USING A GIS FOR SEARCH AND VISUALIZATION OF LITERARY WORKS IN THE DIGITAL HUMANITIES

Ines Schiller, Bastian Entrup, Frank Binder, Sandra Schaarschmidt, Henning Lobin

Abstract: This paper presents challenges and opportunities resulting from the application of geographical information systems (GIS) in the (digital) humanities. First, we provide an overview of the intersection and interaction between geography (and cartography), and the humanities. Second, the “GeoBib” project is used as a case study to exemplify challenges for such collaborative, interdisciplinary projects, both for the humanists and the geoscientists. Finally, we conclude with an outlook on further applications of GIS in the humanities, and the potential scientific benefit for both sides, humanities and geosciences.

Keywords: eHumanities, Digital Humanities, GIS, Historical Maps, Uncertainty, GeoHumanities, Holocaust, Literature

EINSATZ EINES GIS ZUR SUCHE UND VISUALISIERUNG LITERARISCHER WERKE IN DEN DIGITALEN GEISTESWISSENSCHAFTEN


Schlüsselwörter: eHumanities, Digitale Geisteswissenschaften, GIS, Geschichtskarten, Ungenauigkeit, GeoHumanities, Holocaust, Literatur

Authors

M. Sc. Ines Schiller
Dipl.-Inf. Frank Binder
Prof. Dr. Henning Lobin
Justus-Liebig-Universität Gießen
Zentrum für Medien und Interaktivität (ZMI)
ludwigstraße 34
D-35390 Gießen
E: {ines.schiller, frank.binder, henning.lobin}@zmi.uni-giessen.de

M. A. Bastian Entrup
Justus-Liebig-Universität Gießen
Angewandte Sprachwissenschaft und Computerlinguistik
Otto-Beihagel-Straße 10D
D-35394 Gießen
E: bastian.entrup@germanistik.uni-giessen.de

Dipl. Geogr. Sandra Schaarschmidt
Justus-Liebig-Universität Gießen
Institut für Geographie
Sendenbergsstraße 1
D-35390 Gießen
E: sandra.schaarschmidt@geogr.uni-giessen.de
1 INTRODUCTION

"Just as geography, with its focus on space and place, is now actively engaging with the humanities, so are the discourses of the humanities increasingly incorporating place and the spatial dimension" (Richardson et al. 2011, p. 3).

The intersection between humanities and geography that has been visible in disciplines like historiography for a longer time has been rediscovered in the second half of the 20th century. Many humanities disciplines have seen a spatial turn that led to the realization that human actions as the central subject matter of the humanities are fundamentally grounded in space and should be analyzed with regard to their spatial properties. This development is being fostered and fueled today by the broad availability of geographic information systems (GIS).

GIS offer new possibilities to answer research questions in the humanities that were unheard of before. In fact, they help to realize and ask these questions in the first place. As geography and the humanities converge, new topics are being suggested that require a transdisciplinary perspective and a combination of methodologies" (Richardson et al. 2011, p. 3). This development not only urges humanists to adapt new methodologies and to borrow theories from geography, it also demands that the humanities’ approaches to research be incorporated in the geosciences, e.g., when working with uncertain data or interpreting data not solely based on facts, but on methods such as hermeneutics. Accordingly, geosciences and the humanities bear a great potential for collaboration. Both sides can profit from a fruitful cooperation, especially as long as they manage to remain open-minded. As a result, the emerging field of GeoHumanities can be seen as one particular branch of the wider eHumanities, which – from our point of view – are best described as “a cascade of alternating role-taking in enabling each other” in terms of scholarly and scientific challenges and practices (Entrup et al. 2013b, p. 9-4).

In recent years there have been a number of workshops within the Digital Humanities and GeoHumanities. At the main conference of the Digital Humanities (DH) workshops were organized in 2012 and 2014, and at the first conference of the German-speaking branch “Digital Humanities im deutschsprachigen Raum” (DHd) in 2014 the authors organized a GeoHumanities workshop with a focus on developments in the German-speaking community (Binder et al. 2014b).

