individual experiences. We hypothesize that mapping will reveal previously undetected spatial and temporal differences and dynamics within and between regions from late 1939 to early 1945, the period when Nazi policies targeted Jews for murder and then genocide. We hope that the project will add geographical insights to the burgeoning scholarship on Holocaust ghettos.[10]

In this article, we explain the process of interdisciplinary database design that emerged in the course of building the core sections of the Ghettos HGIS. This process foregrounds collaborative design, testing that purposely flushes out paradigmatic differences and ontological problems, and revision to incorporate group decisions and agreed-upon meanings into data structures, field definitions, and instructions for data entry. The result is what mixed methods scholar Pat

Bazeley calls a deeply integrative form of mixed methodology. Our model of database construction is not only based on equal consideration of differing disciplinary priorities and mutual understanding. It also results in data structures designed to hold historically contextualized qualitative and quantitative data, complementary to recent work by Rawson and Muñoz.[11] Hence our inclusion of Bazeley's key term in calling the Ghettos HGIS an *integrative*, interdisciplinary database design.[12]

EXPERIMENTATION IN DATABASE DESIGN

Constructing a set of related tables to capture the wealth of semi-structured information contained in the USHMM Encyclopedia seemed at first a fairly straight-forward enterprise.[13] Our initial assumptions about how to structure historical data came from Knowles's and Jaskot's previous HGIS projects[14] and from other projects that mapped historical places to study individuals' movements to and perceptions of certain places.[15] For the historians and historical geographers in the Ghettos Project team who are trained in GIS, parsing the information in Encyclopedia entries into specified field types was a natural and necessary step to facilitate mapping the information. Some categories raised red flags, however, particularly for the non-GIS practitioners on the team, because the entries contain complex, sometimes contradictory descriptions of events. For example, many entries mention the crucial moment of German occupation but provide no specific date. How could we assign a date in the absence of certain information? The countervailing concern was that leaving German occupation blank where information was vague or uncertain could result in maps that said little, if anything, about regions in Soviet territory where Jewish communities were swiftly overwhelmed by German forces in 1941 and 1942. Some historians on the team were also uncomfortable with the standardization of evidence, such as the ten categories for kinds of labor that Knowles and Jaskot had used in their earlier study of SS concentration and labor camps.[16] Why not include all possible kinds of labor, to be more faithful to the source and individuals' experiences? A tension emerged between the need for expeditious data entry into mappable categories and the desire to capture historically meaningful detail as fully and accurately as possible.

Two ethical issues underlay these and other questions. One was the common Digital Humanities concern that standardized data structures and field definitions asserted certainty where the available source information was uncertain, ambiguous, or contradictory. The other issue stemmed from the widely shared conviction among Holocaust scholars that no mode of representation should do further violence to the victims of genocide.[17] While any careful scholar notes uncertainty and the absence of information in their sources, doing so has a specific ethical dimension in Holocaust studies, where gaps in documentation are often the result of the destruction brought on by war and mass murder. It was mandatory that we find ways to capture as much information as possible while representing it with as little distortion as possible.

Hillebrand's first conceptual design for a ghettos database (see Figure 1) raised other issues for the team. Adhering to the principles of efficient relational design required that Encyclopedia entries be diced into such small pieces that the information became difficult for Knowles, Jaskot, and Walke to read and comprehend. Had we developed this approach further, we also foresaw growing dependency on Hillebrand's technical expertise. This would have been an unacceptable bottleneck and put too much pressure on a younger scholar who had limited hours to give the project. We also tested crowd-sourcing in hopes that it might speed data entry. A trial extracting data on ghetto labor showed that, even for that fairly simple category, data entry was subject to interpretation. For example, team members variously interpreted 'digging up *matzevot* (Jewish tombstones) for paving roads'[18] in one ghetto town as construction, producing building materials, and manual labor. This trial convinced us that all data entry would have to be proofread. We also began to write rules and detailed field definitions to guide data entry. We decided to augment the categories of labor used in the earlier SS camps study, so that we would be able to compare forced labor in camps and ghettos. However, we added many more kinds of labor to the categorical definitions to account for ghetto circumstances. Labor categories, like other lengthy lists of allowed options, would be available as a multi-value field.[19]

As the project team worked to resolve these issues, we developed a routine of weekly meetings by video conference (Hillebrand was in Germany, Jaskot in North Carolina, Knowles in Maine, and Walke in Missouri). Our discussions increasingly focused on the meaning of key terms within our research questions, which in turn led to the definition and redefinition of database fields. For example, recording the movement of people into and out of ghettos began as fields for Transport_in and Transport_out. These terms were rooted in an understanding of the Holocaust that focused on deportations to extermination camps. They fit forced relocations by train but did not seem appropriate for the many instances where people were forced to walk long distances, or the partially self-directed movement of refugees who fled German aggression. The general movement categories became Forced_move_in, Forced_move_out (for involuntary movement directed by non-Jews) and Refugees (where people fled from one location to another). To aid proofing of the trickiest topics we added fields for quotations from the source and notes. We also created certainty fields to enable us to map the degree of certainty of these and other key variables.

