



A Comparative Study of Typing and Speech For Map Metadata Creation

Pei-Chun Lai and Auriol Degbelo (corresponding author)

plai@uni-muenster.de, auriol.degbelo@uni-muenster.de

Institute for Geoinformatics, University of Münster, Münster, Germany

Abstract. Metadata is key to effective knowledge organization, and designing user interfaces that maximize user performance and user experience during metadata creation would benefit several areas of GIScience. Yet, empirically-derived guidelines for user interfaces supporting GI-metadata creation are still scarce. As a step towards mitigating that gap, this work has implemented and evaluated a prototype that produces semantically-rich metadata for web maps via one of two input modalities: typing or speech. A controlled experiment (N=12) to investigate the merits of both modalities has revealed that (i) typing and speech were comparable as far as input duration time is concerned; and (ii) they received opposed ratings concerning their pragmatic and hedonic qualities. Combining both might thus be beneficial for GI-metadata creation user interfaces. The findings are useful to ongoing work on semantic enablement for spatial data infrastructure and note-taking during visual analytics.

Keywords. map metadata creation, web map search, Schema.org, geospatial semantics, visual analytics

1 Introduction

Metadata is vital to progress in several areas of GIScience. They have been identified as a key resource for research on Geospatial Semantics (Hu, 2018), Future Spatial Data Infrastructures (Diaz et al., 2012), Reproducible Research in Geoinformatics (Kray et al., 2019), the Digital Earth Vision (Janowicz and Hitzler, 2012), the Sensor Web (Bröring et al., 2011), and the discovery of Open Geospatial Data (Lafia et al., 2018; Kuo and Chou, 2019). As a result, understanding user interface factors that make metadata contribution efficient, effective and enjoyable to users has the poten-

tial to benefit several areas of research in GIScience. There has been some work that explored automatic approaches to metadata generation (e.g. Olfat et al. (2012); Trilles et al. (2017)). Nonetheless, metadata cannot always be generated a posteriori for geographic resources. For example, there is no way of adding an alternative name for a place in a geospatial application, if this alternative name has not been recorded by a user in the first place. Thus, manual metadata generation is and will remain an important component of metadata generation workflows.

Why metadata for web maps. The importance of semantic descriptions of geospatial resources has been acknowledged in previous work. For instance, Janowicz et al. (2010) indicated that in order to improve discovery in spatial data infrastructures, “semantic descriptions are needed for all types of geospatial data to ensure their correct interpretation”. If the importance of discoverable datasets, sensors, and web services has already been acknowledged in previous work in GIScience, the discoverability of web maps (i.e. static and interactive online maps) as geospatial resources in spatial data infrastructures still deserves more attention. As far as (web) maps are concerned, their uniqueness as geospatial resources can be highlighted from two perspectives: map as a *tool*, and map as a *representation* of geographic space.

First, as a *tool*, they are useful to explore the spatial dimensions and interrelationships between phenomena and activities located in space (see Kent and Klosterman (2000)). Thus, they are helpful to create and communicate (visual) stories about geographic phenomena. This storytelling feature is neither present in (raw) datasets nor web services on the spot. Second, as a *form of knowledge representation*, they index information by location in a plane as opposed to using sentences as primary units to organize knowledge (see Larkin and Simon (1987)). Thus, they enable the re-

trieval of insight hidden in datasets in a more efficient and effective way. Put simply, datasets present ‘raw insights’, maps present ‘refined insights’. Whether approached from the perspective of a tool that enables story construction, or an artifact that stores geographic knowledge in an effective way, maps are sufficiently distinct from raw datasets and web services to deserve efforts aimed to improve their discoverability. In this regard, Hu et al. (2015a,b) presented approaches to improve the discoverability of maps in the context of ArcGIS online.

Opportunities of speech. Speech has been used to control the device interface and fulfill simple tasks in many areas. Smart assistants (e.g. Google Home, Amazon Alexa, and Apple Siri) are notable examples. Accessibility (mentioned in a recent review by Clark et al. (2019)) and broadcasting are additional areas. For instance, speech recognition was used to generate metadata information for a live TV program with a success rate of 82 percent in 2005 (Sano et al., 2005). More related to geospatial data, Lafia et al. (2019) proposed a vision for geospatially-enabled voice assistants to support the discovery and reuse of open government data. Degbelo and Somaskantharajan (2020) argued that speech-based interaction presents some opportunities to improve interaction during digital forms-filling on mobile maps. The same argument can be made for filling out digital forms in general.

Contributions. Given positive results observed in other contexts and the increasing adoption of speech in other application areas, it is timely to investigate the extent to which speech as an interaction modality can benefit metadata generation workflows. This article presents an exploratory study about user interfaces for map-metadata creation. The research question examined is: What is the impact of speech-based interaction on user performance and user experience during map metadata creation? ‘Map metadata’ here refers to both aspects of *context* (e.g. spatial and temporal coverage of the map) and aspects of *content* (e.g. insights gleaned by users during the interaction with the map). Since typing is the primary interaction modality for metadata contribution currently, it is used as a baseline for a comparative study during the work. The key contributions of this article are (i) a prototype that generates semantically-rich metadata for web maps and (ii) insights from an empirical evaluation of this prototype for web map annotation tasks. The empirical evaluation offers a baseline against which future studies on user interfaces for geospatial resource annotations can be compared.

2 Related work

Previous work has developed vocabularies to annotate and techniques to rank digital maps, but user in-

terface factors of semantically-rich metadata contribution have been rarely investigated. In addition, work on visual analytics has long identified annotations as valuable in the sensemaking process, but is yet to provide means to document these annotations in a formal knowledge representation language for sharing and computational reuse. Since this work provides a prototype that addresses the two gaps, work on semantic description of digital maps and the annotation of visualizations is briefly reviewed here.