This paper discusses some recent contributions of geography and GIS science to the humanities in general, and to the fields of literature and history in particular. The observations presented herein are complemented by the point of view of the “GeoBib” project, one of the 24 eHumanities projects currently funded by the German Federal Ministry of Education and Research (Bundesministerium für Bildung und Forschung, BMBF). The paper is structured as follows: Section 2 aims to provide an overview of the application of geography and GIS to the humanities. Section 3, which also contains a brief mission statement of the GeoBib project, shows that despite quite some progress in the (digital) GeoHumanities, many problems remain open and (partly) unsolved. The status quo leaves us with a number of questions that need further investigation and that offer points of contact and engagement for humanists, geographers and GIS scientists alike.

2 GEOHUMANITIES – AN OVERVIEW

2.1 GEOGRAPHY AND/HUMANITIES DISCIPLINES

Humanities are concerned with all kinds of human artifacts, be it art, like paintings or sculptures or music; languages, as in speech and literature, or other man-made objects, as in archeology. Especially History has a long tradition of spatial approaches to their subject of research. Ayers (2010, p. 3) calls history a “bridging discipline [that] deals with the other defining context of human life: time.” But history is more than just “deeply complementary” (Ibid.) to maps. The historiographies of Herodotus or Tacitus were intensely concerned with the space people were living in. Early on, historiography was used to compile maps of the world depicting lands, peoples, stories, and infrastructures; the common origin of both history and geography.

In archeology the spatial allocation of objects plays an important role in communicating the results of research but also in interpreting the findings. Hence, archeology is concerned to a great extend with drawing maps of sites and the artifacts discovered. Recent projects focus on visualizing sites and artifacts using geographical information systems (GIS). For example, the MayaArch3D project (Schwerin et al. 2012) offers a platform designed for sharing and analyzing archeological data in a collaborative way. Single objects, e.g., buildings or statues, are virtually reconstructed, using 3D software, and can be examined using online tools. Furthermore, the data can be queried, e.g., by selecting and showing all artifacts from a certain period or depicting certain motifs.

Linguists in the late 19th century started to consider the spatial distribution of linguistic features. The German linguist Georg Wenker compiled a number of maps of German dialects on different linguistic levels, e.g., morphology, lexicography, or phonology, over the period of many years. Wenker’s maps have recently been digitized and are made available online (Schmidt & Herrgen 2001). The underlying GIS is quite complex and allows to answer a variety of linguistic research questions. Different “linguistic” layers, similar to geographical layers of maps, can be combined and queried. The distribution of certain patterns over time leads to a new understanding and a revision of previous assumptions on language change and linguistic drift. A comprehensive overview of recent findings using GIS in the study of German dialects using the “Digitaler Wenker Atlas” can be found in Schmidt (2010).

Other disciplines, such as literary studies or the social sciences, (re-)discovered the importance of space within the last 30 years. The so-called spatial turn started in the social sciences. In 1974, Henri Lefebvre published his La production de l’espace, in which he states that the social space is a social product (Lefebvre 2012, p. 68 ff.), meaning that every space is a product of human action and interpretation. He defines three levels of space: the experienced space, the conceptual space, and the lived space of the actors.

In the literary sciences the spatial turn introduces a critical perception of the different forms of represented space in literature. Sojas (1996) distinguishes between the actual space, i.e., the first space, the im-
imaginary or fictive space, which he calls second space, and the third space, a coexistence of real and imaginary space. Literary texts recreate space: they absorb and transform natural spaces and construct new worlds: imaginary spaces. Not only fantasy novels construct imaginary worlds, like J. R. R. Tolkien’s Middleearth (Tolkien 1961) or George R. R. Martin’s Westeros and Essos (Martin 1996), but also those genres that are often thought of as being closer to reality, like autobiographies or other kinds of reports, transform the natural spaces into fictional worlds.

