

IN PROGRESS

An open source spatial news web app development project
Masters in GIS&S Thesis

Callie Wentling

Project advisor: Professor Doutor Marco Painho, Ph.D

January 5, 2021

Project summary

This project seeks to develop a proof-of-concept (POC) open source (OS) and free software (FS) web application (Web App) that supports the visualization of the spatial distribution of news story contents (“incidents”), as well as filtering mechanisms for improved temporal, spatial, and thematic investigation of news articles. This geospatial element is expected to provide an additional dimension of understanding that allows users to better contextualize news stories, search repositories, or monitor spatial/temporal trends at a community level (within a city). In addition to the aforementioned improvements of user experience for the public (readers, researchers, and monitors), it is also expected to support publishers via the inference of new insights from their existing internal data, such as the illumination of under- or over-reporting of areas by theme for better investigative coverage. Ideally, this functionality could be expanded to integrate multiple sources, as well as the incorporation of planned events and/or resources to provide a more comprehensive understanding of one’s surroundings in both the planned future and transpired past.

Contents

1	Introduction	1
2	Literature review	1
2.1	Citizen empowerment	1
2.1.1	Smart communities	1
2.1.2	Communication	3
2.1.3	Data visualization	3
2.1.4	Public participation	6
2.2	Spatial Information	9
2.2.1	Location	10
2.2.2	Place	10
2.2.3	GeoIntelligence	13
2.2.4	Cross border mapping	14
2.2.5	Feature types	14
2.2.6	Geoportals/mashups	16
2.2.7	Commercial geospatial platforms	18
2.2.8	Location extraction from text	21
2.2.9	Event extraction	24
2.2.10	Gazetteers	28
2.2.11	Local lexicons	29
2.3	News localization	30
2.3.1	News evolution	30
2.3.2	Personalization	31
2.3.3	Blogs	31
2.3.4	Indie journalism	31
2.3.5	Local journalism	31
2.3.6	Georeferenced News	32
2.3.7	Automated location extraction from published corpora	33
2.3.8	Examples	35
2.3.9	Manual location extraction from published corpora	38
2.3.10	Location assignment	40
2.4	Web applications	40
2.4.1	Design	40
2.4.2	Open Source	42
2.4.3	Data	42
2.4.4	GIS design	44
2.4.5	Web services	44
2.4.6	Graphic user interface	45
2.4.7	Data manipulation	47
2.4.8	Security	48
2.4.9	Products	48
2.4.10	Dashboard design	49
2.4.11	Internationalization and localization	49
2.4.12	Examples	49
3	Intervention Description	50
3.1	Products	50
3.2	Distinction from previous work	51
3.3	Public participation	51
3.4	Nuanced “place”	52
3.5	User experience	52
3.6	GUI	53
3.7	Knowledge facilitation	53
3.8	Features	53

3.9	Evolving gazetteer	54
3.10	Application of results	54
3.11	Structure	54
3.12	Languages	55
3.13	Sustainability	55
3.14	Impact	56
3.15	Concerns	56
4	Methodology	56
4.1	Study area: Portugal	56
4.1.1	Demographics	56
4.1.2	Local news serving Lisbon	56
4.1.3	Geo tools	57
4.2	Model design	57
4.3	GIS design	57
4.4	Dataprep	57
4.5	Architecture	58
4.6	Web Interface	58
4.7	Front end	58
4.8	Backend	58
4.9	Manipulation	58
4.10	Data Collection	59
4.10.1	Corpora	59
4.10.2	Basemaps	60
4.10.3	Gazetteers	60
4.11	Preprocessing	62
4.12	Initialization of development environment	62
4.13	Tool design	62
4.14	Testing	63
4.15	Validation	63
5	Results	63
6	Analysis	63
6.1	Future opportunities	63
6.1.1	Validation	63
6.1.2	Application to GeoIntelligence	63
6.1.3	Sub-article definitions	63
6.1.4	Recommended extraction of place	64
6.1.5	Incorporation of historical stories	64
6.1.6	Statistical analysis	64
7	Conclusion	64
	References	65
A	Relevant terms	68
B	Data visualization	72
C	Preliminary specification	75
D	Example GUIs	78
E	Example Architectures	81
F	User stories	83

F.1	Micronews Portal	83
-----	----------------------------	----

List of Figures

1	Process of data visualization construction and interpretation with major steps in white letters. Types identified in Table 1 are given in italics, and an exemplary US reference system and sample data overlays are given in Right. (Borner2019)	47
2	Preliminary methodology	62
3	Logical mathematical operations permissible for data per Borner2019's DVL-FW .	72
4	Logical mathematical operations permissible, measure of central tendency, and examples for different data scale types (Borner2019)	72
5	Four graphic symbols and 11 graphic variables from full 11 graphic symbols by 24 graphic variables set in ref. 34. Qualitative nominal variables (shapre, color, hue, and pattern) have a gray mark. (Borner2019)	73
6	Preliminary data and information flow	75
7	Preliminary data model fo the spatial database	75
8	Preliminary <i>Input</i> layout	76
9	Preliminary <i>Context</i> layout	76
10	Preliminary <i>Search</i> layout	77
11	Publisher provided data attributes	77
12	Mapped article layout (points) from Gupta2020 [1]	78
13	Annotation via GeoAnnotator from Karimzadeh2019 [2]	79
14	Annotation ambiguities from Karimzadeh2019 [2]	80
15	Geodata Search Interface from Eisl2020 [3]	81
16	general architecture of geoannotator from Karimzadeh2019 [2]	81
17	Full stack deployment map from Makai2'2' [4]	82
18	Progressive georeferencing frameowrk from Cai2016 [5]	82

List of Tables

1	Corpora data	59
2	Corpora attribute collection	60
3	Base map data	60
4	Gazetteer data	61

1 Introduction

The ongoing COVID pandemic has highlighted the value of the visualization of information on a map, not only for specialists to monitor and predict viral outbreaks, but to arm the public with empowering information as well. Of course, the value of geographic information systems (GIS) goes beyond public health services and is already nestled into our everyday activities in the form of daily tasks such as navigation and service selection. Applications like Google Maps, AirBnB, and UberEats allow non-technical users to visualize and filter the distribution of various services through spatial (SA), temporal (TA), and thematic attributes (ThA). For example, a user on AirBnB may filter all apartments with high speed WiFi (ThAs) available in the Estrela neighborhood and within walking distance to a market (SAs) from Aug 1 to Aug 7, 2020 (TA).

Yet, though this type of manipulation is commonplace in the products of many industries, it is glaringly absent from that of news media. When reading about an incident occurring in an unfamiliar place, readers will often need to look up the location. They may have trouble relating the spatial significance of an incident to neighboring occurrences or historical events in the same spot. Many articles define place via textual descriptions, but these can be easily overlooked if searched by keyword, especially if different names or alternate designations are employed by the searcher. This is a problem for researchers who may want to define a study area that does not conform to traditional administrative boundaries or existing points of interest, but also for the casual user or city official. The former might, while perusing headlines, miss an article of interest relating to a place along their commute home from work. The latter could be an elected official who seeks to monitor an issue (such as gentrification or homelessness) but is unable to visualize the subtle distribution of such events throughout his or her district. In these cases, as well as a host of others, there is obvious disconnect between the existence of data and its usability. Though many search engine queries contain geographic keywords [6], news media enterprises have not yet accommodated such spatial associations to their articles that would provide an expected improved user experience and therefore competitive edge in their industry. As such, there is commercial as well as operational and academic value in better understanding the spatial distribution of events within a community, such that additional informative insights can be drawn.

This project seeks to develop a set of functional tools that supports the creation and management of a spatial database of local news stories, a publishing interface (associating place and adding records to the database), a user interface (list and map format search, filter and visualization of results from the database), as well as a story visualization plugin (a map displaying the distribution of a story in a contextual map per story page), see Appendix C. This proof of concept (POC) functionality should demonstrate the value of new spatial products in news media, and provide a basis from which meaningful projects may be developed for mass media applications in the future.

2 Literature review

2.1 Citizen empowerment

2.1.1 Smart communities

- Leverages various sources of data to determine the state of various subsystems and support interventions for improvement.

- “There are four dimensions in which a smart city primarily operates, names: intelligent city (its social infrastructure), the digital city (informational infrastructure), the open city (open governance) and the live city (a continuous adaptive urban living fabric) [7].
- “An active and engaged citizen is indeed the main driving force of a “smart city” [23] [7].
- “As the popularity of social media is growing exponentially we are presented with unique opportunities to identify and understand information dissemination mechanisms and patterns of activity in both the geographical and social dimensions, allowing us to optimize responses to specific events, while the identification of hotspot emergence helps us allocate resources to meet forthcoming needs”. News reporting as a more official or pre-aggregated public pulse can be layered in as well. [7].
- “the need of an enabling geospatial information platform to facilitate data discovery and access in order to support smart cities’ operations.” [8].
- “A smart city is roughly described as a platform or a system of systems, essentially based on three components: Sensors, Networks and Engagement (actuators).” [8].
- “Making a city smarter is neither only a technological infrastructure issue, nor a managerial one. It is also and essentially providing citizens a better and safer way of living in urban areas, in teh places where they live, work, have fun, consume... Therefore, sensing life in those places is a major stake to understand new city dynamics and then to design better living urban environments.” [8]
- “Information and Communication Technologies (ICTs) have given rise to the ideal that cities will become increasingly smart, connected, responsive, and citizen-centric”. [9]
- “Planning and decision making is not all about efficiency, but also about responding to the public interest and community needs.” [9]
- “Communities’ too have internal structures.” [10]
- “the geography of social relations is changing. In many cases such relations are increasingly stretched out over space. Economic, political and cultural social relations, each full of power and with internal structures of domination and subordination, stretched out over the platen at every different level, front eh household to the local area to the internatinal.” [10]
- “The creation of economic wealth, social stability and environmental protection in line with MDGs can be achieved through the development of products and services based on spatial information collected by all levels of government. These goals and objectives can be facilitated through the development of a spatially enabled government and society, where location and spatial information are regarded as common goods made available to citizens and businesses to encourage creativity and product development.” MDG: Millennium Development Goals [11]
- Psychographics: “the prevailing interests of people in the area.” [12]
- “In the big data era, the typical scenario does not start from data but from application. THe research procedure is shifting to a new scenario - ‘search, evaluate, and process’ -, that is, identifying the applications, discovering all data related, adn running a processing model for knowledge generating.” [13]
- “Increased accessibility of technology pushes researchers to consider how these tools might be used in critical scholarship in nontraditional ways.” [14]
- “DURING the last two ecads, urban structures have become more digital and information-based, and there has been a decisive change in the living environment of citizens.” [15]
- Smart city and knowledge society. [15]
- A smart community should (among other things) “Enable a shared understanding of what is happening in the ‘city’.” via “the need to notify citizens about traffic accidents” [15]
- Smart community: “Communities can be large or small, and they may or may not be

ac enter of some noteworthy aspect of human activity; they may be separated from other communities or they may be aggregated into a conurbation of some kind. But any community of any size or significance can be ‘smart’. Communities should also define the services of interest to them, within their territorial boundary or outside it, as required.” [16]

- “there are dimensions of smart which the community pursues.” [16]
- “A smart community may be one of any size or significance, geographically separate or part of some larger urban unit, that employs the IOT to: improve aspects of its operations or other factors within or outside its boundaries that are important to its economic vitality, safety, environmental footprint, quality fo life or other factors deemed significant; respond to the community’s changing needs rapidly and efficiently; engage the community and enable informed understanding of, and where applicable consent to, what it is doing; collaborate with other communities as needed or desired.” [16]

2.1.2 Communication

- “In today’s world, words are moving rapidly, allowing neighbors to share news that was once passed along porches or stoops. Citizens are increasingly sharing information via social networking and virtual reality tools, rather than from the front porch.” [17]
- dIgital tools support engagement beyond physical borders [17]
- “The issue of context-sensitivity may require more attention from planning organizations that try to choose new technologies. Cities are overwhelmed with the availability of new communication technologies and the opportunities that the technologies offer. This issue also exacerbates with the social pressure they receive from citizens who expect the implementation of smarter governance systems.” [9]

2.1.3 Data visualization

- “visually prominent and summary-based information, such as maps and pictures, are participants’ preferred form sof information.” [17]
- “Data visualization is another important functionality for communicating geospatial information to users. A popular method for visualizing data is to use an online map for allowing users to visually evaluate a dataset. Online maps can be interactive allowing panning and zooming and possibly changing the visual appearance of the base map (e.g. satellite image, vector map). Geportals can also provide data download functionality providing either the data directly or through sharing of dataset links.” [13]
- “Maps then illustrate layers of data, visually displaying spatial relationships.” [14]
- “Every publisher and journalist knows the value of charts and wants more of them. Done well, charts throw light on difficult topics - they make stories easy to read by breaking up oceans of text. The reality is different. Busy newsrooms don’t always have the tools or skills to create great charts. And in a world of daily, if not hourly, deadlines, they often don’t have the time.” [18]
- “As well as improving reader understanding of complex issues, the charts are generating 5.6% more page views.” [18]
- bilingual map legend example [19]
- “These interactive data visualizations inform the public and guide decision-makers to save lives”. [20]
- In 2019, “data visualization featured prominently in major new stories and key players in the field created work that didn’t just do well on Dataviz Twitter but all over.” [21]

- “2019 saw the United States President amend a data visualization product with a sharpie. That should have been enough to make 2019 special, but the year also saw the introduction of a data visualization-focused fashion line, a touching book that uses data visualization to express some of the anxieties and feelings we all struggle with, as well as the creation of the first holistic professional society focused on data visualization.” [21]
- “public debates about the presentation of data increase the prominence of data visualization as a meaningful act.” [21]
- “The naive perspective that data visualization is just a final step to help people see the data ignores the importance of subtle steps like showing uncertainty as well as the necessity to design a product that engages the audience”
- “Data visualization as a way of exploring and expressing one’s feelings and traits has always been present in the margins of the field.” [21] Data-driven badges [21]
- “Both SalesForce and Google have already invested significantly in their own in-house data visualization tools but both realized that to compete they needed to rapidly expand their data visualization capacities and were willing to pay top dollar to do so.” [21]
- “Data is increasingly important to all businesses, not just tech, and so much a part of our everyday lives that it makes sense that companies with strong data analysis, data science, and data engineering talent would feel the need to improve their data visualization capabilities.” [21]
- “But we’re now seeing an acknowledgement that if you don’t have good data visualization then your insights are less apparent, resonate less with audiences and are harder to communicate among scientists.” [21]
- “Data visualization is becoming less of a tech company rarity and more a part of everyone’s everyday life... it will only continue to grow more common in the coming years.” [21]
- “Modern data visualization is optimized for producing charts for busy executives. But that’s changing. Now, data visualization is personal stories, small businesses, data science, political campaigns, human resources, community building – in short, data visualization is becoming a part of the fabric that is modern culture.” [21]
- “Exploring information collections becomes increasingly difficult as the volume grows. A page of information is easy to explore, but when the information becomes the size of a book, or library, or even larger, it may be difficult to locate known items or browse to gain an overview.” [22]
- “A picture is often cited to be worth a thousand words and, for some (but not all) tasks, it is clear that a visual presentation - such as a map or photograph - is drastically easier to use than is a textual description or a spoken report.” [22]
- “Visual displays become even more attractive to provide orientation or context, to enable selection of regions, and to provide dynamic feedback for identifying changes”. [22]
- “Abstract information visualization has the power to reveal patterns, clusters, gaps, or outliers in statistical data, stock-market trades, computer directories, or document collections.” [22]
- “Overall, the bandwidth of information presentation is potentially higher in the visual domain than for media reaching any of the other senses.” [22]
- Type by task taxonomy (TTT) [22]
- “The attraction of visual displays, when compared to textual displays, is that they make use of the remarkable human perceptual ability for visual information. Within visual displays, there are opportunities for showing relationships by proximity, by containment, by connected lines, or by color coding.” [22]
- “In the information age, the ability to read and construct data visualizations becomes as important as the ability to read and write text.” [23]

- “The invention of the printing press created a mandate for universal textual literacy; the need to manipulate many large numbers create the need for mathematical literacy; and the ubiquity and importance of photography, film, and digital drawing tools posed a need for visual literacy. Analogously, the increasing availability of large datasets, the importance of understanding them, and the utility of data visualizations to inform data-driven decision making pose a need for universal data visualisation literacy (DVL).” [23]
- “Like other literacies, DVL aims to promote better communication and collaboration, empower users to understand their world, build individual self-efficacy, and improve decision making in business and governments.” [23]
- Insight needs: “DIfferent stakeholders have different insight needs (also called “basic task types”) that must be understood in detail to design effective visualizations for communication and/or exploration.” [23]
- “Most datasets need to be analyzed before they can be visualized” [23]
- Visualization types: “statistical analysis (e.g., to order, rank, or sort); temporal analysis answering “when” questions (e.g., to discover trends); geospatial analysis answering “where” questions (e.g., to identify distributions over space); topical analysis answering “what” questions (e.g., to examine the composition of text); and relational analysis answering “with whom” questions (e.g., to examine relationships; also called network analysis).” [23]
- “Prior research on DVL shows that people have difficulties reading most visualization types but especially, networks.” [23]
- “Controlled laboratory studies examining the recall accuracy of relational data using map and network visualizations have found that map visualization are easier to read and increase memorability”. [23]
- “the time required to read a visual image increases systematically with the distance between initial focus point and the target - independent of the ‘amount of material’ between both points.” [23]
- “However, when subjects had to perform computation while reading a visualization, comprehension became more difficult, showing that the interpretation of graphs is ‘serial and incremental, rather than automatic and holistic’.” [23]
- Because of the “iterative nature of graph comprehension”, “the importance of spatial processes (e.g., the temporal storage and retrieval of an object’s location in memory, allowing for mental transformations, such as creating and transforming a mental image) for the graph comprehension.” [23]
- “the more usage and actionable insights gained, the more important it becomes to empower individuals to properly construct and interpret that visualization.” [23]
- “disparate juxtaposed visualizations are insufficient for understanding nuanced temporal interactions between many data sources.” [24]
- “Data doesn’t have to be scary or intimidating, because, if you think about it, data isn’t even real! Rather it’s abstract, representing the details of our lives and our ideas.” Giorgia Lupi’ [25]
- “Collecting data is a way for us to build records and preserve memories. Reclaiming the human approach to data allows for narrative components that can act as very powerful tools and design materials to create stories we can all relate to.” Giorgia Lupi’ [25]
- “Data projects can be visually beautiful, but their true power lies in their ability to convey meaningful and relatable narratives about parts of the world and people that we can’t necessarily observe.” Giorgia Lupi [25]
- “Data is now recognized as one of the founding pillars of our economy, and the notion that the world grows exponentially richer in data every day is already yesterday’s news.” [26]

- “The phenomena that rule our world are by definition complex, multifaceted and mostly difficult to grasp, so why would anyone want to dumb them down to make crucial decisions or deliver important messages?” [26]
- “We are ready to question the impersonality of a merely technical approach to data and to begin designing ways to connect numbers to what they really stand for: knowledge, behaviors, people.” [26]
- “We can write rich and dense stories with data. We can educate the reader’s eye to become familiar with visual languages that convey the true depth of complex stories.” [26]
- “Data, if properly contextualized, can be an incredibly powerful tool to write more meaningful and intimate narratives.” [26]
- “Let’s just stop thinking data is perfect. It’s not. Data is primarily human-made. ‘Data-driven’ doesn’t mean ‘unmistakably true,’ and it never did.” [26]

Example:

Data visualization literacy framework (DVL-FW)

2.1.4 Public participation

- “The term ‘public participation geographic information systems’ (PPGIS) was conceived in 1996 at the meeting of the National Center for Geographic INformation and Analysis in the United States to describe how GIS technology could support public participation for a variety of applications with the goal of greater inclusion and empowerment of marginalized populations.” [27]
- “PPGIS combines the practice of GIS and mapping at local levels to produce knowledge of place.” [27]
- “Smart urban solutions have to be built on the vision of citizens as active sensors on one hand and on spatial enablement of citizens via social networks on the other hand. These solutions have to be inline with improvement of navigation related spatial skills using geographical information and techniques for annotating spaces with digital information. These kind of solutions have also to be built on the potentials offered both by embedded sensors to crowdsource the process of collecting geo-referenced information about places in the city and social networks to disseminate this information and democratize access to it.” [8]
- “The active engagement of citizens is then a major requirement for smart operations of a city, and this needs for citizens to be spatially (digitally) enable. Capturing the sense of places in then a major stake for urban communities that look for a smarter way of development.” [8]
- “volunteered Geographical Information – VGI - becomes the most prolific sources of information to characterize places.” [8]
- “The importance of encouraging people to act as participative citizens in issues of public concern is essential for a functioning democracy, particularly when researchers are observing that civic engagement (CE) is diminishing in developed countries” [28]
- CE: Civic engagement; “ways in which citizens have a common purpose of preserving and promoting public goods, to improve conditions for others, community or collective benefit.” [28]
- “CE and participation are inherently spatial and, consequently, influence by social relations, time and space. The spatial dimension of CE (e.g., planning decisions or decision-making processes about communal spaces) has been established in administrative boundaries because of the availability of census and socioeconomic data in those areas. However,

teh approach has probably hidden the spatial nature of CE associated with space, place and locality - essential characteristics to determine who is interested in the participatory processes and why.” [28]

- “the understanding of the spatial relationship between SoP, SC and CE establishes novel spatial scenes based on human-city interactions. There is a potential for understanding the spatial relationships of social concepts (i.e., SoP, SC and CE) to provide alternate or completely new units of analysis for citizen participation. The definition and mapping of these individuals’ spatialities allow a more citizen-focused representation of the city.” [28]
- “the role of the geographical perspective in taking another step forward to better understand citizens’ social synergies in the urban context.” [28]

-“There is a shortage of empirical research on the interactions between people and places.”