2.1 Semantic description of digital maps

With the increasing availability of digital maps, work has been undertaken to facilitate their search. Scheider et al. (2014a) argued that linked data can enable the description of both contextual aspects and content aspects of resources, and presented examples of using named graphs in the Resource Description Framework (RDF) to encode both context and content information of digital maps. In addition, several vocabularies have been proposed and used to describe maps in the literature, focusing on distinct properties of maps. For instance, Roula et al. (2010) proposed CartOWL to describe map icons; Gao et al. (2016) formalized concepts related to a map legend in a map legend ontology; and Carral et al. (2013) proposed a design pattern to formally describe cartographic map scaling. Degbelo (2017) used the Schema.org vocabulary to describe the spatial and temporal coverage of online web maps; and Degbelo (2020) proposed the GeoInsight design pattern to formally encode insights gleaned by users during the interaction with online geovisualizations. Besides, there is a body of work on standards for metadata of geographic information (see Brodeur et al. (2019)). OWL ontologies have been derived for ISO GI-metadata standards, for instance, the ISO/TC 211.

Tools to facilitate the annotation of web maps have primarily focused on topographic maps. For example, Simon et al. (2011) proposed the YUMA Map Annotation Tool to facilitate collaborative annotations by scholars studying historical maps. Scheider et al. (2014b) proposed a georeferencing tool that enables spatio-temporal and semantic content descriptions of historical maps, and takes advantage of background knowledge published on the Web (e.g. DBpedia). Contrary to these works where the objects of interest were topographic maps, the prototype presented later in this article focuses on, and enables the annotation of thematic maps.

2.2 Annotation of visualizations

Annotations enable the organization and sharing of knowledge. An annotation, in this work, is defined after Vanhulst et al. (2018) as “an observation, made by exploring a visual representation of data, that is

recorded either as text or visual selection (or both)". Previous work on Human-Computer Interaction and Visual Analytics has documented the importance of annotations during the interaction with visualizations. For instance, Heer et al. (2009) designed and evaluated sense.us, a platform that supports asynchronous collaboration across a variety of visualization types. A key takeaway from their study was that users, when given the opportunity, produce useful annotations of visualizations. Along similar lines, Willett et al. (2011) presented CommentSpace, a visual analysis system that enables analysts to annotate visualizations using a small vocabulary of tags (question, hypothesis, to-do) and links (evidencefor, evidence-against). They reported that the addition of tags and links to a collaborative visual analysis tool can help analysts identify findings in evidence-gathering tasks. Walny et al. (2018) found freeform annotation during active reading of visualizations to improve accuracy when performing low-level visualization tasks. Finally, Mahyar et al. (2012) reported that note-taking is a recurring activity during the interaction with visualizations in a co-located collaborative setting. They distinguished between two types of notes (a.k.a. annotations): *findings* (recorded results, observations, and decisions or outcomes of the analysis process) and *cues* (anything noted by the user that is not directly extracted from the visual representation). They also pointed out that notes have a scope: personal (when it is taken for individual use) or group (when the writer intends to share it with the group).

While annotations have been recognized as valuable, a desirable but missing feature at the moment is the means to record and share the annotations produced by users during their interaction with visualizations on a web-scale¹. By enabling the recording of the annotations produced in a format that is supported by major search engines, the prototype built during the course of this work offers a step in that direction.

3 Prototype for web map annotation

A glimpse at existing metadata creation tools. To get an impression of features of existing metadata management tools, we reviewed tools suggested by the Federal Geographic Data Committee² and the Open Source Geospatial Foundation³. Of the 25 tools available, 10 were working at the moment of the review (March 2021). Three could not be started due to technical issues (e.g. GeoNetwork, Mapbender) or licens-

ing issues (e.g. EPA). Table 1 recaps features of the remaining seven that could be inspected in-depth. The seven include the Greek Inspire Metadata Editor, that is focused on providing compliant metadata with the ISO19139 standard and INSPIRE directive. We document the type of the application (standalone vs web-based), the operating system (OS) on which they are available, the interaction modality for metadata creation supported, the number of fields offered by the interface for metadata contribution, the type of resources supported (e.g. data, web services, sensors), and their licence. The number of fields was counted manually after starting the applications ('NA' implies that we were not able to collect the information during the review). As the table shows datasets are the main resource for which tools for GI-metadata contribution are available, and typing is still the only modality available in several tools. Arguably, the results are not surprising, but they provided empirical evidence to confirm our hunches at the beginning of the work.

Tool	Type, OS	Modality	Nr. of Fields	Resource	Licence
ArcGIS Pro	Standalone, Windows	Typing	7	Data	Proprietary
CatMDEdit	Standalone, Windows	Typing	12	Data	Open-source
EPA	Standalone, Windows	Typing	NA	Data	Open-source
GeoNetwork	Web, Cross-platform	Typing	NA	Data	Open-source
Greek Inspire Metadata Editor	Standalone, Windows	Typing	40	Data	Open-source
Mapbender	Web, Cross-platform	Typing	NA	Data	Open-source
MEtadata Editor	Standalone, Windows	Typing	40	Data	Open-source
Metadata Wizard	Standalone, Windows	Typing	87	Data	Open-source
tkme	Standalone, Windows Linux	Typing	47	Data	Open-source
QGIS	Standalone, Windows	Typing	27	Data	Open-source

Table 1. Features of exemplar GI-metadata applications.

UI features. Annotation can be provided at the general level or the specific level (see Janowicz et al. (2010)). The prototype supports both. Annotation at the general level takes place when users contribute elements of *context* (Figure 2), while annotation at a specific level takes place when users contribute elements of *content* (Figure 3). Also, annotations can be provided as a free-text, use a shared vocabulary, or recorded to uniquely identify named entities (see Hinze et al. (2019)). The annotations of contextual aspects and content aspects of the maps during the work used a *shared vocabulary*.