Literary texts contain a great amount of underspecification and uncertainty (cf. Reuschel et al. 2013). Fictitious places are hard to map since all “narrated spaces don’t have definite borders” (Piatti et al. 2009, p. 184). Hence, Piatti et al. use the term imprecise geography. These characteristics of literary spaces not only result in places and spaces with fuzzy boundaries, but necessarily also in blurred visualizations. Dealing with uncertainty is an inter-disciplinary challenge (Binder et al. 2014a), maybe even more so for cartographers, geographers, and computer scientists. In contrast to the so-called exact sciences the humanities are seldom precise, although the Digital Humanities aim at establishing computer based methods to conduct research in the different humanities disciplines, hermeneutic approaches to interpretation are seldom reproducible or unquestionable truths. Literary scholarship builds on interpretations and not only allows but needs the uncertainty in order to come to conclusions and interpretations of the work of human minds.

In his controversial book Graphs, Maps, Trees, Franco Moretti asks the, in his eyes, rhetorical question: “Do maps add anything, to our knowledge of literature?” (Moretti 2005, p. 35, cf. Goodwin & Holbo 2011). To provide an answer he creates maps of places occurring in different books in order to show how the understanding changes with these visualizations and how they support new ways of interpreting and understanding literature. Piatti et al. demand that literary geography has to be more than an illustration: “literary maps are meant to be tools of interpretation” (Piatti et al. 2009, p. 184).

An interesting example of the additional insight that maps can add to even very well-known old texts, is given in Evans and Jansow (2013): Mapping the catalogue of ships in Homer’s Iliad (see http://ships.lib.virginia.edu/), comprising no less than 190 spatial references that the oral poet had to memorize, they were able to prove theories about how he managed to do so: The order of the towns mentioned in the Iliad follows a path through Greece that the poet could walk or travel in his imagination, memorizing the villages along the way. Moreover, they were able to show how the syntax mirrors the geography.

2.2 DIGITAL AND COMPUTATIONAL METHODS FOR THE GEOHUMANITIES

The humanities deal, to a great extent, with texts or even paintings or sculptures that are not easily translated into numbers or structured data for processing in (geographical) information systems. While economists and sociologists often work with large sets of structured data, such as charts and tables of numbers describing some kind of performance or census data, literary researchers, for example, deal with (mostly) fictional texts, i.e. unstructured data and information that are in need of interpretation. Such textual data need preprocessing and prior analyses, to make them computable or processable by information systems, and more specifically, GIS. The necessary pre-processing can be done either manually or, when working with texts, by applying techniques from natural language processing (NLP). Practically almost every scenario will require a combined approach that integrates manual annotation and automated processing steps.

When working with natural language texts, different problems arise. In natural language, so called named entities (NE) are used to refer to entities by a name, e.g., using a given name to refer to a person, or conventional names for organizations, places, or landmarks. Such entities and their relations and related assertions are the essential pieces of information that many digital humanities projects work with. In order to identify these relevant pieces of information in the context of the GeoHumanities,

- The first step is to identify those named entities that refer to locations.
- The second step is to disambiguate the named entities, i.e. to resolve any ambivalent references.

The third step is to georeference these entities, i.e. to associate them with geographical coordinates. The first step can be achieved using a number of different approaches, e.g. by matching words or phrases with predefined lists of locations, possibly obeying syntactical hints in the text, such as certain prepositions or constructions. The second processing step depends on contextual information: what other tokens or entities are used within a text, paragraph, or sentence, that are known to be connected to one of the available disambiguation candidates? Finally, the third step makes use of knowledge bases, such as gazetteers, to look up each referred entity’s geographical position.

