

RESEARCH ARTICLE

Towards place-based exploration of Instagram: Using co-design to develop an interdisciplinary geovisualization prototype

Linked geovisualization and co-occurrence networks with qualitative analytics for exploring perceptions of cultural routes

Catherine Emma Jones¹, Daniele Guido¹, and Marta Severo²

¹University of Luxembourg, Luxembourg

²Paris Nanterre University, France

Received: November 8, 2017; returned: January 19, 2018; revised: April 3, 2018; accepted: May 30, 2018.

Abstract: An abundance of geographic information is hidden within texts and multimedia objects that has the potential to enrich our knowledge about the relationship between people and places. One such example is the geographic information embedded within user-generated content collected and curated by the social media giants. Such geographic data can be encoded either explicitly as geotags or implicitly as geographical references expressed as texts that comprise part of a title or image caption. To use such data for knowledge building there is a need for new mapping interfaces. These interfaces should support both data integration and visualization, and geographical exploration with open-ended discovery. Based on a user scenario on the Via Francigena (a significant European cultural route), we set out to adapt an existing humanities interface to support social and spatial exploration of how the route is perceived. Our dataset was derived from Instagram. We adopted a thinking by doing approach to co-design an interdisciplinary prototype and discuss the six stages of activity, beginning with the definition of the use case and ending in experimentation with a working technology prototype. Through reflection on the process of tool modification and an in-depth exploration of the data encoding, we were better able to understand the strengths and limitations of the data, the tool, and the underlying

workflows. This in-depth knowledge helped us to define a set of requirements for tools and data that will serve as a valuable contribution for those engaged in the design of deep mapping interfaces for place-based research.

Keywords: Instagram, geovisualization, cultural routes, social media, text mining, spatial analysis, co-occurrence graphs, co-design, interfaces, deep maps

1 Introduction

The notion of place is a complex concept to unravel, but it is of central importance to the knowledge-building process when exploring the everyday lives of people. Many traditional geovisualization or GIS interfaces fail to consider the nuanced notion of place; they are more concerned with the exploration of location through quantitative data. This is at the expense of integrating rich, informal, contemporary or historical textual and multimedia data that is available and is routinely used by humanities scholars. To tackle this challenge, new interfaces are needed, ones that enable more open-ended and exploratory investigations of place while incorporating different data formats and encouraging data visualization at different scales and in different visualization forms. By adopting methodological approaches and tools that are commonplace in the humanities, such as text mining for entity recognition and graph theory and tools for social network analysis, we consider how these can serve as building blocks for new types of interfaces that support place-based exploration that are both socially and spatially enabled. We use standard text mining algorithms to retrieve geographical information from Instagram posts with the aim of creating entities that fit an existing data model/interface (histograph) that was built for historians to perform network analysis based on co-occurrences in large-scale multimedia datasets.

By means of a thinking by doing approach, we discuss the design and adaptation of a graph-based exploration tool from the humanities and consider how it can be adapted in order to integrate an unfamiliar dataset (Instagram). The goal was to facilitate the social and spatial exploration of Instagram data with a view to place-based research. The Instagram resources represent mediated user-generated memories in the form of qualitative and multimedia data, and the resulting prototype exploited photographs, titles, tags, comments, people, and dates. In these resources the geographical information encoded within the Instagram data either consists of existing geographical references from geotags (explicit) or can be captured as part of the text mining process using the textual elements of the posts such as the titles or image captions (implicit). The paper considers how a co-design approach can be used to quickly and parsimoniously develop a working geovisualization prototype that integrates this data. In so doing, we appraise (1) the process of designing interfaces and (2) the use of prototyping to familiarize ourselves with the strengths and weaknesses of unfamiliar Instagram resources and the automatic processing of their geographical information. Our specific use case is based on Instagram posts about the Via Francigena, a significant cultural route connecting the British city of Canterbury with the Italian capital Rome. It was created in the context of the project "*Les espaces du patrimoine culturel numérique: topologies et topographies des itinéraires culturels,*" coordinated and led by Dr Marta Severo.

The paper begins with a review of the literature. We discuss the abundance of implicit or explicit geographic information contained within multimedia resources and consider

the limitations of traditional interfaces as the form for their exploration. We reflect on the perspectives of both humanities scholars and geographers and discuss the need for interfaces that handle different types of geographic data. We then describe the histogram tool, which was the starting point of our thinking by doing approach, and review the co-design methods used to explore the Instagram data and develop the geographic extension of the prototype. The result of this exploration is summarized as a series of observed limitations requiring further work. The paper draws to a close with a short analysis of how well the current tool can be applied to the user scenario about cultural routes, identifying lessons learnt and future requirements.

1.1 Digitized and born-digital (geographic) data and volunteered geographic information (VGI)

Today historians, geographers, and social scientists have an abundance of digital resources and data available at their fingertips. Such resources are often either in the public domain or owned by private/commercial organizations; they may be formal or informal datasets and many contain implicit or explicit geographical information. They are thus ripe for exploitation and use in research. The geographic information within has the potential to enrich our knowledge about the relationships between people and places, bringing us closer to meaningful representations of place and locale [32]. With the advent of text mining and geographical information retrieval (GIR), it is now possible to mine and extract the implicit place-based data expressed within these resources [1,33]. They are structured for analytical purposes using semantic models that are either predefined [32,33] or emergent [3]. Many projects have applied such an approach with a view to investigating archival or text resources, multimedia datasets, and volunteered geographical information (VGI) [1,3,31,33], each with the aim of improving our understanding of how people perceive and experience place. They rethink the role of geographical analysis and traditional GIS by working with mixed data formats. Such work is contributing to the redefinition of how geographical interfaces are used and how they can be developed more creatively [35].

Since the advent of citizens as sensors [6,21], VGI has become particularly interesting for place-based research. Twitter, Flickr, Instagram, and the like collect and curate big data generated and then contributed by users [11], all of which contain spatial-temporal data based on the thoughts, opinions, moods, and activities of people [46]. When these data are mined, enriched, and combined, the resulting analysis can reveal patterns that contribute to new place-based geographical knowledge [19]. By exploiting the information from these sources we can build more nuanced (and personal) representations of place from an individual or collective perspective [20]. However, when working with these data we need to be mindful of their limitations. The data are susceptible to user-generated biases [26] and are the result of the underlying user motivations that led to the user contributions [4]. Additional bias also comes from the fact that only certain societal groups use these social networking tools; many people are yet to adopt these types of technologies into their everyday lives. As such, the VGI data collected contribute to a technology-mediated memory, symbolizing what users wish to communicate about topics, events, and places.

Unlike the somewhat more mature research based on Flickr resources [27,33] or Twitter [2,13,18,46], place-based research using VGI data encapsulated in Instagram posts is still nascent. The few studies that exist focus primarily on the city and encompass a broad range of topics. They include studies that investigate the visual rhythms of the city, city

dynamics and social behavior [42], sensing the city [15], gathering insights into spatial patterns of public green and blue spaces for urban planning [22], evaluating location-based identity [40], and in-place experiences that consider the constructions of narratives to describe museum experiences [45]. Researchers have also focused on population interactions to understand how socio-spatial interactions can provide a deeper sense of the city [10] or measure the attractiveness of a place through the use of explicit geotags [44]. Thus while the potential of using Instagram to provide datasets for place-based exploration is clear, it has yet to be truly realized.

1.2 Interfaces for open-ended exploration of VGI data

VGI datasets containing both implicit and explicit digitized and born-digital (geographic) data demand new ways of thinking about interfaces since they comprise both multimedia and textual geographic content. This is in contrast to the quantitative data that underpins traditional GIS and many web-mapping interfaces. Hence, to work effectively with such VGI data there is a need for new forms of interfaces that integrate, enrich, and visualize mixed data formats. This challenge is not unique to geographers; humanities scholars have been facing this issue for a number of years now and it seems natural that we should look to build bridges between the disciplines and develop new forms of interfaces. Indeed in the humanities, Johanna Drucker [16] noted that interfaces need to be redesigned to better facilitate exploration and knowledge construction for text resources. She stated the requirement to focus more on the critical principles of scholarship rather than the principles of tasks and goals [39]. This thinking led to the conceptualization of deep mapping [34] because of the conflict facing historians with the task-driven approach of traditional GIS/geovisualization interfaces [8,25]. In deep mapping, the foundations were laid for the development of interfaces that focus more on open-ended exploration. The goal of deep mapping is to achieve a depth of understanding about place via the examination of behavior, the material and imaginary worlds, and the relationships that produce scaled conceptions of place [24].

At the same time, in geography, Alan MacEachren made a similar call for innovative thinking about interfaces that deal with place-based concepts [28]. Like Harris [24], he argues that existing technologies do not enable place-based geographic visualization and geographic information systems (GIS) to reach their potential. He notes that traditional GIS and web-mapping interfaces focus primarily on mapping locations and are limited since they denote human behavior as a data point. They are not technology environments designed for use in the humanities and they tend to overly focus on quantitative data. It has been argued that there is a need for exploratory projects to consider the untapped areas of the arts and humanities to determine how they can influence interactive cartography [36]. With this in mind the challenge is to design new tools and interfaces, or adapt existing ones, that work with geographic information derived from mixed formats (e.g., images and texts) to meet the socio-spatial analytical needs of researchers in the 21st century. Interfaces need to adequately model the everyday life and behaviors of people and places.

In this paper we investigate the prototype development of a web-based geovisualization interface to facilitate the open-ended qualitative exploration of places using Instagram data. It presents a form of deep mapping driven by qualitative exploration. The development of this interface relies on graph-based databases combined with web-based mapping built with the co-design process. This was our starting point since it provided an existing system for handling and generalizing patterns in diverse multimedia and text resources. It



also focused on a fluid exploration of qualitative data that better suits the needs of place-based researchers.

2 The starting point for development: histogram, a tool for exploring historical social networks

The point of departure for the development of the prototype for the place-based Instagram explorer was the open-source tool histogram (<http://www.histogram.eu/>), developed by Lars Wieneke, Marten Düring, and Daniele Guido at the University of Luxembourg. Histogram enables the investigation of co-occurrence networks in large diverse multimedia collections containing a total of 17,332 historical resources including treaties, newspaper articles, meeting minutes, letters, photographs, etc. [17]—see Figure 1. The underlying premise is that the multimedia collection is a network of connected material that can be generalized and explored through the use of encoded entities. By mining the co-occurrences in the network through the identification of entities (e.g., people and organizations) it is possible to establish relationships based on the co-appearance of pairs of entities in the resources. For example, if two people such as appear in a photograph together (e.g., Pierre Werner and Jean Claude Junker) they can be connected to each other. The network is then filtered based on the underlying assumption that if entities (e.g., Werner and Junker) are found occurring together many times then there is a more meaningful relationship between them [17]. The histogram tool analyses the textual components of historical resources, for example, the titles and captions (short descriptions) of multimedia objects. It performs text mining and enrichment using named entity recognition and disambiguation (NERD) to identify people, places, and time periods that are mentioned in the resource texts and enrich them using DBpedia or VIAF [17]. A co-occurrence network is created, building up connections between pairs of entities mentioned in the textual attributes of the resource.