The process was slow. The final database design and rules of data entry for the basic attributes table took two years to produce, the second table about eight months (see Supplementary Material: Appendix A and B). The results,

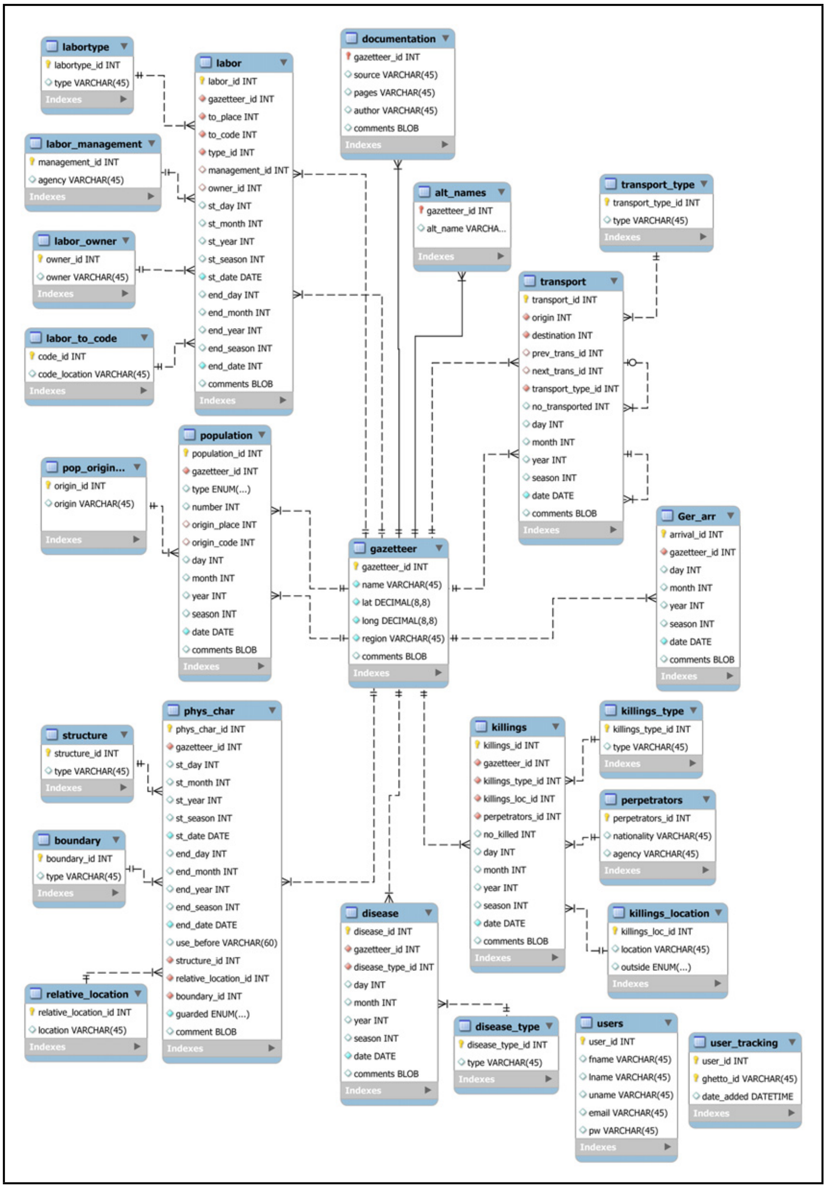

Figure 1. Early relational database design for the Holocaust Ghettos database. This design proved unwieldy and was inadequate for capturing the wealth of information contained in the project's encyclopedic source. Graphic by Justus Hillebrand.

however, were positive in ways we had not anticipated. The rules of data entry anchor the Ghettos Project in the team's expert knowledge of Holocaust historiography and our research questions. The slowness of the process allowed time for issues to surface. Our least technically expert team members gained confidence to speak up when they encountered difficulties, and their issues often flagged problems that the team agreed required further consideration. The rules provide discursive, non-technical metadata, a statement of our agreed standards, a detailed set of instructions for new research assistants, and a valuable reference for all project participants to help ensure the quality and consistency of data entry and proofing. Perhaps most importantly, the process steadily improved the quality and consistency of data entry while cultivating common understanding and cohesion among our diverse team.

FIELDS FOR HUMANISTIC DATA

Testing data entry iteratively made us intimately familiar with the content and interpretive issues in Encyclopedia entries. This part of the process adapts and extends the well-established GIS practice of iterative data exploration[20] by applying iteration to the early stages of project conceptualization and database design. Developing field definitions and deciding what kinds of fields would best accommodate the data proved far more important than the choice of software and ensured some degree of software-agnostic data modeling.[21] Some data fields changed little, others were eliminated or repeatedly reconceived. Each of the seven kinds of fields that the first two tables contain – date, diagnostic, multivalue, numerical, certainty, quotes, and notes – went through a developmental process shaped by the source material, research questions, and extensive discussion.