Some progress has already been made in implementing tools and techniques for the automated identification of spatial and temporal references as well as (geographical) relations between entities in natural language text. Blessing & Schütze (2010) propose a system that makes use of different methods from natural language processing to identify geographical entities and geographical relations in texts. Focusing on such geographical relations and using structured data from Wikipedia infoboxes, that present common relations between occurring entities and the entity that is the topic of the respective wikivpage, such as a birthplace of a person, they identify corresponding pieces of unstructured data, i.e., the sentences that describe these relations in natural language in the wiki text. To automate the extraction process, they train a classifier that extracts these sentences, recognizes the entities, and automatically annotates the learned relations between the entities even in new, unseen texts.

For georeferencing, the third step mentioned above, different gazetteers are available and are in fact applied to different problems in historical and literary studies. The Rigeo.net project (Loos et al. 2013) for example compiles different datasets, data from different gazetteers, and offers access to this data in an integrated, coherent way. Projects such as Pelagios (Simon et al. 2012) or Google Ancient Places (GAP) (Isaksen et al. 2012) make use of gazetteers to (semi)automatically georeference places in texts and visualize the so-called geographical footprint of such
texts as well as the geo-temporal distribution of the places that are referred to in the texts. The Pelagios project aims to “create links between entities on the web based on the places they refer to” (Simon et al. 2012, p. 2). These entities can be any kind of data object, texts, or their elements, as well as records in databases, images etc., that are related to a particular place identified by an Pleiades Gazetteer of the Ancient Worlds URI. The additional benefit over the usage of a regular gazetteer is that a reference to the places themselves is meant to be made available, linking datasets and thereby offering new sources for research on the connections between the entities and on the places themselves.

When the usage of GIS first spread among historians, the so-called “historical GIS” was, in parts, frenetically welcomed. Anne Knowles (2000, p. 451) said that people loved new tools, here the historical GIS, “that enables them to do what they have dreamed of doing”. The major concerns were, and as this paper will show still are, digitizing the historical material at one’s disposal. This means especially scanning and georeferencing maps and other data. This data allows for research questions especially considering “urban, transportation, business, and environmental history” (Knowles 2000, p. 452). Many projects focused on providing digitization of historical maps, extending these with further material for educational purposes. In combination with textual data, projects such as the aforementioned GAP arose. In section 3 a usage scenario for applying GIS to historical, text-based data will be shown and the implication the occupation with historical data has for the GIS scientist.

Within all the different fields of Humanities that GIS are applied in today, a number of methods are adopted from the geo-informatics: Geocoding is the act of annotating spatial information to a medium (e.g. texts, pictures or videos); finding associated geographic coordinates from other geodata or to perform transformations of these information. In this context, spatial information can be postal codes, addresses, coordinates or transformations. Geotagging and georesolving are parts of geocoding (cf. Tobin et al. 2010). In the GeoBib project, the literature is geotagged, e.g. coordinates for places and polygons of regions are added to the texts. In a more generic fashion, the term georeferencing can be used to refer to “the geographic associations of information” and all required processing steps in general (Hill 2006).

As will be shown in the following case study for the application of GIS in the context of historical and literary studies, the availability of maps, vector data, and geographical information in general is one of the most persistent impairments for working with GIS in the humanities. While current maps of the world are available in a number of different ways, such as from Google Maps or OpenStreetMap, ready-to-use historical map data is hardly available at all, and if so, it requires tedious harmonization and integration.

3 THE “GEOBIB” USE CASE – REQUIREMENTS, CHALLENGES, AND STRATEGIES

The GeoBib project – short for GeoBib: An Annotated and Georeferenced Online Bibliography of Early German and Polish Holocaust and Camp Literature (1933-1949) – investigates texts of early holocaust and camp literature in German and Polish language and is constructing an annotated and georeferenced online bibliography of these texts in order to support and stimulate research on narratives of remembrance of the holocaust. It aims at collecting and annotating those texts that were published between 1933 and 1949, and that are mostly forgotten today (Entrup et al. 2013a). As Beorn et al. stated (2009, p. 563), the “Holocaust was a profoundly geographical event, rooted in specific physical spaces, times, and landscapes”.