Elements of context supported include: the place name, the alternative place name (a.k.a. alias), the topic, descriptions, the start time, and the end time. The place name and the alias are names of locations depicted by the map, for instance, 'The United States' or 'USA'. The topic requires users to define a theme for the map

¹The Web Annotation Model (<https://www.w3.org/TR/annotation-model/>) is too generic for visualizations.

²<https://www.fgdc.gov/metadata/geospatial-metadata-tools>, accessed: March 11, 2021.

³https://wiki.osgeo.org/wiki/Metadata_software, accessed: March 11, 2021.

in the form of keywords that can be used for search, for instance, ‘wildfire’ or ‘drought’. The description contains any detail about the map that users wish to record. The start time and end time record the time-frame of the phenomena shown on the map.

As to content, users can select specific portions of a map using one of four options: a rectangle, a circle, a free-drawing pen, or a pin (Figure 3). They can also select one of seven *content statements*: cluster, outlier, correlation, trend, frequency, distribution, and observation (if their recording does not belong to the first six). The seven content statements were taken from the GeoInsight design pattern (Degbelo, 2020): correlation (relationships between data dimensions), frequency (how often items appear), trend (high-level changes across a data dimension), outlier (items that do not match a distribution), cluster (groups of similar items), distribution (extent and frequency of items), and observation (any statement that does not specifically highlight a pattern in the data).

Data format for the annotations. Several formats can be used to save users’ annotations in a machine-readable format. Schema.org was used in this work because it is supported by major search engines (it was founded by Google, Microsoft, Yahoo, and Yandex, and is developed by an open community process). The mappings of terms from the user interface to Schema.org terms were straightforward as Tables 2 and 3 show. Content statements created by users are stored as a *Schema.org:comment*. They have three parts: *Schema.org:termCode* to inform about the type of content statement (observation, cluster, outlier, and so on), *Schema.org:description* to record users’ description of their annotation, and *Schema.org:dateCreated* to record the creation date. The annotations are recorded in JSON-LD to facilitate their reuse by developers. Listings 1 and 2 in the Appendix (Section 7) provide some examples.

UI Term	Schema.org Term	Definition of the Schema.org Term
Place name	spatialCoverage	The place(s) which are the focus of the content.
Alternative place name	alternateName	An alias for the item.
Topic	keywords	Keywords or tags used to describe this content.
Description	description	A description of the item.
Start time/ End time	temporalCoverage	The period that the content applies to.

Table 2. Mapping between the UI Terms and Terms of Schema.org (Context Metadata Creation).

Technologies. The implementation was done using open-source technologies. Vue.js was used as a framework to implement the client-side, Node.js used for the server-side, and the speechRecognition API from Mozilla was used for the speech recognition. The application was deployed on Heroku. Screenshots of the application are shown in the Appendix (Section 7).

UI Term	Schema.org Term	Definition of the Schema.org Term
Pattern	termCode	A code that identifies a defined term within a collection of terms.
Description	description	A description of the item.
-	comment	A comment on an item - for example, a comment on a blog post. The comment’s content is expressed via the text property, and its topic via about, properties shared with all CreativeWorks.
-	dateCreated	The date on which the CreativeWork was created or the item was added to a DataFeed.

Table 3. Mapping between the UI Terms and Terms of Schema.org (Content Metadata Creation). There are no UI terms associated with Schema.org:comment and Schema.org:dateCreated, as they are generated in the background after the annotation has been recorded.

4 User Study

Variables. The experiment had two independent variables: the typing and the speech condition. There were four dependent variables: efficiency (input duration time), effectiveness (self-correction rates), tasks’ difficulty levels, and the overall user experience. With respect to the research question, efficiency and effectiveness are two aspects of performance.

The input duration time indicates the time participants spent on actually filling-in the fields. It does not include the time they spent interacting with the map during the study. Self-correction rates (a.k.a. slips) indicate the number of times users self-corrected themselves during the input of their contributions. Rates for the typing modality were counted using $N = onFocusEvents - 1$, where onFocusEvents is the number of times a html onFocus event has taken place on a given field. Rates for the speech modality were counted using $N = onMicrophoneEvents - 1$, where onMicrophoneEvents is the number of times they performed microphone recording events until they completed a form-filling task. Microphone recording events were implemented as a custom html event during the work. Difficulty ratings were collected using the Single Ease Question (SEQ, see Sauro and Dumas (2009); Sauro (2012)). The SEQ assesses how difficult users find a task on a 7-point Likert Scale (1=‘very difficult’, 7=‘very easy’). The user experience was measured using the short version of the user experience questionnaire (UEQ-S, see Schrepp et al. (2017)).

Procedure. The experiment took place online due to global social distancing restrictions at the time of the study (December 2020 - February 2021). Participants were introduced to the experiment through a few slides presented by the experimenter. If they agreed to proceed with the experiment, their consent was recorded. Since recording the informed consent through signatures was not possible, participants’ consent was col-

lected through a video recording. Afterwards, the experiment started.

The whole experiment is divided into two phases. In each of the phases, participants complete five sub-tasks (see Figure 1). All five are either related to context metadata creation or content metadata creation. Also, they use only one interaction modality (either typing or speech) during the completion of the five sub-tasks belonging to a phase. Since the measurement of the difficulty level intends to capture the difficulty of an annotation task, it follows immediately after the completion of a sub-task. The user experience is measured at the end of each phase. Finally, participants' qualitative feedback was collected in a short interview after the experiment. Task completion time (e.g. the time needed to fill in a field or generate a content statement about a map) was collected through logging of interaction data in the prototype.

A within-group design was used to mitigate learning effects. The ordering of the conditions was counterbalanced using a Greco-Latin square design. The ordering of tasks (context vs content) and interaction modality (typing vs speech) was assigned algorithmically via a scenario controller (see the supplementary material, Section 7 for more details). The instruction for context metadata creation read: "You will see 5 maps with the wildfire and drought theme. Use elements on the left to summarize each map". The instruction for content metadata creation was: "You will see 5 maps with the wildfire and drought theme. Add two annotations for each map. Answer spontaneously". The experiment was approved by the institutional ethics board.