(1) Date fields

In almost no Encyclopedia entry is every kind of date complete and precise. Dates can be expressed to the day (June 26, 1941) or the month (in September 1940) or the year (in 1944), or in vague expressions such as 'late that spring' or 'before 1942' or 'within a few weeks'. Temporal ambiguity and uncertainty are chronic problems in HGIS, particularly when one is trying to construct time series from disparate or inconsistent sources. One common solution is to enter available date information into three fields (month/day/year), so that analysis can be done at whatever temporal resolution includes sufficient data to answer a given question. Alternatively, one can construct artificially precise, complete dates from contextual information and rules based on certain assumptions. Knowles and Jaskot used both of these approaches in the SS camps study, with reference to a table that converted common verbal expressions into dates, such as April 15 for 'spring' (see 'Instructions for Date Fields' in Appendix A).

Typical approaches were inadequate for dating German occupation and the establishment and dissolution of ghettos. We knew from the SS camps study that constructed dates would give misleading emphasis to certain moments in map animations, when, for example, a burst of camp openings in 'spring' appeared on April 15 of a certain year. The bigger issue was that any dates we assigned to ambiguous times could skew temporal analysis in ways that had no basis in historical understanding. Generally, we know that ghettoization followed soon after German occupation in much of Poland and in the Baltic countries, but the process is less well understood and less well documented in regions further south and east. No one rule for assigning dates would be adequate for all regions. A better solution came from Collaborative member Erik B. Steiner, who had used date spans in other projects with complex temporal data. This made good sense, because history is commonly studied in terms of time periods. Framing key events as happening between the earliest and latest possible dates enabled us to relate localized events to historical factors such as regional policy decisions and advancing military fronts. Date spans could also record precise dates (the earliest and latest possible date would be identical), or where occupation happened within 'a few days', 'a week', or 'a month' of a known date.

In addition, the logic of military forces moving across the landscape suggested a method for estimating unknown dates of German occupation. We came up with a temporal adaptation of Tobler's Law, named for its author, Walter Tobler, a pioneering geographer in the development of GIS. Tobler's Law states in part that things that are closer together tend to be more related than things that are farther apart.[22] Extrapolating from that observation, we reasoned that if a ghetto had no known date of German occupation, but other nearby ghettos did, we could estimate a probable date for the unknown instance based on its neighbors (see 'Date of German Occupation' in Appendix A). The extra time and effort this requires is warranted by the importance of having some notion of the date of German occupation for as many ghettos as possible.

Temporal resolution, we came to realize, is not just the unit of time for a given variable. It was something we were actively constructing through database design. In GIS, resolution refers to the unit of spatial data, such as the area of the earth's surface represented by one pixel in a satellite image, or census tracts in demographic data. Geographers are wont to say that analysis can be no more fine-grained than the resolution of the data. By this understanding, spatial resolution is a static, empirical trait that is inherent in data. By entering dates as ranges of event possibility, however, we were actively constructing temporal resolution. When it comes time to map or otherwise analyze these data, we will again construct temporal resolution by choosing from the various options that the date spans make possible.

(2) Diagnostic fields

These typically capture the answers to yes-or-no questions. The name refers to the idea that getting a quick initial sense of the geographic distribution of a variable (a first diagnosis) will help us decide later how to design a table that focuses on themes related to that variable. This simple structure also helps with proofing, because a null value in a diagnostic field means it was overlooked during data entry. The first diagnostic field pertained to whether a Judenrat or other Jewish representative group or individual was put in place during occupation by German authorities or a German ally. To be clear about the certainty of the source information, and to highlight where further research might be necessary, we gave this field a drop-down menu with four values from which to choose: yes, no, uncertain, and unknown. A second diagnostic field tallies the first estimated total of mass murder events, which we defined as a single event in which ten or more people were killed.[23] Although we wrote fairly extensive rules about how to count mass murders (see Appendix A), we expect that the rules, and our understanding of these awful events, will become more refined when we develop a table devoted to the topic. Hence our treatment of the first tally as an estimate, entered for purposes of preliminary visualizations of the relative intensity of killing during ghetto formation and dissolution.

(3) Multivalue fields

The second major table in the ghettos database captures material conditions (see Appendix B). Because the geographical variation of conditions is central to our research, we needed to find ways to capture significant details about matters such as how ghettos were enclosed and what kinds of restrictions were imposed on Jews' everyday activities. We decided on multivalue fields through a process of elimination. The first option we considered was to repeat the approach used in the SS camps study, of entering only what the data enterer deemed to be the most important or primary value into a single-value field. Applying this to ghetto enclosure might yield the values shown in Figure 2. Again, this approach was predicated on a pre-defined list of allowed terms. While the values are correct in this hypothetical case, they provide an incomplete impression of the multiple means used to confine Jewish residents in hypothetical ghettos 1, and 2, and 3.