Accordingly, besides historical and literary editing of these early holocaust texts, the bibliography will also incorporate GIS functionalities to support literary and historical research. Furthermore, educational purposes are considered when making carefully curated geographical and bibliographical information on these early testimonies of the outrageous crimes of the Nazi regime available to research communities and to the interested public.

3.1 REQUIREMENTS FOR SEARCH AND VISUALIZATION

The GeoBib online platform will support its users’ research goals by offering a bibliographic search tool that allows to find texts by thematic queries, by filtering for special authors [or properties of authors, such as their gender], or by conducting geophysical searches. To do so, it leverages various kinds of information. The combination of carefully written abstracts and content analyses, as well as information on scholarly and public reception and history of each holocaust literature text, offers insights on two levels of testimony that these early texts present: the historic events described in the texts and the history of the creation, publication, and reception of the texts. Information on publishers and editors offer insight into the publishing sectors of Germany, Poland, and the exile right at the time when the Nazi regime was committing genocide as well immediately after the dramatic events of the holocaust.

The platform also represents a comprehensive literary lexicon, offering abstracts and plot analyses which allow a more in-depth approach to the texts. By adding information on the authors, it also offers a biographical lexicon for authors that are mostly unknown today. Besides its coverage of – and distinction between – German and Polish texts, the texts are distinguished according to their literary genre, which allows focusing on special areas of interest by selecting only prose texts or only autobiographical works, for example. The combination of all these detailed information on the texts, the authors, and the referenced places allows to answer not only preliminary research questions, e.g., about the role of a certain area in the history of persecution and the holocaust, but also very detailed queries, helping answer questions about, for example, Polish texts by women related to a certain concentration camp.

Georeferencing, or more precisely geotagging the places mentioned in the texts allows searching for and finding texts associated with specific regions or places. The regional distribution of the texts, supposedly resulting from deportation and exile, becomes graspable for the first time.

All these information and the possibility to query this data allows to accurately integrate these texts into specific educational contexts. The GeoBib platform will enable teachers to find holocaust texts of regional relevance, e.g., written by authors that lived or were born nearby, as well as identifying texts with a focus on a special topic.
3.2 CHALLENGES REGARDING GEOGRAPHICAL DATA FOR GEOBIB

Historical maps and geo-data form the data basis of the WebGIS of the GeoBib project, which focuses on Central and East-Central Europe during the period from 1933 to 1949. Hence, the first task within GeoBib’s geodata work package has been to find and collect the necessary data. During the inquiry, 282 institutions or persons were contacted. Of these 101 replied and 14 contacts contributed data to our project (Schaarschmidt 2013). In spite of these responses, the data remains incomplete: We were not able to obtain data for each single country and year of the intended time span.

Much of the data we received are raster data, e.g., raster data of Russia, Ukraine and Belarus. Because of the large workload involved when working with raster data (including geocoding, digitizing boundaries and borders of regions on the maps, and creating the attribute data and metadata), we were aiming to find vector data. But as experience shows, even vector data, just as any geodata of any kind and source usually has to be processed and harmonized in some way before it can be imported into a single WebGIS (Schiller et al. 2013). While processing the authorized data, we noticed inaccuracies, inconsistencies and data heterogeneity. The following sections will provide details on the most pressing issues and the strategies we apply respectively.

Since not one dataset exists that cover the whole of Europe during the time from 1939 and 1949, i.e., a data set that contains all boundaries of all countries for all years, we are forced to forge one, more or less coherent, dataset from different, sometimes contradicting or mismatching datasets. All maps that we employ are only used for visualization purposes so that minor mismatches are not an essential problem. In the following, the three main problems arising when working with such different data sets are presented.