Maps. 10 online maps were selected for the final study. Five interactive maps were used for context metadata creation tasks, and five static maps were used for content metadata creation tasks. All maps were lacking the metadata collected during the study (e.g. place name, topic, description, etc). Exposing participants to maps with unrelated themes could have had the detrimental effect that they would have invested cognitive resources switching between topics during the experiment. For this reason, all maps in the experiment matched a constant theme, i.e. either the topic of 'wildfire' or that of 'drought'. This had the advantage that the maps presented did not appear repetitive while their topic was at the same time transparent to users so that they could focus on the tasks. The online maps were included in the prototype through an iFrame. All maps used during the study are available in the supplementary material (see Section 7).

Pilot study and lessons learned. Three participants (2 Male, 1 Female) were recruited to pilot-test the experiment. All belonged to the age group (30-34) and none was a native English speaker. Also, all had a background in Geography and experience in using a GIS to create maps. Their feedback led to several modifications in the original experimental design. The first

change concerned the number of maps to include in the experiment. The original design planned 10 maps for contextual metadata creation, and 5 for the content metadata task. The pilot study showed that the experiment would have been relatively long and exhausting for participants. Thus, the number of maps in the contextual metadata creation phase was reduced to 5, leading to a study of about an hour on average. Second, the introductory instructions of the experiment were entirely redesigned. We were assuming that subjects will use the landing page of the prototype to self-familiarize themselves with the experiment, but this has not proven to be effective. Instead, using a few introductory slides to the experiment has proven more useful and was adopted in the final study. The slides were shown to all (potential) participants before collecting their consent. They had a short introduction about the experiment, the procedure, the tasks, participants' rights to quit the study at any time. Third, the volunteers in the pilot study spent much time on the first map, then gradually lost interest and rushed to finish the remaining sub-tasks. To mitigate this, we asked participants in the final experiment to make annotations spontaneously. Also, the number of annotations in the content metadata creation phase was set to two, to make the results comparable. At last, a few questions were removed and some were made more precise in their phrasing for the qualitative interviews. The data from these three participants were not included in the final analysis.

5 Results

480 annotations from 12 participants were collected during the user study. The participants produced 360 (= 6 annotations * 5 maps * 12 participants) in the map element condition, and 120 (= 2 annotations * 5 maps * 12 participants) in the map content condition. They made broad usage of the annotation tools for content metadata creation (free-drawing pen: 67 times, pin: 24 times, circle: 15 times, and rectangle: 14 times). They also produced a variety of content statements (cluster: 32, outlier: 22, observation: 21, distribution: 20, trend: 17, correlation: 6, frequency: 2).

The Web Speech API provides a confidence score between 0 and 1 for speech-to-text translations. The score indicates how confident the recognition system is that the recognition is correct⁴. To provide a basis for comparison in future studies, we computed the average confidence score per participant in our study. The confidence scores of the system ranged from 0.01 to 0.9 (Mean: 0.63, sd: 0.32). The study results were analyzed using bootstrap confidence intervals (N=2,000 resam-

⁴See <https://wicg.github.io/speech-api/#dom-speechrecognitionalternative-confidence>, accessed: March 11, 2021.

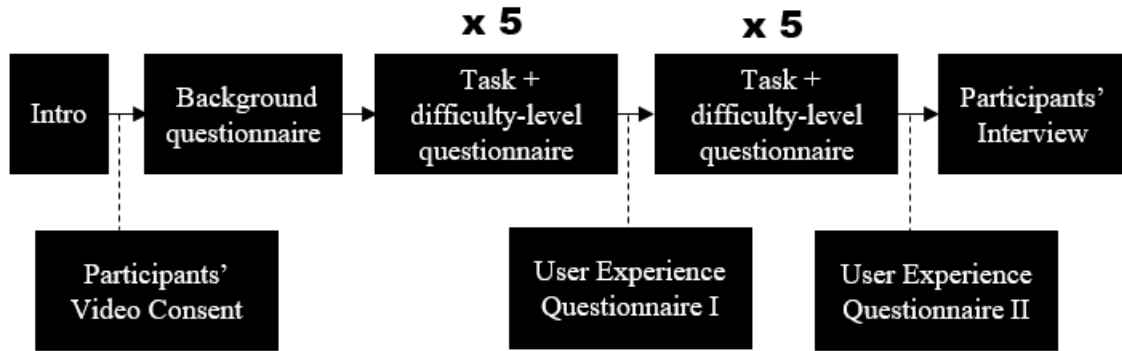


Figure 1. Experimental design.

ples in bootES, see Kirby and Gerlanc (2013)) and linear modelling in R.

Participants' Background. 7 Males and 5 Females, aged between 18 to 34 took part in the final study. Only two were native English speakers. Half of them reported having experience in creating metadata. Two reported having never used speech recognition technologies, and all (10/12) who reported having used speech before the study did so on mobile devices. Only 1/10 has used speech on a desktop computer before the study. All participants were familiar with web maps: 7/12 reported using them every day, 4/12 mentioned using them 1-3 times per week, and the remaining participant reported using them 1-3 times per month.

Efficiency. The average input rates in characters per second were 1.71 char/s (sd: 0.46) for typing and 1.92 char/s (sd: 0.64) for speech during context metadata creation. The input rates during content metadata creation were 1.24 char/s (sd: 0.18) for typing and 2.82 char/s (sd: 0.87) for speech respectively. Considering the task of 'map metadata creation' as the unit for the analysis, participants took on average 94 seconds (sd: 23s) to produce contextual metadata in the typing condition and 104 seconds (sd: 20s) in the speech condition. Typing was thus slightly faster during contextual metadata creation. Nonetheless, the difference between the two modalities was not statistically significant. As to content metadata creation, the users took 92s (sd: 22s) on average in the typing condition, and 74s (sd: 18s) in the speech condition. Here, using speech was slightly faster. The difference between the two modalities was also not statistically significant. We conclude that typing and speech are comparable when it comes to efficiency. Table 4 presents the results. We also learn from the table that annotating a map with six contextual metadata elements takes about 100 seconds, and annotating a map with two content-based statements takes about 80 seconds. These values could be used while planning future user studies on map metadata creation (e.g. to estimate the upper bound of the number of maps to include in the study).