The tool links two views: a gallery wall showcasing the details of the original source material and a searchable and filterable network graph of co-occurrences. Qualitative filters allow users to explore patterns and build historical understanding across the network by revealing connections between people and organizations. The chronological thinking integral to historians' research manifests itself in this interface through a "brushing" interface interaction, allowing users to select a temporal subset of data from the timeline. The results are automatically filtered across the gallery and the graph, enabling historians to evaluate the network and its relationships as they emerge through time.

3 Co-design of the requirements for the prototype geographic extension to histogram—potential uses

Within the world of geographical modeling and GIS there has been an evolving interest in user-centered design techniques developed from human-computer interaction (HCI) [10, 23, 28, 36, 37]. This stems from the notoriously difficult-to-use GIS/geovisualization interfaces of the recent past [14]. There seems to be a clear need to engage non-expert users in design practices for exploratory geographic interfaces, but this is still an under-researched topic. Adding to the complexity is the question of which design approach to use. Typical studies such as Roth and colleagues [37] or Bruggmann and Fabrikant [10]

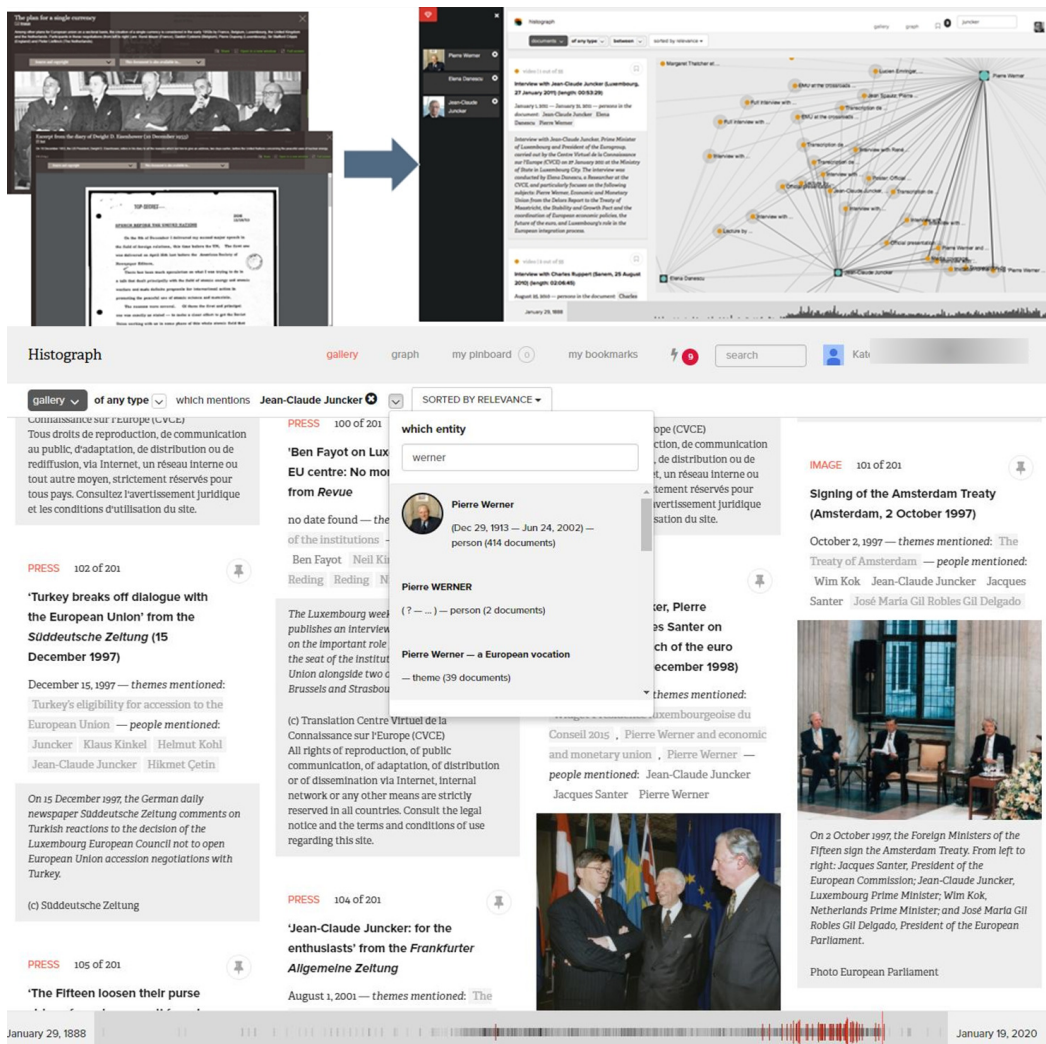


Figure 1: Screenshots of the histogram interface showing original documents prior to processing (top left), graph view of a co-occurrence network (top right), filterable gallery view (bottom).

place the user at the center of the research, defining requirements either through needs assessment studies or focus group research and the like. Whilst these approaches excel at emphasizing the utility and usability of interfaces, they are less likely to foster a sense of user value in the technology or reveal how the users perceive the tool to satisfy their needs or requirements. This is because there is a cognitive distance between the user and the designer. The participant is often being observed and treated as the subject. In view of this limitation there is growing interest in the need to move closer to the future user through the

process of co-design. Making the users partners in the design process fosters collaborative creativity and a sense of value through participation in the co-design team [38].

We adopted a co-design approach, bringing together a variety of users and representatives to jointly produce prototypes that are valued. This approach reduces the distance between the lead users and the designers and is said to create value through the development of more personalized user experiences that more closely match their needs. We achieved this by establishing a small interdisciplinary team composed of the technology stakeholder, three researchers (including the product owner) and the data designer/developer. The interdisciplinary researchers were a geographer (product owner), a historian, and a communication and media specialist (project initiator). The potential end users were represented by the lead users; i.e., the three researchers who are seeking to find collective value in the prototype. The co-design approach involved a six-step process:

1. Definition of the high-level user scenario: to explore Instagram posts for the Via Francigena cultural route.
2. Preparation of the data and independent exploration using different software to explore visualization possibilities.
3. Creation of requirements via a collaborative discussion stimulated by independent exploration of the tool.
4. Rapid development of the extension to create the geovisualization prototype.
5. Experimentation with the prototype:
 - (a) Understanding strengths and weaknesses of the data and its added value.
 - (b) Understanding strengths and weaknesses of the prototype interface and its added value.
6. Identification of user value, future requirements and further work.

Defining the user scenario: Cultural routes are as yet an understudied phenomenon. They embody complex values and represent diverse meanings for an array of users: walkers, residents, institutions, recreational visitors, commercial actors, and the like [7]. The new tool needed to enable open-ended exploration of Instagram posts about the Via Francigena cultural route. The lead users wanted to be able to log in and see a gallery of resources showcasing all the Instagram posts for the Via Francigena cultural route for the selected timeframe, so that they could then drill down to select individual resources. Users would then be able to view social graphs that were built from the network of posts—like in the original histogram tool. These graphs could be used to explore the co-occurrence of topics that were tagged in the original posts to see, for example, if people discussed the concept of pilgrimage and religion alongside topics such as nature and thermal baths or churches. They would be able to discover who posted about what, gaining an understanding of the social space through time. In doing so they would gather information and build knowledge about the usage of the cultural route. Users would then be able to map the occurrences of the different hashtags (stored as themes in the database) and view their associated locations, revealing spatial patterns in the perceptions or activities of the people interacting with each other along the cultural route. It should be possible to filter the maps using more than one theme or entity (place, location, person, time, etc.). Users were also interested to know if people thought of other spiritual places (either imagined places—i.e., were they thinking of other places beyond where they posted the picture from, e.g.,

Lourdes—or other pilgrimage routes such as the pilgrim’s way in North Wales) or people (e.g., saints) whilst they were experiencing the real route and to explore this through time.

Independent exploration using different tools: With the user scenario defined, the initial dataset was prepared for September 2011 to January 2016 (details in Section 4). The data were imported into the existing histogram application to enable exploration via the social graph tool. We were able to explore different concurrence networks showing people discussing various topics and view the images associated with pilgrimage. For geographic exploration Carto was chosen as a simple web-based mapping system with visualization and analysis capabilities. Within Carto we created animated density maps that highlighted the temporal use of the route. We then wrote SQL queries to show hotspots of different activities (religion, pilgrimage, hiking, etc.). This proved a helpful visualization for revealing spatial patterns whilst simultaneously maintaining the underlying linearity that the cultural route signifies. This process allowed us to familiarize ourselves with the data and think more deeply about our requirements which could subsequently be applied to the prototype development.

Identification of requirements via a collaborative discussion: Following these explorations we were then able to discuss the results with the co-design working group and the wider project group in a number of small workshops. This led to the identification of the following co-defined requirements:

- To display occurrences of themes as density heat maps, filter by time, and view the linked co-occurrence graph and gallery.
- To dynamically filter the maps for selected values of different entities:
 - This should update all the linked views (map, gallery, graph) based on the selection.
- To update selectable values of different entities dynamically based on the map window; i.e., the map window becomes a search interface:
 - Applicable to spatial extent—if the user pans the density maps the maps and filters are automatically recalculated.
 - Applicable to spatial scale—if the user changes the zoom level the density maps are displayed accordingly.
- To select and filter a subset of the mapped data using the interactive timeline.
- To be able to differentiate between implicit and explicit geotags (imagined versus real geography).
- To bookmark any visualizations so that they can be viewed later.

4 Development process for (geo)histograph

4.1 Instagram data foraging and extraction

With the requirements gathered, the next step was to adapt and extend histogram to facilitate geographic visualization. After exploring the resources independently in two separate tools the Instagram data was transformed into a set of resource data and imported into

histograph so that we could sort, select, filter, enrich, and explore them. We began by narrowing our scope; our original intention was to include Tweets but during our explorations we discovered that less than 10% of Tweets containing *francigena* or *francigene* had an explicit geotag. So while the few Tweets that did have a geotag were integrated into the app, they were not the focus of the study. Instead, we concentrated on the Instagram dataset due to the rich implicit and explicit geographic information that could potentially be extracted, an approach used to reveal perceptions of place and for modeling the city [10, 27].

Using an Instagram scraping tool developed by the Digital Methods Initiative (DMI) at the University of Amsterdam [9] we extracted relevant multilingual posts for our use case. Posts including the Italian, French or English version of the term "*francigena*" or "*francigene*" (as a hashtag such as *#francigena* or *#francigene*) were included for the timeframe between September 2011 and January 2016. These are the languages of the countries through which the route passes (coincidentally, they were also the languages of the researchers). We extracted 8,834 posts in total. Prior to September 2011 there were few relevant posts, reflecting the adoption curve of Instagram which launched in October 2010.