Boundary Precision

Working with maps from different sources can be cumbersome. Uncertainty or inaccuracy arise from, e.g., the different scales that were used to produce the maps and from different degrees of generalization. Hence, boundaries of maps taken from different sources generally do not match each other (Gregory 2005, p. 37) but result in gaps and overlaps (see Figure 1). Since we are only interested in the visualization, these problems only arise in large-scale views, while they are mostly hidden in small-scale views of the resulting, merged map. The precision of boundaries taken from different datasets is, of course, a common problem. Usually, one would try to circumvent it by using data from only one source. Since this is not available in a case like ours, we will use heuristic approaches to unify the boundaries taken from different sources and to minimize the differences between them.

The resulting map is a compromise of different datasets. For inference and analysis, this data seems inappropriate since the accuracy of the boundaries is dubious. However, in the context of the project presented here, the maps are meant to be a visual aiding tool for the end user. Slight
Inaccuracies of the boundaries seem acceptable.

Conflicting Maps

Major inconsistencies between different datasets when working with historical maps and especially during times of war and conflicts result from the usage of maps as instruments of political power and for propaganda purposes (cf. Pickles 1992, Przybytek & Strauchhold 2012). An example of the resulting mismatches can be seen in Figure 2, which shows (parts of) Belarus. The boundaries shown on the different maps, taken from different sources, apparently do not match. The historical context has to be taken into account when working with such contradictory data. For the visualization of the map on a webpage, showing and clarifying the nature of the discrepancies of the data sets is challenging. While GIS allow in principle to let the user switch between different underlying maps, novice users may quickly become overwhelmed and distracted by a large variety of maps or maps that differ only in small details which are not recognizable by non-historians. However, it is just as undesirable to offer maps that are subject to dispute among historians. Time constraints may also limit our ability to provide different map views, in such cases we will choose and refer to one specific source of data. By applying maps only for visualization purposes and by carefully crafting the map data to support only specific levels of detail, we aim to balance the different interests in that regard.

Data Heterogeneity

The second aspect of data inconsistency refers to separate data of one area supposedly at one point in time, when different datasets show differing boundaries. Figure 3 presents two datasets of Poland from 1939. Although both datasets represent Poland in 1939, the attached metadata sets do not offer information on whether the map presents Poland before or after the territorial reform of 1939. Such questions can only be answered by detailed research and in close cooperation with historians.

Another major source of data heterogeneity lies in attribute data of the vector data. Each project, institute or creator has their own schema and we do have our own requirements within our project as well, so that we have to convert the attribute data of the different sets to our own schema. The implied conversion, integration and editing can only be done manually because of the heterogeneity of the datasets. In our experience it is mandatory to look at every attribute of every layer to verify and if necessary correct them. Fur-
As we have just discussed, any project that uses (digital) historical maps requires close cooperation with data providers, historians and other experts. The following paragraphs show our approach to these problems that commonly arise in the (digital) humanities when working with historical maps and geodata based on literary texts.

The project’s main view consists of a map of Europe, mostly based on the REGIS dataset (Regional European GIS, Flora et al. 2015). National border data will be available for each year between 1933 and 1949. The REGIS dataset does not, however, offer a complete dataset for the whole of Europe during the given time span. Only borders as accepted by international law are covered, which results in ‘white spots’ on the map that we are trying to fill with other datasets. For example, this dataset has no information on the borders of Germany after the Invasion of Poland 1939. Additional data is used which results in an overlap between the REGIS data for Europe, especially Poland and Czechoslovakia, and another data set for Germany as can be seen in Figure 1 (darker blue area).

In this case the mismatch is due to different points in time chosen within the year 1939. Furthermore, our project’s context requires the web maps to be complemented with certain additional information on features of each layer. To meet these requirements we apply our own attribute schema and we have to transform and edit the data of the different data sets to our schema, as discussed in the following section.