Map	Modality	Context	Content	Average
Map1	Type	113.50	95.33	104.42
	Speech	124.33	97.83	111.08
Map2	Type	124.00	115.50	119.75
	Speech	125.50	83.00	104.25
Map3	Type	80.50	109.83	95.17
	Speech	84.67	54.83	69.75
Map4	Type	83.50	69.83	76.67
	Speech	84.83	58.83	71.83
Map5	Type	69.17	67.50	68.33
	Speech	99.83	75.00	87.42
Average	Type	94.134	91.598	92.868
	Speech	103.832	73.898	88.866
Overall		98.98	82.75	90.87

Table 4. Mean Input Duration Time per Map (Seconds).

Effectiveness. Table 5 and 6 shows the mean self-correction rates during contextual metadata and content metadata creation respectively. Typing resulted in lower self-correction rates during context metadata creation (Mean: 3.43, sd: 0.84) than speech (Mean: 5.07, sd: 3.05). The difference between the two conditions was not statistically significant. Also, typing resulted in lower self-correction rates during content metadata creation (Mean: 0.97, sd: 0.55) than speech (Mean: 3.60, sd: 2.04). Here, the difference between the two conditions was statistically significant. Thus, the overall tendency here is that typing resulted in lower self-correction than speech.

Map	Mode	Place	Alias	Topic	Descr.	Start Time	End Time	Sum
Map1	Type	0.33	0	0.5	1.83	0.83	0.17	3.66
	Speech	0.33	0	1.17	2.17	1.33	0	5
Map2	Type	0	0.33	0.33	1.67	1	1.33	4.66
	Speech	0.67	1.5	2.83	3.83	0.67	0.83	10.33
Map3	Type	0	1	0.33	0.5	0.17	0.5	2.5
	Speech	1.83	0.17	0	0.17	0.17	0.5	2.84
Map4	Type	0.17	0	0.5	1.67	0.83	0.33	3.5
	Speech	0.17	0	1.17	1.83	0.33	0.33	3.83
Map5	Type	0.5	0.33	0.33	0.67	0.17	0.83	2.83
	Speech	0.5	0.67	0.5	0.67	0.33	0.67	3.34
Avg.	Type	0.20	0.33	0.40	1.27	0.60	0.63	3.43
	Speech	0.70	0.47	1.13	1.73	0.57	0.47	5.07
Overall		0.45	0.4	0.77	1.5	0.58	0.55	4.25

Table 5. Self-Correction Rates during Contextual Metadata Creation.

Map	Mode	Pattern1	Pattern2	Desc1	Desc2	Sum
Map1	Type	0.67	0.5	0.5	0.17	1.84
	Speech	2.17	1.67	0.17	0.83	4.84
Map2	Type	0.5	0.17	0.33	0.17	1.17
	Speech	1.5	1.17	0.33	0.5	3.5
Map3	Type	0.17	0.33	0.17	0	0.67
	Speech	0.17	0.67	0.33	0.17	1.34
Map4	Type	0.17	0.33	0	0.17	0.67
	Speech	0.83	1	0.17	0	2
Map5	Type	0.33	0	0.17	0	0.5
	Speech	2.5	2	1.17	0.67	6.34
Avg.	Type	0.37	0.27	0.23	0.10	0.97
	Speech	1.43	1.3	0.43	0.43	3.6
	Overall	0.90	0.78	0.33	0.27	2.29

Table 6. Self-Correction Rates during Content Metadata Creation.

Task Difficulty Ratings. Table 7 presents the difficulty ratings of the users during the experiment. Content-metadata creation tasks were rated slightly more challenging than context-based tasks. This is not surprising, as they demand more cognitive resources (e.g. recording of insights that the map actually shows). Though the differences between the modalities were not significant, the tendency here was that the typing condition was rated as slightly easier, especially for the content metadata creation tasks. Based on our observation during the experiment, this could be attributed to the fact that several participants were not native speakers, and as the task of producing metadata became a bit more complex in the content scenario, they needed a few more repetitions (see also Table 6 for the self-correction rates during content metadata creation).

Map	Modality	Context	Content	Average
Map1	Type	5.67	5.33	5.5
	Speech	4.83	4.5	4.67
Map2	Type	5	4.83	4.92
	Speech	4.83	4.83	4.83
Map3	Type	4.33	5.33	4.83
	Speech	5.17	4.83	5
Map4	Type	5.5	5.17	5.33
	Speech	5.67	5	5.33
Map5	Type	5.5	5.33	5.42
	Speech	5	4.67	4.83
Average	Type	5.2	5.199	5.2
	Speech	5.1	4.77	4.93
	Overall	5.15	4.98	5.07

Table 7. Mean Task-Difficulty Ratings (a lower number means more difficult).

User Experience. Table 8 shows the user experience ratings. Following Hassenzahl (2004), there are two dimensions of user experience: pragmatic qualities and hedonic qualities. A product can be perceived as *pragmatic* because it provides effective and efficient ways to achieve behavioral goals (usability). It can be perceived as *hedonic* because it provides stimulation by its challenging and novel character (stimulation function), or identification by communicating important personal values to others (social function). There was a clear tendency here. Annotation tasks completed using the typing modality were rated (pragmatic dimension) as ‘supportive’, ‘easy’, ‘efficient’, and ‘clear’, but neutral

on the hedonic dimension. In contrast, annotations using the speech modality were perceived (hedonic dimension) as ‘exciting’, ‘interesting’, ‘inventive’, and ‘leading-edge’. They were rated as neutral on the pragmatic dimension. Since the study was conducted in a non-collaborative setting, the ratings on the hedonic dimension can be attributed to the stimulating function of the speech modality, rather than its social function. Put differently, typing was perceived as more usable, but less stimulating; speech was more stimulating but less usable.