Unfortunately, the Instagram scraping tool no longer works due to significant changes and restrictions to the Instagram API (introduced after 2016). But it provided us with a set of data relevant to our case study including a photo, date, and time stamp when the post was made, and if available the designated geotag of the user's location at the time of posting (longitude/latitude in WGS84 format). Other metadata includes the caption title of the post, all the corresponding hashtags, the user ID of the poster posting, users mentioned (with their user IDs), comments made on the post, and the number of likes and shares.

4.2 Data preparation: deconstructing Instagram posts into useful entities that are part of the histograph data model

Structuring an Instagram post to import it into histograph required its transformation into a resource comprising different entities (see Table 1 and Figure 2). We worked with the original data model of histograph to see if the tool was flexible enough to handle different data sources. A histograph resource contains the following entities (which form the nodes in the graph network): person, date, themes, title, location, places, and multimedia object (photograph). Building the resource involved cleaning and structuring the data by: (1) reverse geocoding of geotagged images to assign a place name that could be used to build the co-occurrence graph; (2) converting each user ID, including the sender and any others mentioned in the caption, into an entity ("person") to build the co-occurrence graph of people; (3) creating themes by transforming all of the hashtags in the post; and (4) extracting dates of posting into a date entity to facilitate the temporal organization of the data. The object (photograph) and its title did not require any preprocessing and were assigned directly to the data model as *object* and *title* entities.

Step 1: Place names were assigned to the longitude and latitude of the location stamps (geotags) in the original Instagram post. Geonames and the Google API were queried for the nearest village, town, or city to our geographic coordinate pair for the original post. We did not compare the results between the two providers but simply used the aggregated result. Coordinates were transformed into a named entity corresponding to the nearest place where the post was sent. In the data model this is called *place*.

Attribute entity	Person	Place*	Location*	Theme	Object	Title	Date
Element of the Instagram post	User ID	Location picture was taken plus corresponding place name	Places mentioned in resource context	Hashtag	Photo	Caption included with post	Time of posting
Description	Denoted with or persons mentioned in hashtags such as saints	Geo-tagged at time of posting	Places referred to in the title or hashtags	All text denoted with a # (can be people, places or organizations)	Picture posted to Instagram	Free text field where users can add a short title to their picture	Date field that time- and date-stamps the post
Side note	*We are aware that the labeling of these concepts in the system is confusing. In future versions of the interface/data model they will be updated.						

Table 1: Structure of a resource ingested into (geo)histograph.

Step 2: User IDs were converted into a histogram data model entity called *person* by taking all IDs denoted with the prefix which represented the individual user ID of the original poster or those that are mentioned in the caption. There was no differentiation between people or organisations due to time and resource constraints—although in the histogram data model there is an existing entity called organization so this would be feasible in the future.

Step 3: Since the histogram tool was originally designed for exploring historical resources, the *theme* entity corresponds to a subject or topic reoccurring across resources. In this case, themes were created from all the unique hashtags. We took all words and phrases denoted with a hashtag and populated the *theme* entity. No further processing was done to produce more effective machine-readable texts, for example through topic modeling, as we did not have any means to validate the output. In future, such topics could be validated and augmented with the help of the crowd (for example, using CrowdFlower).

Step 4: The creation of a *date* entity required directly transferring the date-stamp of the post into the entity, to facilitate temporal organization and filtering.

4.3 Enriching the dataset and building the tool

With the Instagram posts transformed into resources, the next step was to extend histogram to facilitate geographic visualization and exploration through interactive maps and qualitative filtering. We first had to enhance and enrich the resources with entities not explicitly tagged by the original user—to do this we used a series of discovery scripts. The

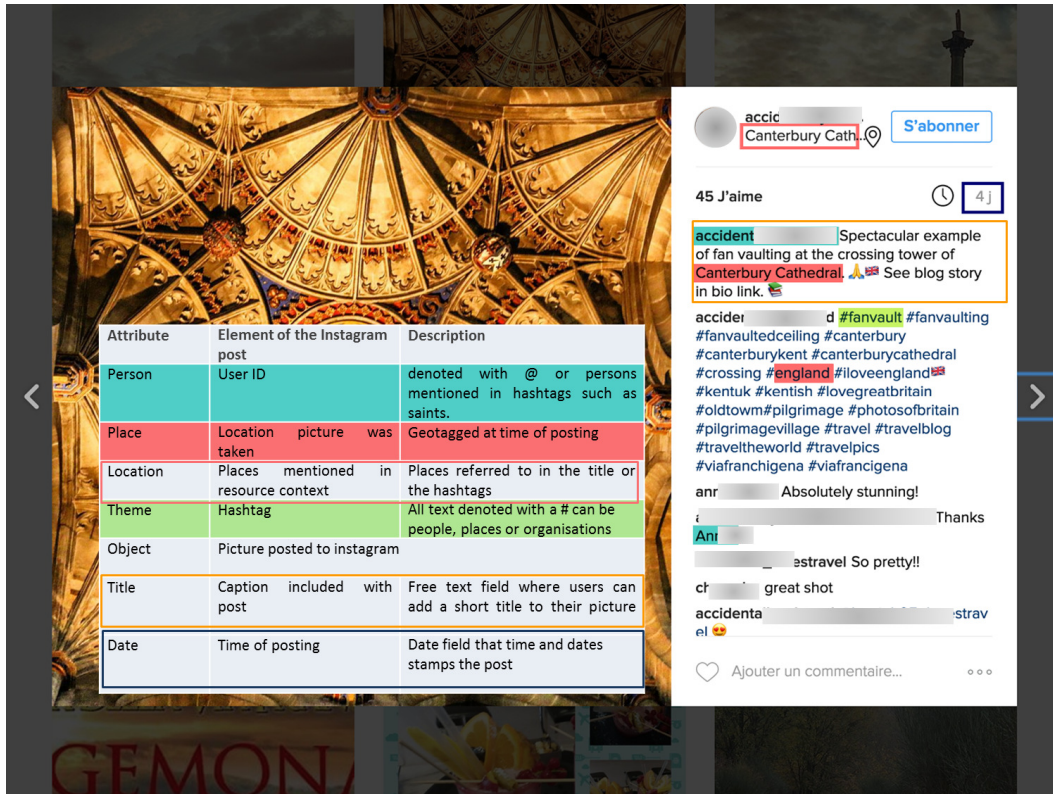


Figure 2: Structuring of an Instagram post into a resource to be ingested into (geo)histograph.

architecture of the tool and its workflow is described in Figure 3. At the heart of the tool is a Neo4J graph database which stores the Instagram resources together with their entity attributes (Table 1). The graph data transforms and optimizes our resources as a set of connected data which forms an initial social graph of the Via Francigena, including places, people, dates, and themes. The different resource entities such as theme, place, person, and date were the nodes in the graph and the co-occurrences of entities in posts were used to build the network of relationships between them.

Once the resources had been ingested into the graph database, a series of text analyses were undertaken to discover “hidden” data in the title, tags, or caption of the original posts, identifying further locations, people, and themes mentioned in textual elements. We applied a process known as multilingual named entity recognition and disambiguation (NERD) across all of the textual content (see Figure 3), according to its language. The resources were in Italian, English, and French, so the multilingual capabilities of the Text Razor API were preferred over services such as YAGO. The text mining process took the title, caption, and comments of an original post and searched and highlighted additional people, places, and themes so that we could enrich the original entities. The result was a set of candidate entries for people, locations, and themes.

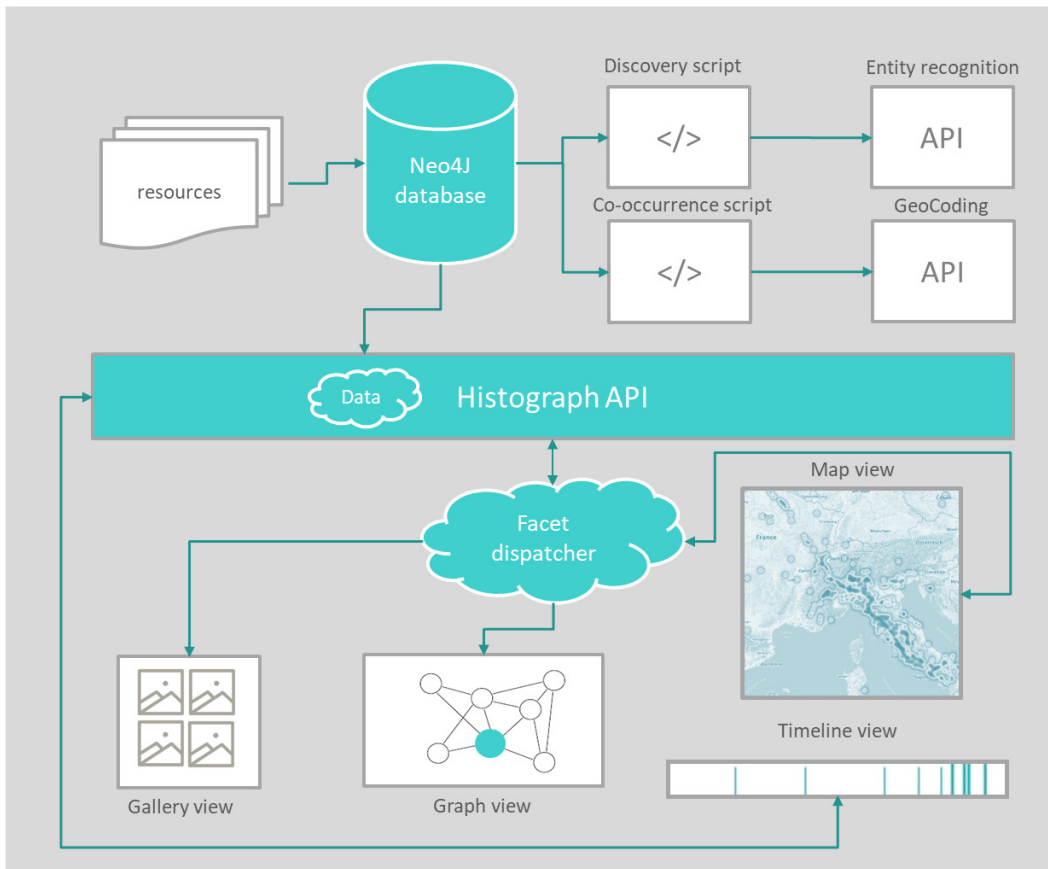


Figure 3: Simplified outline of the (geo)histograph architecture.