Justification of project [28]

- “in spite of all the technological developments in recent years, one of the biggest barriers to public participation in urban policies remains unsurpassable: the difficulty that people have to understand how the planning proposals are projected in space, how they redefine it, and how they impact the use of urban space.” [29]
- “it is assumed that the use of techniques and tools derived from the Neo-Geography, in particular VGI, helps to develop an understanding of the complexity of geographical space and place, through the analysis of geographical phenomena and that this change may influence the attitudes of citizens to the discussion of the problem facing society.” [29]
- “Today, technology allows for an entirely new generation of forms and practices of public participation that promise to elevate the public discourse in an unprecedented manner while providing an interactive, networked environment for decision-making.” [17]
- Digital tools can increase public engagement in politics and urban planning without citizens even actively intending to contribute. [17]
- “loca-l governments are increasingly designing their websites from the standpoint of ‘what does the citizen need to know’ versus ‘what information do we have to provide,’ offering the potential for better engagement. However, very few municipal websites facilitate on-line public dialogue or consultation.” [17]
- “Often-cited benefits of participation include increasing the education and awareness levels of the citizenry, civic engagement, government responsiveness, and citizens’ commitment to implementation”. [17]
- “Participation helps to build social capital in a community, which in turn strengthens the community.” [17]
- “The highest level of participation opportunities hold that all citizens must be equally empowered and fully informed to ensure that they can exert influence in decisions that affect them” “The highest level of participation opportunities hold that all citizens must be equally empowered and fully informed to ensure that they can exert influence in decisions that affect them” [17]
- “The rate of utilization and willingness to accept new methods may be based on teh demographics of a community. In some communities, there is a growing expectation on the part of citizens that there will be online participation opportunities.” [17]
- “An active particpatory environment that uses Internet technology has the potential to engage the public, and may therefore facilitate knowledge retention and use by the public.” [17]
- “Public participation has been a hallmark of the planning process for over 30 yeras, with each generation trying to improve access and interactivity to ordinary citizens.” [17]
- “Such new and creative participation strategies [hybrid: partially in-person and online experiences] could ameliorate the potential harmful effects of an increasingly digital di-

vide.” [17]

-“Active citizenship, promoting citizens’ self-organization and engagement in urban development are high on the political agenda. As a result, many cities are experiencing a surge in place-based and technology initiatives, both government-initiated and grassroots activism, and collective networked action to foster civic engagement in urban and neighbourhood contexts.” [30]

-Citizens engagement: “invites coproduction of content without necessarily engaging contributors in dialogue”. - “ONline participatory tools (OPTs) refer to two types of technologies: (1) web-based tools that are particularly designed for public engagement (e.g. MySideWalk, PlaceSpeak, CitySourced, CrowdBrite); and (2) social networking sites (e.g. Facebook, NextDoor) that are not designed for public engagement but can be used for participatory planning.” “smart-city approaches should contribute to innovation and enhance democratic decision making and transparency through public participation.” [9]

-“These processes focus on responding to public interest and promoting open-ended interactions to provide opportunities for participants that constantly redefine the “what” and “how” of the issues they address. These processes can provide opportunities for consensus building or learning among diverse stakeholders, democratic decision making, mobilizing actions, engaging local knowledge, or responding to regulations or community norms.” [9]

-“organizations’ attitudes towards managing and controlling the online environment can influence engagement by encouraging or discouraging particular behaviors. For example, responding to participant questions in a timely manner may result in a higher ongoing participation rate.” [9]

-“Using online tools in public meetings can facilitate the participation of a more diverse community.” [9]

-“Common to all types of PPGIS data capture is the need to symbolically represent the spatial attribute of interest on a map.” [27]

-“Citizens’ characteristics and skills, and their attitudes towards using technology for participation should be also considered in choosing online participatory tools. Communities that will be directly or indirectly affected by plans are indispensable components of participatory processes.” [9]

-“Before using an OPT an organization must clearly define the goals for the tool’s use. Is the main goal of using the OPT to inform and educate citizens, follow up with citizens about particular aspects of the plan, engage citizens in a consensus-building process, resolve tensions between conflicting ideas, or build trust in communities? Is the goal of using OPTs to attract those who usually do not or cannot attend public meetings, or is the purpose of using OPTs to encourage excitement about a project in a community?” [9]

-“Community norms can also influence the organizations’ decision towards using new technologies. Cities’ efforts towards creating open data portals or developing transparency advisory boards are examples of such activities.” [9]

-Examples: MySideWalk, SeeClickFix, nextdoor, citysourced, placespeak, mindmixer [9]

-“The technical capacities of online participatory tools may facilitate specific types of decision-making processes by providing opportunities for deep dialogue, quick information sharing, or social mobilization. The type of technology may not directly influence the entire decision making process; but, it may promote certain types of processes by guiding how people can share or discuss ideas.” [9]

-“Mapping technology has also been included in community engagement projects and sites of activism, leading to a new form of visual politics. Scaled representation of the community has helped stakeholders understand political implications for their community in a broader context.” [14]

- “Minimizing cognitive complexity and enhancing ease of the use for hte PPGIS participant appears essential to increasing participation rates. And yet, there may be a contrary relationship between the simplicity of the PPGIS data capture methods and the resulting quality of the spatial data.” [27]
- “Perhaps the one constant factor that needs to be kept in mind as these choices are made is that technology will be alienating without participation.” [16]

2.2 Spatial Information

- “Location data is now commonly regarded as the fourth driver in the decision-making process.” [8]
- “Location-based services and information are the basic components needed to dynamically describe and represent places’ life” [8]
- “the spatial aspects of data is being largely underutilized in processing, although the importance of spatial decision-making is now widely accepted”. [31]
- “a city could not be smart without spatially enabled citizens.” [8]
- “From a practical point of view, and in the smart city context, “spatial enablement” refers to the individuals’ (or collective) ability to use any geospatial information and local technology as a means to improve their spatiality, that is to say, the way they interact with space and other individuals on/in/through space. Spatiality is the dynamic component of place making.” [8]
- “SDI and spatial technologies are now used routinely in decision making to support city planning and forecasting and at the same time to address some of the world’s most pressing societal problems.”
- “Location is involved with everything, hence a spatial system is vital for better urban information management and spatial data infrastructure (SDI) creation.” [31]
- “Nowadays it is increasingly imperative to capture data at all spatial scales, local to global, and extract useful information from it ubiquitously and regularly.” [31]
- “Almost every decision that an individual or organization makes has some geospatial component.” [31]
- “the concept of geo-data democratization is gaining momentum and is being described as the next big disruption.” [31]
- “Ready and timely access to spatial information - knowing where people and assets are - is essential and is a critical tool for making any informed decisions on key economic, environmental and social issues.” [11]
- “Spatial informaiton is an enabling technology /infrastructure for modern society.” [11]
- “in modern society, spatial information is an enabling technology or an infrastructure to facilitate decision making.” [11]
- “Spatial information can be a unifying medium in which linking solutions to location.” [11]
- “Traditionally, when we think of connecting information to a geospatial location, we think in terms of maps.” [32]
- Georeferencing data content: “users to add locations or location-related data content onto a map”. [32]
- “Geographical datasets include maps and feature-based datasets, such as cadastral maps, national and regional boundaries, and other GIS-type data and information.” As opposed to earth observation datatype (in-situ or remote sensing). [13]
- “Evolution of, and increased access to, GIS technology provides an opportunity to visually create and share experiences, information, and perspectives across a broader audience, breaking geographic barriers for users through multimodal construction of meaning.” [14]

- “Geographic information systems (GIS) are computerised systems that allow users to electronically tag geographic locations with various information. Maps can then be constructed for visual representation.” Definition of GIS and applications [14]
- “The spaces in which we live, work, and play affect our lives both positively and negatively, and GIS provides a way to visually convey relational meaning of spaces.” [14]
- An opportunity for maps to “contradict the dominant narrative” and “increase understanding”. [?]
- “Though GIS studies have the potential to represent a fluidity through constant access to technology that written reports often lack, they are still bounded by restrictions of language and symbols, as well as technical and financial resources.” [14]
- “Decades of technology R&D, including, not least, US Government investment in relatively ‘open’ data made available via Global Positioning Satellite (GPS) infrastructure have unleashed the potentials of real-time location tracking and the myriad services this data can be used to enhance, via machine learning techniques.” [33]

2.2.1 Location

- “Location plays a vital part by helping understand relations between datasets.” [31]
- “In context-aware computing, location is a fundamental component that supports a wide-range of applications.” [34]
- “Locality is used to describe a more precise area.” [35]
- Hyperlocality: “a spatiality that is endemic - i.e., locationally specific - to the individual, real-time positionalities of digital platform users.” Ex: “one of the most immediate examples of hyperlocality is the ‘blue dot’ on the mobile Google Maps interface, which visually centers the map data on the real-time location of the user.” [36]

2.2.2 Place

- “Place is often uses in the sense of community or neighbourhood, implying an informal relationship to an area surrounding the individual’s place of residence” (refers to Goodchild [8]) [8]
- “But more important for a smart city is its capability to capture the sense of places. A city is not a machine, bur rather made by people local actions and feelings.” [7].
- “Indeed with the exponential growing of location-based social networks (geosocial), Geoweb 2.0 and geoinformation crowdsourcing, citizens are increasingly involved in the the production of geographic information. This kind of information, voluntarily produced and diffused by people, mainly refers to the places they live/use. Indeed, people whom live in a place are often the “experts” of this place.” [8]
- “Spontaneous and localized contributions of individuals (localized tweets, Facebook Places or foursquare chek-in) are most often materialized by Points of Interest (POI). This POI could be considered as new forms fo spatial projections of social relationships and human spatiality.” [8]
- “We find that the sense of place is significant and positively correlated with social capital, while the latter also significantly explains civic engagement at the individual level.” [28]
- “the citizens’ perception of pre-established administrative boundaries can differ from the “real” one nd, consequently, whole administrative boundaries might not cover the SoP, SC, and aCE of all its dwellers.” [28]
- NeoGeogrphy: Location based services (LBS) and volunteer geographic information

(VGI) “This change of paradigm of functioning of the Internet, combined with strong social networks diffusion, changed the concept of community.” [29]

- “How, in the face of all this movement and intermixing, can we retain any sense of a local place and its particularity? An (idealised) notion of an era when places were (supposedly) inhabited by coherent and homogeneous communities is set against the current fragmentation and disruption.” [10]

- Globalization, “speeding up and spreading out” = Time-space compression. “Movement and the communication across space, to the geographical stretching-out of social relations, and to our experience of all of this.” [10]

- “What is it that determines our degrees of mobility, that influences the sense we have of space and place?” [10]

- “A ‘sense of place’, of rootedness, can provide - in this form and on this interpretation - stability and a source of unproblematical identify.” [10]

- “We need, therefore, to think through what might be an adequately progressive sense of place, on which would fit in with the current global-local times and the feelings and relations they give rise to, and which would be useful in what rae, after all, political struggles often inevitably based on place. The question is how to hold on to that notion of geographical difference, of uniqueness, even of rootedness if people want that, without it being reactionary.” [10]

- Places don’t have “single, essential identities”, nor are they based on “internalized origins”. [10]

- “While [a place] may have a character of its own, it is absolutely not a seamless, coherent identity, a single sense of place which everyone shares. It could hardly be less so. People’s routes through the place, their favourite haunts within it, the connections they make (physically, or by phone or post, or in memory and imagination) between here an the rest of the world vary enormously. If it is now recognised that people have multiple identities than the same point can be made in relation to places. Moreover, such multiple identities can either be a source of richness or a source of conflict, or both.” [10]

- “On the one hand communities can exist without being in the same place - from networks of friends with like interests, to major religious, ethnic or political communities. On the other hand, the instances of places housing single ‘communities’ in teh sense of coherent social groups are probably - and, I would argue, have for long been - quite rare. Moreover, even where they do exist this in no way implies a single sense of place. FOr people occupy different positions within any community.” [10]

- “what gives a place its specificity is not some long internalised history but the fact that it is constructed out of a particular constellation of social relations, meeting and weaving together at a particular locus.” [10]

- “If places can be conceptualised in terms of the social interactions which they tie together, then it is also the case that these interactions themselves are not motionless things, frozen in time. They are processes.” [10]

- “places do not have single, unique ‘identities’; they are full of internal conflicts.” [10]

- “The specificity of place is continually reproduced but it is not a specificity which results from some long, internalised history. THere are a number of sources of this specificity - the uniqueness of place. There is the fact that the wider social relations in which places are set are themselves geographically differentiated. Globalisation (in the economy, or in culture, or in anything else) does not entail simply homogenisation. On the contrary, the globalisation of social relations is yet another source of (the reproduction of) geographical uneven development, and thus of the uniqueness of place.” [10]

- “There is specificity of place which derives from the fact that each place is the focus of a

distinct mixture of wider and more local social relations. There is the fact that this very mixture together in one place may produce effects which would not have happened otherwise. And finally, all these relations interact with and take a further element of specificity from the accumulated history of a place, with that history itself imagined as the product of layer upon layer of different sets of linkages, both local and to the wider world.” [10]

- “It is a sense of place, an understanding of ‘its character’, which can only be constructed by linking that place to places beyond. A progressive sense of place would recognize that, without being threatened by it. What we need... is a global sense of the local, a global sense of place.” [10]

- Psychogeographies: “the prevailing interests of people in the area” [12]

- “Understanding that meaning shifts through an interplay of social, linguistic, and material interactions, GIS has the unique ability to show fluidity yet connectedness across geographic borders, providing intriguing opportunities as a research tool.” [14]

- “the potential to build understanding of spatial context as both socially and politically constructed, but also the tendency to see a map as bounded data in finality.” [14]

- “Place is constructed meaning within [physical spaces]. These places of meaning are a crossing of ‘discursive, interpretive, lived and imagined practices. Spaces are made up of places with socially created, accrued meaning. Places support the formation of identity, security, and belonging within spaces. Connecting with ideas of new materialism, spatial researchers see this formation as fluid rather than fixed... Humans impact and are impacted by the social relations in those spaces.” [14]

- “Looking at the politics of spaces helps policy makers, community members, educators, and students understand oppressive cycles of policies imposed on communities.” [14]

- “spatial theory offers a glimpse into the complicated relationships between humans and their surrounding contexts. GIS could be first, a snapshot of those relationships and second, a conversation starter for further investigation and action.” [14]

- New materialism: “the material world is not lifeless soil but something more dynamic.” [14]

- Maps “are objects of analysis that are linguistically and materially [hypercontextually] entangled... resulting in a continual state of becoming. In other words, the material does not construct our reality any more than the discursive defines us socially – instead, it is how the relationship morphs, depending on the context.” [14]

- “When objects, words, people, or places come together, they make something altogether new.” onto-epistemology: “the ability to use maps as meaning-making of marked experiences in time and space, with the understanding that this is not a static representation of meaning but a point in an evolving state.” [14]

- GIS is a tool to powerfully represent data and effect change by demonstrating how the spatial affects people, and new materialism forces us to concede GIS visual data as already obsolete. The relational ontologies of new materialism disrupt the static representation of a fixed point on a map, bringing the geography and biology of social life into conversation ‘in a discipline which for a long time intentionally decided to ignore the issues arising from this connection.’ “ [14]

- Distinguishing characteristics of local knowledge: “It is based on experience. It is developed over time by people living in a given community, and is continuously developing. It is embedded in community practices, institutions, relationships, and rituals. It is held by individuals or communities. It is dynamic and changing. Based on these characteristics, we may anticipate that local knowledge is unique from place to place. Therefore, the gazetteer used for geo-referencing local newspaper articles should be place-specific.” [5]

2.2.3 Geointelligence

- Disaster response: “66% of enterprises rank Location INtelligence as either critical or very important” [37]
- Deriving and delivering intelligence: extraction/collection, integration/fusion, iltering/cleaning, enrichment/analysis, distribution/consumable [37]
- “Geospatial intelligence, or the frequently used term GEOINT, refers to the discipline of exploitation and analysis of satellite imagery and other forms of earth observation data to describe, assess, and visually depict physical features and geographically referenced activities on the earth. The term GEOINT is typically used for the defense and internal security domains and offers the capability of monitoring, predicting and countering threats, while helping strategize and support various field operations.” [38]
- Intentionally georeferenced data can contribute to GEOINT as a geolocated data source in addition to other imageries and info fontes. [38]
- Trends: AI and automation, social media and mobile data, analytics-as-a-service, drive for cloud, short shelf life of technologies, geospatial information science [38]
- “Smartphones in every hand, increasing use of social media, and sensors in transport vehicles have turned human beings into sensors. And this is what is driving the next wave of strategy and innovation for the GEOINT community. With the convergence of new sensors and social analytics, it is possible for defence organizations to access multi-source data for enhanced decision making. As more terrorist/criminal organizations take to social networks like Twitter and Facebook to communicate their efforts, we have seen defence and intelligence organizations are using the right data analytic tools for mining this data to gain a better operational picture of enemy activity.” [38]
- Geospatial Information Science: “Different people behave in different ways and culture and location plays an important role in it. With terrorist/criminal activities rising, it is becoming paramount for defense and intelligence communities to learn how humans impact their environments and vice versa; how often cultural norms differ in relation to various environments. This will also enable the GEOINT community to effectively use technology to capture, manage and analyze geographic information that supports leaders in making informed decisions in complex environments.” [38]
- “Who? What? Why? When? Where? These questions form the basis of human exploration. They are fundamental to knowing and understanding our world. Their answers are essential to information gatherings, storytelling, and problem solving.” NGA SHOW THE WAY video [39]
- “Everything on Earth, from its watery depths to its hughes peaks, can be measured in space and time, whats known as Geospatial Intelligence or “GEOINT.” [39]
- Situational awareness. “Tell me everything that is happening” prior to forming specific questions [40]
- “Hot spot analysis refers to the use of geostatistics and time series analysis to find out the most significant spatial hot spots of a certain indicator and their respective temporal trends.” [40]
- ORACULO peacekeeping leverages remote sensing imagery (IMINT), relationship development (HUMINT), reconnaissance, and open sources (OSINT). OSINT: “Thus, though every other way of collecting events promises more reliable results, if we want fast, better than nothing results, we have to turn to open sources, such as news websites.” [40]
- “To whom does the information flow? The Media, of course, which has continued to operate throughout the conflict.” [40]
- “Extract events from open sources / manage event geodatabases / mine event space time patterns / improving situational awareness for a given context.” [40]

- Merge multiple reportings of the same event based on temporal, spatial, and thematic proximity [40]
- “Without location context, first responders are unable to decide where or how to respond to information they receive. This is especially true during natural disasters when geographic content is necessary for dispatching appropriate emergency response.” [41]

2.2.4 Cross border mapping

- “cross-border mapping can contribute to cross-border cooperation.” [19]
- “The crucial point of transboundary spatial analysis is data heterogeneity.” [19]

2.2.5 Feature types

- Massey 1991: “highlights the problem of recurrently trying to draw boundaries to the conception of place and place-related concepts that, in inherently, distinguishes between an inside (e.g., us) and an outside (e.g., them). She also supports that there is no need to conceptualize boundaries in order to define place, advocating that place is a process of social interactions.” [28]
- Polygons (vs. points): ease of implementation, “better encompass of a high range of spatial scales, (from an armchair to the whole earth)”, “better performance of polygon features when there is a limited spatial dataset.” [28]
- “A particular problem with this conception of place is that it seems to require the drawing of boundaries. Geographers have long been exercised by the problem of defining regions, and this question of ‘definition’ has almost always been reduced to the issue of drawing lines around a place.” [10]
- “Instead then, of thinking of places as areas with boundaries around, they can be imagined as articulated moments in networks of social relations and understandings, but where a large portion of those relations, experiences and understandings are constructed on a far larger scale than what we happen to define for that moment as the place itself, whether that be a street or a region or even a continent. And this in turn allows a sense of place which is extroverted, which includes a consciousness of its links with the wider world, which integrates in a positive way the global and the local.” [10]
- “places do not have to have boundaries in the sense of ivisions which frame simple enclosures... Definition in this sense does not have to be through simple counterposition to the outside; it can come, in part, precisely through the particularity of linkage to that ‘outside’ which is therefore itself part of what constitutes the place.” [10]
- “Borderland is a natural transition and convergence area where people, goods, services, and ideas flow across boundaries or seas from state to state. Such cross-border commonalities, which cannot be divided by politically dictated and artificial boundary lines, potentially contribute to sustainable development within the world.” Definition of borderland to describe areas that don’t convene to administrative boundaries. “A borderland region generally refers to the land area adjoining and outside state boundary lines, or hte ocean area among maritime neighbors, and also has different characteristics or geographic conditions than the inner or central parts of the neighboring nations.” [32]
- “Any static method used to represent reality, whether linguistically, discursively, or culturally construed, distorts reality in some way. Researchers acknowledge the material experiences of participants while anticipating that these realities flow and change.” [14]
- “the same PPGIS attributes identified by points and polygons will converge on a collec-

tive spatial ‘truth’ within the study area provided there are enough observations, however, the degree of spatial convergence varies by PPGIS attribute type and the quantity of data collected.” [27]

- “The use of points for mapping PPGIS attributes and aggregating areas through density mapping constitutes a conservative approach to spatial inference about place significance, but the data demands for point collection are considerably higher than for polygon features.” [27]
- “the spatial feature chosen for soliciting spatial information may influence both the empirical results and the inferences that can be made.” [27]
- “Precision is a measure of the exactness in placing the PPGIS marker on the map... The precision of market placement on the map depends on a number of variables including marker size and map scale as well as participant characteristics such as visual acuity and physical dexterity. Flexible mapping environments that provide multiple map scales and marker sizes such as Google Maps can, in theory, enhance the precision of marker placement.” [27]
- “For PPGIS attributes used in regional planning applications conducted at a larger scale, concern with mapping precision is relatively small compared to the accuracy of the geographic area represented by the marker. Accuracy reflects how well the marker approaches the true spatial dimensions of the attribute being mapped. Accuracy in PPGIS is influenced by a number of variables including the nature of the PPGIS attribute being mapped (i.e., clarity in operational definitions and instructions enhance accuracy), the quality of the mapping environment (e.g., what base map features are included), and respondent characteristics such as map literacy.” [27]
- “The nature of the PPGIS attribute being identified, in particular, affects accuracy.” What does a point representation of a line or area or other fluid definition mean? “the level of accuracy may be indeterminant”. How hard are polygon boundaries? [27]
- “Limiting PPGIS attributes to landscape features where accuracy can be objectively verified is not a solution to the accuracy problem because some of the most relevant PPGIS attributes for land-use planning and management are respondent perceptions whose accuracy is not easily determined.” [27]
- “In mapping a PPGIS attribute with a point, the spatial attribute of interest is presumed to extend outward from the point in some unknown distance in some unknown direction. For PPGIS attributes identified as polygons, the participant is required to create boundaries that necessarily bifurcate the PPGIS attribute on the landscape, many of which are best viewed as continuous. For polygons, one can argue that the inaccuracy of a single point to represent the spatial attribute is replaced by the inaccuracy of an infinite number of points along the polygon’s edge. Alternatively, one can argue that some areal boundaries are better than none and increase the accuracy, even if indeterminate, of the attribute being identified.” [27]
- “Because point-based PPGIS attributes can have highly variable spatial distributions, density heuristics may need to be developed for the specific PPGIS attributes being measured.” [27]
- “The simplicity of point placement for the participant in PPGIS necessarily results in greater complexity in spatial interpretation for the analyst.” [27]
- “The trend line [agreement] has a poor fit when there is such large variation in the size of the polygons and the potential spatial error (amount of polygon area outside the point area) can be up to 18 times greater than the point hotspot area.” [27]
- “The quasi-experimental results presented herein demonstrate that the same PPGIS attributes identified by points and polygons will converge on a collective spatial ‘truth’ within

- a region provided there are enough observations in the study area. However, points and polygons do present different data information demands and spatial error trade-offs.” [27]
- “The use of points for identifying PPGIS attributes and aggregating areas through density mapping constitutes a conservative approach to spatial inference. Using point densities to create polygon areas results in a smaller probability of accepting spatial areas as collectively significant, when in fact they are not.” Higher probability of misattributing importance to non-important place with polygons. [27]
 - Requires fewer participants (by an order of magnitude) to draw polygons than points to achieve spatial agreement. “However, it is essential that the sample size be sufficiently large to claim regional representativeness regardless of whether points or polygons are the chosen method.” [27]
 - “PPGIS participants found the placement of points less ambiguous than identifying polygons and thus participants are more likely to complete the PPGIS mapping activity.” [27]

2.2.6 Geoportals/mashups

- “There is a big need for spatially referenced data creation, analysis and management.” [31]
- “The launch of Google Earth in 2005, and the availability of Google maps programming interface (API), as well as other initiatives, have transformed the way Internet users relate to geographic information (GI). The transfer of the information-creation processes from the specialized domain of Geographic Information Science to the field of action of the non-experts, the fact that other citizens, in addition to the geographers, cartographers or GI specialists, can create their maps with their own content, is to radically change the domains of interest and application of these mechanisms and to impact the criteria for the collection, analysis, implementation and the standard of truth of the information, with implications for information access, participation, power balance and nature of the data.” [29]
- GIS: “the opportunity to open spatial information to all stakeholders (presumably leading to better policy-making) and the idea that spatial analysis and outputs (.e. maps) can persuasively convey ideas.” [30]
- “innovative and imaginative geo-visualization interfaces such as Google Maps or Open Street Map - made possible by Web 2.0 technologies – have created low-key opportunities for almost any citizen with an Internet connection to generate and publicize their own maps and geographic information.” [30]
- “Organizations have different capacities to use new participatory methods. For example, private organizations or planning consultants may be better equipped than local governments to launch an online tool or collect and analyze the data.” [9]
- “Internal and external organizational collaborations influence the successful application of new technologies and methods within institutional systems. Collaboration with outside organizations can facilitate the incorporation of new technologies and data sources.” [9]
- “There has been a need to develop automated integral spatial system to sense and categorize events and issue information that reaches users directly.”
- “It has been proven that the spatial analysis of data gives more meaning to the information extraction and hence enables easier assimilation of large volumes of data. Presently the systems that implement such processes are limited in effect by not utilizing all the data due to their standalone nature, offline or disconnected design, lack of spatial capabilities, unintegrated approach, and temporally disjoint.” [31]
- “The solution could be addressed through integrating: data source, spatial data platform, data understanding, knowledge base, inferencing and visualization into single, well-

connected online real-time system. Such a spatial expert system (ES) with knowledge base (KB) will not only serve the critical research of spatializing developmental works but do so to any research relying on real time data capture and analysis with spatial domain of data being the unique enabler.” [31]

-“A new paradigm of more open, user-friendly data access is need to ensure that society can reduce vulnerability to spatial data variability and change, while at the same time exploiting opportunities that will occur.” [31]

-“Although research scientists have been the main users of these data, an increasing number of resource managers need and are seeking access to spatial data to inform their decisions, just as a growing range of policy-makers rely on spatial data to develop spatial change strategies. With this gravity comes the responsibility to curate spatial data and share it more freely, usefully, and readily than ever before.” [31]

-“The interactive graphical user interface (GUI) allows for data visualization manipulation and sharing.” [31]

-“The ability to harness and render this information in a location context is a major challenge.” [31]

-“GeoWiki, a geographical semantic wiki system that was introduced in [10], can parse and store multi-source geographical knowledge, mash up with Google Maps and serve geospatial decision-making.” [32]

-“A mashup is a Web application that aggregates multiple services to achieve a new purpose. It is an important feature of Web 2.0 and can be used with software provided as a service. With various mashup techniques, it is convenient for developers to obtain data from a variety of data sources on the web and to integrate these data to build new applications.” [32]

-“A map mashup therefore combines at least one map data source or services with added information, often geo-referenced to the map data, to create a new map.” [32]

-“Geoportals are a consolidated web-based solution to provide open spatial data sharing and online geo-information management.” [13]

-“Geoportals usually provide access to distributed data systems, offering maps, data discovery, and data downloads. Some of them are also capable of offering online analysis and processing service, enhanced semantic search engines, and dynamic visualization tools.” [13]

-“There is an ongoing challenge to effectively manage and communicate the vast and growing amounts of spatial data and geo-information.” [13]

-“Over the past few decades, the concept of geoportals has emerged as one of the key solutions for spatial data and geo-information accessing and sharing.” most recently through the internet. [13]

-“Geoportal: “a point of access to spatial data and geo-information. It is able to provide a geospatial data inventory linking to an inclusive collection of spatial data, geographic information, online services, and data processing tools. In this article, the use of the term ‘geoportal’ refers to the human-to-machine interface performing as a single point-of-access to spatial data and geo-information systems, offering sharing capabilities and connecting between geospatial data providers and end users. It is typically employed as a web-based graphical user interface (GUI) equipped with functionalities for accessing Earth observation data and geographical information.” [13]

-“Drivers of geoportal advances: “(i) scientific geospatial projects and applications, (ii) international organizations, (iii) governmental agencies, and (iv) commercial purposes”. [13]

- International organizations: Sharing data from “heterogeneous data sources”. [13]