Item	Speech	Typing	Negative	Positive	Scale
1	-0.5	1.5	Obstructive	Supportive	PQ
2	0.3	1.6	Complicated	Easy	PQ
3	-0.3	1.3	Inefficient	Efficient	PQ
4	0.1	1.8	Confusing	Clear	PQ
5	1.4	0.3	Boring	Exciting	HQ
6	1.8	0.5	Not interesting	Interesting	HQ
7	1.1	-0.7	Conventional	Inventive	HQ
8	1.3	-0.4	Usual	Leading-edge	HQ

Table 8. User Experience Ratings.

PQ-Pragmatic Quality, HQ-Hedonic Quality. A value higher than 0.8 implies positive, less than 0.8 is negative, between 0.8 and -0.8 is neutral.

Users’ Qualitative Feedback. We asked participants for their preferences with respect to the two modalities and the two tasks, and the reasons for their preferences (‘if you were to create metadata for your project which interaction modality would you choose?’). Most reporting typing as their first choice: (11/12, for contextual metadata creation), and (10/12 for content metadata creation). The users liked typing because it is accurate, gives more time to organize ideas before entering them into the system, and easy to modify. On the other hand, they pointed out that speech is unfavourable because of its inaccuracy and the fact that it leaves no time for organizing ideas. Also, some expressed that they would have been more comfortable with speech-to-text translation in their native languages than in English. For example, when a participant said: ‘wildfire’, the result became ‘why do fire’. Without further evidence, it is unclear whether these ‘misunderstandings’ are due to inherent limitations of the speech recognition system, or the level of English of the participants, or both. This is an issue to investigate in future work.

Impact of Participant’s Background. The impact of participants’ background was assessed for all results. We checked if gender, (not) being a native speaker, and previous experience in creating metadata have a significant influence on the performance (Typing), performance (Speech), difficulty ratings (Typing), difficulty ratings (Speech), and self-correction rates. This was not the case in our study. Thus, the key takeaway here is that both interaction modalities were relatively robust to participants’ prior experience during the study.

6 Discussion

As mentioned in Section 2, work on semantic descriptions of maps has so far missed considerations of user interface factors, while work on annotation of visualizations has not gone as far as discussing the recording of these annotations in a formal knowledge representation language. By providing a prototype that helps investigate user interface factors of metadata contribution, and uses the JSON-LD encoding of Schema.org as a data model for the interchange of visualization annotations (Listings 1 and 2), our work contributes to addressing these two gaps. The experiment is an exploratory study about user interfaces for map-metadata creation. Below, we revisit the research question (Section 6.1), discuss the scope and implications of our results (Sections 6.2 and 6.3), and comment on limitations and future work (Sections 6.4 and 6.5).

6.1 Key takeaways

What is the impact of speech-based interaction on user performance and user experience in the process of map metadata creation? The pieces of evidence collected during the experiment converge. As to performance, both modalities are comparable regarding input duration time. When the fields to fill-in were very simple (context metadata), typing was slightly faster. As content became a bit more complex, speech was slightly faster (content metadata). Typing has also proven more effective because it resulted in fewer edits and was easier to edit after the entries. Data from the qualitative interviews have confirmed this as well. In a study comparing typing and speech on mobile phones, Ruan et al. (2018) reported that speech (in English) had an input rate nearly three times greater than typing, made fewer errors during entry, but left slightly more errors after entry was complete. We did not observe exactly the same magnitude of differences in input rates, but observed that speech left more errors after entry was completed (Tables 5 and 6).

As to user experience, typing was perceived as more usable, but less stimulating; speech was more stimulating but less usable. We attribute users' perception of usability to the fact that they were more familiar with the modality, and that it presents more facilities for editing. The fact that speech was rated as more stimulating may be due to two reasons: either it truly adds something to the user experience, or participants simply found it interesting because they are not used to seeing it (i.e. a 'novelty' effect). Contrasting our results with those from previous work may be useful here. In a study exploring geodata contribution on mobile devices via speech, Degbelo and Somaskantharajan (2020) reported that the speech modality was rated as usable, but *not* stimulating. This speaks against the argument that speech would be systematically rated by

users as stimulating because it is used in a new application scenario. There are thus reasons to hypothesize that the speech modality truly adds something to the user experience during metadata contribution. Further studies are needed to confirm this.

6.2 Scope of the findings

The experimental setting involved thematic maps (as opposed to topographic maps), young users, a small number of fields (i.e. six), and a desktop computer. The scope of the findings is thus limited to these settings. In addition, Schmidt et al. (2021) pointed out that annotations can support different steps in visual analytics: data preprocessing (i.e. generate a structured and consolidated data set based on a raw dataset), data cleansing (add missing values, delete or changed erroneous data values), and data exploration (i.e. generate findings and insights by exploring the cleansed data). Since our study attempted to produce contextual metadata for maps that had none, and help users record insights, the findings apply to the data cleansing and data exploration stages.

6.3 Implications

The results above suggest that combining typing and speech might be useful to improve the overall user experience of user interfaces for GI-metadata creation. How this combination should be designed and implemented to yield optimal results remains to be seen, but given that existing tools are still uni-modal (see Table 1), the results suggest that looking further into means to introduce speech in metadata workflows could be a fruitful research avenue. The user experience ratings in Table 8 can serve as a baseline to evaluate gains in user experience from newer interfaces as they become available. Furthermore, as mentioned in Section 2, the our contributions could be useful to research semantic description of maps, and annotation of visualizations. Concerning semantic description of maps, the results suggest that designing interfaces supporting both modalities could benefit crowdsourcing efforts aimed at producing semantically-rich metadata. Regarding, the annotation of visualizations, the prototype is a proof-for-concept that annotations recorded during the interaction with visualizations can be made interoperable. Others may reuse our mappings (Tables 2 and 3) and expand on them to fit their own scenarios.