The second option we considered involved adding as many single-value fields as necessary to capture all variations (Figure 3). This approach would allow full capture of descriptive terms, and could assign a hierarchy of significance or suggest temporal sequence. But the number of fields required for some complex variables (such as restrictions, which includes 31 terms), would make the data entry table or form unwieldy. The majority of fields for the variable would be empty, producing many null values that did not necessarily indicate the absence

ID	Name	Enclosure
1	Ghetto 1	wall
2	Ghetto 2	barbed wire
3	Ghetto 3	barbed wire

Figure 2. Example of a single, or primary, descriptor entered for each ghetto's mode of enclosure. Graphic by Justus Hillebrand.

ID	Name	Enclosure1	Enclosure2	Enclosure3	EnclosureX
1	Ghetto 1	wall	barbed wire	guard(s)	
2	Ghetto 2	barbed wire	sign(s)		
3	Ghetto 3	barbed wire			

Figure 3. In this approach, every possible descriptor can be entered for each ghetto's various enclosures, provided enough columns are added to accommodate every type. Graphic by Justus Hillebrand.

of information. Queries would have to include many fields, also a cumbersome process.

A third option was to return to classic relational database design, using a unique identifier for each term and linking back to the uniquely identified ghetto in each case (Figure 4). This method reduces redundancy and null values and it allows entering an unlimited number of values per object (here, per ghetto). It would also allow linking further, co-dependent values, such as how long each enclosure type existed. We have already noted the issue of the more specialized training required to implement and manage this approach, and the problem that the connections between bits of information only become visible to the user after they are assembled by command. This is also true of using related tables in ArcGIS, our chosen GIS software, and querying across related tables to narrow down results. For us, the decisive problem was that assembled tables would hold multiple rows for each ghetto, which in ArcGIS would be visualized as multiple points overlying one another wherever ghetto enclosure had multiple values. This was unacceptable.

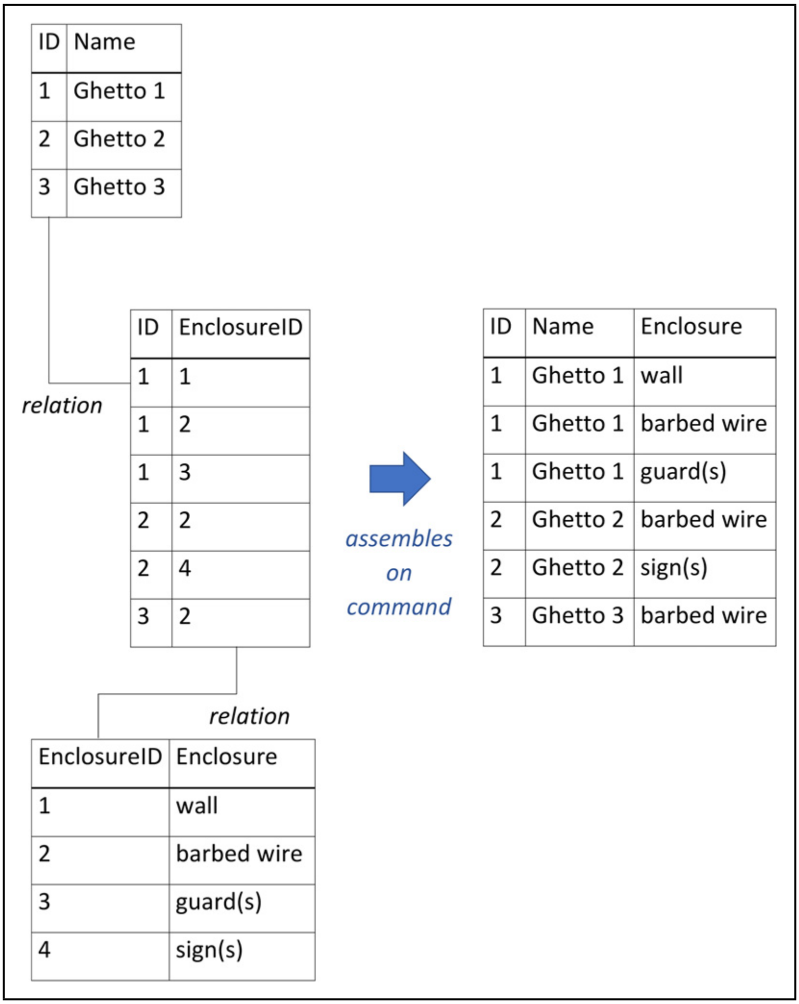

Figure 4. In a classic relational design, each ghetto would be linked to each of its uniquely identified enclosures, which in turn would be linked to the description of each enclosure type. Graphic by Justus Hillebrand.