- Governmental drivers: open government policies. “These portals, which were built by

answering the open government call (or named the e-government), act as the gateways, anchors, or major starting sites for governmental data, no matter whether the data is spatial or non-spatial in nature.” “Openness and transparency are fundamental to ensuring citizens’ trust in their governments. Thus the objective of a government geoportal is to foster greater transparency and accountability, providing information available to the public from digital technologies.” Example of Geospatial OneStop (GOS) in the USA. [13]

-Commercial geoportal: connect users to digital resources, including geospatial datasets. [13]

-“Data is the content that geoportals provide to end users. A single geoportal may provide heterogeneous geospatial datasets coming from multiple data sources.” Geographical and earth observation dataset types, additional products, models, and aggregation, data harmonization [13]

-“The geoportals can facilitate researchers, government officers, and ordinary users in helping them to find the data they needed, with basic searching services equipped in the geoportal.” [13]

-“a geoportal is normally a general-purpose tool to discover spatial data and geo-information, but often there are community-specific needs that require customized geoportals with dedicated tools for scientists, policy-makers, and the general public.” [13]

-“GEoportals should also play a role as the interface of so-called social spatial data infrastructure. New system architectures (e.g. Linked data and Semantic Web) to establish a shared information space relying on URLs and Resource Description Framework (RDF) can also provide another solution for accessing geoportal data.” [13]

-“Geoportal is considered as the entry point and the human-to-machine interface of the data and information management system.” [13]

-“Promote a common way of accessing data for an enterprise and any communities that need to make use of that data” [37]

2.2.7 Commercial geospatial platforms

-The definition of the search criteria by the user is usually done based on a form, defining the spatial, temporal and thematic coverage of the searched data. In the context of geo-data search the use of maps on which searched areas can be defined have become standard – here the region of interest is selected by identifying a point, a rectangle, a polygon or some pre-defined region (e.g. a country). In most cases alternatively geographical names or coordinates can be searched. Especially the latter approach is less intuitive, but also the use of geonames requires some care as very often the way in which names are written differs depending on the language and special national character sets are used. [3]

-Platform ecosystem: “technologies, platforms, data points, user interfaces, relational infrastructure and application programming interfaces (APIs), service providers, investors-financiers and customers” [33]

-“the platform’ is not only a particular computational structure but also a discursive innovation, which works on behalf of its business owners to recast conditions of value-sharing between companies and their marketplace as ostensibly ‘open’. These ecosystem-relationships make possible conditions of algorithmic control, subtly shaping, nudging and ‘steering’ urban behaviors from infinitesimal to global scales, in ways that point to new asymmetries of urban governance in a world of connected devices, things, people and infrastructures.” [33]

-GPS enabled smart phones enables the superior experience of ridesharing to traditional taxis. [33]

- “No local knowledge is required: the app provides the requisite information to fulfill the transaction”. [33]
- “the value harvested by platforms is in the data generated through digitally mediated transactions and relationships, which informs the terms dictated by platform companies for wider data-driven applications.” [33]
- “The computational architecture put in place across a platform ecosystem hinges fundamentally on the role of the Application Programming Interface, or API, which enables different kind of software to connect.” [33]
- “Platform architecture can thus be read as highly recombinatory, ingesting and harvesting data from a highly porous, intermediated platform ecosystem of multisided relationalities and transactions, while underpinned by the protocols and modes of value-exploitation that are closely controlled by platform owners.” [33]
- “The tactics adopted by platforms firms to entice users to choose one service over another are described as ‘steering tactics’. For economists, steering tactics are central to the operations of successful platforms, which devote significant attention towards and investment in strategies that can enhance the overall value of the platform to its users.” [33]
- “geolocation organized five ‘platform affects’ for digital platform ecosystems. (i) affective space-times of hyperlocality and real-timeness; (ii) experiential affects of smoothness; (iii) affects of connectedness; (iv) affects of trust; and (v) affective value.” [36]
- “THe merging of digital location with web content, services, and interfaces is pervasive to the extent that digital location - or simply ‘geolocation’ - now features as a native design, logistical, and organizational component of digital platform interfaces, utilities, and affordances.” [36]
- Social media updates (geotag), ridehailing (geologistics), dating (search radius definition), accommodation (interactive map-based filtering) [36]
- “Initial work by geographers has begun to preliminarily attend to the factors atively underwriting the integration of geolocation with digital economies. Yes this work has been largely market-oriented, engaging geolocation as first and foremost an economic commodity.” [36]
- Platform affect: “informs an understanding of digital platform ecosystems as affectively oriented and as expressing a capacity for orienting affects. Platforms generate, accumulate, and cycle affects within ecosystems; they also draw on the affective capacities of other techno-material formations, such as geolocation, to organize affects for digital ecosystems.” [36]
- “as geolocation is put into circulation, it expresses a capacity to affect - perturb, animate, align, mobilize, organize, dis/assemble - other things, both humans and other-than-human alike. Affect is useful for grappling with what geolocation ‘does’, how it ‘does’ what it does, and to what ends.” [36]
- “geolocation organizes affects for platforms by investing users within platform ecosystems through individual and collective attunements towards using, participating within, contributing to, remaining within, and returning to these ecosystems.” [36]
- “The integration of geolocation in the form of positional awareness as a native component of digital platform design invests individuals in highly structured yet simultaneously ontogenetic hyperlocal microgeographies that solipsistically spatially and temporally centre the user.” [36]
- “These space-times invest users within platform ecosystems by literally positioning them at the centre of their own platform-mediated universe, affectively attuning users towards returning to the platform to regain access to a service brokered solipsistically to their hyperlocalities and real-times.” [36]

- “Certainly, the instantaneousness and hyper-locality of platform mediation is significant to the competitiveness of platforms such as Uber in increasingly crowded platform economy marketplaces.” [36]
- “beyond understanding these space-times in terms of the business strategies that they may underwrite affect theory also capture the ways in which once users are affectively invested in these ecosystems via geolocation-organized space-times of the ridehailing platform, these platform-mediated affects ‘turn back’ on themselves... organizing an affective frame of reference which conditions user’s expectations of future experiences of the platform: a potential ride available frome very possible location within minutes of being requested.” [36]
- Amazon prime: “the ability of platform enterprises to operationalize these space-times for consumers through drone delivery is contingent on a complex of geolocation affordances, data, hardware components, and software technologies operationalized in concert,... also expands the range and possibility of modalities through which Amazon may organize affect that retain consumers within its ecosystem... through the leveraging of geolocation to organized the desired hyperlocal rea-times that align both online and offline consumption experiences with the expectations of consumers.” [36]
- Geolocation as “an affective capacity as an instrumentality that organizes and aligns experiences of platform ecosystems. This especially pertains to consumptive experiences (of goods and services), which are affectively smoothed through the leveraging of geolocation as a coordinating force that produces seamless digital consumption experiences not only within but also across platform ecosystems.” [36]
- “Streamlining the retail consumer experience across all stages of a purchase, from shopping online through to last mile fulfillment... here, geolocation is leveraged as a real-time coordinating force that links and streamlines multiple stages of an online-to-offline retail transaction so as to render the consumption process as experientially seamless as possible for the end consumer.” [36]
- “Geolocation is also leveraged to organize affects of smoothness across platform an application ecosystems through practices of digital aggregation.” [36]
- “By integrating multiple urban transportation asset providers, infrastructures, and operators within a single app ecosystem, Transit smooths urban mobilities by investing users within a singular, cohesive, integrated digital platform experience that mitigates the impracticalities of switching between multiple branded apps of mobility actos both across transportation modalities... and within a modality class.” [36]
- “Inarguably, digital location has become central to quotidian experiences of digital platforms, be it through the ways in which content is accessed and interacted with (location-aware search, spatial interfaces); the ways in which we generate and contribute content (geotagging, personal location data mining); the ways in which we rely on location-aware devices such as sensors and digital assistant to produce ‘smart’ environments for us (smart cities, smart homes); and also the ways in which we interact with the world (augmented and virtual realities)... Digital location is also an affective instrumentality that organizes how platforms and platform ecosystems are encountered and experienced by end users, and where thesmoothness of these experiences - as underwritten and facilitated by geolocation - is central to the ways in which users themselves become affectively invested in platform ecosystems, coming to see platforms as indispensable to how they get around cities. “ [36]
- “geolocation also functions as a techno-discursive instrumentality that underwrites the affective expression and enactment of social, familial, and romantic relations of connect-edness performed through the sharing and monitoring of personal locations via digital platforms and their affordances of connectivity.” [36]

- “digital location is itself also being leveraged as an instrumentality of securing trust in platform ecosystems.” [36]
- “Mobilizing an understanding of location as implicated in arranging transparency and security form platforms informs a nuanced understanding of how it is that users become invested in platform ecosystems despite the ways in which platforms themselves undermine user confidence in digital systems through practices such as psychometric profiling and microtargeting..., covert user experimentation..., and regular data security breaches, including those involving the theft of personal location data”. [36]
- “As an affective capacity to invest users in digital ecosystems, geolocation is leveraged to build trust in platforms by organizing technics and discourses of authorization (digital transactions), and of securitization (against misinformation).” [36]
- Geolocation “invests developers, marketers, and entrepreneurs in the speculative potential of geolocation - in the form of embedded interactive map objects and interfaces - to organize annexation economies for end users. These attention economies are predicated on investing users themselves in digital platforms by ‘attracting eyeballs’, expanding opportunities for user engagement by providing additional interactive elements (interactive map objects), and extending the duration of user visits (how long users remain within a digital ecosystem).” [36]
- “geofilters are a brand growth strategy realized mobilizing real-world location... Here again digital location is being discursively mobilized to affectively attune digital developers and advertisers to the capacities of geolocation to both organize spatio-temporally contingent user attention economies and to ultimately convert that attention into both online and offline consumption.” [36]
- “Geolocation expresses a capacity to align and attune corporealities and inorganic materialities in ways that work to invest users in platform ecosystems by organizing and orienting affects for platforms.” [36]

- Uber
- AirBnB

2.2.8 Location extraction from text

- Identify place name: Stanford Core NLP, OpenNLP; Identify placeal nd disambiguate: CLAVIN, Yahoo placemaker; Geography3 [42]
- “Extracting location names from informal and unstructured social media data requires the identification of referent boundaries and partitioning compound names. Variability, particularly systematic variability in location names challenges the identification task.” [34]
- Location name contraction problem: “pragmatic influences on writing style shorten names to reduce redundant content in social media.” [34]
- Nameheads: “complex phenomenon of alternate name forms”. “1) Appellation formation, 2) explicitly metonymy, 3) category ellipsis, and 4) location ellipsis.” Appellation formation: removing the designator of a specific thing (O PONTE vs. PONTE DE 25 DE ABRIL). Explicit Metonymy removes the thing and leaves only the designator (NOVA vs. UNIVERSIDADE DE NOVA). “Both Appellation formation and metonymy pose disambiguation problems, and require context such as the author’s location to resolve.” “Category Ellipsis and Location Ellipsis pose delimitation problems that can be resolved with a statistical language model.” Category Ellipsis: “the author strips words related to the location category” (LISBOA vs. CIDADE DE LISBOA). “Location Ellipsis occurs when an author drops the specific location reference in the location name”. Entity delim-

itation: “to identify the boundaries of a location mention in the text”. Previous efforts invoke the application of heuristics (syntactic and semantic) with fallible results, challenge linking to gazetteers. Collocations: “a sequence of ordered words... neither strictly compositional nor always atomic.” “Given a region-specific gazetteer, which retains the same location-context as the text, we can construct a statistical model of the token sequences it contains.” [34]

-“Tokenize all location names in the gazetteer to construct the n-gram model and then save the resulting lists of unigrams, bigrams, and trigrams.” From the gazetteers, try to predict all the different ways one might reference location (assumed shortening of the official names) and create a new list. [34]

-“The field of geolocation extraction collectively involves many different tasks and analyses to be performed over text. The three main tasks among these are: (i) Location named entity extraction; (ii) Location named entity resolution; (iii) Event’s location extraction.” [35]

-Even though several geoparsers such as Cliff-Clavin, Mordecai, and Stanford-CoreNLP have been developed to automatically extract named locations from unstructured English text, location extraction from a text is still a challenging task due to the complexity, diversity, and ambiguity of location information in different languages. However, these tools cannot extract the focus location with good accuracy, and most of them cannot differentiate between different locations in the text - i.e. focus locality versus non-focus locality - and are not language agnostic.” [35]

-“While traditional named entity taggers are able to extract geo-political entity and certain non geo-political entities, they cannot recognize precise location mentions such as addresses, streets and intersections that are required to accurately map the news article.” [1]

-Named Entity Recognition (NER) taggers: “contain tags to identify organizations, geo-political entities (GPE) and certain non-CPE locations such as mountain ranges and bodies of water from text”; “they are not able to extract precise locations mentions such as addresses, streets or intersections in their entirety.” [1]

-“Tools trained on conventional NER models.... Have been successful in identifying common named entities. However, challenge comes when high level of granularity is of interest in extracting location entities such as specific addresses, streets, or intersections.” [1]

-“Fine-tuning pre-trained language model fo domain-specific machine learning tasks has become increasingly convenient and effective”. [1]

-Process: “named-entity tagger for the task of precise location extraction involves fine-tuning an existing neural network on a target dataset. [1]

-“Geocoding is the process of taking input text, such as an address or the name of a place, and returning a latitude/longitude location on the Earth’s surface for that place.” [1]

-Geotagging: 1. Entity feature vector extraction, 2. Gazetteer record assignment, 3. Geographic name disambiguation/toponym resolution, 4. Geographic focus determination [43]

-“Unlike the spatial information used in a Geographic Information System (GIS), spatial information obtained from web documents is often incomplete and fuzzy in nature. A GIS user can formulate data retrieval queries specifying complex spatial restrictions, while a search engine targets a wide variety of users who only provide simple queries. However, these users are also in need of retrieval mechanisms for queries with geo-spatial relationships. In order to support this, a first step concerns with assigning geographical scopes to web resources, so that the same resources can latter be retrieved according to geographical criteria.” [6]

-“Machine learning provides effective techniques for text classification, involving the au-

tomatic generation of classifiers from manually annotated training data. However, with very few exceptions, most work in automated classification has ignored the presence of hierarchically structured classes and/or features.” [6]

- “Much of the contextual information that could be used to disambiguate the geographical scopes in natural language texts is absent or external to the texts. The amount of training data per feature is also so low that there are no repeatable phenomena to base probabilistic methods on. For instance, the frequency of location names is in itself not sufficient for a good classification, as the same location name will usually not be repeated, even if the name is important.” [6]

- “Disambiguating geographical references in the text and assigning documents with a corresponding geographical scope are two crucial steps in building a geographical retrieval tool.” [6]

- “Previous studies have demonstrated that recognising geographical place names in text (usually called geo-parsing) is a crucial precondition for geo-referencing web documents. In language processing, the task of extracting and distinguishing different types of entities in text (i.e. names of people or organizations, dates and times, events, geographic features or even ‘non entities’) is referred to as Named Entity Recognition (NER).” [6]

- “Ambiguity is the main problem associated with geographical references in text... ambiguity in geographical references is bi-directional, as the same name can be used for more than one location (referent ambiguity), and the same location can have more than one name (reference ambiguity). The former has another twist: the same name can be used for locations as well as for other class of entities, like persons or company names (referent class ambiguity).” [6]

- “To be useful, NER systems focusing on geographical concepts should handle the complex issues related to how people use geographical references. Place names lack precision in their meaning, and often vary with time, from person to person, and with the context in which they are used. Many times place names are simply temporary conventions, and people’s vernacular geography is also often vague, as they may also be interested in the vicinity of a place without knowing its exact name. Not only spelling variation are common on geographical names, but also the places those names reference change in shape and size.” [6]

- “About 60% of all data (textual and otherwise) are geospatially referenced”. [44]

- Geographic information retrieval: GIR [44]

- Overall, some recurrent issues persist in geoparsing research: methods are not evaluated or evaluated on non-public datasets or using proprietary systems (inaccessible, or behind paywalls): demonstration and evaluation rely on small tasks, or small gazetteers, or small geographic scope: and/or focused harvested corpora are used in evaluations that greatly simplify the toponym resolution task.” [44]

- In unstructured, textual data: “the spatial data is specified using text (called toponyms) rather than geometry, which means that there is some ambiguity involved.” [45]

- “The ambiguity has an advantage in that from a geometric standpoint, the textual specification captures both the point and spatial extent interpretation of the data (analogous to a polymorphic type in parameter transmission which serves as the cornerstone of inheritance in object-oriented programming languages). On the other hand, the disadvantage is that we are not always sure which of many instances of geographic locations with the same name is meant” [45]

- Geotagging: “The process of identifying and disambiguating references to geographic locations (i.e., toponyms), known as geotagging, consists of two steps: toponym recognition, where all toponyms (e.g., “Paris”) are identified, and toponym resolution, where

each toponym is assigned to the correct geographic coordinates among the many possible interpretations (e.g., “Paris” which can be one of over 140 places including France and also Texas). Geotagging is difficult because the first step involves understanding natural language, while the second step requires a good understanding of the document’s content to make an informed decision as to which of the many possible locations is being referenced.” [45]

-Toponym recognition: “the most common strategy is simply to find phrases in the document that exist in a gazetteer, or database of geographic locations, and many researchers have used this as their primary strategy.” [45]

-Section 2 for shortcomings of current geotagging strategies [45]

-Fuzzy geotagging: “does not fully resolve toponyms in a single article, instead returning sets of possible interpretations for ambiguous toponyms.” [45]

-“Geocoding is the process of parsing places and addresses written in natural language into canonical geocodes, i.e., one or more coordinates referring to a point or area on earth.” [46]

-“Geographic description in texts reflect human conceptualization and experiences of space and places. Different from other forms of geographical data, text-based spatial description are subject to all sorts of ambiguities that prevent effective use.” [5]

-“Geospatial referencing textual documents refers to the task of discovering location phrases and creating unambiguous representation (or footprints) of the meaning of those textual references.” [5]

-“To create unbiased corpora that capture the complexity of natural language and place name ambiguity and annotate with ground-truth toponyms from high coverage and detailed gazetteers, documents used for testing and training should be manually ‘geo-annotated’ by human annotators, i.e., place names should be recognized (segmented) and manually resolved to toponyms in gazetteers. The process of manually tagging (segmentation) and annotating place names in text with entries (toponyms) from a geographic gazetteer, here called ‘Geo-Annotation’, is laborious, costly, time-consuming and error-prone. The scarcity of publicly available geo-annotated corpora can partially be attributed to the lack of available efficient software infrastructure capable of facilitating this laborious task.” [?]

-“Extracting the ‘correct’ location information from text data, i.e., determining the place of event, has long been a goal for automated text processing.” [47]

2.2.9 Event extraction

-“Event extraction is the process of recognizing defined event types in text (e.g. “attack” or “protest”) and extracting and classifying the actors involved in the events.” [48]

-“To be useful in subnational research, these events require information on the location where they occurred. A second related information extraction task is ‘geoparsing’, the process of recognizing place names in text (‘toponym recognition’) and resolving them to their coordinates or gazetteer entry (‘toponym resolution’). [48]

-Tokenize a sentence. Potential for multiple events per sentence, potential for zero, one, or multiple tokens per event. Tokens labeled as 1 (if location of event) or 0 (otherwise). Assumes “that events are ‘anchored’ by a verb, is a common assumption in semantic role labelling, a closely related task to event-location linking”. [48]

-“Many existing open source geolocated event datasets, including GDELT and Phoenix, make no effort to explicitly link events and locations, simply returning a top location from a sentence, without using information on the extracted event to inform the geolocation step, which has also been used in NLP”. [?]

-“Primary focus location” of Imani (2017): “makes the simplifying assumption that docu-

ments have one single, fixed ‘focus location’ that is invariant to different potential events in the document.” [48]

- “While humans are able to pick up on nuance and deal with grammatical complexity that machines still cannot handle, humans are also unsuited to the tedium of labeling thousands of sentences and may be susceptible to drift in their definitions or understanding of the task. Not only is the automated method vastly cheaper and faster than the human process, it does so with accuracy at least as good.” [48]
- “Text also holds a great deal of factual information and new techniques are needed to allow researchers to extract political information from text.” [48]
- “Many quantitative studies of conflict rely on event data. Recently, these studies have also retreated from the country-year framework and have focused on disaggregating the event flows both in terms of space and time. Disaggregating temporality - even to the daily level - is a straightforward task. But figuring out precisely where an event actually occurred is a difficult and uncertain task that has been perplexing for most contemporary event data efforts”. [47]
- “Most data in the conflict realm comes from non-official sources. For many that means some form of data collected from historical and journalistic sources. This need is often filled by event data, which are typically collected on a daily basis, and can be aggregated temporally to the level required by the analysis. Even data can also be aggregated to the geographical region that is appropriate. Given the increasing demands from event data, the scientific community has recently devoted significant efforts to automate the data collection process. Having humans read and code a large set of archive documents sometimes limits reproducibility, and hence hinders scientific research. It is also expensive and limits the currency of the data. Further, ensuring inter-coder reliability is challenging, especially over global events that span decades.” [47]
- “These automated event data allow researchers to observe and extract information on politically relevant events around the world in near real-time.” [47]
- “Outstanding issues [in automated data collection] include machine translation of texts in foreign languages (wherein great progress is being made in both Chinese and Arabic), duplicate reports from multiple sources, and the relatively low accuracy in determining the event location.” [47]
- “Named entity recognition in the context of geolocation involves determining which words in the given sentences are location names. In principle, the task of capturing location names from texts can be done easily by using a dictionary. In practice, however, developing a dictionary that is sufficiently comprehensive for such a task may be challenging. To begin, the geographic boundary of the texts being analyzed may be unclear, given that the domain of many event data is the entire world. Further, because conflict events often spread to new and rural places, texts may include location names not defined in the gazetteer. Still further, a location name may be written in multiple forms, requiring the dictionary to comprise every variant for each location.” [47]
- Ground truth: hand coded set of actual locations for training and verification [47]
- “The task of determining event locations involves three steps, each non trivial.” 1. named entity recognition (NER): “all location names are identified and extracted from an appropriately preprocessed text. This step is a prerequisite for the other steps because to determine the location of an event in a news article, capturing the exhaustive list of location names is required.” 2. Ambiguation/resolution: “which involves identifying the actual location of the recognized name string. Once this is accomplished, it is possible to extract the ontologically defined meaning from the text in terms of who does what to whom, and when and where”. 3. Determine if disambiguated location names are the event

location. [47]

- “geolocating events (identifying the location of the event described in a document) is an objective for many scholars, particularly those who intend to collect and build original databases from text corpora, be they news articles, congressional records, campaign speeches, party constitutions, or twitter feeds. While automating this task will aid many, the research avenue in this topic is still under development.” [47]
- “Given that a substantial number of location words are incorrect event locations, the automated event data community needs a better coding scheme that can reduce the error rates.” [47]
- Event-relevant: “those locations that are part of the main description of the event of interest, i.e., all locations that are key to the narrative of the event of interest.” [47]
- Event-occurring: “all locations where events occurred regardless of whether the event is the event of interest.” [47]
- Eventirrelevant and event occurring location: “such could occur when the raw texts contain news summaries of events that are not of interest.” [47]
- “An N-gram is a sequence of N words. Collections of N-grams are known to provide valuable information about each word in a phrase, taking into account the complexity and long distance dependencies of languages... Given that the collocation patterns in which the event-occurring location words appear differ from those of the non-event-occurring collocations, teh N-gram patterns are able to provide the contextual information of even-occurrence to our classifiers.” [47]
- “Geolocation inference at the event level estimates the location of events mentioned in text. This level of inference predominantly relies on geoparsing the process of identifying geolocations in text and disambiguating between multiple toponym references... event geolocation inference might not reflect the actual location of individual tweets.” [41]
- User geolocation: “user locations can be predicted by utilizing toponym references within their tweets a well as user metadata such as friend networks and time zones.” [41]
- Tweet level (article level): “estimates the location of individual tweets. THis differs from user-level prediction in that a tweet might be posted in a separate location from where they live, such as during a vacation or work hours.” [41]

Examples:

- Cliff-Clavin
- Mordecai
- Stanford-CoreNLP
- BERT: BERT-based language model. Google’s Bidirectional ENcoder Representations from Transformers, “powerful, pre-trained deep learning based language models” [1]. -BERT: “a general purpose language representation model pre-trained on millions of articles on English Wikipedia an BookCorpus.” [1]
- Wheba-where: Web-a-Where: “This small size imposes a serious limitation on Web-a-Where’s practical geotagging capabilities, as it is unable to recognize the small, highly local places that are commonplace in articles from local newspapers.” [45]

Tumba

- Tumba: a web search engine for Portugal (Gomes & Silva 2003) [6]

- “We propose to further improve the quality of search systems, by integrating the geographical knowledge that can be inferred from web resources.” [6]
- Scope: “the region, if it exists, whose readers find the page more relevant than average. A geographic scope is specified as a relationship between an entity on the web domain (a HTML page or a website) and an entity in the geographic domain (such as a location or administrative region). The geographic scope of a web entity has the same footprint as the associated geographic entity.” Geographic scope definition for reader engagement (not generically organizing location for research or general applications later [6]
- “Geographic information is pervasive on the web. An analysis of 3,775,611 pages 8,147,120 references to the 308 Portuguese municipalities (administrative division of the territory corresponding to populated sub-regions of the Eurstat NUT 3 area), an average of 2.2 references per document.” [6]
- Geo Tumba provides a retroactive organization and characterization of existing web-pages, versus specifically defined ruing writing. [6]
- “The Portuguese ontology included more place names, but results in terms of recall are inferior to experiments using hte smaller global ontology... results indicate that recall does not improve considerably with the amount of available place names.” [6]
- “Our framework differs on the emphasis put on geographic name entity recognition, the use of a graph ranking method for assigning a single scope to each document, the extensive use of names instead of coordinate information, and the availability of ontologies associating entities to geographic scopes.” [6]