6.4 Limitations

Given that typing is a ubiquitous modality, it might be said that it had an 'unfair' advantage during the study, since users will always be more familiar with typing than any other modality. The familiarity bias is acknowledged. Nonetheless, its impact on the con-

clusions is minor because the typing familiarity bias implies that if participants received more training on speech, they would have got even better results in the speech condition during the experiment. In addition, since our aim in this study was primarily to learn about user interface factors, we did not analyze the quality of the annotations produced in-depth. Finally, our tool left several elements of context and content out, to keep the study manageable. Elements of context not included, but valuable include keywords, abstract, purpose, and usage (see Ahonen-Rainio (2006)). Elements related to content, but not included are those that Mahyar et al. (2012) called cues, e.g. reminder, evidenceFor, evidenceAgainst, hypothesis, or questions that came up during the interaction with the map.

6.5 Future work

This work is an exploratory study about user interfaces for GI-metadata creation. Moving to the point where empirically-derived guidelines for these user interfaces can be confidently formulated necessitates much additional work. We sketch here three possible directions (the ideas are not mutually exclusive).

Deployment in the wild. The current prototype could be extended for a large-scale study collecting semantic annotations for web maps lacking metadata through crowdsourcing. The newer version of the prototype could then mix both modalities, and learn empirically which one users use most, when and why. Furthermore, a missing feature of the recording of the annotations in the current prototype is a documentation of the level of spatial detail (i.e. zoom level) at which the data patterns were identified. This is an aspect the GIScience literature is well-aware of: patterns are often detected at a spatial granularity. Extending the data model to incorporate this feature would be relatively easy, and more importantly, it could enable the design of intelligent zooming: users will be directed first to ‘where the action is’ (i.e. start data exploration at zoom levels where others have highlighted interesting patterns first). This feature could contribute to advance the vision of better user activity support in intelligent geovisualizations (Degbelo and Kray, 2018).

Summary of graphical annotations. 120 annotations were collected in this relatively simple study, and thousands of these can be produced in a large-scale study. This has raised the question of how graphical annotations could be summarized. For example, several circles and rectangles were drawn on the map that overlap and could indicate the same data pattern. Getting rid of these redundancies to show new users a neat interface is an open question. A key issue here is that the annotations are produced with different screen coordinate systems. Developing both computational and visual means to summarize these annotations thus presents an opportunity for visual analytics research.

Cross-validity. Finally, an interesting direction for future work would be to test the extent to which the findings are valid in other contexts, e.g. metadata contribution on mobile devices, data contribution on topographic maps, different age groups, increased number of fields, and different scenarios (e.g. annotations of resources in the context of the Sensor Web, and computational reproducibility).

7 Conclusion

Metadata generation is and will remain an important task, and positive results of using speech-based interaction to generate metadata in other contexts (e.g. broadcasting, conversational assistants) suggest that speech might present opportunities for metadata generation in geospatial applications. Nonetheless, the impact of speech-based interaction has not been investigated as such in the literature. This work has implemented and evaluated a prototype to generate semantically-rich (i.e. Schema.org-compliant) annotations for web maps. The lessons learned in a controlled experiment are twofold. First, typing and speech were comparable as far as input duration time is concerned. Second, they exhibited distinct properties from the user experience point of view. The participants rated typing as more pragmatic (supportive, easy, efficient, clear) while rating speech as more hedonic (exciting, interesting, inventive, leading-edge). The software and the lessons learned during the experiment can serve as building blocks for the design of intuitive (geospatial) interfaces for metadata contribution.

Acknowledgements. The work has been funded by the Taiwanese Ministry of Education (MOE) and conducted as part of the Erasmus Mundus Master in Geospatial Technologies (Erasmus+/Erasmus Mundus Mundus program, grant agreement FPA-2012-0191, <http://mastergeotech.info/>). We thank Marco Painho and Sven Casteleyn for their feedback during the work.

Software and Data Availability. The code of the prototype is open-source and available on GitHub (<http://doi.org/10.5281/zenodo.4739992>). The data collected during the study as well as the code used for the analysis are available at <https://doi.org/10.6084/m9.figshare.14207735>.

Appendix

Screenshots of the application are presented below. Further material presenting the architecture of the prototype, all questionnaires, and all maps used during the study are available at <https://doi.org/10.6084/m9.figshare.14207735>.

Listing 1. Example of Contextual Metadata produced by a participant during the study (and its encoding by the prototype in JSON-LD)

```

1  \\User Input:
2  {"Place": "United States",
3   "alternateName": "USA",
4   "Topic": "drought",
5   "Description": "this map may show the factors that lead to drought events
6   across the USA"
7   "Time": "2020-11-17/2020-12-17"}
8  \\Result
9  {
10   "@type": "http://schema.org/Map",
11   "http://schema.org/alternateName": "USA",
12   "http://schema.org/description": "this map may show the factors that
13   lead to drought events across the USA.",
14   "http://schema.org/keywords": "drought",
15   "http://schema.org/spatialCoverage": "United States",
16   "http://schema.org/temporalCoverage": "2020-11-17/2020-12-17"
17 }