Multivalue fields that contain lengthy drop-down menus of defined vocabularies for these variables enable the data enterer to select however many terms apply to each ghetto (Figure 5). This kind of field is not entirely compatible with GIS, because querying a multivalue field as it stands could return a bewildering number of qualifying returns, depending on how many

ID	Name	Enclosure
1	Ghetto 1	wall, barbed wire, guard(s)
2	Ghetto 2	barbed wire, sign(s)
3	Ghetto 3	barbed wire

Figure 5. Multivalue fields enable one to record all relevant values in a single field. Graphic by Justus Hillebrand.

combinations of terms were selected. So-called wildcard queries could resolve this difficulty.[24]

(4) Numerical fields

The simplest numerical field was Jewish population as stated in a pre-war census. For settlements where a census was conducted, Jewish population is usually given as an integer. Ghetto population is a much more fraught variable. Historical information is scarce for many ghettos, particularly in regions that German forces claimed during the invasion of the Soviet Union. People of all identities in Belarus, Ukraine, and other nations in this zone suffered repeated invasion, arrests, atrocities, bombings, fires, displacement, and the destruction of governments and records in the 1930s and early 1940s.[25] In areas that the Germans quickly controlled, such as Lithuania and Latvia, mass killings followed so swiftly on the heels of German invasion that some ghetto populations may never have been counted. In larger, longer-lasting urban ghettos, such as those in Warsaw and Łodz, Poland, Kovno, Lithuania, and Theresienstadt in the German-allied country of Czechoslovakia, the population rose and fell as Jews arrived as refugees or on transports, then large numbers were shot or deported to killing centers.[26] Epidemic diseases such as typhus and tuberculosis claimed many lives when ghettos became severely overcrowded and people had little access to food or medical treatment, especially in winter. Residents of larger ghettos also tended to be able to sustain some semblance of community life and religious practice for a time, and to mount resistance.[27] We eventually decided to record ghetto population in a series of fields (see Appendix B).

(5) Certainty fields

Historical information can be full of varying degrees of certainty or uncertainty. Admitting, during data entry, what we do or do not know for sure, based on our source, has become a core principle of our humanistic practice. In the ghettos database, 'certain' means that the Encyclopedia entry describes information as factual or verified, without qualification. 'Uncertain' means the entry provides some information on the point in question, but it is too ambiguous or contradictory to allow a confident decision; or the information is described as uncertain, probable, likely, perhaps true, and so forth. 'Unknown' means that the entry includes no information or says that none was found. Additional variants of certainty emerged from our discussions of particularly thorny issues. For ghetto population, we added 'approximate' for cases where an entry includes some information but not a definite number; for example, the ghetto had less than 2,000 Jews, or it held about 300 people. Imperfect as our numbers are, they will enable us to get a rough sense of the geography of ghetto population. As more research is done on prewar and wartime Jewish communities, thanks to the opening of archives in Eastern Europe and Russia since the fall of the Soviet Union, we hope that more population data will become available.

In regard to numerical, temporal, and spatial uncertainty, the challenges of applying computational methods to the Holocaust are no more extreme – perhaps less so – than the obstacles researchers have faced while trying to map, for example, the origins, identities, and life trajectories of enslaved Africans[28] or the locations of ancient sites in the Greek and Roman world.[29] Whereas the locations of the ghettos we are studying are relatively clear, there is considerable uncertainty with regard to other aspects of ghettoization. Noting the gaps not only supports scholarly standards for honesty and transparency; it also meets an ethical demand for noting absences caused by the Holocaust while flagging topics and locations that need further research.

(6) Quote fields

While the content of this kind of field is self-apparent, its importance in a humanistic database may not be. We include direct quotations from Encyclopedia entries for fields, or groups of closely related fields, for which data entry often requires considerable interpretation. The dates bounding German occupation and ghetto start and end dates are examples. The physical characteristics table has a quotation field for the text that says when non-Jewish authorities or actors first imposed restrictions on the local Jewish population, whether officially or in practice. While we were prototyping data entry, these topics proved particularly nuanced. Including the source language in the database gives the proofreader convenient access to the key textual excerpt. It also provides the team easy references in case the fields in question generate further debate.

(7) Note fields

Doing relatively complicated data entry made it helpful to add notes fields for a number of variables, so that those entering data could jot down issues and details that were either especially difficult or warranted further discussion. Note fields provide a familiar pause in the flow of data entry for students accustomed to humanities courses: Here you can write a little, a natural complement to reading. Along the way, the note fields collect observations that may become part of project metadata and raise questions for future research.

CONCLUSION

In this essay, we have pulled back the curtain that usually conceals decisions that go into database design because we regard those decisions as anything but mundane. They are consequential on many levels – for determining what questions can be answered, what data can be represented in visualizations and analysis, and how much of the content in source material is available for analysis because it has been included or shorn away. Discursive metadata provides context, guidance, and technical explanations for team members and others who may use the data in future research. Group engagement demands time and patience, but we found it to be an effective way to keep our focus on research questions and Holocaust historiography throughout the process of developing digital infrastructure.