LocateXT

- Search unstructured data for spatial locations [49]
- Geneartes point features representing locations [49]
- Attributes include file, context related text, dates, keywords [49]
- Not addresses [49]
- Custom locations: “When documents are scanned, they are examined for place names specified in a custom locations file. The custom locations file associates a place name with a spatial coordinate. A point is created in the output feature class to represent each location found.” [49]
- fuzzy matching: applies to custom locations. 70% of alphanumerics match [49]
- Require word breaks setting “when word breaks are required, text is considered a word when it is bounded by whitespace or punctuation characters as in European languages.” [49]
- Require word breaks setting “when word breaks are required, text is considered a word when it is bounded by whitespace or punctuation characters as in European languages.” [49]
- Custom location: “is defined by associating text with a spatial coordinate. Current and historical places, and natural features and structures can all be defined as custom locations in a custom location file (.lxtgaz).” [49]Input your own gazetteer (custom or otherwise). Again, only points (not polygons).
- Integrated Crisis Early Warning System (ICEWS) [47]
- Open Source Event Data Alliance [47]

GeoTxt:

- a scalable geoparsing systems [44]
- SPIRIT: spatially aware information retrieval on the internet [44]
- Geographic footprint of documents [44]
- Heuristics for toponym disambiguation: spatial proximity (“spatial minimality”), co-occurrence (relation based on spatial hierarchy levels) [44]
- Appropriate for hyperlocal applications: “advantageous on a corpus with a smaller known geographic footprint” [44]

DeepGeo [41]

2.2.10 Gazetteers

- “The locale name for an object within a language area is termed an endonym. Outside this language area, the same object may have other names according to the respective language. These variations are called exonyms.” [19]
- “The United Nations Group of Experts on Geographical Names (UNGEGN) is run by specialists from the fields of linguistics, cartography and history. The UNGEGN requests national gazetteers (alphabetical lists of names, with coordinates and other data) to promote the use of nationally standardised names on maps and in written documents.” [19]
- “An exonym is a place name that isn’t used by the people who live in that place but that is used by others.” [50]
- “a locally used toponym - that is, a name used by a group of people to refer to themselves or their region (as opposed to a name given to them by others) - is called an endonym. [50]
- Gazetteers assist in geographical entity identification performance [6]
- “The larger the gazetteer (in terms of coverage and detail), the harder the task of toponym resolution, since the gazetteer will contain a higher number of ambiguous names. However, a large gazetteer such as GeoNames enables resolving more places to toponyms.” [44]
- Progressive geo-coding environment: “this configuration of geo-referencing system ensures that a local gazetteer is incrementally enriched through local community use and the demand for the effort of human geo-coding is kept to the minimum.” [5]
- “Gazetteers are well-structured and maintained, but they have limited coverage on those less known places at local levels. GIS databases have geographical features in greater geometric precision and details, but they tend to have limited textual metadata for matching to place names. Open Source geographical information now comprises and ever growing part of geographical knowledge, and they frequently include vernacular descriptions of locations, as well as references to imprecise areas. The problem with using CGI is that their local coverage is uneven from locality to locality, and they tend to have various degree of quality and trust. Ideal VGI is supposed to be produced entirely by individual citizens that reside in their locality, and therefore potentially represent the best knowledge about their local.” [5]
- “Majority of gazetteer entries use points (coordinates) to represent areas (such as a city), which is considered inadequate for local features. [5]

Gazetteer examples:

- OpenStreetMap: open data, licensed under the Open Data Commons Open License (ODbL)

by the OpenStreetMap Foundation (OSMF): [51]. Relies on OSM “performed the best” ... “DBpedia is not focused on geographical information; therefore, it does not contain the metadata useful for the system’s future use (e.g., extents and full addresses). Also, OSM has more fine-grained locations and more accurate geo-coordinates than Geonames”. [34].

Use [34]

- Geonet Names Server (GNS): US NGA maintained: GeoNames. “The GEOnet Names Server (GNS) is the official repository of standard spellings of all foreign geographic names, sanctioned by the United States Board on Geographic Names (US BGN). The database also contains variant spellings (cross-references), which are useful for finding purposes, as well as non-Roman script spellings of many of these names. All the geographic features in the database contain information about location, administrative division, and quality. The database can be used for a variety of purposes, including establishing official spellings of foreign place names, cartography, GIS, GEOINT, and finding places.” [52]. Use [34]; “The GeoNames gazetteer was chosen for GeoText because of its extensive coverage, quality, inclusion of metadata (such as alternate names and geographic hierarchical information), and frequent updates”. [44]; Official sources include U.S. National Geospatial Intelligence Agency (NGA) and the U.S. Board on Geographic Names, “assuring extensive coverage and quality throughout the world.” [44] “GeoAnnotator uses GeoNames as its default gazetteer because of its frequent updates, ease of ability by years to correct information, extensive coverage and quality and inclusion of metadata items necessary for geoparsing/geo-annotation, such as alternative names and spatial hierarchies. All GeoNames toponyms have a unique identifier, enabling creation of corpora useful for linked data applications.” GeoNames: “the richest gazetteer available at the time of this writing.” [2]

-DBPedia [34]

Requires premium subscription for polygon access. -Getty THesaurus of Geographic Names (TGN): http://www.getty.edu/research/conducting_research/vocabularies/tgn/ “TGN is a structured vocabulary including names and associated information about both current and historical places around the globe.” [6]

2.2.11 Local lexicons

-Automatic inference of local lexicons: “1) Stability: a Local lexicon is constant across articles from its new source. 2) Proximity: Toponyms in a local lexicon are geographically proximate. 3) Modesty: A local lexicon contains a considerable but now excessive number of toponyms.” [45]

-“a spatial lexicon can be classified as a local lexicon if and only if the toponyms within it are geographically proximate.” [45]

-“Associating a single local lexicon with each data source allows for a variety of applications. However, it may be possible to fine-tune the use of spatial lexicons in situations involving different types of content. For example, a blog may track several different topics simultaneously, and use different spatial lexicons for each topic. Furthermore, individual authors may write for specific audiences as well, as in the case of journalists stationed in certain geographic areas and concentrating on stories in that area. It thus might be beneficial to determine separate spatial lexicons assumed by different authors, and further improve geotagging performance. Or generally, we might associate a particular spatial lexicon with any type of entity found in each document, be they authors, persons, organizations, or particular keywords. For example, upon finding a mention of ‘Robert Mugabe’, we might assume a spatial lexicon including Zimbabwe and nearby locations, even without specific mentions in the text.” [45] Complicated yet potentially effective alternative to

manual association. [45]

- “It would also be interesting to detect and observe evolving spatial lexicons over time for data sources with evolving geographic interests, thus further improving geotagging on these sources. For example, the first few articles of an ongoing, prominent news story will often fully specify the toponyms relevant to the story. Later articles in the series, however, will often underspecify the same toponyms, since they have already been introduced into the audience’s spatial lexicon and been fully resolved in early articles.” [45]

- “As newspapers and other data sources continue to move into the virtual space of the INternet, knowing and using spatial lexicons will be ever more important. Previously localized newspapers will cater to a broader, global audience, and thus will adjust their notion of their audiences’ spatial lexicons, perhaps limiting or ding away with an assumed local lexicon altogether. On the other hand, as more and more people publish highly individual and geographically local content, inferring individual local lexicons will be a necessity for correct geotagging. Geotagging with knowledge of local lexicons will thus continue to play a large role in enabling interesting geospatial applications. [45]

2.3 News localization

- News is a distribution of information to humans which in turn affects their actions. We have manners of measuring the physical world via sensors, and now are layering in emotional responses. The news is a feedback loop that is at once an input to human decision making and a report of some of those outputs.
- “Everything happens somewhere and some- whence. Spatial and temporal aspects of data lead to critical insights into the information contained in it.” [31]
- “in the current hypermodern context, a smart city must also be able to identify the main components of an event (where, when, who, what, how), to analyze it, to provide location-aware (and contextual) sense to it and, to react (actuate) properly (and in real-time - at least compliant with the nature of this event). [8]
- “Spatio-temporal aspects of data lead to critical information.” [31]
- “A media or news desert is an uncovered geographical area that has few or no news outlets and receives little coverage. Mapping locations mentioned in news articles is the primary step in identifying news deserts.” [1]

2.3.1 News evolution

- “There’s a lot of opportunity for innovation across the news ecosystem, and this is also true for local publishers... They’re connecting the dots with their communities and are trying to build innovative and more engaging experiences using local data.” LB [18]
- “Having the tools to access and analyse this data is now critical for effective reporting... There are many powerful public interest stories out there that will only be discovered if traditional investigative techniques are combined with technology” Megan Lucero [18]
- Linked in provided a business skillset “which you don’t often get in journalism because we’re, you know, so adamant about separating the editorial and commercial side of things.” Now this is more present in journalism (as of the past 5 years). [53]
- “I really believe in the community element of journalism. So I’m really interested in not just broadcasting but hearing back and bringing people together. It not where I want it to be, but I think by calling it a community I’m stating that intention.” [53]
- Matching product to audience: “You have to throw a lot of lines out in the water and see

which ones catch. So you end up having a lot of platforms.” [53]

2.3.2 Personalization

- “While personalization of content is recognized as a powerful advantage for publishers and readers alike, it can mean that people aren’t often exposed to a range of viewpoints on a particular subject.” [18]

2.3.3 Blogs

- “Blogs, as social software, are designed for rapid content creation, archiving, and syndication within online communities. Due to the community intelligence, blogs have lately become a valuable source of information. Blogs, used as collaboration environments, support people in associating tags with content that they generate, share, or consume within a community. More generally, by tagging content in the web, one makes this content shareable to other web participants, as well as links it to other web contents.” [32]

- “There has been very little research on geoparsing microblog text.” [44]

2.3.4 Indie journalism

- Indie journalism: “independent side-projects” as entrepreneurial endeavors of journalists. [53]

2.3.5 Local journalism

- “The news industry may be in flux, but it’s vitally important that good local journalism doesn’t disappear.” Megan Lucero, Director at The Bureau Local [18]

- “Local journalists play an important role in democratic society by holding power to account. However, newsrooms have fewer resources to dedicate to much-needed investigative reporting.” [18]

- “While almost every reader survey concludes that local news is the most highly valued of all content, making it pay is another matter entirely.” [18]

- “The appetite for hyperlocal news is vast” [18]

- “The core problem in local and hyperlocal news is that there is no sustainable business model,” says Gabriel Kahn. “These areas have no voice, they have no way of discussing problems in the community and, whether it’s governments or companies, they have no accountability.” [54]

- “Local news is a medium for communication among residents, and they reflect the nature of local geographical knowledge. THe residents’ local knowledge concerning their living environment is often invisible, descriptive and vague, and thus difficult to collect.” [5]

- “Local newspapers can be considered as one important tool to support the accumulation and sharing of local knowledge.” [5]

- “Because local news serves the important role of being repositories of political, social, and cultural knowledge in the community of practice, it is important to make such knowledge accessible to community members who can maintain the awareness of what the community have known or done in the past when facing a new or repeating problem.” [5]

- “citizens’ attention to local news promoted political participation.” [5]

- “Local news are textual artifacts of human experience on the places they live and interact with, and they play an important role in the making of place meaning. They reflect what local people think and are written for the local people” [5]
- “Hyperlocality is a concept that has emerged in both online journalism and in mobile advertising circles to designate community-oriented digital content endemic to small geographies such as neighbourhoods, streets, or single postcodes.” [36]

2.3.6 Georeferenced News

- “Once a blog system organizes the entries with geographical semantics and display localities, spatial analysis can be performed based on the user-tagged information. In order to develop such system, the domain specific ontologies that take geographical knowledge into account should be built.” [32]
- “In the era of web 2.0, blog systems such as GeoBlog provide an approach for individuals to input various geographical data and the asynchronous functions to meet the asynchronous interaction requirements. This apparently widens the data source of geographical information applications. [32]
- “Considering the map as an object that conveys meaning, we had to recognise not only the agency of the participants but the agency of the object itself. Yes, this mapping of experiences has the potential to be a powerful source of information, and it also has the power to label places as safe, dangerous, racist, oppressive, accepting, and so forth.” [14]
- “Participants and researchers come with assumptions wrapped in powerful collective memories. Acknowledging them is important, but we also need to look for disturbances and consider the impact of those disturbances.” [14]
- “If you are a journalist for a news outlet where location is a key part of the story (such as a local or hyperlocal news site), it is worth geotagging your stories to help search engines deliver local results.” [55]
- “Location-based feeds have huge potential. They may further develop as a way Google and other search engines return results based on a person’s location; they may be used by social newsreader apps as a way of delivering [?]
- “Political news reports are populated all over the world in various languages. It has a great value to automatically detect the geolocation from these reports for better understanding of the associated events.” [35]
- “Primary focus location as the actual location where the event occurred amongst other focus locations mentioned in the report.” [35]
- “determining crime pattern locations, predicting the place of protests and political unrest, and identifying the geolocation of natural disasters. Such applications can largely benefit from identifying precise geolocation information in a timely manner to provide better support for decision making.” [35]
- “News articles contain a wealth of implicit geographic content that if exposed to readers improves understanding of today’s news.” [43]
- “Given that much of the interest in news is motivated by location-related attributes of readers (e.g. where readers are situated, where from, aspire to be), it is somewhat surprising that they cannot deal easily with the two most common types of spatially-related queries: 1. Feature-based - ‘Where did story X happen?’ 2. Location-based - ‘What is happening in Location Y?’” [43]
- “The traditional newspaper layout “is linear and static, whereas the app interface is dynamic, in that the articles associated with a particular location can vary over time without disturbing the positioning of other articles.” [43]

- “an audience’s local lexicon plays a key role in how news authors write for their audiences.” [45]
- “geo-referencing local news can contribute to improved accessibility of community knowledge.” [5]
- “For newspapers targeted at local communities, geographical locations are likely to be in finer granularity and demand gazetteer-based approach” (vs. statistical language model approach) [5]
- “Due to uniqueness of each locality, there is a lack of local gazetteers that reflect the richness and details of local spatial language.” [5]
- “Location ambiguity in the local has its own unique nature that presents both challenges and opportunities to resolve them.” [5]
- In this study, only 16 percent of references could be mapped without ambiguity, 59% were discernible with the applied heuristics, and the remaining 12% “rely on human annotation to create footprints.” [5]
- Therefore, the gazetteer used for geo-referencing local newspaper articles should be place-specific.” [5]
- “What kinds of spatial language are used in the locality? Answers to this question will make clear the nature of gazetteers needed for geoparsing local newspaper.” [5]
- “Local news articles are an important source of knowledge about local events, place-specific culture, and peoples’ thoughts about their environment. Reliable geocoding of such articles is the first step towards unlocking such local knowledge for community engagement and development.” [5]
- “existing geo-referencing methods and tools do not work well for local news because they do not reflect the ways local people encode and communicate geographical knowledge.” [5]
- “For human coders, locating events by reading a news article may be time-intensive, but straightforward. This is not the case for machine-coding: many news articles contain multiple location names, such as the location of the journalist writing the story, the birthplace of a person being interviewed, or the place of a similar event that occurred several decades ago; at times, human names are identical to geographic names;; and location names are transliterated into English in a variety of potentially confusing ways. All these sources of noise in the data increase the difficulty in automatically locating events.” [47]
- “Further complicating matters, new articles often use nearby landmarks to indicate the location, in lieu of using the official names.” [47]

2.3.7 Automated location extraction from published corpora

- “The networks are global and spatial proximity is no longer a determinant of relations established in communities, as it used to be. These new communication models allow us to be in the world, however we run the risk of not reaching our street, or our city, and in it to see the reflection of the image of what we desire our urban environment to be.” [29]
- “To map local news coverage, it is important to extract precise location mentions from textual news content.” [1]
- The traditional newspaper layout “forces readers to perform a brute force sequential search (i.e. read the various articles while looking for mentions of the locations which interest them).” [43]
- “Determining the geographic focus of a document can be challenging, as not all documents have an easily identifiable focus, and not all locations referenced in a document may be related to its focus. For example, news articles often contain the address of the newspaper that published the articles.” [43]

- “One related issue is that stories may continually change and be updated, even after they have been ‘published’ in an RSS feed.” [43]
- “As retrieved, the story webpage is unsuitable for article processing, as it was meant to be read by humans and contains extraneous formatting markup and rendering scripts. . . . Furthermore, the extraction must be independent of the source website, as it is infeasible to create custom extraction rules for each individual website.” [43]
- “The successful execution of location-based and feature-based queries on spatial databases requires the construction of spatial indexes on the spatial attributes. This is not simple when the data is unstructured as is the case when the data is a collection of documents such as news articles, which is the domain of discourse, where the spatial attribute consists of text that can be (but is not required to be) interpreted as the names of locations. [45]
- “news articles (and more generally, documents on the Internet) are written to be understood by a human audience, and therefore geotagging will benefit from processing (i.e., reading) the document in the same way as an intended reader).” [45]
- “the reader’s spatial lexicon - those locations that the reader can identify and place on the map without any evidence - is very limited. In fact, even more importantly, this inherent limitation means that a common spatial lexicon shared by all humans cannot exist, which is one of the key principles used by systems such as MetaCarta and Web-a-Where.” [45]
- “the existence of a reader’s local spatial lexicon or simply local lexicon that differs from place to place, and that it is separate from a global lexicon of prominent places known by everyone.” [45]
- “The local lexicon is even more necessary when geographically indexing locations with smaller spatial extent which correspond to address intersections. . . , since street names are even more ambiguous than regular toponyms.” [45]
- “While statistical NER methods can be useful for analysis of static corpora, they are not well-suited to the dynamic and ever-changing nature of the news” [45]
- “The main idea behind our geotagging framework within an article, to make it easier for human readers in the author’s intended audience to recognize and resolve toponyms. Authors create this framework by using linguistic contextual clues that we can detect using heuristic rules. Furthermore, readers are expected to read articles linearly, so article language has a contextual and geographic flow. Toponyms mentioned in a sentence will establish a geographic framework for subsequent text.” [45]
- “Certain phrases in article text denote relative geography, which is language that defines a usually imprecise geographic region in terms of distance from or proximity to another geographic location. These imprecise regions are important because they usually target the geographic areas where the events in an article took place, and therefore are useful for resolving the article’s toponyms. . . We refer to the toponyms in such phrases as anchor toponyms, and we term the resulting regions as target regions.” [45]
- “Recall drops “reflects the fact that most gazetteers are still rather incomplete or at least not in sync with the frequency of use of location descriptions that do not have formally defined boundaries, such as ‘New England’ and ‘Upper West Side.’” [45]
- “Event extraction from news articles is a commonly required prerequisite for various tasks, such as article summarization, article clustering, and news aggregation. Due to the lack of universally applicable and publicly available methods tailored to news datasets, many researchers redundantly implement event extraction methods for their own projects.” [46]
- “The extraction of a news article’s main event is an automated analysis task at the core of a range of use cases, including news aggregation, clustering of articles reporting on the same event, and news summarization.” [46]
- “Beyond computer science, other disciplines also analyze how news outlets report on

- events in what is known as frame analyses.” [46]
- “Explicit event descriptors are properties that occur in a text to describe an event, e.g., the phrases in an article that enable a reader to understand what the article is reporting on.” [46]
 - “State-of-the-art methods for extracting events from articles suffer from three main shortcomings. First, most approaches only detect events implicitly, e.g. By employing topic modeling. Second, they are specialized for the extraction of task-specific properties, e.g., extracting only the number of injured people in an attack. Lastly, some methods extract explicit descriptors, but are not publicly available, or are described in insufficient detail to allow researchers to reimplement the approaches.” [46]
 - “Similar to temporal phrases, locality phrases are often heterogeneous, i.e., they do not only contain temporal NEs but also function words.” Nomanatim: geocode leveraging OSM [46]
 - “For ‘when’ and ‘where’ questions, we found that in some cases an article does not explicitly mention the main event’s date or location. The date of an event may be implicitly defined by the reported event, e.g., ‘in the final of the Canberra Classic.’. The location may be implicitly defined by the main actor, e.g., ‘Apple Postpones Release of [...]’, which likely happened at the Apple Headquarters in Cupertino. Similarly, the proper noun ‘Stanford University’ also defines a location.” [46]
 - “The journalistic 5W1H questions are capable of describing the main event of an article, i.e., by answering who did what, when, where, why, and how.” [46]
 - “Answering the 5W1H questions is at the core of understanding any article, and thus an essential task in many research efforts that analyze articles.” [46]
 - Class imbalance: “often degrades machine learning approaches, which skews the classification capacity in favor of the most crowded classes.” (applicatin to tags) [15]
 - Most previous studies are english based [15] innapropriate for local investigations in other langauges

2.3.8 Examples

GLOBAL

GDELT Review Some projects are already mining place (as well as other attributes) from existing data lakes of publication data to provide geospatial and temporal distributions. One such effort is [The GDELT Project](#), which extracts place as well as actors, sentiment, and event connection (among other elements) from journalistic media across the globe, including publications from as far back as 1979. This and similar projects are powerful and hugely informative, especially as they apply to existing published data. The proposed project should leverage such tools for the inclusion of historic data into the developed database for investigation into the past (already published) incidents. However, the existing automated extraction includes several challenges:

1. It is not yet perfect, and places may be misattributed (Lisbon, Ohio in the USA may be accidentally attributed to Lisbon, Portugal).
2. It does not support the subtlety of incidents occurring in non-conforming places (an incident may not apply to a single administrative boundary but really fall into a subsection of one or several).
3. It requires technical prowess and tools to explore the data. A user is unable to define a spatial area of interest (such as their route to work with a half mile buffer or some

other irregular shape) and search for all spatially related results, nor it is easy to apply temporal or thematic attributes without prior experience querying results.

Therefore, this project offers a functionality specific to the defined user types of news publication services and provides an appropriate user experience to these.

5W1H -“Since events according to our definition occur at a single point in time, we only retrieve datetimes indicating an extract time, e.g., “yesterday at 6pm”, or a duration, e.g., “yesterday”, which spans the whole day.” [46]

Rivera2020

-Tasks: “1. Get a corpus of news published by the local newspaper through an RSS (Really Simple Syndication) feed, 2. Get a vector characterization based on the well-known Bag-of-Words (Bow) representation, 3. Select features through a mutual information-based method, 4. Train a supervised-learning model, 5. classify online news reports in real-time; the interest is in the ‘traffic accident’ class, 6. process the text of the RSS reports to retrieve the location where the accidents happened, and 7. notify users about the events on a map.” [15]

Lee2019 -Automatic POS extraction: Stanford named entity recognizer, apache open NLP algorithm, MIT information extraction (MITIE), OEDA [47]

-Leveraged location lists from GEonames, Wikipedia, and Google map
-Preprocessing: building gazetteers and import of data
-Follow for verification and comparison (Validation) type steps if there is time after

LOCAL

Crosstown

-“Our MIssion: Crosstown seeks to deliver community-level data and analyses to the people of Los Angeles who want to make their neighborhoods and city safer, healthier and more connected.” [56]
-“We want to turn those numbers into stories that people can use to understand what’s happening around them.” [56]
-Organized by neighborhood boundaries (LA Times), different from administrative boundaries. [56]
-Leverages public health, crime, administrative inputs, etc. [56]
-Not geolocating news, but building stories based on location [56]
-“Crosstown’s focus is not on increasing revenue but decreasing costs of producing local news. It does this through “mass customization” of public datasets in a three-tier system”. [54]
-Geocoded and aggregated by neighborhood [54]
-“We have created hyperlocal news at scale, and also allowed people in the city to see how their neighbourhood fits in with others and understand their experience in context” (- Gabriel Kahn) [54]
-“Tier one: data collection informing news stories”. Collecting and geo-locating “quality of life” data. “Journalists can then access the data and spot newsworthy trends at a glance.” “Tier two: data visualisation and dashboards” “There are a lot of different opportunities to take that data and not simply turn it into a classic news story, but turn it into other opportunities for the audience to engage with it.” “Tier three: neighbourhood

newsletters”. Avoid the need to have a dedicated reporter in each area. [54]

- NewsStand

- “NewsStand extracts the ‘interesting’ phrases that are most likely to be references to geographic locations and other entities by using NER methods. LOCATION phrases are stored as geographic features of the entity feature vector. Then, it uses a Gazetteer to find those geographic features in the entity feature vector that are names of actual locations. It also employs Gazetteer to identify the hierarchical information for each location (i.e. country and administrative subdivisions).. After that, it extracts geographic focus (or fous location) based on the frequency of the locations in the news.” Description of NewsStand used specifically for identifying primary locations of news articles? [35]
- “NewsStand monitors RSS feeds from thousands of online news sources and retrieves articles within minutes of publication. It then extracts geographic content from articles using a custom-built geotagger, and groups articles into story clusters using a fast online clustering algorithm. By panning and zooming in NewStand’s map interface, users can retrieve stories based on both topical significance and geographic region, and see substantially different stories depending on position and zoom level.” [43]
- NewsStand: Spatio-Textual Aggregation of News and Display [43]
- Based on “transactional database technology” [43]
- “It places markers representing story clusters on an interactive map interface, thereby allowing meaningful, visual exploration of the news.” [43]
- “The interplay between significance and zoom level is an important feature of NewsStand, and differentiates it greatly from existing spatially-referenced news reading systems (e.g. the Reuters News Map, the maps locations found in stories using MetaCarta).” [43]
- Global scalability: obscures articles not of interest to a global audience. [43]
- Geographic information extraction: “1. Provider scope, the publisher’s geographic location; 2. Content scope, the story content’s geography; and 3. Serving scope, based on the readers’ location.” [?]
- A global solution [43]
- “In NewsStand we are also interested in the geographic focus of a collection of news articles about the same subject/topic, rather than just one article, and this is done with the aid of a document clustering algorithm.” [43]
- “After a new article has been introduced to the system, NewsStand must locate and extract the geographic content from the article. This process, described earlier as geotagging, unifies the explicit textual article content with the implicit geography, and enables spatial exploration of the news.” [43]
- “Our main goal in designing NewsStand’s user interface was to convey as much geographic and non-geographic information about current news as possible. The interface consists of a large map on which stories are placed, and the viewing window serves as a spatial region query on the geotagged news stories. Users interact with

NewsStand using pan and zoom capabilities to retrieve additional news stories. As users pan and zoom on the map, the map is constantly updated to retrieve news stories for the viewing window, thus keeping the window filled with stories, regardless of position or zoom level.” [43]

- “NewsStand also features a smaller map that shows the geographic span of the selected story. This minimap allows users to easily see the selected story’s geographic focus, without having to leave the area of interest on the main map.” [43]
- To avoid geographically clustered visualization of uneven news coverage “Marker selection is therefore a tradeoff between story significance and spread. “To achieve a balance in marker mode, NewsStand divides the viewing window into a regular grid, and requires that each grid square contains no more than a maximum number of markers. The markers to display are selected in decreasing order of story significance and story age. This approach ensures a good spread of top stories across the entire map. [43]

Suntimes Crime

- mapped instances of crime within a city [57]
- “Find out about crime in your neighborhood and your city. The best and most timely analysis of crime trends in the city of Chicago.” [57]
- No map or spatial exploration of the news. No textual/tag organization of place.