3.3 APPROACHING THESE CHALLENGES WITHIN GEOBIB

As we have just discussed, any project that uses (digital) historical maps requires close cooperation with data providers, historians and other experts. The following paragraphs show our approach to these problems that commonly arise in the (digital) humanities when working with historical maps and geodata based on literary texts.

The project’s main view consists of a map of Europe, mostly based on the REGIS dataset (Regional European GIS, Flora et al. 2015). National border data will be available for each year between 1933 and 1949. The REGIS dataset does not, however, offer a complete dataset for the whole of Europe during the given time span. Only borders as accepted by international law are covered, which results in ‘white spots’ on the map that we are trying to fill with other datasets. For example, this dataset has no information on the borders of Germany after the Invasion of Poland 1939. Additional data is used which results in an overlap between the REGIS data for Europe, especially Poland and Czechoslovakia, and another data set for Germany as can be seen in Figure 1 (darker blue area).

In this case the mismatch is due to different points in time chosen within the year 1939.
due to the historic period that is covered by the GeoBib project.

One of the main difficulties is the accuracy and ‘validity’ of the borders. This is caused by many border changes in the period from 1933 to 1949, most of which, of course, do not accord with international law. The REGIS dataset forms the basis of the visualization maps. Unfortunately, this data does not cover the whole of Europe between the time of 1933 and 1949: only those borders recognized by international law are covered. This results in the fact, for example, that no data is available for France during the years from 1940 to 1945. During this time France was de facto split in two parts (the northern zone occupied and the so called southern zone libre). De jure these border changes were never made according to international law.

By combining different vector data, we will choose a visualization which makes the differences of the datasets graspable: controversial borders and overlapping regions from different datasets will be displayed as hatched areas.

The main map, that has been described so far, is used for standard visualization according to the time span. Furthermore, additional maps will be made available. These are taken from different datasets and might visualize certain views on Europe that are not necessarily in accordance with the borders accepted by international law. Furthermore raster data for certain time periods and regions of Europe are available and will be made accessible to the user to combine and contrast them with other maps.

Another source of uncertainty in working with literary texts and references to historical places is the process of georeferencing itself. Literary places (second space, cf. above) do not necessarily accord to real life places (first space). They form a third space (cf. above). This also holds for our collection of mostly autobiographical texts and historical reports: some place names can hardly be georeferenced. Sometimes a fictional name is used while a reference to a place on a map can be inferred. Sometimes the naming of places is uncertain. In other cases, the exact geographical location of historical locations, cf. for example, some smaller concentration camps or mass graves, is at least controversial. The geocoding, but also the temporal classification of place references made in the texts, raise the question of certainty again: How meaningful is a geographic point used to indicate a place when it comes to literature? How certain are borders of places? As mentioned before, Piatti et al. (2009) call this imprecise geography.

For GeoBib we are using geographical points to mark places but we visualize these in the fashion of heat maps that indicate the center of each point but also include a region around it. Figure 5 shows this kind of visualization using a preliminary dataset of only 67 texts and their annotated place references. The darker regions on the overview map refer to places with many references, while the lighter colored ones occur less frequently. The smaller figure shows Berlin and a nearby place reference: instead of using singular points to mark places, rather fuzzy visualizations are applied.

4 DISCUSSION AND OUTLOOK
Certainly, the application of GIS or of maps in general offers new insight into data that was not considered ‘spatial’ before, or whose spatial distribution had mostly been ignored. GIS can help answer new, previously unasked questions. They do so by offering an intuitive view on data. A more critical view on GIS is expressed by Bodenhamer (2010, p. 16 ff.) who criticizes that GIS would “suggest that the world is indeed flat [...] by offering a view of the physical environment seemingly stripped of its cultural assumptions”. Although GIS offer the possibility to visualize a great amount of data, Bodenhamer argues that it also treats the data uncritical. Humanists usually have to handle “[a]mbiguity, uncertainty, nuance, and uniqueness, all embedded in the evidence typically available to humanists” (Bodenhamer 2010, p. 23).