Making the process of database development more explicit can benefit Digital Humanities scholarship more broadly. More complete articulation of database design in project planning, metadata, and publications helps explain the intellectual work involved in digital scholarship to non-practitioners, including senior scholars on promotion and tenure committees.[30] Frank, probing discussion of data structures fosters communities of practice and promotes development of data standards that reflects the values of humanistic scholarship. Our approach could also improve communication between humanists and collaborators in other fields, providing stronger foundations for interdisciplinary scholarship.

As in other digital projects that require manual data extraction, the combination of integrative, interdisciplinary database design and trial data entry was an intensive form of close reading.[31] One unexpected outcome from this part of the process was a dawning discontent with received definitions of the terms ‘ghetto’ and ‘ghettoization’ in Holocaust historiography. Holocaust ghettos have mainly been defined as places imposed on Jews by the Nazis (places where German authorities ‘concentrated the Jews’ by moving them to a designated area where ‘only Jews were permitted to live’[32]) or by how they differed from camps (unlike camps, ghettos were not ‘totally cut off from urban life’ and they retained some semblance of prewar community and family relations[33]).

While these definitions highlight important common elements among Holocaust ghettos, they omit the specifically geographical nature of ghettos as weaponized places. We define ghettoization as the imposition and enforcement of spatial forms of restriction that were intended to control the location and mobility of Jewish bodies. The exercise of control had three interlocking spatial aspects: the declaration of one or more holding places, restriction of movement, and means of enclosure. This conceptualization is closer to the granularity of action and experience that we have come to recognize through the very process of building the database. It provides a stronger basis for explaining how Nazi practices profoundly changed physical and social space for Jews and their non-Jewish neighbors. We will explore the evidence and implications of this view in a future essay.

Integrative, interdisciplinary database design recognizes the iterative, qualitative, quantitative, and analytical components of many HGIS projects. Together, these elements constitute a model of dynamic, deeply informed standardization that is quite different from normalization. Rather than paring away complexity and nuance, this process seeks ways to capture those qualities. Rather than excluding data that is uncertain, ambiguous, or unknown, we propose methods that acknowledge the grey dimensions of sources and render them accessible for analysis. The key is for humanists to apply their content expertise at every stage of database creation.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the many contributions of our research assistants and consultants on the Holocaust Ghettos Project: at University of Maine, Jennifer Cashin, Dakota Gramour, and Sarah Treadwell; at Duke University, Eve Duffy and Antonio LoPiano; and at Washington University at St. Louis, Nicci Mowszowski and Christopher St. Aubin. We also thank the very helpful comments of three anonymous readers, and suggestions from Tim Cole, Jürgen Matthäus, Daniel Patrick Miller, and Levi Westerveld. We bear full responsibility for the results. Funding provided by National Endowment for the Humanities Digital Humanities Advancement Grant HAA-261290.

END NOTES

[1] D. J. Bodenhamer, J. Corrigan, and T. M. Harris, eds, *The spatial humanities: GIS and the future of humanities scholarship* (Bloomington, IN, 2010); M. Dear, J. Ketchum, S. Luria, and D. Richardson, eds, *GeoHumanities: art, history, text at the edge of place* (London, 2011); A. Burdick, J. Drucker, P. Lunenfeld, T. Presner, and J. Schnapp, *Digital humanities* (Cambridge, MA, 2012); J. Flanders and F. Jannidis, eds, *The shape of data in the digital humanities: modeling texts and text-based resources* (London, 2019).

[2] T. Presner, 'The ethics of the algorithm: close and distant listening to the Shoah Foundation Visual History Archive', in C. Fogu, W. Kansteiner, and T. Presner, eds, *Probing the ethics of Holocaust culture* (Cambridge, MA, 2016), 175–202.

[3] D. J. Bodenhamer, 'Narrating space and place', in D. J. Bodenhamer, J. Corrigan, and T. M. Harris, eds, *Deep maps and spatial narratives* (Bloomington, IN, 2015), 7–27. Cited here on p. 10.

[4] A. K. Knowles, 'GIS and history' in A. K. Knowles, ed., *Placing history: how maps, spatial data, and GIS are changing historical scholarship*. (Redlands, CA, 2008), 1–26.

[5] D. J. Bodenhamer, 'History and GIS: implications for the discipline', in A. K. Knowles, ed., *Placing history: how maps, spatial data, and GIS are changing historical scholarship* (Redlands, CA, 2008), 220–33.

[6] M. Cope and S. Elwood, eds, *Qualitative GIS: a mixed methods approach* (London, 2009).

[7] T. Presner, D. Shepard, and Y. Kawano, *HyperCities: thick mapping in the digital humanities* (Cambridge, MA, 2014); M. Yuan, J. McIntosh, and G. DeLozier, 'GIS as a narrative generation platform', in D. J. Bodenhamer, J. Corrigan, and T. M. Harris, eds, *Deep Maps and Spatial Narratives* (Bloomington, IN, 2015), 179–202.

[8] A. R. H. Baker, *Geography and history: bridging the divide* (Cambridge, 2003).