In Your Area

- InYourArea lets you follow the latest local news, information, events and more in your area. InYourArea covers local news and information for all Towns, Cities, Villages and Hamlets in the United Kingdom. As well as the latest news and information for where you live, you can also connect with other members of your community, submit events, promote a local business and more.” [58]
- “InYourArea includes news from all the top local and national news publishers and blogs. It includes updates from local councils, police and public services.” [58]
- “InYourArea covers news for all UK Cities, Towns, Villages and Hamlets including London, Birmingham, Manchester, Leeds, Liverpool, Sheffield, Cambridge, Bristol, Norwich, Reading, Cardiff, Edinburgh and more. So download it today and see what’s happening in your area. [58]
- Post code based (not searchable by specific locations) [58]
- Method for sharing relevant information by postcode [58]
- Signup: input post code, select three areas of interest [58]
- Includes: state of transit (roads, rail, bus, TFL); weather; notices; recent updates (comments and news stories); all news vs. my news; ability to interact (like, comment, share, follow); no map searching options; whats happening (discussions, about my area, planning applications, funeral notices, council notices) and find (homes near you, items for sale, deals and offers, things to do, local services) [58]

2.3.9 Manual location extraction from published corpora

- Finding 1 “There are significant portion of place names in local news that are vernacular, vague places, or finer granularity places that were not found in gazetteers.” [5]
- Finding 2: “The majority of those place names found matches from gazetteers are found

to be ambiguous.” [5]11 -Finding 3 “For those place names that were not able to find proper footprints in gazetteers, humans can create footprints with ease when supported well.” “Because GeoAnnotator helped human in locating the area and providing good spatial context, drawing the footprint in the map view is relatively straightforward.” [5] -Finding 4: “There are high degree of verbatim repetition of place references across local news articles over time.” potential “to end up with a local gazetteer that is enriched over time”. [5]

- “heuristic rules that human coders use in disambiguation of place references in local newspaper are very different from those used in resolving ambiguities in global news.” [5]
- “generic gazetteers are not adequate for geoparsing and geomatching for local news.” [5]
- Suggestive geocoding: “Geocoding... requires deep, human-level knowledge to evaluate a potentially large list of candidate matches and make a choice. THese tasks re best done by bringing human and computer to work collaboratively. A simple way to do this is to have a computer doing step 2 (generating all the candidate matches) and uman take care of step 3 (disambiguation). Alternatively, computer can further reduce human efforts in step 3 by ranking the candidates such that those with higher degree of likelihood are placed on the top of the list, and those candidates that are deemed impossible after considering the context are eliminated from the list.” [5]
- “While a system that follows the progressive geospatial referencing framework is expected to enrich its gazetteer and improve its performance over time, ti does require human effort to deal with uncertainties, and complex geographical descriptions that have no gazetteer match (e.g. a long prepositional phrase representing a place). THe key challenge here is to reduce the human effort to the minimal and accelerate the process of incremental improvement.” [5]
- “Geo-annotation requires either local knowledge of the place or general geographic knowledge and the initiative and ability to further research the existence and location of places.” [2]
- “Presenting place names to users with pre-annotation hastens the geo-annotation process but comes with a risk. It is likely, especially for larger pieces of text, that users might miss place references that are not already highlighted in text by the pre-annotation process, and therefore, these entities are not added to the map view and might be omitted due to visual absence.” [2]
- “the quality of a geoparser’s output annotation is domain-dependent. Depending on the kind of text (e.g., news story, social media posts or academic articles), the geographic coverage area of documents (e.g., North America, Middle East, Eastern Europe), and the geographic level of toponyms (e.g., country level, state level, physical features, landmarks or building names or addresses), the geoparser may underperform and therefore create additional work for users performing annotation.” [2]

EXAMPLES:

GeoAnnotator Workbench

- Protocol: “1. Open an article and read through the whole article to understand the story; 2. Geoparsing: identify all place names (or prepositional phrases) and highlight them in the article; 3. Geomatching and disambiguation: for each identified location reference, search the best gazetteers to find all matches, compare candidates, and finally make a choice from the candidates, and finally make a choice from the candidates. In case no candidate match is found or all candidates are rejected by the human analyst, he or she can create a new footprint and add it to the local gazetteer. 4. Coding: depending on the gazetteer matching outcomes and the disambiguation strategies used by the analyst, each

location reference is coded by a case number.” GeoAnnotator workbench [5]

- Uses a local gazetteer (subset of nominatim), nominatim global, and google maps [5]
- “1. Start with a VGI-based gazetteer that is best for the target locality. The quality of this gazetteer does not matter so much, as it only provides a starting point for us to bootstrap the system. In the current implementation of GeoANnotator+, we bootstrap the system by taking Nominatim as the initial source of gazetteer. Nominatim provides a search engine API for using OpenStreetMap data as a proxy gazetteer to match place names. The quality of the footprints in Nominatim is quite good, as it normally returns matches of local features by good approximation of their real shape (points, lines, or polygons). We choose Nomination as the initial gazetteer because it is considered the best VGI-based geographical data sources with good coverage in most localities worldwide.
- 2. Progressive geocoding to enrich gazetteer incrementally. Provide a semi-automated workbench for human analysts to evaluate all matches and make a choice. If none of the matches are appropriate, analysts will create the footprint in the workbench. The system not only remembers the result of this human-generated geocode, but also adds a new entry to its gazetteer as enrichment. In this way the system is able to leverage real local and community conversations and outsource this geocoding task to local community members. Hence, new gazetteer entries learned from human reflect local spatial language and local knowledge.
- 3. Smart footprint recommender ranks the matched candidates by their likelihood of being correct. Based on our understanding of how humans disambiguate multiple interpretations of place names, a computer can play the role of a smart recommender by automating a set of heuristic rules. By presenting the candidate list in the order of likelihood, human annotators are more likely to find the answer by exploring only the top few candidates. This creates savings to human cognitive efforts.” 1 and 2 are critical (but with two- how should they be saved/named when hyper specific?), and 3 is gravy. [5]

-Smart recommendation: “prioritize previous annotations”, “exploit local gazetteer”, “rank search results with regard to a moving focus”. See p7 for more. [5]

2.3.10 Location assignment

During publication

- Geo My WP: Integrates GoogleMaps, Leaflet, and OpenStreetMaps, includes proximity search, point location definition [59]
- CG Geo Plugin: Focused on customer geolocation integration: geotargeting, legal requirements, etc. [60]

2.4 Web applications

2.4.1 Design

- “planning organizations should choose a participation platforms based on the capacities of their organization, the characteristics of the communities that are going to use the tool, user-community norms and rules, and the tool’s capabilities.” [9]
- “Discussing these factors with planners and decision makers, technologists, and communities who use the tools will add insights on identifying new considerations for implanting online participatory technologies” [9]
- “The geospatial Web 2.0, or Geoweb for short, is a collection of online location-enabled services and infrastructure that engages a wide range of stakeholders in mapping processes.” [32]

- Basic components: web participants, user groups, geoblog, blog components, blog profile, blog component profiel, blog description, time points and intervals, blog metrics. Relationships: generalization, composition, collaboration, possession. [32]
- “a general mashup architecture consists of three different participants that are logically and physically disjoint. THe generic architecture has the mashup client sitting on the top and the data sources and services sitting on the bottom. The middle tier is where the mashup logics reside. It should be noted that the mashup logics for generating mashed content could be either executed on the server or within the web browser. [32]
- Tagging: used GeoBLog with spatial extensions (server side scripting using JavaScript, Visual Studio 2010 2ithin the asp.net code). Requires secure user authentication (login), access to post and transfer messages, and management tools programmes in asp.net. [32]
- Geoportal design: i. spatial database and data server; ii. geospatial web server; iii. Geospatial metadata catalogue server; iv. other catalogue server; v. download tool. [13]
- “deconstructing the user-designer dichotomy... the boundary between user and designer is fluid and configured during the design process, and that users can have multiple identities. In addition to being users, they can perform activities traditionally ascribed to designers by dynamically participating in the ongoing design process.” “a user-centered design that deconstructs the traditional power relationship between designers and users through role hybridization by creating an environment where users or designers are able to shift from one role to another, effectively belonging to otherwise two distinc groups.” “designers can take the role of users, and users can participate in the design process, with constant interactions within and between the users and designers in order to combine and reconcile design ideas from the wto groups of users and designers.” “designer-user interaction models as a combination of interaction of designers, interaction of users and mutual interaction between designers and users with constant communication to increase knowl-edge sharing, crossing the social boundaries and deconstructing the design-user dichotomy to produce novel design artifacts and increase usefulness.” [2]
- “With inter-user communications capability built into the platform (in addition to the periodic face to face meetings and email exchanges), both users and designers were able to articulate their needs, comment on particular linguistic and geographic difficulties, deliberate together and suggest solutions from implementation.” [2]
- “In fact, code writers can be considered designers because they create software and algorithms that support evaluations and decisions directly impacting people’s lives. The filters that determine the selection and analysis of large and complex data sets on topics ranging from food access to health and nutrition originate from their coders’ specific world-views and reflect their implicit bias, inevitably shaping the outcome of research and surveys, which in turn influence policy decisions based on those enquiries.” [61]
- “The dramatic rise of data visualization could be traced to hardware factors such as widespread use of high-resolution large desktop displays tied to powerful computers. Other important trends are the increased availability of vast data resources, familiarity with data management software, and innovative web-based software that support rapid display and update of visual information.” [20]
- “The datavisualization process (also called ‘work flow’) starts with the identification of stakeholders and their insight needs. Just as a verbal math problem needs to be reformulated into a numerical math problem, the verbal or textual description of a real-world problem presented by a stakeholder must be operationalized (i.e., reformulated into a data visualization problem so that appropriate datasets, analysis and visualization workflows, and deployment options can be identified).” [23]
- “The task hierarchy provides insights about the ‘bigger-picture’ of each atomic task - a

task that does not contain any subtasks - helping ensure that we always consider higher-level goals.” [24]

- “As tasks often have dependency relationships with each other, there can be a required task sequence. E.g., one has to view data before reasoning about it, and an overview visualization needs to be presented before selecting and examining outliers. Therefore, visualization tools should not simply fulfill discrete tasks without providing for a natural flow between consecutive steps. Integrating task sequence information can help mitigate the user’s mental burden and provide a better overall user experience.” [24]

- “While many abstract tasks are similar (e.g. discover patterns and trends, reason about outliers), visualization tools often are used in a domain-specific context... We cannot simply peel off the domain-specific context, abstract to common data visualization tasks, and then design for those tasks. Context is key.” [24]

- “When designing data visualization tools it is crucial to choose visual and interaction designs that solve the real-world problems of our target users. Task abstraction is widely used to support this process by characterizing user tasks as domain- and interface-agnostic abstract tasks for which design best practices are known.” [24]

- Three step hierarchical task abstraction: “(1) understand the problems and data of domain users; (2) perform hierarchical task analysis to understand user tasks, goals, and the relationships among the tasks as part of a larger analysis process; and (3) abstract the user domain tasks.” [24]

- “Once constructed, the hierarchical task and data type abstractions should inform visual encoding and interaction design.” [24]

2.4.2 Open Source

- “To develop on an open source platform is extremely vital when huge databases are to be created and consulted regularly for region planning at different scales particularly satellite images and maps of their locations”. [31]

- OGC Standards

2.4.3 Data

Volunteer generated information sources

- VGI and geoweb: “there is mounting evidence that institutions can use VGI as a mechanism to build a local capacity to support collaboration, supplement traditional data sources, and inform decision making.” The VGI paradigm “Is enabled by the GeoWeb, location-aware devices, and citizens acting as sensors, and their tools and resources for collecting and processing geographic information from volunteers are readily available.” [32]

- “VGI involves the creation and dissemination of geographic data provided voluntarily by individuals and overlaps with PPGIS in that both involve the investigation and identification of locations that are important to individuals.” [27]

- “PPGIS projects are often implemented to inform planning and policy issues while VGI systems may have no explicit purpose other than participant enjoyment.” [27]

- “Given well-defined insights needs, relevant datasets and other resources can be acquired. Data quality and coverage will strongly impact the quality of results, and much care must be taken to acquire the best dataset with data scales that support subsequent analysis and visualization.” [23]

Data harmonization - Data harmonization: “Some geoportals not only provide data as

they come from the original source, but they are also able to provide spatial data and geo-information coming from different sources into a common (i.e. standard) format... Data harmonization can help to receive, process, and exchange data, and to ensure interoperability, because the harmonized data and information are accessible to end users at their demands.” Federated approach: [13]

-Federated approach: “participants to agree on common specifications in terms of metadata, data models, and service interfaces, typically based on international de-jure or de-facto standards.” Brokered approach “leveraged middleware between the client and the server tiers by addressing heterogeneity through mediation (i.e. of metadata and data models) and adaptation (i.e. of interfaces). In the brokered approach, the brokers as key component are dedicated to mediation and haronization. The brokered approach has been particularly appealing for those wanting to build Systems-of-systems, e.g. distributed infrastructures connecting systems, which keep their autonomy at a certain degree.” “Generally, the brokered approach is more effective in addressing variety and heterogeneity when central authority cannot be easily achieved.” [13]

-Basic functionality: “metadata registry, data discovery through a catalogue service, data visualization, and data access.” See section 2.3.4 for more info [13]

-“A geospatial metadata catalogue provides data descriptions in terms of metadata (e.g. contributor, data type, language, contact point, keywords, and dataset identifier for data localization and indexing). In addition, the metadata catalogue is often used for implementing harmonized data discovery. Since users are typically interested in finding datasets matching specific constraints, the data discovery functionality is one of the basic functions that geoportals offer. SSpecifically, geoportals providing data discovery generally allow searching datasets along the who, when, where and what axes, that is, by geo-location (where), data provider (who), time range (when), thematic layer, and keywords (what). The user interface provides graphical tools, like a bounding box on a map, to set spatial and temporal constraints. Moreover, users can be directed to a gazetteer, a thesaurus, or other knowledge bases for better scoping their query. Various approaches have been developed to enhance geoportal search capabilities, e.g. the use of thesauri, ontologies, and semantic text matching algorithms”. [13]

-“Big data are posing challenges that are usually referred as V challenges: large Volume, high Velocity, and wide Variety. On the geoportal user interface side, they have an impact on discovering, visualizing, processing, and storing big Earth data. Specific challenges come from teh continuity of data updating, methods of data harmonization, and mulitple functionalities fro professional users.” [13]

-Top down approach: “administrative approach”. Support “incorporate large amounts of spatial datasets and geo-information products, multi layered and multi types of data for earth sciences.” [13]

-Top down approach: “administrative approach”. Support “incorporate large amounts of spatial datasets and geo-information products, multi layered and multi types of data for earth sciences.” [13]

-Top down approach challenge: sustainability and homogeneity, insufficient searchability for nuanced use. Open Geographic Modeling System (OpenGMS) to combat this. [13]

-“The data systems, where the geoportals connect, are leveraging different methods for coordination, in terms of federated coordination (top-down) and broker coordination (bot-tom.up). Both coordination solutions are able to connect distributed data servers and information systems, and to facilitate complex systems for solving big data storage, management, and assessment problems. The federated approach generally shows better performance, but it requires that participants agree on participating in the overall system

bylaw or their own interests. If this is not available, a brokered approach is the only one that can be adopted. In addition, the brokered approach is better, if the requirements cannot be fully defined allowing one to select the best federal agreement. Therefore, a hybrid data system architecture for a complex data system is recommended for meeting the challenges posed by big data.” [13]

-Data integration: “The discipline of data integration comprises the practices, architectural techniques and tools for achieving the consistent access and delivery of data across the spectrum of data subject areas and data structure types in the enterprise to meet the data consumption requirements of all applications and business processes.” (Gartner) [37]

-“Performing temporal inference using disparate applications for each data source is challenging. Despite several advances in data interaction and visualization, device interoperability and data integration remains problematic.” [24]

Data access

-Geoportal design: Spatial metadata catalogue server: “The geospatial metadata catalogue provides metadata in accordance with OGC and ISO standards and is retrieved by the geoportal” [?]

-Geoportal design: other catalogue server “Other catalogues could be supplemented by metadata catalogues for a data registry. A well-known catalogue is the Gazetteer Service which can host a location-based feature dataset” [13]

2.4.4 GIS design

-“the design of the event model is composed of two parts: Spatiotemporal and Attribute information. Spatiotemporal information is polymerized by Feature, Temporal and Geometry. Among them, Temporal is composed of the time instant and time interval. Geometry is described by Position coordinate and Coordinate system, which belongs to Point, Line and Polygon. Attribute information is composed of Border type, Feature type, and Rss:Item; the model provides an extension of the elements by Rss:Item.” [32]

-Geoportal design: Spatial database and data server “Spatial databases store and maintain the data that will be delivered. The database could be on the local server or distributed servers.” [13]

-“data storage and management is one of the critical technical components for addressing the volume challenges of big data. Currently, relational databases (e.g. PostgreSQL), NoSQLs (e.g. MongoDB and HBase), and distributed file systems (e.g. Hadoop) are widely used for big data storage and management. These technologies are also important for geoportals, although the technical component for big data storage and management still requires more theoretical and practical developments.” [13]

2.4.5 Web services

-Standards: Geoportals are usually compliant with several OGC standards, with some of the most popular ones being the Web Coverage Service (WCS), the Web Processing Service (WPS), the Web Coverage Processing Service (WCPS), the Web Map Service (WMS), the Web Feature Service (WFS), and the Catalog Service for Web (CSW). GeoServer [13]

-Geoportal design: Geospatial web server: “Geospatial web services provide geographic data to the user through forms of map data services under the guidance of OGC and ISO standards. Open-source options for web services are capable of hosting these web services,

e.g. GeoServer and MapServer [13]

Example:

- Twitter map [62]
- COVID visualization [20]: Good design and professional practice is to show the sources of data, name the contributors to the visualization, and provide an email address for comments, corrections, and questions. [20] Example of Johns Hopkins University visualization of COVID data. References of John Snow and William Playfairs addition to geodisplays to distill information and draw new spatial conclusions [20]. Use of spatial data to “persuade, understand current conditions, and predict future outcomes based on behaviors”. [20] Many references to additional display options (weather.com, microsoft power bi, esri, visual action, etc. “These widely used COVID-19 visualizations depend on maps, bar charts, line charts, and scrolling lists, however designing them to fit on small mobile-device displays requires skillful programming.” [20]

2.4.6 Graphic user interface

- “The interactive graphical user interface (GUI) allows for data visualization manipulation and sharing.” [31]
- “The six areas should be covered on the geoportal interface are including metadata, data thumbnail, data description document, data sample, data entity (i.e. download link), and connections and groups. The goal of mode is to provide a more detailed description of the data on the geoportal and enable users enough information for faster data positioning and discovering.” Main elements of a geoportal. [13]
- “A well-designed GUI is beneficial for interacting with the user and providing an easy-to-use experience and powerful visualized interactivities.” [13]
- “Geoportals tend to prefer adopting an atlas-like user interface (using OpenLayers and Leaflet). THis may result in the requirement of visualizing geographic distributions of the data and of the functionality for visualized data selection. An atlas-like user interface, as the metaphor, is recommended for geoportal design.” Validation of openlayers and leaflet. [13]
- “In order to address some Volume, Variety, and Velocity challenges on the user interface side, the next generation geoportals should specify the time-series investigation, graphs and charts for visual analytics, and 3D and 4D visualization widgets through dynamic visualization tools.” [13]
- “For an effective presentation, the news must be shown in an informative, aesthetically pleasing manner, but this must not overwhelm the viewer.” [43]
- “Because it is inevitable that multiple news stories will mention the same location, we also need a strategy for dealing with occlusion, since we do not want to place markers on top of each other.” [43]
- Dynamic map labeling: “placing text labels tied to geographic coordinates on a map, usually requiring that some corner or edge of the label’s bounding box touch the coordinates associated with the label.” Leverage a precomputed conflict graph (“labels correspond to nodes, and edges exist between labels that overlap”). [43]
- “some occlusion of markers is tolerable, as long as the more significant story markers are placed above less significant markers. One exception to this rule is when markers exactly coincide - that is, when several stories mention the same geographic location. It is unacceptable to place markers at the exact same coordinates on the map, as users cannot infer that many stories refer to that location. This is often a problem with large cities,

as they are a part of geographic focus of many news articles.” Spiral strategy: “this allows significant geographic locations to have more of their stories visible, at the expense of accuracy in marker placement. However, due to their regular shape, these spirals are usually easy to identify and do not contribute significantly to user confusion.” Ideally the inclusion of custom reference points/areas will alleviate this issue, though more popular POIs will likely include several concurrent news stories. [43]

- “Many of the currently popular visualization follow the well-established Information Visualization Mantra: ‘Overview first, zoom and filter, then details-on-demand.’” [20]
- Visual Information Seeking Mantra: overview first, zoom and filter, then details-on-demand [22]
- Data types: 1D, 2D, 3D, temporal and multi-dimensional data, tree and network data). Includes detailed descriptions of these. [22]
- Tasks: overview (“gain an overview of the entire collection”), zoom (“zoom in on items of interest”), filter (“filter out uninteresting items”), details-on-demand (“select an item or group and get details when needed”), relate (“view relationships among items”), history (“keep a history of actions to support undo, replay, and progressive refinement”), extract (“allow extraction of sub-collections and of the query parameters”). [22]
- Multi-dimensional: “the interface representation can be 2-dimensional scattergrams with each additional dimension controlled by a slider.” [22]
- Multi-dimensional: “the interface representation can be 2-dimensional scattergrams with each additional dimension controlled by a slider.” [22]
- Data Scales: “Data variables may have different scales (e.g., qualitative or quantitative), influencing which analyses and visual encoding can be used.” [23]
- Visualize “This step can be split into two main activities: pick reference system (or base map) and design data overlay. The First activity is associated with selecting a visualization type, and the second activity is associated with mapping data records and variables to graphic symbols and graphic variables.” [23]
- Deploy “Different interface controls make diverse interactions possible: buttons, menus, and tabs support selection; sliders and zoom controls let users filter by time, region, or topic; hover and double click help users retrieve details on demand; and multiple coordinated windows are connected via link and brush.” [23]
- Interpret: “Finally, the visualization is read and interpreted by its author(s) and/or stakeholder(s). This process includes the translating the visualization results into insights and stories that make a difference in the real-world application.” [23]

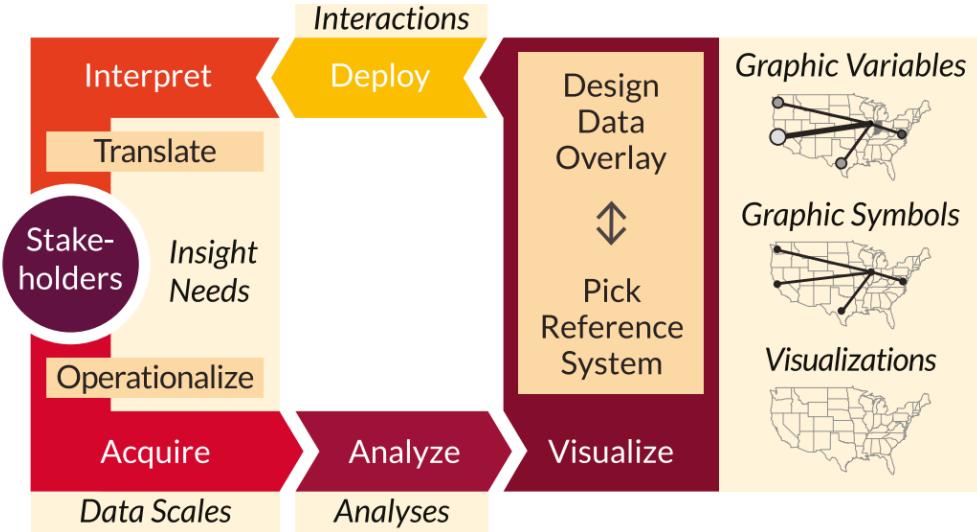


Fig. 6. Process of data visualization construction and interpretation with major steps in white letters. Types identified in Table 1 are given in italics, and an exemplary US reference system and sample data overlays are given in Right.