The discovery script enriched our entities by searching for and identifying additional people mentioned in the title or caption. For example, the Via Francigena is a route of pilgrimage so we wanted to identify entities of historical significance, such as the *Archbishop of Canterbury* or *Sigeric the Serious*, or the names of saints such as *Saint Mary*. Identifying these relevant person entities goes beyond simply creating a network of people using Instagram; it includes an aspect of the “imagined”—people that are envisioned by users of the route. The motivation for doing this was based on the use case: we wanted to observe co-occurrence networks of conversations about historical figures and what other topics they discussed as well, having the option to visualize these occurrences, although they were not treated as separate entities. Secondly the discovery script searched for themes that were not originally assigned a hashtag but might be contained in the title, such as mentions of weather (e.g., the sun) or of the physical landscape (e.g., thermal water). This helped identify themes that users discussed as they interacted with the cultural route.

The third role of the discovery script was to search for additional places discussed in the titles or captions or from within hashtags. The NERD process examined the texts and

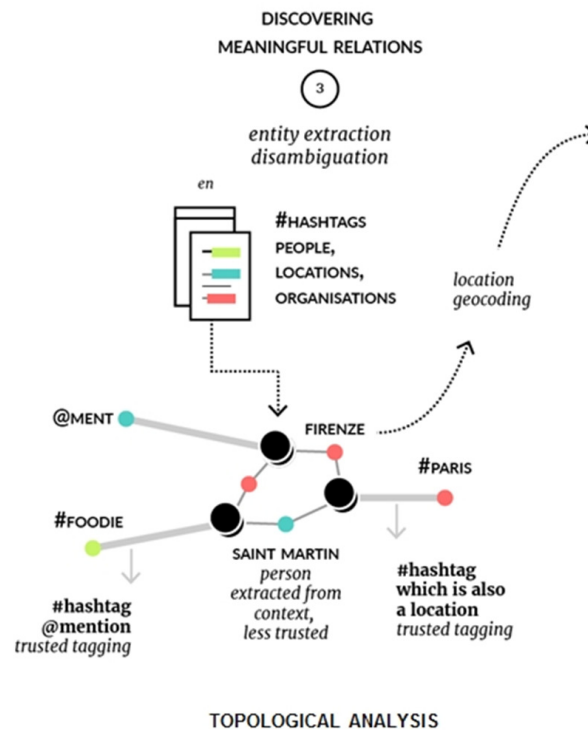


Figure 4: Topological analysis of entities to discover co-occurrence relationships.

produced a set of *location* names to add to the place where the original picture was taken. We created a new type of entity in the data model (see Table 1) called *locations* which was populated with these additional names, such as *Jerusalem* and *Camino di Santiago*. Conceptually the results were assigned into two categories: trusted tagging or less trusted tagging (see Figure 4). A trusted tagging of a place name was considered as one derived from processing the hashtags. The tags are *trusted* as they have been explicitly marked by the original Instagram poster and represent volunteered geographic information, for example #Toscana or the mention of *Camino di Santiago* within the picture title. *Less trusted tags* are considered those location names extracted as named entities from titles and captions. They are less trusted because they were not explicitly marked by the original poster so do not represent their explicit intention to symbolize geographic meaning, instead they are machine-generated tags.

For this prototype the trusted and less trusted places were handled together and not differentiated beyond the conceptual level. A future iteration of the tool might well introduce measures of uncertainty and translate this into a more nuanced visualization. Working with tags is complex and the development of a detailed model that deconstructs and validates all entities was beyond the scope of the project but it is certainly a future requirement. We were interested in the results coming directly out of the discovery script. These could be considered as “imagined places.” For those entities encoded as location names, the discovery script used geonames and Google Geocoding APIs to obtain the latitude and

longitude. If the two geo APIs were in agreement then the candidate entity was accepted and assigned a coordinate pair. In a future version of the tool we expect to streamline the process by using DBpedia Spotlight to look up the coordinates of the locations returned by TextRazor and only using the geo APIs for any locations that are missing.

Once the discovery script has completed its processing and the resources have been enriched, the co-occurrence script generates all the links between the entities, building a social network of co-occurrences. On the server side, this script builds the node-link relationship diagram using the Jaccard similarity measure. The result is a social graph of co-occurrences of people, themes and locations in our resources (see Figure 4). (geo)histograph then uses its internal histogram API service to serve all the data as JSON files to the client side and in doing so builds the resource gallery, the co-occurrence graph interface and the prototype occurrence mapping interface. For the mapping interface the service translates the JSON into a geoJSON of geometry type point. Mapbox was then used to dynamically create the density maps and serve the vector tiles for the base map. We used Leaflet JS on top to create a simple map interface that can be queried using qualitative filters. The timeline component was created using D3. The histogram API service includes a filtering function which is embedded into the mapping interface. This facilitates dynamic co-occurrence text based filtering according to (1) whether the map zoom or the spatial extent of the map window changes, and/or (2) whether the user selects one or more of the filter facets associated with a theme. If no facets are selected then the co-occurrence graph that was created on the server is used to generate the node-link diagram and the density map. If the filters are activated then the histogram API service calculates co-occurrences on the fly.

4.4 Description of the user interface

In addition to the resource gallery and the social network graph of co-occurrence, a number of new components were required. These included the development of: (i) a map window to visualize the occurrence of the entities, (ii) a timeline component to support the dynamic filtering of the occurrence map, (iii) a qualitative filtering panel for querying entities based on individual or multiple values, and (iv) the updating of the menu bar to facilitate switching between the linked map, gallery, and graph views.

Map component for occurrence visualization: The component calculates the number of instances of an individual or multiple entity and then uses interpolation to estimate the values across the area. For example, it can highlight concentrations of use of different themes such as *pilgrimage* or can be used to view imagined locations that are other places of pilgrimage through the location entity, e.g., *Lourdes*, *Rome*, or *Jerusalem*. The mapping results show raw counts of occurrence per cell, not normalized. Typical heat map cartography is used to highlight more occurrences as red hotspots. The maps are dynamic and make use of the Mapbox interpolation feature to create an appearance of depth to the heat map, the visualization changes its appearance based on the zoom level of the map, see Figure 5. A heat map was preferred over other forms of choropleth mapping to avoid imposing arbitrary administrative boundaries as the #hashtags have little underlying relationship to the social-cultural phenomena that is the cultural route. We must be mindful that using such administrative boundaries for this data visualization and modeling of the data entities might give rise to the ecological fallacy and lead to the misinterpretation and misrepresentation of the data.

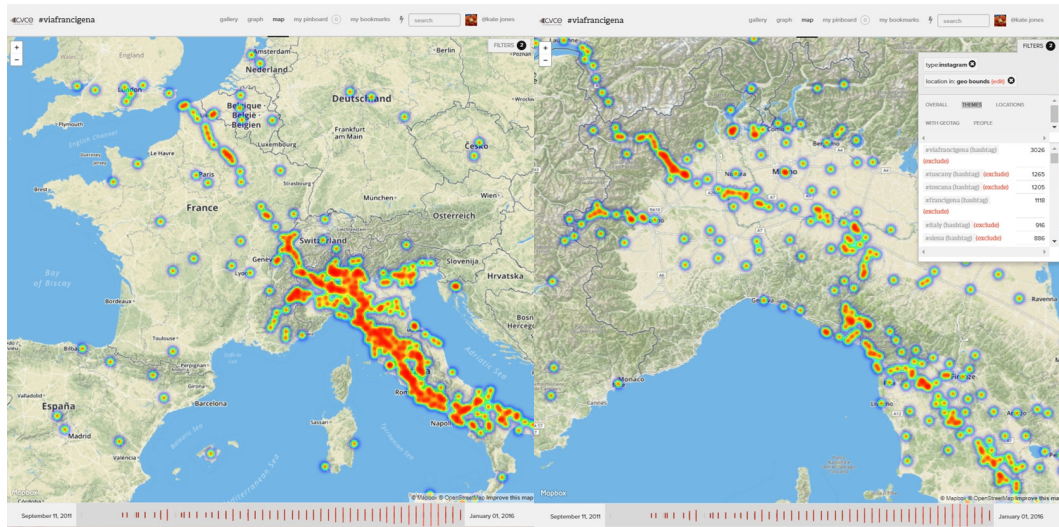


Figure 5: Screenshots of the linked map at different scales.

The map component performs as a spatial search, so the filtering component and timeline component are adjusted to show results according to the map extent and zoom level. The map view is linked to the graph and gallery views and serves to filter their display results (Figure 6). When a map is zoomed in on a particular area, the gallery view is updated automatically so when a user switches between the map and the gallery they see images only for the extent of the map window. The geographical extent of the map window are also recorded and used within the URL so that explorations can be stored and retrieved for later use.

Timeline component: With the timeline component the tool builds a chronological bar graph which is located on the horizontal axis at the bottom of the interface window. It extends full width and is embedded within the gallery and graph views. It shows the number of resources per day by the length of the vertical bars. The width of the vertical bar varies according to the number of entity instances per day. The start date and end date of the resources are noted on the far left and right respectively. By dragging a bounding box around the time period of interest users can dynamically filter the displayed results and recalculate the density map. It adjusts according to spatial filtering via the map window and qualitative filtering via the query panel.

Qualitative filtering component: This component enables users to query resources using qualitative facets permitting the filtering of entities that restrict the map display according to the selected entities (see Figure 7). When data are filtered, the density map and timeline are dynamically updated. So, if the theme of hiking and pilgrimage is selected, the map highlights the locations where the uses of these terms are concentrated. The following qualitative dimensions are available for complex query creation: (1) the list of all themes, (2) the resource type (e.g., Instagram or Twitter), (3) places where resources were

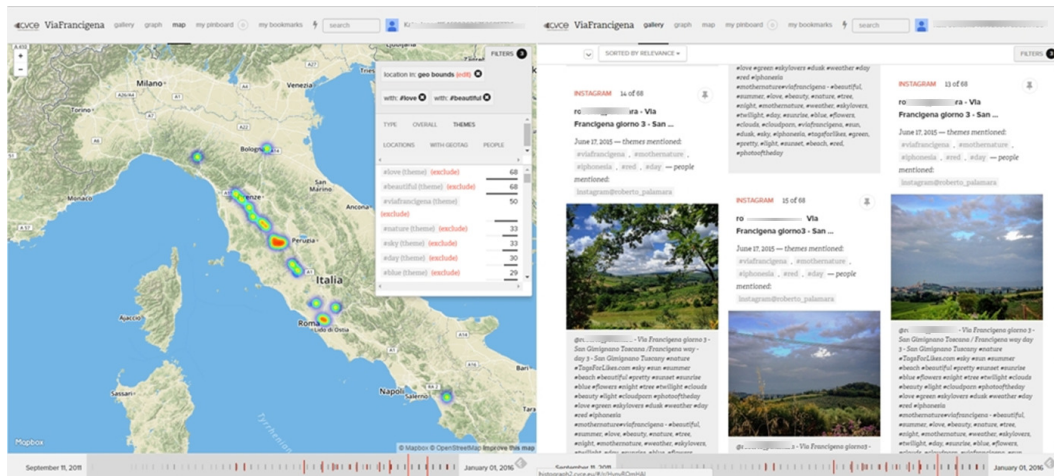


Figure 6: Screenshots of the linked map and gallery view with active filters for the themes of *love* and *beautiful*.

originally geotagged, (4) enriched locations identified from within the textual elements of the resources, and (5) people. It seamlessly filters the data and recalculates occurrences based on the geographical extent of the map window. In this tool, qualitative entities are the central component. They provide access to spatial and non-spatial concepts that are embedded in the resources. Qualitative filters are used consistently across all the views (gallery, map, and graph). Such interface mechanisms offer ease of use to non-experts and ensure that cognitive load is spent on exploring the data rather than on constructing queries and complex sub-queries. Users can dynamically refine the results by selecting or excluding the entities. If particular entity values are selected then the dates corresponding to the resources are highlighted in red on the timeline component.