[9] G. P. Megargee and M. Dean, eds, *The United States Holocaust Memorial Museum encyclopedia of camps and ghettos 1933–1945*, vol. II, *Ghettos in German-occupied Eastern Europe*. (Bloomington, IN, 2012); G. P. Megargee and J. R. White, eds, *The United States Holocaust Memorial Museum encyclopedia of camps and ghettos 1933–1945*, vol. III, *Camps and ghettos under European regimes aligned with Nazi Germany*. (Bloomington, IN, 2018).

[10] G. Miron, ed, with Shlomit Shulhami, co-ed, *The Yad Vashem encyclopedia of the ghettos during the Holocaust* (Jerusalem, 2009); D. Michman, *The emergence of Jewish ghettos during the Holocaust*. (Cambridge, 2011); W. Z. Goldman and J. W. Trotter, Jr., eds, *The ghetto in global history 1500 to the present* (London, 2018). On ghettos more generally, see M. Duneier, *Ghetto: the invention of a place, the history of an idea* (New York, NY, 2016) and D. B. Schwartz, *Ghetto: the history of a word*. (Cambridge, MA, 2019).

[11] K. Rawson and T. Muñoz, 'Against Cleaning', in M. Gold and L. Klein, eds, *Debates in the digital humanities* (Minneapolis, MN, 2019), chap. 23.

[12] P. Bazeley, *Integrating analyses in mixed methods research* (Los Angeles, 2017).

[13] While the length of entries ranges from a few short paragraphs to several large pages, they all contain what information researchers could find about prewar Jewish life, German occupation, establishment of the ghetto and related institutions, prejudice and violence against Jews, forced labor, Jewish population before and during ghettoization, mass murder, deportations, the final days of the ghetto, number of survivors, and post-war trials. Manuscript and published sources conclude each entry.

[14] A. K. Knowles, *Mastering iron: the struggle to modernize an American industry* (Chicago, IL, 2013); A. K. Knowles, T. Cole, and A. Giordano, eds, *Geographies of the Holocaust*. (Bloomington, IN, 2014); P. B. Jaskot and I. van der Graaff, 'Historical journals as digital sources: mapping architecture in Germany, 1914–24', *Journal of the Society of Architectural Historians,* 76, 4 (2017), 483–505.

[15] D. Cooper and I. N. Gregory, 'Mapping the English Lake District: a literary GIS', *Transactions of the Institute of British Geographers*, new ser. 36 (2011), 89–108; G. Ceserani, G. Caviglia, N. Coleman, T. De Armond, S. Murray, and M. Taylor-Poleskey, 'British travelers in eighteenth-century Italy: the grand tour and the profession of architecture', *American Historical Review,* 122, 2, (2017), 425–50.

[16] A. K. Knowles, P. B. Jaskot, and C. Harvey, "Mapping the SS Camps," in A. K. Knowles, T. Cole, and A. Giordano, eds, *Geographies of the Holocaust* (Bloomington, IN, 2014), 18–50.

[17] S. Friedländer, ed., *Probing the limits of representation: Nazism and the "final solution"* (Cambridge, MA, 1992); C. Fogu, W. Kansteiner, and T. Presner, eds, *Probing the ethics of Holocaust culture* (Cambridge, MA, 2016).

[18] J. Śliwa, 'Leżajsk', in G. P. Megargee and M. Dean, eds, *The United States Holocaust Memorial Museum encyclopedia of camps and ghettos 1933–1945*, vol. II, *ghettos in German-occupied Eastern Europe* (Bloomington, IN, 2012) 537–8. Cited here on 537.

[19] This decision influenced our final choice of software for the basic attributes and physical characteristics tables. Microsoft Access allows selection of multiple values in a single field. Data exported is then separated by commas, a format that is not ideal but can be managed in GIS queries.

[20] J. Dykes, A. M. MacEachren, and M. Kraak, *Exploring geovisualization* (Oxford, 2005).

[21] We began using the relational database systems MySQL and PostgreSQL with a custom-made PHP data entry form but soon found that this solution required unnecessarily advanced programming and SQL skills. We switched to Microsoft Access which provided all the basics of database design as well as easy design of data entry forms and multivalued fields.

[22] D. DiBiase, 'Spatial Dependency', in *The nature of geographic information: an open geospatial textbook* (State College, PA, n.d.), 1:17. Online at e-education.psu.edu/natureofgeoinfo/.

[23] The number of people who must be killed for an event to count as mass murder is a contentious issue. A planned table for mass murder will capture the range of murderous actions, from single shootings to round-ups of local leaders and much larger actions.

[24] In a wildcard query, bracketing search terms with particular symbols (for ArcGIS, a wildcard query would look like "enclosure LIKE '%wall%'" or "enclosure LIKE '%barbed wire%'") signals to the computer to ignore text on either side of the queried value. Thus one can search on one term, such as "barbed wire," and get results that include every ghetto where that term is included as a kind of enclosure.