Figure 1: Process of data visualization construction and interpretation with major steps in white letters. Types identified in Table 1 are given in italics, and an exemplary US reference system and sample data overlays are given in Right. (Borner2019)

2.4.7 Data manipulation

- Geoportals can also offer online processing functionalities, ranging from basic transformations (e.g. coordinate reference system re-projection, sub-setting, format mapping) to complex algorithms for classification, and statistical analysis. That may require an underlying infrastructure providing a workflow engine to orchestrate the multiple actions required to implement the required processing.” [13]
- Spatial queries require indices (such as R-trees or quadtrees) for efficient execution. [45]
- Zoom: “smooth zooming helps users preserve their sense of position and context”. [22]
- Filter: “By allowing users to control the contents of the display, users can quickly focus on their interests by eliminating unwanted items.” [22]
- History: “It is rare that a single user action produces the desired outcome. Information exploration is inherently a process with many steps, so keeping the history of actions and allowing users to retrace their steps is important.” [22]
- “Dynamic queries can reveal global properties as well as assist users in answering specific questions.” [22]
- “dynamic visualizations: overview, zoom, filter, details on demand, relate (viewing relationships among items), history (keeping a log of actions to support undo, replay, and progressive refinement), and extract (access subcollections and query parameters)” [23]
- “Keim distinguishes zoom filter, and link and brush as well as projection and distortion techniques as a means to provide focus and context.” [23]
- “Brehmer and Munzner covers two main abstract visualization tasks. The first is ‘why’, which includes consume (present, discover, enjoy, produce, search (lookup, browse, locate, explore), and query (identify, compare, summarize). The second is ‘how,’ which consists of encode, manipulate (select, navigate, arrange, change, filter, aggregate), and introduce

(annotate, import, derive, record).” [23]

-Other setup: “data and view specification (visualize, filter, sort, derive), view manipulation (select, manage, coordinate, organize), and process and provenance (record, annotate, share, guide).” [23]

2.4.8 Security

-“some people might be concerned about sharing their identify in online environment or making their profile information visible, as they are worried about organizational and social threats”. [9].

-“For implementing these common functionalities, additional functionalities are often required. For example, a geoportal should implement a subset of functionalities concerning privacy and security aspects, e.g. authentication, data access control, logging. Moreover, the geoportal administrators need dedicated management functionalities. [13]

2.4.9 Products

-Recommended services for geoportals: catalogue, preview, access, and online analysis and processing [13]

RSS

-“Publish/Subscribe has emerged as a communication paradigm able to facilitate the development of complex distributed applications in open network environments. THe strong decoupling it introduces between communication parties enables applications to publish information without being aware of the identities of potential receivers or even of their existence. Similarly, it enables receivers to issue subscriptions that express their interests in messages with a given content regardless of the identity of their publishers.” [32]

-“GeoRSS is an emerging standard for RSS feeds to be described by location or geo-tagged in a standardized way in which the location is encoded. Many map APIs support GeoRSS feeds with either coordinate (lat/long) data or address information specified in XML items. “ [32]

-Geoportal design: Download tool: “A download tool incorporates the capability of obtaining data and metadata. Metadata could be downloaded by XML structured files, and data could be downloaded as raster or vector based data outputs through various ways, such as FTP” [?]

-GeoRSS feed support [55]

API

-“it is suggested to provide tools (e.g. API, widgets, configuration) to create community portals, and applications tailored to specific users. They would allow developing, discovering, and running of applications to support different online and cloud based scenarios in particular for big earth data processing... Geoportals could involve online spatial analysis functionality for addressing geospatial analysis tasks, e.g. online computing environments and geospatial processing web, online analysis, and cloud computing.” [13]

-“Geoportals could also provide APIs to enable developers to create alternative user interfaces to data systems, including community portals and mobile/desktop apps.” Incorporate API functionality. [13]

-“System components communicate using open application programming interfaces (APIs) so that each component can be replaced by a different technology if necessary.” GeoAn-

notator [2]

2.4.10 Dashboard design

- “Viewing multiple folded and aligned units of event sequences simultaneously and in their entirety would require considerable screen space. However, aggregation can obscure rapid changes and outliers. Overview and detailed approaches are beneficial in these cases.” [24]
- “The benefit of superimposing time series from multiple days is to allow direct comparison without changing the view. However, these line charts may cause significant visual clutter and conflicts. Superimposing distinct views can sidestep this problem. [24]
- “Our participants appreciated and were able to effectively use superimposed distinct visualizations of multivariate data in the overview and detail panels.” [24]

2.4.11 Internationalization and localization

- “Localization refers to the adaptation of a product application or document content to meet the language, cultural and other requirements of a specific target market (a ‘locale’). [63]
- Localization: l10n [63]
- “Internationalization is the design and development of a product, application or document content that enables easy localization for target audiences that vary in culture, region, or language.” [63]
- Internationalization: i18n [63]
- “Retrofitting a linguistically- and culturally-centered deliverable for a global market is obviously much more difficult and time-consuming than designing a deliverable with the intent of presenting it globally.” [63]
- “So ideally, internationalization occurs as a fundamental step in the design and development process, rather than as an afterthought that can often involve awkward and expensive re-engineering.” [63]

2.4.12 Examples

Geo my WordPress

- “Geotag any of your post types, buddypress members and other components” [59]
- “Create unlimited advanced, proximity search forms to search and find any of the geo-tagged components of your site” [59]
- “Add geographic location to any of the registered post types of your site. Display post location on a map, and create proximity search forms to search and find posts based on address, distance categories and more.” [59] -Based on Google Maps API, also supports Leaflet and OpenStreetMaps [59]
- Reviews: 4.6/5 stars. [59]
- Appears to only be for point data (not area) [59]

NewsPack

- A wordpress model for news sources (mostly american) [64]
- The backoffice 60% of global news rooms [?]

- Doesn't allow for customization [?] Opportunity to create a plugin for wordpress to accomplish this
- Easy to work in [64]
- Mostly tailored to US local news [64]
- Doesn't incorporate or recommend a geolocation element [65]

3 Intervention Description

The goal of this project is to explore if there is a value to plotting the spatial distribution of news story contents, as well as provide users a portal through which to interact with the data to extract the desired information. The project will use a study area of new story data from a selection of applicable public and private news media sources in Lisbon, Portugal during the month of October 2020. By building a tool specific to Lisbon, the project seeks to accommodate the culture and the business processes of the local community, providing a platform that is useful and valuable to users (whether they be citizens, local officials, researchers, or publishers).

It is expected that the association of specific place (potentially non-conforming to existing administrative boundaries or defined points of interest) to traditional news articles will provide an added dimension of understanding to communities at a local level. Users will find additional insights from the ability to view or filter spatial attributes, especially in conjunction with thematic and temporal attributes. This type of data preparation, though it is initially cumbersome to establish and requires adjustment of publishers' processes to maintain, will provide a powerful foundation from which future economic (improved publisher products elevating their offering and attracting/maintaining a customer base), societal (illumination of local trends requiring intervention, improved community engagement of readers with their surroundings, or improved city resources), and academic (improved research functionality) benefits may stem. If this type of functionality and improved user experience are well-implemented by a handful of productive news services, it will force a shift of the industry standard towards integration of spatial attributes and spatially related products.

- Both a tool and the beginning of a database that can be used and applied in the Lisbon area. Focus of both "producing a geo-annotation software platform, but included a proof-of-concept effort to build a manually geo-annotated corpus along the way." [2]
- Integration of visualization into existing products is an important part of relevance and commercial value propositions [21]
- of the drivers of geoportal advances (scientific geospatial projects and applications, international organization, governmental agencies and commercial purposes), governmental agencies and commercial purpose are most applicable to this project [13]

3.1 Products

1. A spatial database of incidents that supports the association of spatial, temporal, and thematic attributes. See Appendix C/Figure 7: data model.
2. A POC Input tool for publishers that allows users to define the place(s) (via search for existing administrative boundaries and points of interest [POIs] through existing gazetteers or definition of new polygons or points via drawing) as well as time of

occurrence of incidents. It shall also, of course, preserve or potentially improve upon the association of traditional thematic attributes and keyword search. See Appendix C/Figure 8: *Input* layout.

3. A POC Context map (visualization of an incident on a local map) for integration into each article page. See Appendix C/Figure 9: *Context* layout.
4. A POC Search tool for researchers that allows users to filter by spatial (one or multiple defined places or via drawn definition of the study area), temporal, and/or thematic attributes. The results should be displayable via both map and list views, as well as support CSV export functionality. See Appendix C/Figure 10: *Search* layout.

3.2 Distinction from previous work

-Unlike many of the example provided in the literature, this effort seeks to extract a variety of articles instead of tweets. Tweets or other specific data types include structured organizations and data APIs. Instead, this method seeks to develop a novel structure standard for distinguishing attributes of a particular event [41]

- The movement to incorporate citizens as sensors is powerful, but we have jumped over public and commercial sources of information that can not only contribute but contextualize citizen feeling towards a place. Public and private data sources of events are being underutilized – the content exists but needs to be georeferenced in order to be better accommodated by citizens, public management, or private enterprises. Citizen knowledge of a place is drawn from both anecdotal experience as well as learned (read) information from 3rd person sources (news papers, reports, etc.)
- “we will investigate human-in-the-loop methods to improve the geolocation prediction. For instance, it might be possible for users to explicitly provide city labels or non-geotagged tweet, or correct predicted city labels.” [41] this project expands upon this concept

3.3 Public participation

-The tool is a PPGIS system in that those assigning spatial definitions will not be trained GIS users and will be assigning place based both on more objective place definitions (addresses or existing boundaries) as well as less-conforming definitions (areas in which something happens that does not conform – sub areas of boundaries or custom boundaries of multiple places). [27]

-The tool would contribute to public participation by providing additional insights to citizens about the areas that affect them – where they live, work, and play. [17]

-Return to the concept that the tools is a local government hosted tool that facilitates better understanding of place within the community. It aggregates and georeferences the city’s own public information onto a navigable map, but can invite and accommodate commercial information (online publications) as well as private citizens micro-blogs (less structured and potentially long form versions of “Na Minha Rua”) to gain further citizen feeling or commentary about their physical surroundings. [17]

-Much like facebook and other informative platforms, the tool would provide a platform for dissemination of knowledge and providing geo-referenced articles from which further information development is required for various purposes, and additional tools are required for further action. [17]

-“using social media and mobile technologies as tools to increase two-way interaction be-

tween citizens and (local) governments will not reduce the workload of professionals.” [30] additional workflow to manage information and incorporate ideas from citizens. [30] Spatial data agreement only applies to drawing inferences (statistical analysis to derive information) from the system, considering journalists/publication and officials/public institutions as ‘public participators’ [27]

3.4 Nuanced “place”

- Geographical cue words. “City of X”, ‘just outside of X’, ‘X-based’, and ‘X Ocean” [45] These often do not accurately describe the location of a news story. “Just outside of X” is particular area near X, but likely not a buffer of the place (more likely an area to one side), X Ocean shouldn’t necessarily associate a point to the center of the entire body of water, but is rather relative based on the context of the story/place (an area of ocean off the coast of a particular place, for example). In these cases, there isn’t an appropriate automatic method to distil this from textual description. [45]
- Where something is happening may be best associated with existing gazetteers of information. These are expected to be point data (addresses, POIs, etc.) or polygons (administrative boundaries). These can be left as-is (incorporating both types of data). The tool should allow the publisher the flexibility to define a point or polygon (more likely) location based on the “exact” occurrence of instance. Any data analysis from there can be transformed on the fly as necessary. [27]
- “A map, relied upon for accuracy in terms of direction and location, is a definitive symbol of capital “T” truth. As such, geographic boundaries, both physically and politically constructed, powerfully separate and shape the identity of those bounded by being located or represented on a map.” Validating incorporation of custom place definition [14]
- Having the ability for contributors to decide which is more convenient with quick assignment tools may be critical for user buy-in “we suggest tha polygon methods appear better suited for structured interviews, group-administered surveys, or workshops that provide face-to-face support for completion, whereas points may be a better choice for self-administered surveys.” [27]
- “Further complicating matters, new articles often use nearby landmarks to indicate the location, in lieu of using the official names.” [47] Though for my purposes this is better, the sub-specific is helpful in sub-city level definitions. Of course, this does lead to challenges in places being defined by near-by POIs when they occur within, encompassing, or nearby (not exactly to the extent of the POI itself). [47]

3.5 User experience

- Aim to create a tool that is easy to navigate and understand without prior GIS knowledge. Similar level as standard applications such as AirBnB, Uber, Glovo, etc. Additional API and dashboards options for more advanced users. [17] -The tools can be used offline to demonstrate to non-digital citizens the distribution of issues during planning meetings. [17]
- “Asking participants to provide personal information (e.g. age, zip code, gender) as a part of their registration process may discourage their participation, but such inquiries can provide valuable information for planners in understanding the extent to which the online community represents community demographics.” [9] The information on the tool is open and free. Users need only create a profile if they wish to save searches or create dashboards, export info, etc. What information should be requested from users to do

so? [9]

- The tool must be easy enough for non GIS trained people to define place easily and efficiently with the given tools. Leverage well-utilized UIs and tools (google maps or similar) to ease the usability and increase precision [27]
- Value of geolocation in platforms to user experience via hyperlocalization and centering platform services to the literal user [36]

3.6 GUI

- determining spatial patterns from news sources supports GEOINT and therefore the application and dashboard views for decision makers. [35]
- Hovering on a story indicator yields a bubble of story content and attributes [43]
- Layout: documentView, CandidateListView, and FootprintView [5]
- Include textual tag in reference map, include different map symbols for each tag [2]

3.7 Knowledge facilitation

- The tool is meant to support insights within or about a community from mapping events/instances drawn from the contents of news stories. [14]
- Geographic information Science (GISc) can be enhanced by the location of incident reports (news articles). Beyond colloquial data (social media), geolocation of official reports can contribute to both how an event is felt within and outside of a community. The differences between styles of reporting can be evaluated from place to place, as well as the relative importance given to a place by various communities. [38]
- Use reporting of news (both the reporting of an event as well as the time and place of the contents) with continuous sensor data (such as physical sensors or VGI) to see impacts and trends. Ex: reporting of an election's impact on societal activity or environmental accelerometer data pre-coverage of an earthquake [31]
- the tool should support situational awareness via dynamic dashboards [40]
- Geopolitics and immigration: “subjects that 2020 has pulled into sharp focus” Importance of and interest in geopolitics for everyday users [53]
- “The geoportals can facilitate researchers, government officers, and ordinary users in helping them to find the data they needed, with basic searching services equipped in the geoportal. [?] General theoretical support for geoportals supporting multiple types of users, so as in the publication geoportal tool. [13]

3.8 Features

- The system should ultimately have filters that equate to verified contributors (those who have been trained to use/the tool is a part of their process), such as commercial or government organizations. Individuals or smaller users can also contribute to the map, but may be filtered from these views as “recreational contributors” (or similar) that can allow viewers/researchers/readers the ability to remove individual blog or potential garbage from cluttering the view (in addition to searching for specific sources, or checking/unchecking sources from a list to customize the contributors). [27]
- Instead of the spread method of balanced marker modes, I prefer the clustering of ALL stories with a number indicating how many can be found in an area that is resolved by

zooming in. [43]

- “Focus’ tag... users can assign it to indicate that a particular toponym is the focal place of a story, i.e., a place that a document is ‘about’. A document might name many places in its textual content, but might primarily tell a story about only one or a few of them.” [2]

3.9 Evolving gazetteer

-Evolving spatial lexicon: Any particular article may not be fully specified. Just as events don’t necessarily associate exactly to administrative spatial boundaries, they won’t necessarily all initiate at the beginning of or within the study timeframe. Just as a reader’s lexicon develops over time, so should the associated gazetteer. Manual specification allows the user the opportunity to name/address and define these flowing areas as their and general understanding of an event’s location changes over time. A military base may change and move over time, the same named thing within a city could come to mean different locations at different times. The flexibility to adjust on the fly is critical to the success of this type of initiative. [45]

3.10 Application of results

The project results will be licensed as free and open source such that these can be accessible and leveraged by other individuals or organizations for further development or related projects. Wherever possible, the project will leverage existing open source tools, platforms, and data. Thus far, all sources of data are already publically available, which mitigates concerns of distribution of proprietary materials. -The tool can support the aggregation of public, private, and citizens knowledge to be used for more nuanced applications. Foundation for future development and sophistication of tools in an increasingly spatialized data context. [9]

- he tool is not an SDI but could contribute to a local SDI, like the one described in this paper, to support “geomatics for sustainable societies”. [31]
- Results should be sharable via GeoRSS. Format for distributing search results, easy search and linkin to other open projects. [32]
- Results should also be shareable via downloadable CSV, shapefile, etc.
- Extract: value in research and monitoring to be able to save datasets (or better yet support integration in other platforms) [22]

3.11 Structure

- The tool should support only enough information per article to validate interest of the party (they can be read from the article sources themselves... this is not a blog hosting site), the rest should be left for nuanced filtering and discovery. [13]
- Even in blog type scenarios... the tool should leverage existing blogsites to host individual content, adding only a plugin for a location element. The tool does not host blog, only summaries and attribute data that links to the original source. The content is included within the spatial database only as an attribute such that it can be searched for keywords [9]
- This tool is not intended to be an automatic extraction, yet rather a beginning of an industry pivot towards including spatial definition such that it does not need to be extracted

in the future. Those describing the content should have the ability to identify precisely the location (point or area) of interest at the point of publication (leaning on existing addresses, boundaries, or points of interest as appropriate, if custom defined areas aren't necessary). [31]

-The tool may leverage different coordination methods (top down and bottom up). Historical data should be mapped via broker coordination, whereas anything included via the input portal (white labeled on no) can be considered "federated" coordination. Focus on the broker to improve sustainability and open application of the tool. [13]

-Expected use of 1D, 2D, Temporal, Multi-dimensional, tree (related articles) data types [22]

3.12 Languages

The Web App should support the definition of use in English and Portuguese (leveraging a platform for expansion to other languages via internationalization and localization techniques) for all elements of the user interface, such as lproejct description, instructions, filters, units, etc. All data incorporated from external sources (such as news article contents, publisher tags, gazetteer names, etc.) may remain in their original forms/languages. If possible, alternate forms will be supported if provisioned by the original source. The langauge opptions of English and Portuguese should support the international use and cross investigation of a wider user base. -Allow users to select several languages (English and Portuguese) [20]

3.13 Sustainability

The project is a foundation for future development in the geospatial and temporal distribution of news story contents. The proof of concept should demonstrate the value of such filtering and may be built upon in one or more of the following ways:

1. as a free tool;
2. as the base of a new online news journal product;
3. incorporated into existing online databases to incorporate the temporal spatial dimension into and enhance their own thematic tools; or
4. to be incorporated into municipalities as a public participation platform / community empowerment tool to better understand incidents that are spatially relevant.

This last option is especially interesting if future planned events and city data are layered in. It is also the direction of most interest to me and future efforts may involve collaboration with one or more cities to design a public participation tool. Additional functionality may include additional languages, additional study areas, development of a smart phone application, an option for automatic localization (such as for geo-tagging news stories or proximal searching), incorporation of historical datasets, incorporation of future events, additional data visualization options, APIs for integration with other applications, white-labeling options for commercial applications, etc.

At minimum, its documentation and codebase will be available under an open license from which anyone may develop in the future.

3.14 Impact

1. 1 webapp, freely and openly accessible, available in English and Portuguese languages
2. 1 webapp development code, open licensed for further or related future development by any individual or organization.

3.15 Concerns

-What kind of potential themes may emerge (accurately or otherwise) of places in Lisbon? How should these be dealt with? This project doesn't have any visibility, yet it is meant to identify trends – both actual (because such events are happening in a particular pattern) and contrived (only events of a certain type are being covered in certain places by journalists). [14]

*Incorporate verbage from section 1.3 to 2 of proposal

*Organize from here

4 Methodology

4.1 Study area: Portugal

- DNI Fund Portugal has taken an interest in innovative journalism: João Palmeiro (president of the Portuguese Publishers Association) chairs the DNI Fund, which both supports portuguese projects (Alberto Pereira, Cofina Media, Empresa jornalística, região de leiria, global notícias, global notícias publicações, impresa, impresa publishing, INESC TEC, João Antunes, Agência de Notícias de Portugal [LUSA]) and includes Portuguese partners (Plataforma de media privados, Público, Público: comunicação social, Ricardo Lafuente, Universiety of Porto, and Visapress). [18]

-Independent Radio “The end of the 70s marked the appearance of several pirate radios in Portugal beginning a process that would lead to the liberalization of the radio sector at the end of the following decade. For eleven years, several hundred small radios have broad-casted without a license bringing the voices of local people.” [66]; “Os principais impulsionadores do fenómeno da radiodifusão local portuguesa referiam frequentemente a importância da existência de um discurso descentralizado e da partilha do processo de decisão sobre a coisa pública.” The main catalysts of the local portuguese pirate radios frequently refer to the importance of the existence of decentralized speech and the sharing of the public decision making process. [66];

4.1.1 Demographics

4.1.2 Local news serving Lisbon

- Commercial media
 - Público
 - Diário de Notícias
 - Jornal de Notícias
 - Correio da Manhã

- Lisboa Euronews
- Exame
- Jornal Finanças
- Expresso
- Agência Lusa
- TSF
- Destak
- Sapo
- Notícias ao Minuto
- Observador
- Jornal I
- Public media
 - Boletim Municipal Câmara de Lisboa
 - CML Notícias
 - Juntas de Freguesia: notícias, newsletters, agendas, etc.
- Independent
 - **identify**

4.1.3 Geo tools

4.2 Model design

See Appendix C

-Limit classification complexity. [13]

4.3 GIS design

- Database: PostgreSQL [7, 31, 43, 67]
- PostGIS [31, 67]
- Geonode: open source platform for collaborating geo-spatial data; Geonode/GeoNetwork/Django/GeoExt: “provide a platform for sophisticated web browser spatial visualization and analysis” [31]

4.4 Dataprep

- QGIS [67]

4.5 Architecture

- Choose dynamic hosting for an application driven site [68]
- Drupal as content management system (CMS) [68]
- Shared hosting is presumably sufficient: cheap, easy to manage, low traffic volume [68]
- Confirm external database linking [68]

4.6 Web Interface

- API [13]
- OGC WMS, CSW [13, 31]
- GeoServer [13, 31, 67]

4.7 Front end

- Leaflet [59, 67]
- Openlayers [?, 67]
- GoogleMaps API [59]
- OpenStreetMaps [59]

4.8 Backend

- GeoExt [31]
- “Open Geogrpahic Modeling System (Open GMS) provides specific data process services (including data mapping services, refactoring services, and visualization services) beyond traditional data services to help modeler to prepare suitable data for geographic simulation the open web environment. [13]
- OpenSearch: navigation of gazetteers [13]
- Google address autocomplete [59]
- Python [67]
 - Shapely [67]
 - Geopandas [67]
 - Gdal [67]
 - Pyproj [67]
- Geodjango [67]

4.9 Manipulation

- Leaflet Boundary Canvas
- Free draw
- GeoAnnatator: supports human annotators [2]

4.10 Data Collection

4.10.1 Corpora

The following corpora is selected as a representation of event reports for the month of October, 2020. It incorporates public organization (Municipality and selected Freguesias) as well as private newsmedia sources to inform the organisational rubric and data model.

Source	Data	#	Contents' dates	Scale
CML	Boletim 1392	36	Mar - Oct20	address - municipality
	Notícias	42	Oct20	sub-bairro
JFC	Notícias	21	Sep - Dec20	sub-bairro - freguesia
	Newsletter			
JCE	Agenda	4	Jun-Dec20	freguesia
	Notícias	10	Oct20-Jan21	address - freguesia
Público*	ípsilon ímpar Público Fugas p3 (uncategorized)	83	Oct20-2030	address - inter-municipality
Total		196	Mar20-2030	address - inter-municipality

Table 1: Corpora data

*Público data was procurred via the website's search feature, which includes results from all of their products. The filters used were keyword "LISBOA" published during October 2020.

All attributes are copied directly from the source materials, with the exception of times and places. These are extracted from the corpora, as these elements may not be explicitly stated ("yesterday" or "near the road" or "on his birthday"). Required fields indicate those that must be extractable from articles to be included in the corpora. Priority fields are ideally included, and efforts to extract this data will be made if not immediately obvious. If not applicable, these fields may remain blank. Other fields are helpful but not critical for inclusion.