In the filtering pane, a visual cue is provided to users showing the number of occurrences. On the right-hand side of the pane (see Figure 7, right), there is a dynamic horizontal bar whose length is adjusted according to the filters that are applied. If users explore the map without any filters, for example, the length of the gray bar corresponds to the total number of occurrences of all the themes in the dataset. For each individual theme the black bar shows the number of times it has been used, giving us an idea about the commonality of different themes. This is dynamic and adjusts if users change the map scale, apply a temporal filter (i.e., make a selection from the timeline), or include/exclude any entities. As users build up a set of complex sub-queries they can immediately explore the textual elements of the datasets.

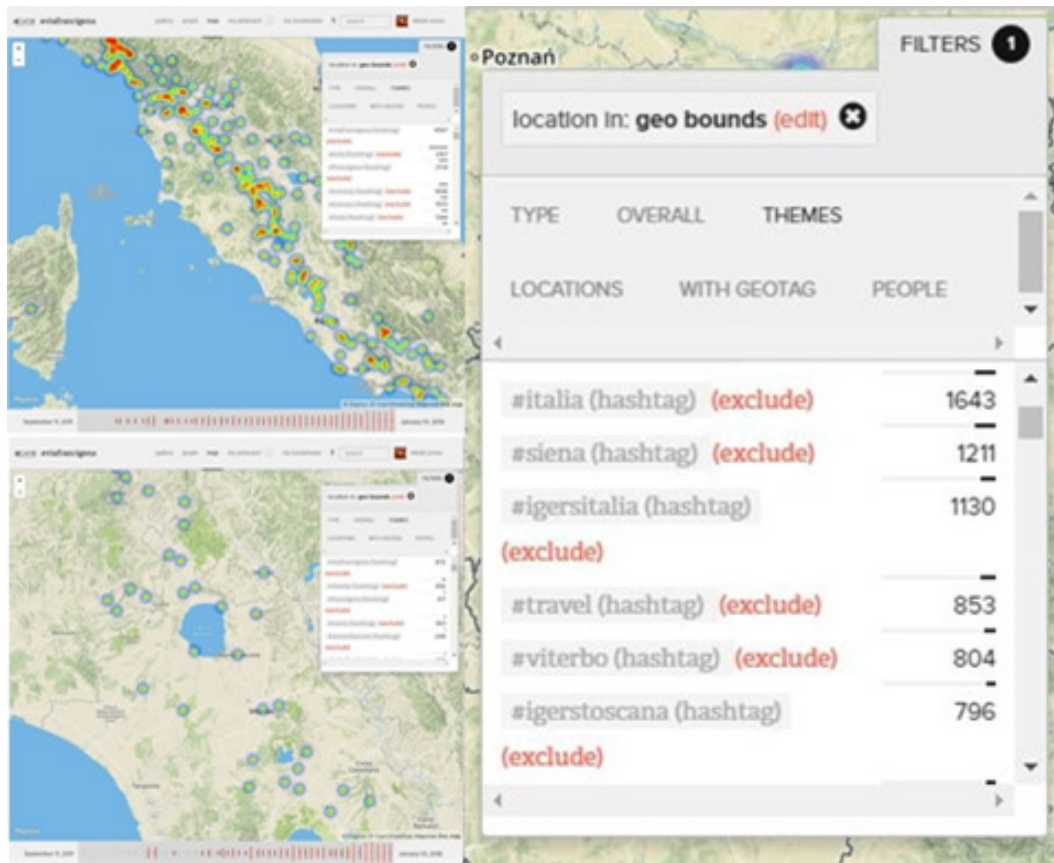


Figure 7: Screenshots showing dynamic recalculation of co-occurrence based on map scale (top and bottom left) and the qualitative filtering pane for themes (right).

Bookmarking component: As a query is developed, a unique dynamic URL is created that stores the query information including its spatial extent, meaning that queries can always be stored and retrieved¹.

5 Experimentation with Instagram via (geo)histograph

With the interface and back-end developed, the lead researcher used it to experiment, beginning with a detailed exploration of the data. There were 8,834 Instagram posts (including reposts), of which 2,702 had an original geotagged place at the time of posting. 30% of the resources therefore had an explicit spatial reference from the original user sharing their geotag. Of the two-thirds of resources without, the NERD processes for text mining

¹NB: this link only works internally at the University of Luxembourg and is shown as an example: <http://10.240.4.107:5000/#/geo?bounds=52.908902047770255;18.08349609375;37.84015683604136;-6.52587890625&type=instagram>

enriched the resources so that at least 4,001 resources had one location identified and 2,017 had two or more locations mentioned. A total of 6,018 implicit locations were identified from the enrichment process (see Table 2), indicating the added value of the text-mining procedures for geographical information retrieval for this dataset.

Type	Description	Total number
Dataset as a whole	Total number of Instagram posts	8,834
	Total number geotagged in original post (place entity)	2,702
	Total number not geotagged	6,132
Locations mentioned in geotagged posts	Geotagged Instagram posts with one location mentioned	2,849
	Geotagged Instagram posts having two or more locations mentioned	3,009
Post enrichment with NERD for location entity of non-geotagged posts	Total post enrichment with NERD for location entity of non-geotagged posts	6,018
	Non-geotagged Instagram posts having one location mentioned	4,001
	Non-geotagged Instagram posts having at least two locations mentioned	2,017

Table 2: Geotagging and location extraction results.

5.1 Spurious results, precision, and relevance

The tool helped us to explore the limitations in the dataset by visualizing the results of our discovery processes. As part of the co-creation process the in-depth exploration of the data using the prototype provided the researchers (lead users) with a hands-on approach to understanding the strengths and weaknesses of the workflow, data, and underlying processes. By using the interface to filter the data in map or gallery view and simultaneously exploring the co-occurrence of other related tags (by applying filters) it was possible to rapidly and dynamically compare subsets of entities at different socio-spatial scales through time. This allowed us to inspect the results of the human and machine processes that were undertaken to codify the data (see Table 3).

A number of issues, which can broadly be categorized as spatial or aspatial, arose from the machine encoding processes of entity extraction, the geocoding of entities or the human encoding processes. Spatial errors can be defined as: (1) spurious encoding of locations, (2) questionable relevance of trusted tagged locations, (3) limitations from scale of locations encoded, (4) assigning points for more complex spatial phenomena, and (5) geocoding duplication of location entities. Aspatial limitations can either be derived from user generated ambiguities or machine generated error. They include (1) relevance of themes and the long tail of tags, (2) limitations of tags as they are not true texts, and (3) mis-assignment of themes to people entities.

5.1.1 Limitations of the spatially encoded place and location data

Spurious encoding of locations—people to locations or themes to locations

Spurious location results from automatic enrichment led to geocoding errors (see Figure 8). The world map shows how resources that were attributed to unexpected places such as the

Source of error	Ambiguity/Error category	Description
Machine	Spatial	Spurious encoding of locations—people to locations or themes to locations
Machine		Geocoding duplication of location entities
Machine/human		Limitations from scale of locations encoded
Human		Questionable relevance of trusted tagged locations
Human		Assigning points for more complex spatial phenomena
Human	Aspatial—user generated	Relevance of themes and the long tail of tags
Human	Aspatial—user generated	Limitations of tags as they are not true texts
Human	Aspatial—machine generated	Mis-assignment of entities (e.g., people to themes, etc.)

Table 3: Types of error observed.

Americas, Africa, and beyond. Some of these errors arose because the text mining process assigned people such as saints to locations. For example, San Francisco was assigned to the city in North America when it should have been assigned to a convent along the route named after Saint Francis of Assisi. Ambiguity between alike person names and location names led to many mismatches, without inspecting individually every resource it is difficult to measure the true extent of the issue. In future we would use a retrieval process that includes context-based geocoding or topic-based refinement.

Duplication of location entities—geoambiguity

We also observed a phenomenon that seemed rather peculiar until closer inspection. The same thirteen resources with the same location entity value of “Dover” explicitly marked in the posts as #Dover were assigned to both Dover, Kent in the United Kingdom and Dover, Delaware State in the US. The prototype permitted exploration of this issue by (i) exploring the resource gallery (see Figure 9) and (ii) creating the co-occurrence graph between locations (implicit: imagined) and places (explicit: geotagged). In terms of machine-based geocoding the result is not strictly incorrect; the ambiguity arises when discrete places have the same name [5]. In a future prototype we would incorporate context-based information retrieval together with a front-end interface to enable validation of the results via annotation methods across multiple resources.

Questionable relevance and meaning of trusted tagged locations

The imagined locations led to questionable results because their relevance and meaning of the location was difficult to interpret. By way of an example, location entities marked within Tokyo in Japan might first seem like a geocoding or an enrichment error but the graph of co-occurrence between locations and persons filtered by Tokyo highlights that it is one individual mentioning many locations (London, Russia, Bolsena, Tokyo, etc.) (Figure 10). It is a valid *location* but one that is not necessarily relevant. The spurious results likely derive from the cultural production of tags [4] giving rise to this geographical and

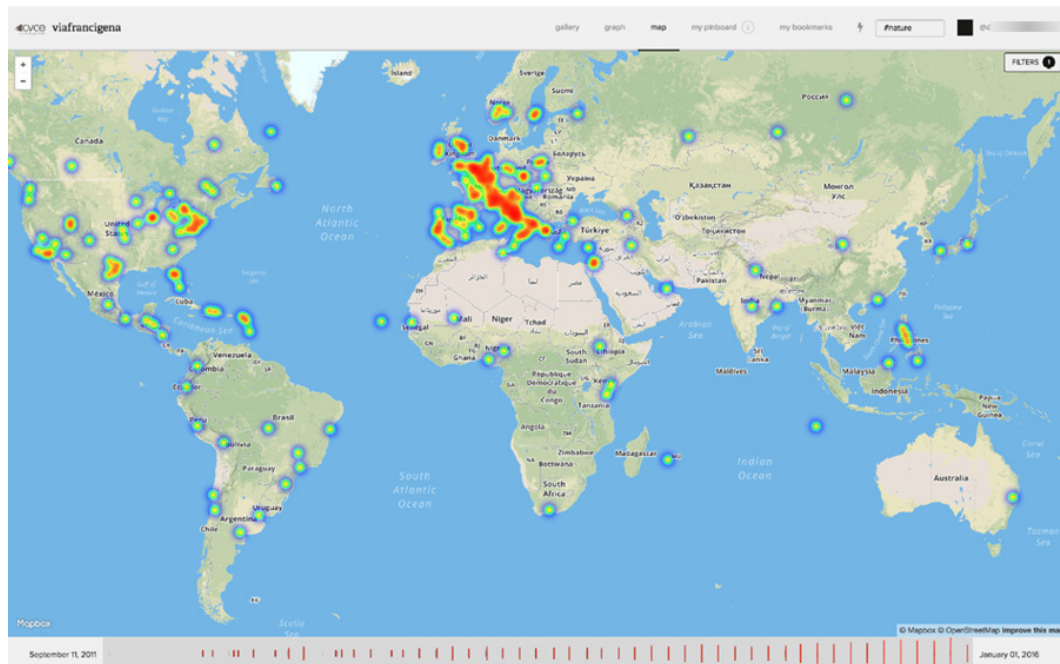


Figure 8: Spurious locations identified in our resources.

interpretive ambiguity. Should these locations be considered relevant if the relationship is unknown and their meaning is unclear? Future versions need an improved strategy for managing *irrelevant* entity values, which could be calculated using measures from the co-occurrence graph.