Efficient methods to treat uncertainties in GIS (but also in other digital tools) are missing and have to be constructed with effort as has been shown. Since many projects are concerned with visualizations of (uncertain) data, more research and effort is required in this regard. Besides the creation of persuasive visualizations, their ease of use for non-geoinformation scientists is an important criterion for the application of such tools in the digital humanities.

Regarding the problems resulting from boundary precision and validity in heterogeneous map data, one has to weigh the amount of work needed to correct the datasets against the problem arising from poorer map concordance in higher scale where details become visible. Furthermore, maps representing different views on the world result in contradicting boundaries and borders. In our case we will not be able to compile and present a complete dataset for Europe between 1933 and 1949. We refer to an international project that aims at providing clearly defined European bound-

Figure 5: Overview of GeoBib map with heatmap of literary locations
aries between the late 19th century and present times that has been working towards that goal for years and has not yet released a complete dataset [Flora et al. 2015]. This situation illustrates the high workload associated with such endeavors.

Regarding the fundamental ability of the humanities to integrate and discuss contradicting sources and assertions, GIS – like any other digital tool – are no magic wands per se. Regarding the GeoBib project, for example, many known challenges in mapping the Holocaust remain as hard to solve as before [cf. Bode & Renz 2012]. However, due to their potential interactivity, GIS applications allow in principle to communicate contradictions, mismatches, and cartographic decisions to the user or recipient, making them visible by including interactive elements, by allowing to select different views or layers, or by switching between conflicting maps. Compared to the limited groups of scholars and interested experts to whom this sort of knowledge has been accessible so far through in-depth intellectual work, interactive GIS, with their prospective availability to the interested public, enable open communication of more advanced historical, political, geographical, and cartographical knowledge to wider audiences than ever before.

Both, the criticism expressed above, but also most projects in history and literature, have so far ignored the more sophisticated scientific methods of geoinformatics. GIS are more than just visualization tools. Besides such mere visualizations (cf., for example, different, contradicting propaganda maps), they offer methods to calculate intersections of different boundaries and might help to gain insight into the expansion or decline of states and countries over time. Geocoded texts furthermore offer possibilities to calculate similarity between texts with regard to their spatial footprints. Some research in these directions has already started, especially in the context of [geographic] information retrieval, but the humanities have not yet taken up these emerging possibilities more widely.

Such efforts, however, will only thrive if they are backed by a reliable geographical information infrastructure that also provides carefully curated historical geographical data and historical maps. Current Pan-European initiatives such as INSPIRE should therefore be complemented by branches that focus on historical geographical maps, both as scans and vector data as well as other related data. The community of interested scholars has recently formed a special interest group on GeoHumanities under the Alliance of Digital Humanities Organisations (ADHO) [Grossner & Weiher 2013], which is just one starting point when looking for potential fellow campaigners and co-workers in this regard.

As can be seen from the overview given in this paper, GIS are being successfully applied to answering research questions in the humanities. Still, the humanities have specific needs and requirements, which need to be addressed by geographers and computer scientists alike. Furthermore, these latter disciplines have methods and approaches to offer that have not yet been applied to research in the humanities. More fruitful cooperation might lay ahead, if all sides come together, learn from each other, and adapt methods to the needs of the other.

5 ACKNOWLEDGEMENTS

GeoBib is funded by the German Federal Ministry of Education and Research (grants no. 01UG1238A/B).

We thank the editors of gis. Science for the opportunity to compose this article which is largely based on the previous conference talk and paper by Schiller et al. (2014) as well as the impressions from a workshop on GeoHumanities in Passau (Germany) in March 2014, which was organized by GeoBib (Binder et al. 2014b). We would also like to thank the reviewers for their helpful comments.

References


We would also like to thank the reviewers for their helpful comments.