Attribute	Type	Priority	Visibility	Note
Title	Text	Critical	Yes	
Summary	Text	Low	Maybe	Keyword search priority
Contents	Text	High	No	Keyword search
Photo	Web link	Low	Maybe	
Section	Text	High	Yes	Thematic filtering
Themes	Text	High	Yes	Thematic filtering
Times	Text list	High	Yes	Temporal filtering
Places	Text list	High	Yes	Geospatial filtering
Referenced articles	Web link	Low	Maybe	Suggestion
Related articles	Web link	low	Maybe	Suggestion
Author	Text	High	Yes	
Source	Text	Critical	Yes	
Publication date	Date	Critical	Yes	Default temporal filtering and extraction
Link	Web link	Criticl	Yes	

Table 2: Corpora attribute collection

4.10.2 Basemaps

Source	Name	# Records	Geometry	SRS
CML Geodados	Quarteirões	1086	Area	EPSG:4326
	Grandes parques e jardins de lisboa	190	Area	EPSG:4326
	Rede viária	3763	Line	EPSG:4326
	Limite do conselho	1	Area	EPSG:4326
	Rede ferroviária subterrânea	1	Line	EPSG:4326
DGTerritório	CAOP 2019 - Continental	3,223	Area	EPSG:3763
Bairros				
Address				
Total		8,224		

Table 3: Base map data

4.10.3 Gazetteers

-Include feature codes from GeoNames to help with disambiguation of human annotators selecting toponyms [2]

Source	Name	# Records	Geometry	SRS
GeoNames	PT	37327	Point	WGS84
OSM PT	Waterways	49,376	Line	WGS84
		54,041	Area	WGS84
	Transport	18,696	Point	WGS84
		1,078	Area	WGS84
	Traffic	84,756	Point	WGS84
		35,339	Area	WGS84
	Roads	1,035,765	Line	WGS84
	Railways	8,310	Line	WGS84
	POIs	100,214	Point	WGS84
		71,335	Area	WGS84
	POWs	2,133	Point	WGS84
		7,358	Area	WGS84
	Places	23,594	Point	WGS84
		1,971	Area	WGS84
	Natural	89,483	Point	WGS84
		1,369	Area	WGS84
	Landuse	209,330	Area	WGS84
	Buildings	1,060,745	Area	WGS84
Total		2,892,220		

Table 4: Gazetteer data

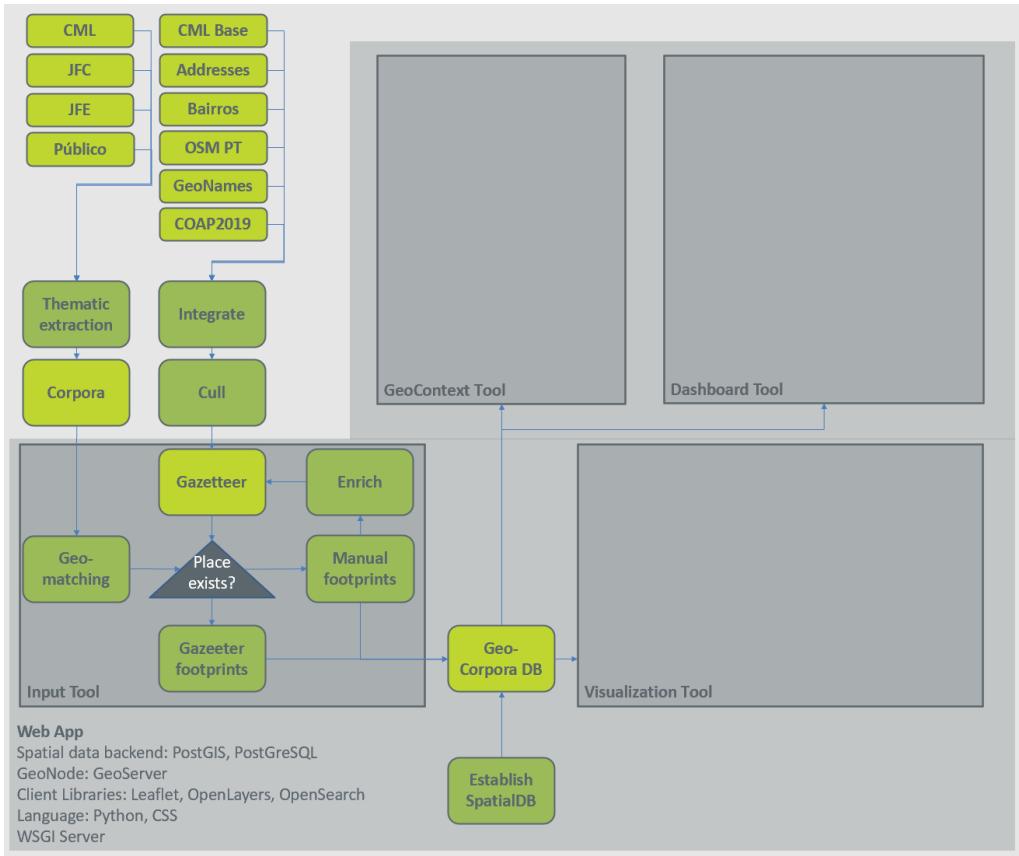


Figure 2: Preliminary methodology

4.11 Preprocessing

Load all gazetteer data into QGIS. Establish georeference (note in data breakdown), transform as necessary Manually code all article data. Using multiple sources to ensure continuity across local data sources.

4.12 Initialization of development environment

Selection and integration of resources and tools:

- Database: PostgreSQL

Establish geodatabase structure, accommodating multiple language options, load gazetteer(s) and relevant administrative boundary data Develop and test Input tool: wordpress plugin? Develop and test search tool Develop and test Context tool Translate web app content to portuguese and load translations Migrate site to the server Test system Compare results against mined location results Document results

Future development: dashboard tool

4.13 Tool design

-Define design requirements and views to achieve them [24]

4.14 Testing

- Test in different browsers for feature functionality. [20]

4.15 Validation

1. GDELT: compare Lisbon Oct 2020 to results 2. Automatic extraction over sample corpora to compare results leverage Rivera2020 study - spaCy - “Large scale infomration extraction tasks” [69]
 - Free online classes to learn [69]
 - Utilize NER to extract exact and complete places. As a baseline comparison to the input tool? Used together? Feed through the system with a “check point” (verify the identified place before input)? [1]
 - Use local datasets (Público, CML, Freguesia) from 1 month of each to build and test a backoffice plugin for defining localization of news articles. Ask MAGG or similar to apply the plug in for upcoming articles for a 1 month period and test against the same dataset with NewsStand. Is it richer/more accurate? Is the journalist satisfied with the result? Is the reader? Load the same local base gazetteer (OSM Portugal and Global) [45]
 - Use method for extraction on input articles to extract and geolocate place for input comparison to automated methods?? Python and Restful API

5 Results

6 Analysis

6.1 Future opportunities

6.1.1 Validation

Though testing of these hypotheses through rigerous comparison to the status quo (traditional online news sources without a spatial element) and emerging product performing automatic extraction of place (such those of the GDELT project) are not included in this endeavor, the resulting tools should provide a basis from which future projects may develop and evaluate.

6.1.2 Application to Geointelligence

- Potential future use of this toolset: GeoINT source. [38]
- Potential future focus/further development for dashboard monitoring OR the system could feed into such systems, providing high confidence georeferencing as it has been manually defined by the author. [43]

6.1.3 Sub-article definitions

- Future research directions: “spatial role labelling... ‘the task of identifying and classifying the spatial arguments of spatial expression mentioned in a sentence’... spatial role labelling is key not only in geographic information retrieval but also in domains such as text-to-scene conversion, robot navigation or traffic management systems.” [2]

- “We should experiment with how to visualize uncertainty, possible errors and imperfections in our data. And most importantly, we should keep in mind how data can be a powerful tool for all designers, bringing stories to life in a visual way and adding structural meaning to our projects.” [26]

6.1.4 Recommended extraction of place

- Avoid the issue of high volume evaluation all together. Those who write the story can apply the location which is the most accurate option. In the absence of the author assigning place, automatic extraction (perhaps such an LSTM model) is an appropriate tool for historical articles. [48]

6.1.5 Incorporation of historical stories

GDELT

6.1.6 Statistical analysis

- Online clustering: “a clustering algorithm for the news domain should group together all news articles that describe the same news event into groups of articles termed story clusters. Broadly, a news event is defined in terms of both story content and story lifetime - articles in the same cluster should share much of the same important keywords, and should have temporally proximate dates of publication.” [43]

- Future development: “We will consider ways to use clustering to determine the news provider’s geographic scope (i.e. the geographic location of the newspaper), and use it to improve both geotagging and local news coverage.” leveraging local knowledge/gazetteers [43]

7 Conclusion

References

- [1] Sarang Gupta and Kumari Nishu. Mapping Local News Coverage : Precise location extraction in textual news content using fine-tuned BERT based language model. In *The Fourth Workshop on Natural Language Processing and Computational Social Science*, pages 155–162. Association for Computational Linguistics, 2020.
- [2] Morteza Karimzadeh and Alan M. MacEachren. Geoannotator: A collaborative semi-automatic platform for constructing geo-annotated text corpora. *ISPRS International Journal of Geo-Information*, 8(4), 2019.
- [3] Markus Eisl. Searching European Data, 2020.
- [4] Matt Makai. Deployment.
- [5] Guoray Cai and Ye Tian. Towards geo-referencing infrastructure for local news. *Proceedings of the 10th Workshop on Geographic Information Retrieval, GIR 2016*, pages 1–10, 2016.
- [6] Mário J. Silva, Bruno Martins, Marcirio Chaves, Ana Paula Afonso, and Nuno Cardoso. Adding geographic scopes to web resources. *Computers, Environment and Urban Systems*, 30(4):378–399, 2006.
- [7] Tiago H Moreira De Oliveira and Marco Painho. Open Geospatial Data Contribution Towards Sentiment Analysis Within the Human Dimension of Smart Cities. In A Mobasher, editor, *Open Source Geospatial Science for Urban Studies.*, pages 75–95. Springer International Publishing, 2021.
- [8] Stéphane Roche and Abbas Rajabifard. Sensing places’ life to make city smarter. *ACM SIGKDD International Workshop on Urban Computing (UrbComp 2012)*, 2012.
- [9] Nader Afzalan, Thomas W. Sanchez, and Jennifer Evans-Cowley. Creating smarter cities: Considerations for selecting online participatory tools. *Cities*, 67(April):21–30, 2017.
- [10] Massey D. A Global Sense Of Place. *Marxism Today*, June(June):24–29, 1991.
- [11] A. Rajabifard. Realizing Spatially Enabled Societies – A Global Perspective in Response to Millennium Development Goals. *Eighteenth United Nations regional Cartographic Conference for Asia and the Pacific, Bangkok, 26-29 October*, page 9, 2009.
- [12] Chris Chiappinelli. Democracy: Fueled by Pizza, nov 2020.
- [13] Hao Jiang, John Van Genderen, Paolo Mazzetti, Hyeongmo Koo, Min Chen, Hao Jiang, John Van Genderen, Paolo Mazzetti, Hyeongmo Koo, Min Chen, and Hyeongmo Koo. Current status and future directions of geportals. *International Journal of Digital Earth*, 13(10):1093–1114, 2020.
- [14] Jayna McQueen Baker, Gabriel Huddleston, and Erin Atwood. The Map as Object: Working Beyond Bounded Realities and Mapping for Social Change. *Educational Research for Social Change*, 8(1):138–152, 2019.
- [15] Gilberto Rivera, Rogelio Florencia, Vicente García, Alejandro Ruiz, and J. Patricia Sánchez-Solís. News classification for identifying traffic incident points in a Spanish-speaking country: A real-world case study of class imbalance learning. *Applied Sciences (Switzerland)*, 10(18), 2020.
- [16] Peter Williams. What, Exactly, is a Smart City?, 2016.
- [17] Jennifer Evans-Cowley and Justin Hollander. The new generation of public participation: Internet-based participation tools. *Planning Practice and Research*, 25(3):397–408, 2010.
- [18] Elevating quality journalism Digital News Innovation Fund Report 2018. Technical report, Digital News Innovation Fund, 2018.
- [19] Sabine Witschas. Cross-border mapping-Geodata and Geonames. In *Association for Borderland Studies European Conference*, Graz, 2004. Borders in a new Europe.
- [20] Ben Shneiderman. Data Visualization’s Breakthrough Moment in the COVID-19 Crisis, apr 2020.
- [21] Elijah Meeks. 2019 Was the Year Data Visualization Hit the Mainstream, dec 2019.
- [22] Ben Shneiderman. The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. Technical report, University of Maryland, College Park, jul 1996.
- [23] Katy Börner, Andreas Bueckle, and Michael Ginda. Data visualization literacy: Definitions, conceptual frameworks, exercises, and assessments. *PNAS*, 116(6):1857–1864, feb 2019.
- [24] Yixuan Zhang, Kartik Chana, and Cody Dunne. IDMVis: Temporal Event Sequence Visualization for Type 1 Diabetes Treatment Decision Support. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):512–522, 2019.

- [25] Giorgia Lupi.
- [26] Giorgia Lupi. Data Humanism, The Revolutionary Future of Data Visualization, 2017.
- [27] Greg G. Brown and David V. Pullar. An evaluation of the use of points versus polygons in public participation geographic information systems using quasi-experimental design and Monte Carlo simulation. *International Journal of Geographical Information Science*, 26(2):231–246, 2012.
- [28] Albert Acedo, Tiago Oliveira, Mijail Naranjo-Zolotov, and Marco Painho. Place and city: toward a geography of engagement. *Heliyon*, 5(8):e02261, 2019.
- [29] Marco Painho and Isabel Pina. The invisible cities-can PPGIS connect citizens to urban policies? *Revista Internacional de Ciencia y Tecnología de la Información Geográfica*, 1(13):1–4, 2013.
- [30] Reinout Kleinhans, Maarten Van Ham, and Jennifer Evans-Cowley. Using Social Media and Mobile Technologies to Foster Engagement and Self-Organization in Participatory Urban Planning and Neighbourhood Governance. *Planning Practice and Research*, 30(3):237–247, 2015.
- [31] Devanjan Bhattacharya and Marco Painho. Location intelligence for augmented smart cities integrating sensor web and spatial data infrastructure (SmaCiSENS). *GISTAM 2018 - Proceedings of the 4th International Conference on Geographical Information Systems Theory, Applications and Management*, 2018-March(Gistam):282–289, 2018.
- [32] Hanfa Xing, Jun Chen, and Xiaoguang Zhou. A geoweb-based tagging system for borderlands data acquisition. *ISPRS International Journal of Geo-Information*, 4(3):1530–1548, 2015.
- [33] Sarah Barns. Joining the dots: Platform inermediation and the recombinatory governance of Uber’s ecosystem. In Mike Hodson, Julia Kasmire, Andrew McMeekin, John G. Stehlin, and Kevin Ward, editors, *Urban Platforms and the Future City: Transformations in Infrastructure*, chapter Joining th. Routledge, 2020.
- [34] Hussein S. Al-Olimat, Krishnaprasad Thirunarayan, Valerie Shalin, and Amit Sheth. Location Name Extraction from Targeted Text Streams using Gazetteer-based Statistical Language Models. pages 1986–1997, 2018.
- [35] Maryam Bahojb Imani, Latifur Khan, and Bhavani Thuraisingham. Where did the political news event happen? Primary focus location extraction in different languages. *Proceedings - 2019 IEEE 5th International Conference on Collaboration and Internet Computing, CIC 2019*, (Cic):61–70, 2019.
- [36] Agnieszka Leszczynski. Platform affects of geolocation. *Geoforum*, 107(May):207–215, 2019.
- [37] Dean Hintz and Craig Hantke. How government agencies are integrating & delivering data for emergency response, 2020.
- [38] Anusuya Datta. Top six GEOINT trends, 2018.
- [39] National Geospatial-Intelligence Agency, 2020.
- [40] Artur Jorge Abreu Varanda. Project "ORACULO": Extracting events from news streams and mining their spatiotemporal patterns to support UN operations in the Central African Republic, 2020.
- [41] Luke S. Snyder, Morteza Karimzadeh, Ray Chen, and David S. Ebert. City-level Geolocation of Tweets for Real-time Visual Analytics. *arXiv*, pages 0–3, 2019.
- [42] James. An attempt to extract geo-location from text, 2020.
- [43] Benjamin E. Teitler, Michael D. Lieberman, Daniele Panizzo, Jagan Sankaranarayanan, Hanan Samet, and Jon Sperling. NewsStand: A new view on news. *GIS: Proceedings of the ACM International Symposium on Advances in Geographic Information Systems*, (January):144–153, 2008.
- [44] Morteza Karimzadeh, Scott Pezanowski, Alan M. MacEachren, and Jan O. Wallgrün. GeoTxt: a scalable geoparsing system for unstructured text geolocation. *GIS*, 23(1), 2019.
- [45] Michael D. Lieberman, Hanan Samet, and Jagan Sankaranarayanan. Geotagging with local Lexicons to build indexes for textually-specified spatial data. *Proceedings - International Conference on Data Engineering*, (May):201–212, 2010.
- [46] Felix Hamborg, Corinna Breitinger, and Bela Gipp. GiveMe5W1H: A universal system for extracting main events from news articles. In *Proceedings of the 7th International Workshop on News Recommendation and Analytics*, Copenhagen, Denmark, 2019. INRA.
- [47] Sophie J. Lee, Howard Liu, and Michael D. Ward. Lost in space: Geolocation in event data. *Political Science Research and Methods*, 7(4):871–888, 2019.
- [48] Andrew Halterman. Geolocating political events in text. In *Proceedings of the Third Workshpo on Natural Language Processing and Computational Social Science*, pages 29–39, Minneapolis, Minnesota, 2019. Association for Computational Linguistics.

- [49] ESRI. Adjust how locations and attributes are extracted.
- [50] Richard Nordquist. Exonyms and Endonyms, jan 2018.
- [51] OpenStreetMap.
- [52] NGA: GNS Home.
- [53] Jacob Granger. Isabelle Roughol of Borderline podcast, on the pros and cons of 'indie journalism', nov 2020.
- [54] Jacob Granger. "Mass customisation" of neighborhood data can help hyperlocal news become more sustainable, sep 2020.
- [55] Sarah Marshall. #Tip of the day for journalists: Geotag your content using a WordPress plugin, oct 2012.
- [56] Crosstown: About Us, 2020.
- [57] Crime, 2020.
- [58] InYourArea, 2020.
- [59] Eyal Fitoussi. GEO my WordPress.
- [60] Infinitum Form. CF GEO Plugin.
- [61] Fabio Parasecoli and Mateusz Halawa. Rethinking the global table. In May Rosenthal and Catherine Flood, editors, *Food: Bigger than the plate*, pages 80–89. Victoria and Albert Museum, London, 2019.
- [62] Eric Fisher. Making the most detailed tweet map ever, dec 2014.
- [63] Richard Ishida and Susan K. Miller. Localization vs. Internationalization, 2005.
- [64] Diogo Queiroz de Andrade and Catarina Carvalho. Interview, 2020.
- [65] Newspack, 2020.
- [66] Luís Bonixe. As primeiras experiências de radiofusão local em Portugal (1977-1984). *Os média no Portugal Contemporâneo*, 19(2):183–195, 2019.
- [67] Ramiz Sami. Tools I recommend for building Geospatial Web Applications — by Ramiz Sami — The Startup — Medium, oct 2019.
- [68] Jerry Low. How to Host Your Own Website: Complete Beginner Guide, sep 2020.
- [69] Exposition AI. spaCy, 2020.

Appendix

A Relevant terms

Abstract place (AP): A point in space or area non-conforming to current or historical ABs or recognized POIs.

Administrative boundary (AB): A geographical area limit managed by an entity; ex: the municipality of Lisbon, Portugal or the 2nd congressional district in Colorado

.

Aliasing: “multiple names refer to the same geographic location, such as ‘Los Angeles’ and ‘LA.’” [43]

Attribute: an informative element of data stored in a data field.

Spatial attribute (SA): a description relating to location; ex: ‘where did something happen’ or ‘where was it logged’.

Temporal attribute (TA): a description of when; ex: ‘at what time did it happen’ or ‘which day was it published’.

Thematic attribute (ThA): a description of what, why, or how; ex: ‘what happened’ or ‘who published it’).

Ambiguity

Referent ambiguity: “the same location can have more than one name” [6]

Referent class ambiguity: “the same name can be used for locations as well as for other class of entities, like persons or company names” [6]

Referent ambiguity: “the same name can be used for more than one location” [6]

Civic engagement (CE): [define!](#) [28]

Comma separated value (CSV): text file of data records (features) in which each record is stored as a new line and its attributes (fields) are delimited by a comma.

Content management system (CMS): [define](#) [68]

Contributor: [define!](#)

Data model: a graphical representation of the data structure and relationships definitions.

Data visualization literacy (DVL): [define](#) [23]

Endonym: “a locally used toponym” [50]

Even-occurring: “all locations where events occurred regardless of whether the event is the event of interest.” [47]

Even-relevant: “all locations where events occurred regardless of whether the event is the event of interest.” [47]

Exonym: “a externally used toponym” [50]

Gazetteer: A geographical index relating descriptors to location; ex: [GeoNames](#), which related names of places to geographical coordinates.; Gazetteer: “a geographical dictionary, most commonly containing place names and associated properties such as geographic coordinates, type of place, and population, among others.” [44]

Geographic information retrieval: [44]

Geographic name ambiguity: “a given name might refer to any of several geographic locations.” [43]

Geocoding: “the process of taking input text, such as an address or the name of a place, and returning a latitude/longitude on the Earth’s surface for that place [1]; “Geocoding is the process of parsing places and addresses written in natural language into canonical geocodes, i.e., one or more coordinates referring to a point or area on earth.” [46]

Geographic information system (GIS): A framework for the manipulation and analysis of geographic data.

Geoinformatics: [define](#)

Geoparsing: [the recognition of place names in text](#)’ [6]; Geoparsing: “to enable the use of unstructured text in GIS, place references mentioned in text must be automatically recognized and resolved to the geographic coordinates of those places.” [44]; “the process of recognizing place names in text (‘toponym recognition’) and resolving them to their coordinates or gazetteer entry (‘toponym resolution’)” [?]; “the process of automatically resolving place reference in natural language (unstructured text) to toponyms in a geographic gazetteer with geographic coordinates. Geoparsing enables the extraction of textual information about places for use in geographic information systems (GIS) and other applications. [2]

Geoportal: A user (usually a journalist for commercial publications, a government official for public organizations, or an individual for a blog contribution) that assigns spatial definitions to an incident via the *Input* tool. ”A consolidated web-based solution to provide open spatial data sharing and online geo-information management” [13]

Geotagging: “The process of identifying and disambiguating references to geographic locations (i.e., toponyms), known as geotagging, consists of two steps: toponym recognition, where all toponyms (e.g., “Paris”) are identified, and toponym resolution, where each toponym is assigned to the correct geographic coordinates among the many possible interpretations (e.g., “Paris” which can be one of over 140 places including France and also Texas).” [45]

Ground truth: hand coded set of actual locations for training and verification [47]

Incident: Defined within the project as any content of a news article that has spatial and temporal dimensions. These can be past, present, future, or related to multiple instances in time. Likewise, each can occur in a single place or in multiple places, as a point in space or as an area (polygon), and be associated with a recognizable place (such as an AB or a POI) or over areas not commonly recognized (an AP).

Internationalization (i18n): “the design and development of a product, application or document content that enables easy localization for target audiences that

vary in culture, region, or language.” [63]

Localization (l10n): “the adaptation of a product application or document content to meet the language, cltural and other requirements of a specific target market” [63]

Location: [define](#)

Location based services (LBS): [define](#)

Named entity recognition (NER): “which is concerned with the identifying entities such as person, location, and organization names.” [43]; “ the task of extracting and distinguishing different types of entities in text (i.e. names of people or organizations, dates and times, events, geographic features or even ‘non entities’) “ [6]

Natural language processing (NLP): [define](#)

NeoGeography: [define](#) [29]

Open source (OS): a development methodology, the product of which is free of any restrictions of use, permits access to (for the study or modification of) the source code as well as the distribution of original or modified copies to third parties.

Online participatory tool (OPT): [define](#) [9]

Ontology: “a formal, explicit specification of a shared conceptualization” [32]

Place: [define](#)

Place attachment (PA): [define](#) [28]

Point: [define!](#)

Point of interest (POI): any entity (natural or artificial) with a well-defined location; ex: Praça do Comércio or Garden of the Gods.

Polygon: [define!](#)

Proof of concept (POC): functional or demonstrative of the basic project concepts. [28]

Reader's spatial lexicon: “those locations that the reader can identify and place on the map without any evidence” [45]

Really simple syndication (RSS): [define](#)

Scope: [define](#)

Content scope: “the story content's geography” [43]

Provider scope: “the publisher's geographic location” [43]

Serving scope: “based on the reader's location” [43]

Sense of place (SOP): [define](#)

Smart city: [define](#)

Social capital (SC): [define](#) [28]

Spatial data infrastructure (SDI): [define](#) [8]

Tag: content, section, or descriptive designations defined by the media publisher; ex: ‘política’, ‘primeiro-ministro’, ‘governo’ (from Público), or ‘coronavirus’, ‘denver’, ‘homelessness’ (from The Denver Post).

Toponym: a textual reference to geographic location [45]

Type by task taxonomy (TTT): [?]

User interface (UI): the method of interaction between a user and the program.

Volunteered geographic information (VGI): Define [8]

Web application (Web App): a program running on a web server that is accessible via a web browser with internet connectivity.

Wireframe: a design mockup of a website to demonstrate functional logic.