Limitations from scale of locations encoded—spatialization or assigning points to represent more complex spatial phenomena

We did not carry out further cleaning or structuring of the geocoding results, processes that are generally recommended in geographical information retrieval literature [32], due to resource and time constraints and the exploratory nature of the project. Therefore, all locations and place entities were treated and visualized in the same way. Consequently, we have a representation that encompasses different granularities of scale, e.g., *France* or *Switzerland* versus *Canterbury* or *Rome* or *Kent*. This must be addressed in the future by creating different visualization forms for the different geographic scales and new forms of cartography for the phenomena being represented (e.g., buildings, points of interest, mountain ranges, etc.) [10,31,32].

5.1.2 User-generated tags and their limitations

Relevance of themes, the long tail of tags and tags as texts

There was a long list of entity values for the themes produced from the tags, taking the

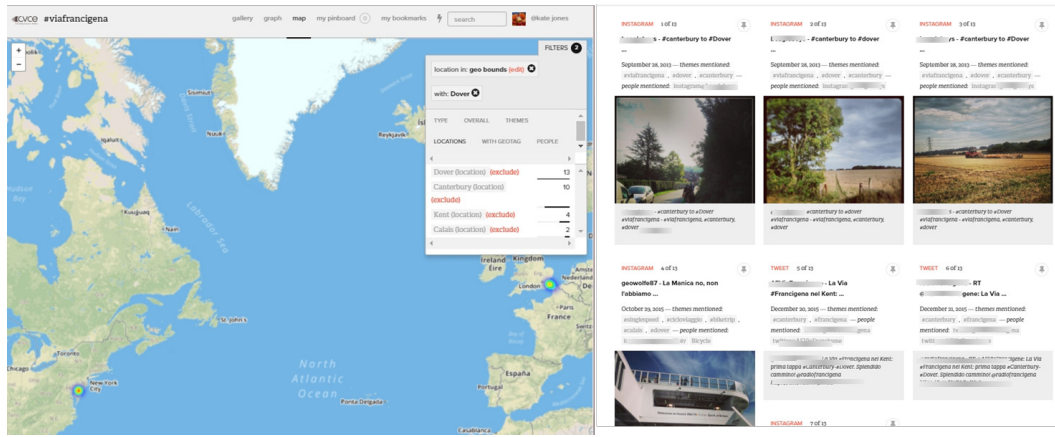


Figure 9: Gallery view of resources marked with location entity *Dover* view alongside map of geocoded location entities known as *Dover*.

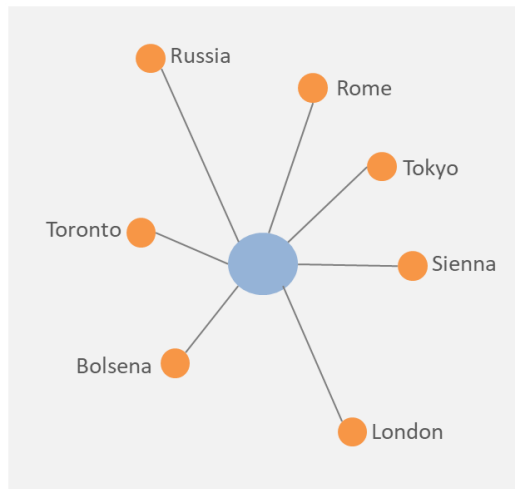


Figure 10: Graph of location (orange nodes) to person (blue node) filtered by Tokyo.

shape of a long tail. Most users used only the top five—*#viafrancigena*, *#francigena*, *#tus-cany*, *#toscana*, *#italy* (which are very broad)—or tags more closely associated with the production of Instagram likes, e.g., *instagood* (3%), *tagsforlikes* (1%), *picoftheday* (4%) and *photooftheday* (2%) (see Figure 11). The result is a form of user-generated bias [26,31], which skews the results and tells us more about the social practices of users interacting with a specific technology than people’s perceptions of place. It would be helpful to classify these or create an entity called social media cultural practices, which could then be easily excluded/included in analysis depending on the form of the research questions.

@ [redacted] - #iphone #italia #iphone4 #igmasters #instagood #instamood #instadaily #iphoneonly #igersitalia #instatalent #italianlife #instacollage #iphoneography #italianeography #me #mobilphoto #master_pics #popular #photo #people #picoftheday #amici #friend #viafrancigena #pellegrini - #pellegrini, #people, #photo, #italianeography, #iphoneography, #amici, #instadaily, #instamood, #me, #viafrancigena, #igmasters, #igersitalia, #iphone4, #iphoneonly, #instatalent, #mobilphoto, #instagood, #instacollage, #italia, #iphone, #master_pics, #popular, #picoftheday, #friend, #italianlife

@ [redacted] - Ha detto che questo mondo tanto orribile a lei sembrava solo addormentato. E che a risvegliarlo non sarebbe stata la logica dei sapienti, ma l'energia degli innamorati: gli unici ancora capaci di coniugare i verbi al futuro. Massimo Gramellini #tramonto #sunset #sunsetlovers #sunsetporn #lagodiviverone #summermemories #viafrancigena #sky #skylovers #skypainters #skyporn #cloudporn #cielomania #cielo #amoremio #tweegram #love #us #igers #latergram #like4like #❤️ - #love, #skyporn, #summermemories, #sunsetporn, #tramonto, #amoremio, #igers, #skylovers, #cielomania, #❤️, #cielo, #cloudporn, #lagodiviverone, #viafrancigena, #sunsetlovers, #sky, #us, #skypainters, #sunset, #latergram, #like4like, #tweegram

Figure 11: Example of cultural practices of tags for likes.

During the processing of the tags they were treated as texts, even if they are not always *real words*, e.g., “skylovers” or “sunsetlovers.” Whilst these terms tell us something about the perceptions of the materiality of the place, they are difficult to process with the discovery scripts. Thus we need a better way of handling such tags. Also, because we have no underlying semantic ontology or thesaurus, the values contained duplicates were very messy and created a lot of noise. For future iterations of the tool there is a need to develop a process to clean, structure and classify these themes. Furthermore, it would be helpful to build upon the work of semantic classification of places to include both the vernacular (local terms) and official names of places (perhaps by creating emergent gazetteers [3]) as well as analyze and enrich both the text and indeed the images with information about the geographical footprints of places [3, 29, 32, 33]. In so doing we would expect to have a cleaner and improved structure for the tags (themes) which should mean they will be more useful for exploring the nuances of the locale.

6 Testing the validity of the (geo)histograph interface—the Via Francigena use case

The interface provided to the co-design team enabled them to explore with relative ease the Instagram resources related to the Via Francigena cultural route. It served as a lens through which the team could become accustomed with a new dataset. It gave us the tools to ask open-ended questions of the data from a qualitative perspective and helped us understand the relevance and reliability of the enrichment processes and the general usefulness of Instagram as a data source. The next stage was to discover if we could demonstrate the capability of the prototype and determine if it could support our user scenario. Using the tools we were able to explore what people choose to memorialize as personal representations captured via their phone—albeit as technologically-mediated representations subject to underlying cultural practices that lead to particular perspectives [20, 30].



We started by considering the terms related to the traditional use of the route as a path of pilgrimage. We first filtered the dataset using the term *pilgrim*, creating a co-occurrence graph that connects locations (places mentioned in the textual elements) to themes. We observed that the theme *pilgrim* co-occurs with other locations of pilgrimage such as Camino and Jerusalem. Connections were revealed between pilgrim and the places of *Santiago*, *Assisi*, and *Umbria* via the practice of walking. Terms such as *fountains*, *lights*, and *city* provided a sense of the materiality that connects people with places. The places also evoked a sense of *beauty*, *love*, and *enjoy[ment]*. The map shows concentrations of these entities on the route between Switzerland and Rome.

Next we searched for the presence of the person entity *Sigeric the Serious* (see Figures 12 and 13), a historically significant figure for the Via Francigena, with a view to finding out more about people's engagement with historical figures associated with the route. A small set of co-occurrences and mentions (approximately 40) were returned connected to meaning and practice such as *pellegrino* (pilgrim), *walk*, and *trekking*. There were only a few references to the materiality of the place such as *nature* and *sun*. All were distributed along the stretch between Siena and Milan, the most popular part of the route. The gallery view reveals users capture the material structures of the route for this theme through a predominance of images of historical buildings or symbols of the route, perhaps demonstrating an interest in stopping and exploring the cultural heritage along the route.