[25] Timothy Snyder, *Bloodlands: Europe between Hitler and Stalin* (New York, 2012); J. Tingler, 'A vortex of violence: the Holocaust and mass atrocities in Chełm, Poland, 1939–1944,' paper delivered at Lessons and Legacies (St. Louis, MO, 2018).

[26] B. Engelking and J. Leociak, *The Warsaw Ghetto: a guide to the perished city*, trans. from the Polish by Emma Harris (New Haven, CT, 2009); Megargee and Dean, *Encyclopedia*, vol. II; Megargee and White, *Encyclopedia*, vol. III; W. Gruner, *The Holocaust in Bohemia and Moravia: Czech initiatives, German policies, Jewish responses*, trans. from the German by Alex Skinner (New York, NY, 2019).

[27] This is particularly well documented for the Warsaw ghetto, the largest in Eastern Europe and one of the longest-lived. See S. D. Kassow, *Who will write our history? Emanuel Ringelblum, the Warsaw ghetto, and the Oyneg Shabes Archive* (Bloomington, IN, 2007).

[28] H. B. Lovejoy, 'Mapping uncertainty: the collapse of Oyo and the trans-Atlantic slave trade, 1816–1836', *Journal of Global Slavery,* 4 (2019), 127–61.

[29] R. J. A. Talbert, ed., *Barrington atlas of the Greek and Roman world* (Princeton, NJ, 2000). See esp. map key following Map 1.

[30] American Historical Association, 'Guidelines for the professional evaluation of digital scholarship by historians', 2015, online at https://www.historians.org/teaching-and-learning/digital-history-resources/evaluation-of-digital-scholarship-in-history/guidelines-for-the-professional-evaluation-of-digital-scholarship-by-historians; College Art Association, 'Evaluating digital scholarship in art and architectural history', 2016, online at https://www.collegeart.org/pdf/evaluating-digital-scholarship-in-art-and-architectural-history.pdf.

[31] J. Jenstad, 'Building a gazetteer for early modern London, 1550–1650', in M. L. Berman, R. Mostern, and H. Southall, eds, *Placing names: enriching and integrating gazetteers* (Bloomington, IN, 2016) 129–45; A. K. Knowles, 'A more humane approach to digital scholarship', *Parameters: Knowledge Under Digital Conditions*, Social Science Research Council. Posted online August 3, 2016 at http://parameters.ssrc.org/2016/08/a-more-humane-approach-to-digital-scholarship/.

[32] M. Dean, 'Introduction', in G. Megargee and M. Dean, eds, *The United States Holocaust Memorial Museum encyclopedia of camps and ghettos 1933–1945*, vol. II, *Ghettos in German-occupied Eastern Europe* (Bloomington, IN, 2012), xliii-xlviii. Cited here on xliii.

[33] G. Miron, 'General introduction,' in Guy Miron, ed., with Shlomit Shulhami, co ed., *The Yad Vashem encyclopedia of the ghettos during the Holocaust* (Jerusalem, 2009), xl-lxiii. Cited here on xl, xliii.

Your short guide to the EUP Journals Blog http://euppublishingblog.com/

A forum for discussions relating to Edinburgh University Press Journals

1. The primary goal of the EUP Journals Blog

To aid discovery of authors, articles, research, multimedia and reviews published in Journals, and as a consequence contribute to increasing traffic, usage and citations of journal content.

2. Audience

Blog posts are written for an educated, popular and academic audience within EUP Journals' publishing fields.

3. Content criteria - your ideas for posts

We prioritize posts that will feature highly in search rankings, that are shareable and that will drive readers to your article on the EUP site.

4. Word count, style, and formatting

- Flexible length, however typical posts range 70-600 words.
- Related images and media files are encouraged.
- No heavy restrictions to the style or format of the post, but it should best reflect the content and topic discussed.

5. Linking policy

- Links to external blogs and websites that are related to the author, subject matter and to EUP publishing fields are encouraged, e.g.to related blog posts

6. Submit your post

Submit to ruth.allison@eup.ed.ac.uk

If you'd like to be a regular contributor, then we can set you up as an author so you can create, edit, publish, and delete your *own* posts, as well as upload files and images.

7. Republishing/repurposing

Posts may be re-used and re-purposed on other websites and blogs, but a minimum 2 week waiting period is suggested, and an acknowledgement and link to the original post on the EUP blog is requested.

8. Items to accompany post

- A short biography (ideally 25 words or less, but up to 40 words)
- A photo/headshot image of the author(s) if possible.
- Any relevant, thematic images or accompanying media (podcasts, video, graphics and photographs), provided copyright and permission to republish has been obtained.
- Files should be high resolution and a maximum of 1GB
- Permitted file types: *jpg, jpeg, png, gif, pdf, doc, ppt, odt, pptx, docx, pps, ppsx, xls, xlsx, key, mp3, m4a, wav, ogg, zip, ogv, mp4, m4v, mov, wmv, avi, mpg, 3gp, 3g2**.***