B Data visualization

- “quantitative data can be converted into qualitative data (e.g., one may use thresholds to convert interval data into ordinal data). Ordinal rankings can be converted to yes/no categorical decisions (e.g., to make funding decisions). The reverse is possivel as well [e.g., multidimensional scaling converts ordinal into ratio data]” [23]
- “quantitative data can be converted into qualitative data (e.g., one may use thresholds to convert interval data into ordinal data). Ordinal rankings can be converted to yes/no categorical decisions (e.g., to make funding decisions). The reverse is possivel as well [e.g., multidimensional scaling converts ordinal into ratio data]” [23]

DVL-FW	Bertin	Description
Nominal	Qualitative	“support equality check”
Ordinal	Ordered	“assumes some intrinsic ranking but not at measurable intervals”
Interval	Quantitative/Numerical	“the zero point is arbitrary”
Ratio		“there exists a unique and non arbitrary zero point”

Figure 3: Logical mathematical operations permissible for data per Borner2019’s DVL-FW

Data Scales	Logical Math Operations				Measure of Central Tendency	Examples
	= ≠	< >	+ -	× ÷		
Nominal	✓				mode	  
Ordinal	✓	✓			median	  
Interval	✓	✓	✓		arithmetic mean	
Ratio	✓	✓	✓	✓	geometric mean	

Fig. 1. Logical mathematical operations permissible, measure of central tendency, and examples for different data scale types.

Figure 4: Logical mathematical operations permissible, measure of central tendency, and examples for different data scale types (Borner2019)

		Geometric Symbols		Linguistic Symbols	Pictorial Symbols
		Point	Line		
Retinal	Spatial Position	X 			
	From	Size 			
		Shape 			
	Color	Value 			
		Hue 			
	Texture	Saturation 			
		Granularity 			
	Optics	Pattern 			
		Blur 			
Motion	Speed				

Fig. 3. Four graphic symbols and 11 graphic variables from full 11 graphic symbols by 24 graphic variables set in ref. 34. Qualitative nominal variables (shape, color hue, and pattern) have a gray mark.

Figure 5: Four graphic symbols and 11 graphic variables from full 11 graphic symbols by 24 graphic variables set in ref. 34. Qualitative nominal variables (shape, color, hue, and pattern) have a gray mark. (Borner2019)

Temporal folding

- “With long, rapidly changing time series it can be difficult to see the subsets relevant to an event of interest. Many visualization excelat revealing periodic or cyclic phenomena through carefully chosen folded time scales, but non-periodic patterns can be obscured by a fixed time scale. When considering multiple event paris across sequences with varied interspersed gaps, it can be difficult to see the overall pattern of relationships between co-occurring events. THese problems are compounded by issues o data quality such as missing data, uncertainty in sensor or manual logs, inconsistency between sources with variou temporal granularities and level of accuracy, adn incorrect timestamps.” [24]

- “Temporal folding, or splitting, a sequence into periodic units like hours, weeks, months or years can be used to find cyclic phenomena. Folding can reduce pattern variety facilitating visual analysis.” [24]
- “Aligning sequences by sentinel events of interest helps users identify precursor, co-occurring, and aftereffect events... When aligned by a single event we can maintain a consistent time scale between folded or reconfigured units of event sequences. However, it can be valuable to explore the sequence between two separate sentinel events.” [24]

C Preliminary specification

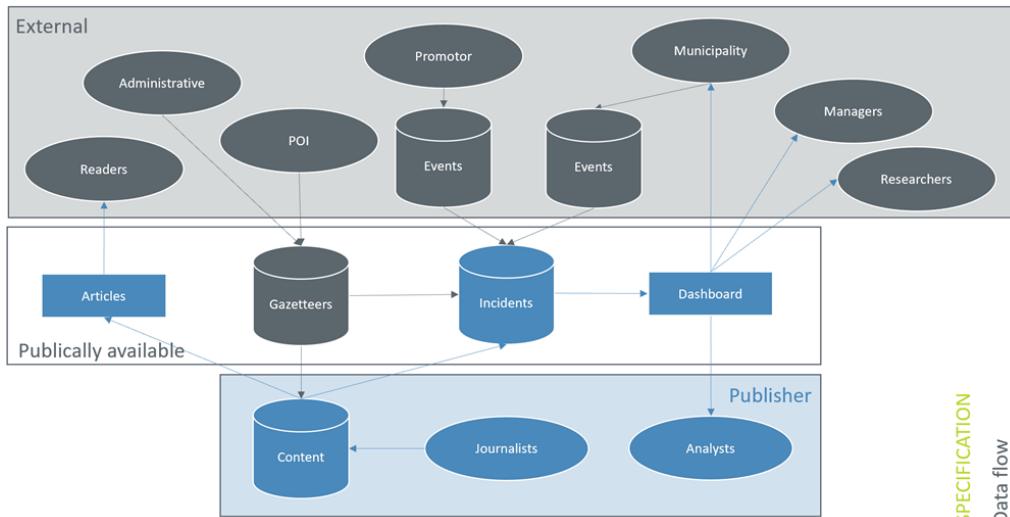


Figure 6: Preliminary data and information flow

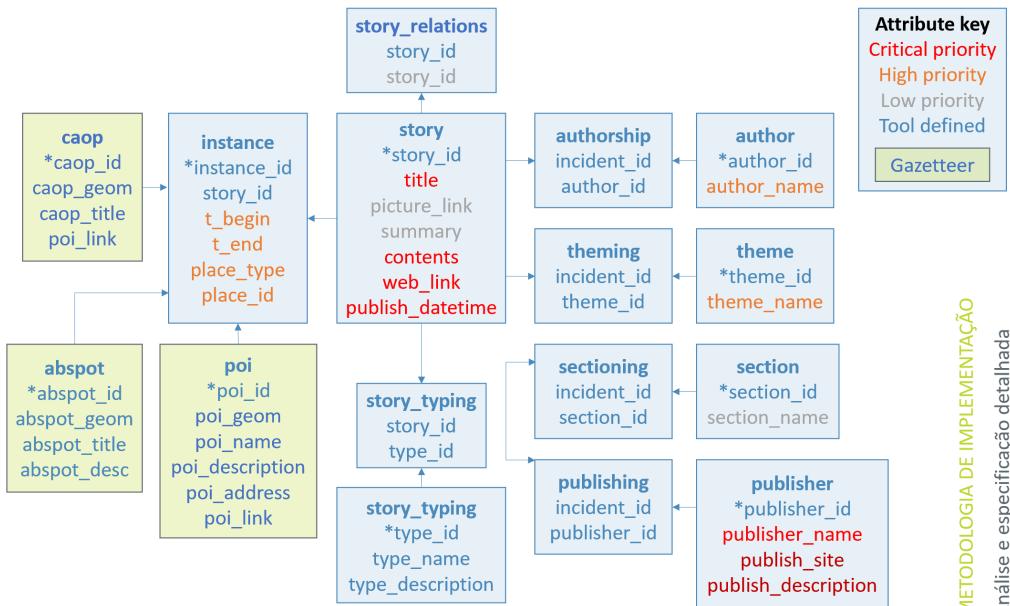


Figure 7: Preliminary data model for the spatial database



SPECIFICATION
Input UI

Figure 8: Preliminary *Input* layout

SPECIFICATIONS
Context Layout

Figure 9: Preliminary *Context* layout

RESULTS

13 out of 265

Bikes theft is on the rise in Denver. Here's how to prevent it and what to do if your bike is stolen.
10% more bikes stolen this year than last, police say
Bike Month: Nov. 1-30, 2019

[Go to Article](#)
[View tag: bike, The Local](#)

It only took a few seconds for Paul Custer's bike to be stolen when he was walking his dog in the 16th Street Mall area of Denver. "I just heard a bang — the sound of the thief taking my bike," he said. "I turned around and saw him running away with my bike. He probably had a gun or something."

[Switch to List](#)

[Export Results](#) [Save Results](#) [Build Dashboard](#)

SEARCH THE MAP

- + Date published
- + Date of action
- + Author
- + Section
- + Tags
- + Keywords
- Location
- Search
- Draw

[Clear](#)

SPECIFICATIONS Search UI

Figure 10: Preliminary *Search* layout

BANCA 5 BPI recua no pagamento de dividendo ao Caixabank 1

Banco estava isolado na intenção de distribuir dividendos em plena crise econômica provocada pela pandemia de covid-19. Depois da pressão do Banco de Portugal e de Marcelo Rebelo de Sousa, instituição prepara recuo.

Crédito Ferreira - 6 de Abril de 2020, 15:55

PAULO PIMENTA

[https://www.publico.pt/2020/04/06/banco-recua-no-pagamento-de-dividendo-ao-caixabank-1011189.html](#) 4

A informação é uma vacina contra o vírus

O PÚBLICO decidiu disponibilizar, aberta e gratuitamente, toda a informação útil sobre a evolução do novo coronavírus. O jornalismo saiu a ponderar contudo para onde ir com a crise sanitária. A informação é mais capaz de superar ansiedades, desafios e dificuldades como as que enfrentamos. Todas as histórias relevantes, de todos os países, de todos os meios de papel ou de qualquer outro bemeto no seu acesso. O PÚBLICO tem publicado vários artigos jornalísticos e de opinião sobre o tema e fez do sujeito da crise 19, assim como quais os impactos que terá na economia e a sua evolução iminente e inédita. Com responsabilidade e sentido de serviço público — porque a informação é a melhor vacina contra o medo. E contra os vírus.

Pense bem, pense PÚBLICO

Consulte este guia para melhor perceber o que está a acontecer.

[ASSINE JÁ](#)

TOPICS ECONOMIA BANCA CORONAVÍRUS 6

[ASSINE JÁ](#)

DATA ATTRIBUTES

1. Title
2. Author
3. Subtitle / summary
4. Permanent link
5. Section
6. Tags
7. Time and date of publication
8. Article text content
9. Main pictures
10. Related stories

SPECIFICATIONS Publisher provided data attributes

Figure 11: Publisher provided data attributes

D Example GUIs

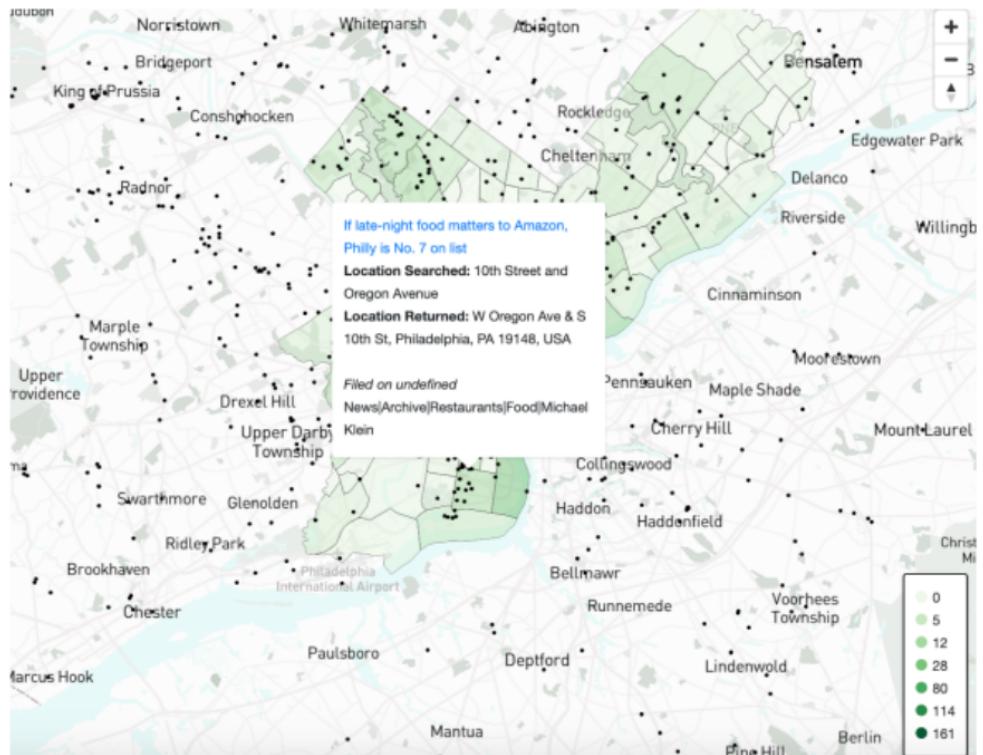


Figure 6: 1000 articles from Philadelphia Inquirer plotted on the map. The green polygons represent Philadelphia neighborhoods with color indicating number of times they have been referenced. Dots represents the locations referenced. The tooltip shows the metadata for the article that references one of the locations.

Figure 12: Mapped article layout (points) from Gupta2020 [1]

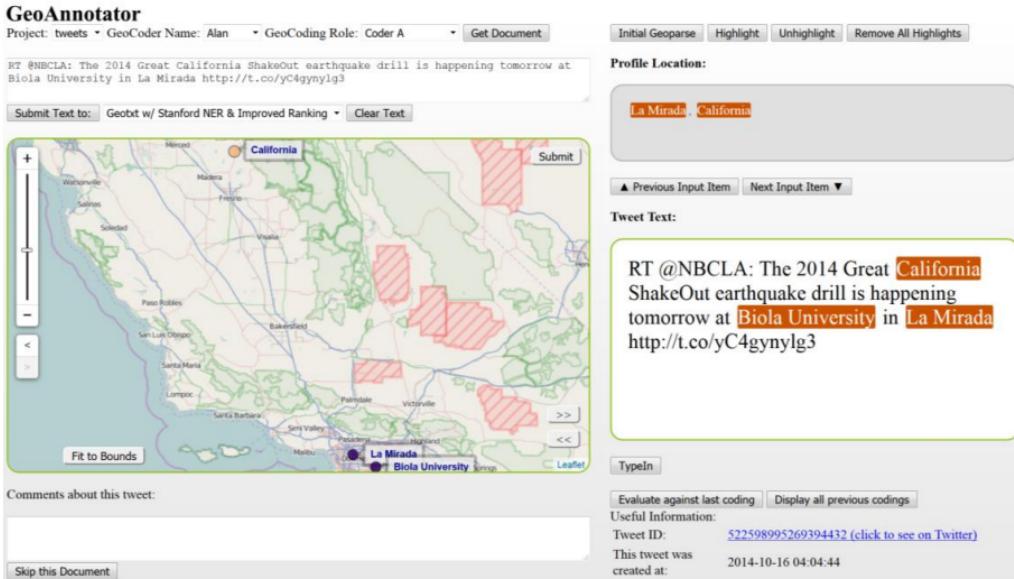


Figure 1. GeoAnnotator’s user interface (UI) main window during the initial stages of development. The panel on the right shows the text that is already annotated for named entities and fed into the platform. Place names are highlighted in orange. The panel on the left shows the map interface where highlighted named entities are mapped using a pre-generated list of toponyms through pre-processing. If the default toponym is not the correct one, users can click on it, retrieve a list of “alternative toponyms” and select the correct one. There is often some context that makes it possible for a human annotator to identify the correct toponym, either within the tweet itself, from the profile information and/or from URLs that may be included in the tweet. The interface provides a link to the original tweet on Twitter so that the annotator can search for more context if necessary.

Figure 13: Annotation via GeoAnnotator from Karimzadeh2019 [2]

Table 1. Special tags in GeoAnnotator to demarcate special cases of toponyms. For each special case, an example is provided with the place name underscored.

Tag Name	Application	Example	Notes
Uncertain semantics	When it is unclear if a name is in fact a place name, attribute, reference to an organization, or a boundary case with mixed word sense. The uncertain semantics tag enables corpus end users to include, exclude or isolate such cases for different research studies.	The rising violence by <u>Rikers Island</u> correction officers ...	Rikers Island can be interpreted to refer to the island or the Rikers Island Correctional Center Facility (both of which are “places”) or to the prison as an organization, as well as a noun adjunct modifying “correction officers”.
Vague boundaries	When the place name refers to an area or region whose boundaries are not clearly agreed upon.	Temperatures in <u>Hudson Valley</u> ...	Sources indicate that there are differences in opinion on the exact bounds of Hudson Valley.
Not in gazetteer	When the place name in text does not exist in the gazetteer yet, or is so vaguely defined that addition to gazetteer is not justified.	Headed to the <u>West Coast</u> .	Explained in more detail in the following paragraph.
Overlapping ambiguous (always including human annotator assigned surrogates list, enforced by the system)	When human annotators cannot confidently determine which one of multiple candidate toponyms that overlap in space is being referred to. GeoAnnotator allows users to assign multiple toponyms (i.e., surrogate toponyms) to the place name, and apply the “overlapping ambiguous” tag to indicate that these toponyms can interchangeably be used as the resolved toponym for that mention of Lagos (or any other similar situation).	A man just died of Ebola in <u>Lagos</u> .	GeoNames lists three toponyms for “Lagos” in Nigeria: Lagos State (administrative region), Lagos (section of populated place—the city that is within Lagos State) and Lagos Island (within Lagos City, which is within Lagos State). These entities have overlapping geospatial positions and all can be correct assignments.
Non-overlapping ambiguous (with surrogates list)	When human annotators cannot determine which one of multiple candidate toponyms that do not overlap in space is being referred to. Users can assign a surrogate list of potential candidate toponyms to a place name and apply the “non-overlapping ambiguous” tag to indicate that these toponyms can interchangeably be used.	Washington's changing demographics.	Washington may refer to “Washington D.C.” or “Washington State”, for example. These toponyms do not overlap and it is unclear which one the text author originally meant to refer to.
Non-overlapping ambiguous (without surrogates list)	When human annotators cannot determine which one of numerous candidate toponyms (that do not overlap in space) is being referred to, and there are too many potential candidates to assign as surrogates. Users can select a potential toponym and apply the “non-overlapping ambiguous” tag without providing a surrogates list (making such cases distinguishable to corpus users, who may exclude or use the cases for special studies).	<u>Springfield</u> feels like spring!	Without additional context, Springfield may be referring to numerous toponyms in different geographic regions.

Figure 14: Annotation ambiguities from Karimzadeh2019 [2]

Geodata Search Interface (1)

- **Interface**
 - Language selection
 - Information on use of search tool
 - Clear indication of a required format for data input
 - Quick searches / Extended searches (Expert mode)
 - Map representation
- **Presentation of results**
 - User defined ranking of results
 - Reasonable order of results
 - Information on data source / data volume / use restrictions etc.

ZGIS

CENTRE FOR GEOINFORMATICS – UNIVERSITY OF SALZBURG
© 2020 UNIGIS 6/18

Figure 15: Geodata Search Interface from Eisl2020 [3]

E Example Architectures

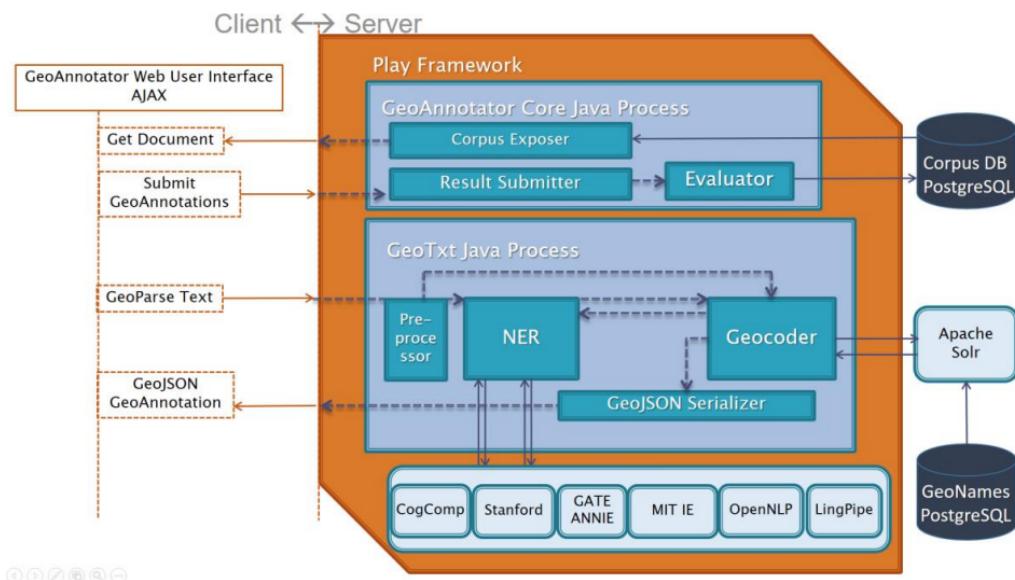


Figure 15. General architecture of GeoAnnotator. The Play framework is used to expose Java application programming interfaces (APIs) as HTTP web-service endpoints.

Figure 16: general architecture of geoannotator from Karimzadeh2019 [2]

-Client server model: server implemented in Java, and client (UI) implemented as web-pages (HTML5, CSS and JavaScript). “to be scalable and accessible on a web browser

to many annotators.” Bootstrap to be for display size flexibility, leaflet for mapping, jquery/jQuery UI/Rangy for functional operations. Server: PostgreSQL database, corpus exporer module, result submitter module, evaluator module [2]see section 4.2 for architecture description and inspiration

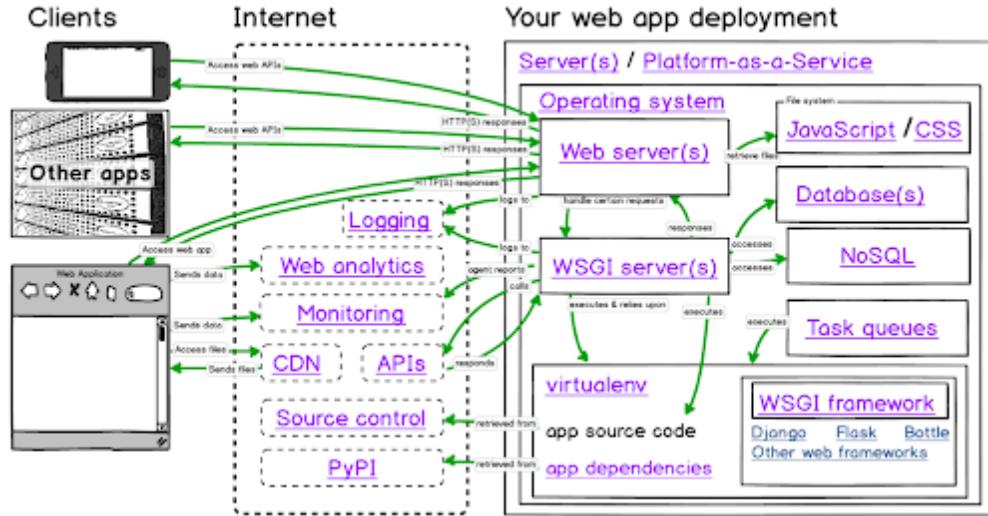


Figure 17: Full stack deployment map from Makai2'2' [4]

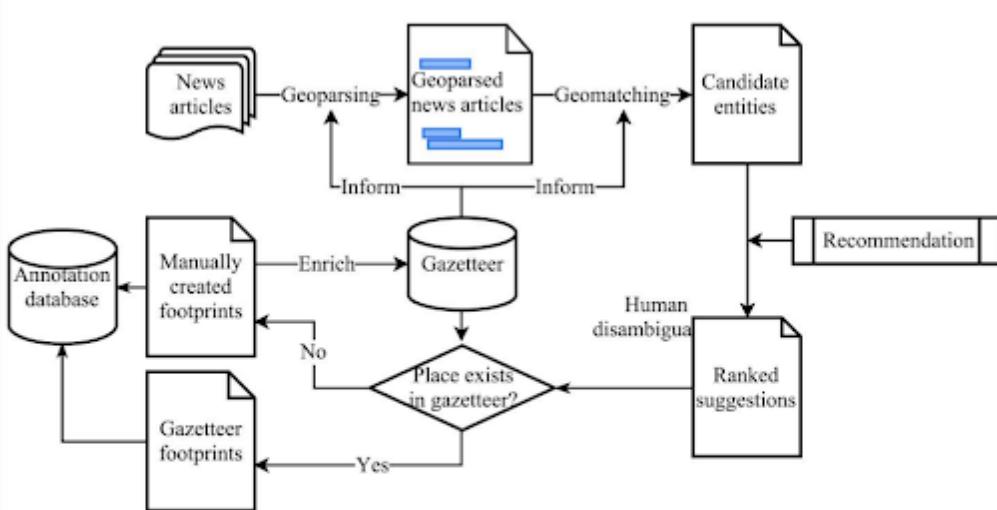


Figure 5: Progressive Geospatial Referencing Framework

Figure 18: Progressive georeferencing frameowrk from Cai2016 [5]

F User stories

F.1 Micronews Portal

- [64] - Uses Wordpress and NewsPack to publish news stories
- Require a plugin for incorporating main brands of maps (google maps, infogram, mapquest, storymaps, etc.)
- Uses geogrpahical data thatalready exists and is accessed. Want to understand which areas in Lisbon have more traffic crossing, number of stories, super markets, etc.
- Want to pull in elements from Maps.Me
- Have concerns about exposing which areas are covered versus not (news desserts)
- How do you connect with people in different areas: news letters vs. push alerts. Notifying of incidents in areas of interest/proximity. Frustrations with push alert communication.
- Seek to map geographic infomration in the website: a visualization of incidents occurring on a map. All content georeferenced in and around Lisbon.
- Have three jouranlists in the field and data from the municipality
- References:
 - functionality of In Your Area to apply to communication of covid cases per area
 - Oriental: orienting people in the middle east
- Lisbon data
 - Access of juntas de freguesias. All data comes from City hall. Covid data come from health minister
 - Usability: many people don't know the covid situation in their neighborhood. Are there 2 or 200 cases? TOOL DEVELOPED INFO APPLICATION STUDY. COVID specific: map cases AND news AND nursing homes AND layered data.
 - Lisboa data: Programa renda acessível