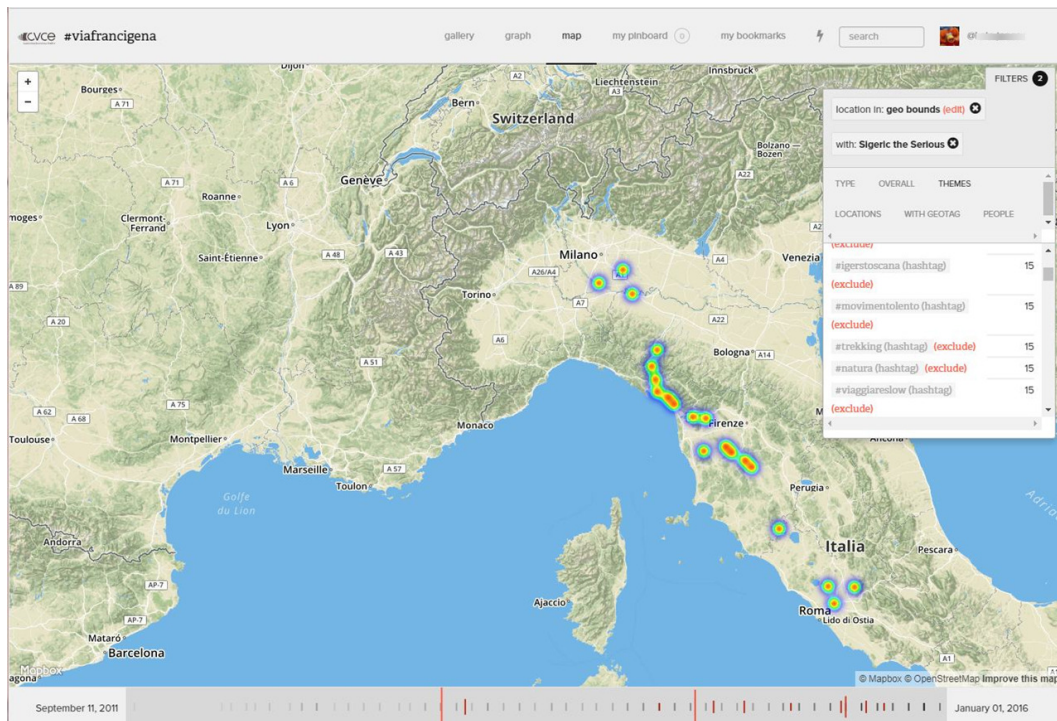


Figure 12: Spatial distribution of the person entity *Sigeric the Serious*.

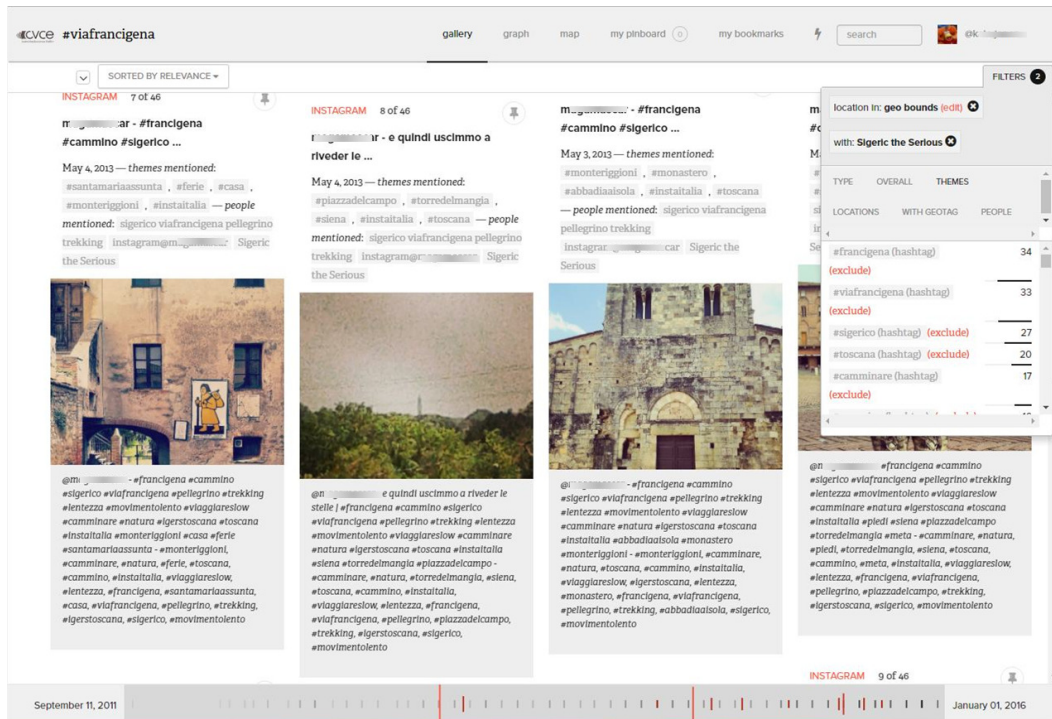


Figure 13: Gallery view of posts associated with person entity *Sigeric the Serious*.

The Via Francigena is a “path,” and with the tool we can visualize its coordinates as a linear route joining Canterbury to Rome and see the material representation of the route through its memorialization. Historically the route embodied practices of pilgrimage associated with the meaning of the Christian religion, but more recently there has been a renewed interest in cultural routes for other reasons linked to culture, tourism, wellness, the environment, or spirituality [10]. The tool enables us to build complex queries that help us to understand who uses the route, when they are using it and how, while providing insights into what people perceive as important to the sense of place that they have constructed [12,31].

7 Discussion and conclusion

With the tool it was possible to conduct open-ended queries and free-form exploration—in line with the goals of deep mapping interfaces [8,34,43] and calls for different types of interfaces [16,28]. The qualitative querying aided foraging using the textual elements and the map helped us make sense of the different relationships and patterns within the resources, although meaning-making was hampered by a lack of structure and classification of the entities. A deeper, more structured classification for the qualitative facets—using either a predefined set of semantics [32] or one that is emergent [3]—is needed. The text

mining should also be extended to identify entities that are associated with the physicality of places through geographical footprints such as bridges, mountains, thermal water, etc. [1,31], and incorporate consideration for the local vernacular expressions of place. The tool enabled us to develop queries and ask questions about who, what, where, and when from different perspectives. The ability to drill down to examine the relevance and validity of the text processing results helped our inquiry processes. We could examine and question the original data in detail, first by inspecting an overview of the dataset, then by zooming in and drilling deeper down to discover more detail. This process implicitly reproduces Shneiderman's rules of information-seeking [41] and meets one of the expectations of deep mapping interfaces [38].

The advantages of this type of framework are its adaptability and its potential to integrate implicit and explicit geographic information. In a short space of time and with minimal resources it was possible to add new functionalities and adapt an existing data model that was originally built for historical documents. We were able to extend it to integrate less formal social media data to help us get to grips with this new type of data. The user-led, co-design approach enabled us to gather invaluable hands-on experience to identify future requirements of both the geovisualization interface and its cartography and the design of the geographic data. These included: (1) the need to develop processing and visualization methods to ascertain relevance, precision, and uncertainty in location- and place-based entities; (2) the need for improvements to the geographical information retrieval process through context-based coding and use of semantic modeling and classification; and (3) the need to develop alternative methods for visualizing the data according to different geographical scales and types of open-ended exploration, so that occurrence maps reflect different forms of research questions (for example, direction flow mapping to determine paths used or text maps for highlighting themes). There remains a wealth of implicit geographic information locked within the pictures that, with the emergence of APIs for image analysis (such as Vision API or Image Recognition API), is ripe for retrieval.

These new tools have vast reuse potential, reducing the costs involved in developing new products as part of a parsimonious and pragmatic approach. Co-design processes and adaptable technologies provide opportunities for rapid experimentation and enable new ways of thinking about geo-technology development: users are given the role of participants in design, and technology prototypes themselves become mechanisms for reflecting on needs and requirements. This process also allows the visualization and exploration of unknown/new datasets to understand the limitations and advantages of the tool and assess its general fitness for purpose with respect to our use case. As the histogram tool was originally designed to work with qualitative resources, it provides a counterbalance to the more traditional GIS and web-based quantitative mapping. The prototyping process provides a framework for critical and creative thinking through collaboration.

By bringing users into contact with designers, we were able to develop an interface through end-to-end participation. We started with the user scenario, then carried out an independent exploration of the datasets and the encoded geographical information using visualization methods, which led to the design of the geovisualization tool and finally to our experimentation with the interface and the data. We explored how an existing humanities tool could be adapted to suit the needs of place-based exploration for understanding users of a cultural route, by integrating explicit and implicit geographical information contained in Instagram posts. Through co-design and rapid collaborative prototyping tools

we explored how the design process for geographic interfaces can be enriched by drawing on user participation to better source iterative requirements.

The development of the prototype served four purposes. First, it provided a co-design workflow that fostered user value for geographic interfaces. Second, the process of rapid prototyping including tool adaptation and reuse was a method for creatively thinking through development and design decisions for thoughtful iterative requirement gathering. Third, the use of technology prototyping provided a pragmatic solution that enabled lead users to deconstruct the results of automatic text processing of geographical information so as to understand the limitations and implications of placed-based meaning-making and explore its value. Fourth and finally, the use of prototyping and co-creation provided a means for understanding the strengths and weaknesses of unfamiliar Instagram data, ascertaining its fitness for purpose and identifying the benefits and limits of its geographic visualization. Overall participation in the co-design process proved to enhance understanding and reflection and to improve requirement gathering and iterative development for geovisualization interfaces. Such an approach contributes to the development of new interdisciplinary open-ended mapping interfaces suitable for place-based research and we hope provides a foundation for the deeper humanistic/heritage-driven exploration.

Acknowledgments

The case study presented in this paper has been carried out within the research project Digital Itineraries, funded by the PEPS University of Lille/CNRS et by the FR CIST. The geo-component of this tool builds on the original histogram tool which was developed at the former CVCE Digital Humanities Lab (now part of the C²DH, University of Luxembourg) by Lars Wieneke, Daniele Guido, and Marten Düring. histogram builds on a demonstrator application developed for the EU-funded collaborative project CUBRIK (grant agreement number 287704, 2011-14).

References

- [1] ACHESON, E., WARTMANN, F. M., AND PURVES, R. S. Generating Spatial Footprints from Hiking Blogs. In *Proc. of Workshops and Posters at the 13th International Conference on Spatial Information Theory (COSIT 2017)* (2017), Lecture Notes in Geoinformation and Cartography, Springer, Cham, pp. 5–7. doi:10.1007/978-3-319-63946-8_2.
- [2] ACHREKAR, H., GANDHE, A., LAZARUS, R., YU, S.-H., AND LIU, B. Predicting Flu Trends using Twitter data. In *2011 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)* (2011), pp. 702–707. doi:10.1109/INFOCOMW.2011.5928903.
- [3] ADAMS, B., MCKENZIE, G., AND GAHEGAN, M. Frankenplace: Interactive Thematic Mapping for Ad Hoc Exploratory Search. In *Proc. of the 24th International Conference on World Wide Web* (2015), WWW '15, International World Wide Web Conferences Steering Committee, pp. 12–22. doi:10.1145/2736277.2741137.
- [4] AMES, M., AND NAAMAN, M. Why We Tag: Motivations for Annotation in Mobile and Online Media. In *Proc. of the SIGCHI Conference on Human Factors*

- in Computing Systems* (New York, NY, USA, 2007), CHI '07, ACM, pp. 971–980. doi:10.1145/1240624.1240772.
- [5] AMITAY, E., HAR'EL, N., SIVAN, R., AND SOFFER, A. Web-a-where: Geotagging Web Content. In *Proc. of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (New York, NY, USA, 2004), SIGIR '04, ACM, pp. 273–280. doi:10.1145/1008992.1009040.
- [6] BAMPATZIA, S., ANTONIOU, A., LEPOURAS, G., VASILAKIS, C., AND WALLACE, M. Using social media to stimulate history reflection in cultural heritage. In *2016 11th International Workshop on Semantic and Social Media Adaptation and Personalization (SMAP)* (2016), pp. 89–92. doi:10.1109/SMAP.2016.7753390.
- [7] BELLENS, A., BANC, N. V. L., ELOIRE, F., GRABAR, N., KERGOSIEN, E., AND SEVERO, M. Social media and European cultural routes: Instagram networks on the Via Francigena. In *Proc. of the 8th International Conference on Management of Digital EcoSystems* (2016), ACM, pp. 122–128. doi:10.1145/3012071.3012082.
- [8] BODENHAMER, D., CORRIGAN, J., AND HARRIS, T., Eds. *The Spatial Humanities: GIS and the Future of Humanities Scholarship*. Indiana University Press, Indiana, 2010.
- [9] BORA, E. Instagram Scraper. <https://tools.digitalmethods.net/beta/instagram/>, 2015.
- [10] BRUGGMANN, A., AND FABRIKANT, I., S. How does GIScience support spatio-temporal information search in the humanities? *Spatial Cognition & Computation* 16, 4 (2016), 1542–7633. doi:10.1080/13875868.2016.1157881.
- [11] CAPINERI, C., HAKLAY, M., HUANG, H., ANTONIOU, V., KETTUNEN, J., OSTERMANN, F., AND PURVES, R. *European Handbook of Crowdsourced Geographic Information*. Ubiquity Press, London, 2016.
- [12] CRESSWELL, T. *Place: A Short Introduction*. Wiley, Oxford, UK, 2004.
- [13] CROOKS, A., CROITORU, A., STEFANIDIS, A., AND RADZIKOWSKI, J. Earthquake: Twitter as a Distributed Sensor System. *Transactions in GIS* 17, 1 (2013), 124–147. doi:10.1111/j.1467-9671.2012.01359.x.
- [14] DAVIES, C., AND MEDYCKYJ-SCOTT, D. GIS users observed. *International Journal of Geographical Information Systems* 10, 4 (1996), 363–384. doi:10.1080/02693799608902085.
- [15] DOMÍNGUEZ, D. R., DÍAZ REDONDO, R. P., VILAS, A. F., AND KHALIFA, M. B. Sensing the city with Instagram: Clustering geolocated data for outlier detection. *Expert Systems with Applications* 78 (2017), 319–333. doi:10.1016/j.eswa.2017.02.018.
- [16] DRUCKER, J. Reading Interface. *PMLA* 128, 1 (2013), 213–220. doi:10.1632/pmla.2013.128.1.213.
- [17] DÜRING, M., WIENEKE, L., AND CROCE, V. Interactive Networks for Digital Cultural Heritage Collections - Scoping the Future of HistoGraph. In *Engineering the Web in the Big Data Era*, P. Cimiano, F. Frasincar, G.-J. Houben, and D. Schwabe, Eds., no. 9114 in Lecture Notes in Computer Science. Springer International Publishing, 2015, pp. 613–616. doi:10.1007/978-3-319-19890-3_41.

- [18] ELETA, I., AND GOLBECK, J. Bridging languages in social networks: How multilingual users of Twitter connect language communities? *Proc. of the American Society for Information Science and Technology* 49, 1 (2012), 1–4. doi:10.1002/meet.14504901327.
- [19] ELWOOD, S. Geographic Information Science: new geovisualization technologies—emerging questions and linkages with GIScience research. *Progress in Human Geography* 33, 2 (2009), 256–263. doi:10.1177/0309132508094076.
- [20] ELWOOD, S., AND MITCHELL, K. Technology, memory, and collective knowing. *Cultural Geographies* 22, 1 (2015), 147–154. doi:10.1177/1474474014556062.
- [21] GOODCHILD, M. F. Citizens as voluntary sensors: spatial data infrastructure in the world of Web 2.0. *International Journal of Spatial Data Infrastructures Research* (2007), 24–32. doi:10.1.1.162.2017.
- [22] GUERRERO, P., MOLLER, M. S., OLAFSSON, A. S., AND SNIZEK, B. Revealing Cultural Ecosystem Services through Instagram Images: The Potential of Social Media Volunteered Geographic Information for Urban Green Infrastructure Planning and Governance. *Urban Planning* 1, 2 (2016), 1–17. doi:10.17645/up.v1i2.609.
- [23] HAKLAY, M., Ed. *Interacting with Geospatial Technologies*. John Wiley & Sons, West Sussex, 2010.
- [24] HARRIS, T. Deep Geography—Deep Mapping: Spatial Storytelling and a Sense of Place. In *Deep Maps and Spatial Narratives*, D. Bodenhamer, J. Corrigan, and T. Harris, Eds. Indiana University Press, Bloomington, Indiana, 2015, pp. 28–53. doi:10.1111/j.1467-9787.2012.00780_3.x.
- [25] HARRIS, T., BERGERON, S., ROUSE, J., DEAR, M., KETCHUM, J., LURIA, S., AND RICHARDSON, D. Humanities GIS. In *GeoHumanities: Art, History, Text at the Edge of Place*. Routledge, Oxon, UK, 2011, pp. 226–240.
- [26] HECHT, B. J., AND STEPHENS, M. A Tale of Cities: Urban Biases in Volunteered Geographic Information. In *Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media* (North America, 2014), vol. 14, pp. 197–205.
- [27] HOLLENSTEIN, L., AND PURVES, R. Exploring place through user-generated content: Using Flickr tags to describe city cores. *Journal of Spatial Information Science* 2010, 1 (2010), 21–48. doi:10.5311/JOSIS.2010.1.13.
- [28] MACEACHREN, A. M. Leveraging Big (Geo) Data with (Geo) Visual Analytics: Place as the Next Frontier. In *Spatial Data Handling in Big Data Era, Advances in Geographic Information Science*. Springer, Singapore, 2017, pp. 139–155. doi:10.1007/978-981-10-4424-3_10.
- [29] MCKENZIE, G., JANOWICZ, K., GAO, S., YANG, J.-A., AND HU, Y. POI pulse: A multi-granular, semantic signature-based information observatory for the interactive visualization of big geosocial data. *Cartographica: The International Journal for Geographic Information and Geovisualization* 50, 2 (2015), 71–85. doi:10.3138/cart.50.2.2662.

- [30] MENDELSON, A. L., AND PAPACHARISSI, Z. Look at us: Collective narcissism in college student facebook photo galleries. *The networked self: Identity, community and culture on social network sites 1974* (2010), 1–37.
- [31] PURVES, R., EDUARDES, A., AND WOOD, J. Describing place through user generated content. *First Monday* 16, 9 (2011). doi:10.5210/fm.v16i9.3710.
- [32] PURVES, R. S., AND DERUNGS, C. From Space to Place: Place-Based Explorations of Text. *International Journal of Humanities and Arts Computing* 9, 1 (2015), 74–94. doi:10.3366/ijhac.2015.0139.
- [33] RATTENBURY, T., AND NAAMAN, M. Methods for Extracting Place Semantics from Flickr Tags. *ACM Trans. Web* (2009), 1:1–1:30. doi:10.1145/1462148.1462149.
- [34] RIDGE, M., LAFRENIERE, D., AND NESBIT, S. Creating deep maps and spatial narratives through design. *International Journal of Humanities and Arts Computing* 7, 1-2 (2013), 176–189. doi:10.3366/ijhac.2013.0088.
- [35] ROBINSON, A. C., DEMŠAR, U., MOORE, A. B., BUCKLEY, A., JIANG, B., FIELD, K., KRAAK, M.-J., CAMBOIM, S. P., AND SLUTER, C. R. Geospatial big data and cartography: research challenges and opportunities for making maps that matter. *International Journal of Cartography* 3, sup1 (2017), 32–60. doi:10.1080/23729333.2016.1278151.
- [36] ROTH, R. E., ÇÖLTEKIN, A., DELAZARI, L., FILHO, H. F., GRIFFIN, A., HALL, A., KORPI, J., LOKKA, I., MENDONÇA, A., OOMS, K., ET AL. User studies in cartography: opportunities for empirical research on interactive maps and visualizations. *International Journal of Cartography* 3, sup1 (2017), 61–89.
- [37] ROTH, R. E., ROSS, K. S., AND MAC EACHREN, A. M. User-Centered Design for Interactive Maps: A Case Study in Crime Analysis. *ISPRS International Journal of Geo-Information* 4, 1 (2015), 262–301. doi:10.3390/ijgi4010262.
- [38] SANDERS, E. B.-N., AND STAPPERS, P. J. Co-creation and the new landscapes of design. *CoDesign* 4, 1 (2008), 5–18. doi:10.1080/15710880701875068.
- [39] SCHOFIELD, T., WHITELAW, M., AND KIRK, D. Research through design and digital humanities in practice: What, how and who in an archive research project. *Digital Scholarship in the Humanities* 32, suppl_1 (2017), i103–i120. doi:10.1093/llc/fqx005.
- [40] SCHWARTZ, R., AND HALEGOUA, G. R. The spatial self: Location-based identity performance on social media. *New media & society* 17, 10 (2015), 1643–1660.
- [41] SHNEIDERMAN, B. The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. In *The Craft of Information Visualization*, Interactive Technologies. Morgan Kaufmann, San Francisco, 2003, pp. 364–371. doi:10.1016/B978-155860915-0/50046-9.
- [42] SILVA, T. H., MELO, P. O. S. V. D., ALMEIDA, J. M., SALLES, J., AND LOUREIRO, A. A. F. A Picture of Instagram is Worth More Than a Thousand Words: Workload Characterization and Application. In *2013 IEEE International Conference on Distributed Computing in Sensor Systems* (2013), pp. 123–132. doi:10.1109/DCOSS.2013.59.

- [43] TAYLOR, J. E., DONALDSON, C. E., GREGORY, I. N., AND BUTLER, J. O. Mapping Digitally, Mapping Deep: Exploring Digital Literary Geographies. *Literary Geographies* 4, 1 (2018), 10–19.
- [44] VUKMIROVIC, M., AND LAZAREVIC, E. V. Place competitiveness expressed through digital data—measuring the place attractiveness tracking the geotag data visuals. In *Places and technologies* (2014), pp. 914–925.
- [45] WEILENMANN, A., HILLMAN, T., AND JUNGSELIUS, B. Instagram at the Museum: Communicating the Museum Experience Through Social Photo Sharing. In *Proc. of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2013), CHI '13, ACM, pp. 1843–1852. doi:10.1145/2470654.2466243.
- [46] WIDENER, M. J., AND LI, W. Using geolocated Twitter data to monitor the prevalence of healthy and unhealthy food references across the US. *Applied Geography* 54 (2014), 189–197. doi:10.1016/j.apgeog.2014.07.017.