```

Listing 2. Example of Content Metadata produced by a participant during the study (and its encoding by the prototype in JSON-LD)

```

1  \\User Input:
2  [{"Pattern": "Cluster",
3   "description": "there is a bigger hotspots close to this cluster however
4   they are separated and there is a distinct difference in the
5   values before it becomes a cluster"},
6   {"Pattern": "Distribution",
7   "description": "there seems to be a corridor that is allowing for this
8   that explains the rainfall deficiency however without more
9   information about the elevation or other variables not sure what
10   is the cause of this rainfall deficiency corridor"}]
11 \\Result
12 {
13   "@type": "http://schema.org/Map",
14   "http://schema.org/comment": {
15     "http://schema.org/dateCreated": {
16       "@type": "http://schema.org/Date",
17       "@value": "2020-12-18T10:59:18",
18       "http://schema.org/description": "there is a bigger hotspots close to
19       this cluster however they are separated and there is a distinct
20       difference in the values before it becomes a cluster",
21       "http://schema.org/termCode": "Cluster"},
22     "http://schema.org/dateCreated": {
23       "@type": "http://schema.org/Date",
24       "@value": "2020-12-18T11:0:47",
25       "http://schema.org/description": "there seems to be a corridor that is
26       allowing for this that explains the rainfall deficiency however
27       without more information about the elevation or other variables
28       not sure what is the cause of this rainfall deficiency corridor",
29       "http://schema.org/termCode": "Distribution"}
30   }
31 }

```

References

Ahonen-Rainio, P.: Metadata for geographic information, *Journal of Map & Geography Libraries*, 2, 37–66, https://doi.org/10.1300/J230v02n01_03, 2006.

Brodeur, Coetzee, Danko, Garcia, and Hjelmager: Geographic information metadata—An outlook from the international standardization perspective, *ISPRS International Journal of Geo-Information*, 8, 280, <https://doi.org/10.3390/ijgi8060280>, 2019.

Bröring, A., Echterhoff, J., Jirka, S., Simonis, I., Everding, T., Stasch, C., Liang, S., and Lemmens, R.: New Generation Sensor Web Enablement, *Sensors*, 11, 2652–2699, <https://doi.org/10.3390/s110302652>, 2011.

Carral, D., Scheider, S., Janowicz, K., Vardeman, C., Krisnadhi, A. A., and Hitzler, P.: An ontology design pattern for cartographic map scaling, in: *The Semantic Web: Semantics and Big Data*, 10th International Conference (ESWC 2013), edited by Cimiano, P., Corcho, O., Presutti, V., Hollink, L., and Rudolph, S., pp. 76–93, Springer Berlin Heidelberg, Montpellier, France, https://doi.org/10.1007/978-3-642-38288-8_6, 2013.

Clark, L., Doyle, P., Garaialde, D., Gilmartin, E., Schlögl, S., Edlund, J., Aylett, M., Cabral, J., Munteanu, C., Edwards, J., and R Cowan, B.: The state of speech in HCI: trends, themes and challenges, *Interacting with Computers*, <https://doi.org/10.1093/iwc/iwz016>, 2019.

Degbelo, A.: Linked data and visualization: two sides of the transparency coin, in: *Proceedings of the 3rd ACM SIGSPATIAL Workshop on Smart Cities and Urban Analytics - UrbanGIS'17*, edited by Vo, H. T. and Howe, B., pp. 1–8, ACM Press, Los Angeles, California, USA, <https://doi.org/10.1145/3152178.3152191>, 2017.

Degbelo, A.: An ontology design pattern for geovisualization content description, in: *Advances in Pattern-based Ontology Engineering*, edited by Blomqvist, E., Hahmann, T., Hammar, K., Hitzler, P., Hoekstra, R., Mutharaju, R., Povedaf, M., Shimizuc, C., Skjaeveland, M., Solanki, M., Svátek, V., and Zhou, L., IOS Press, 2020.

Degbelo, A. and Kray, C.: Intelligent geovisualizations for open government data (vision paper), in: *26th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, edited by Banaei-Kashani, F., Hoel, E. G., Güting, R. H., Tamassia, R., and Xiong, L., pp. 77–80, ACM Press, Seattle, Washington, USA, <https://doi.org/10.1145/3274895.3274940>, 2018.

Degbelo, A. and Somaskantharajan, S.: Speech-based interaction for map editing on mobile devices: a scenario-based study, in: *Mensch und Computer 2020*, edited by Alt, F., Schneegass, S., and Hornecker, E., pp. 343–347, ACM, Magdeburg, Germany, <https://doi.org/10.1145/3404983.3409996>, 2020.

Diaz, L., Remke, A., Kauppinen, T., Degbelo, A., Foerster, T., Stasch, C., Rieke, M., Schaeffer, B., Baranski, B., Bröring, A., and Wytzisk, A.: Future SDI - Impulses from Geoinformatics research and IT trends, *International Journal of Spatial Data Infrastructures Research*, 7, 378–410, <https://doi.org/10.2902/1725-0463.2012.07.art18>, 2012.

Gao, S., Janowicz, K., and Zhang, D.: Designing a map legend ontology for searching map content, in: *Advances in Ontology Design and Patterns* [revised and extended versions of the papers presented at the 7th edition of the Workshop on Ontology and Semantic Web Patterns], edited by Hammar, K., Hitzler, P., Krisnadhi, A., Lawrynowicz, A., Nuzzolese, A. G., and Solanki, M., vol. 32, pp. 119–130, IOS Press, <https://doi.org/10.3233/978-1-61499-826-6-119>, 2016.

Hassenzahl, M.: The interplay of beauty, goodness, and usability in interactive products, *Human-Computer Interaction*, 19, 319–349, https://doi.org/10.1207/s15327051hci1904_2, 2004.

Heer, J., Viégas, F. B., and Wattenberg, M.: Voyagers and voyeurs: supporting asynchronous collaborative visualization, *Communications of the ACM*, 52, 87–97, <https://doi.org/10.1145/1435417.1435439>, 2009.

Hinze, A., Heese, R., Schlegel, A., and Paschke, A.: Manual semantic annotations: User evaluation of interface and interaction designs, *Journal of Web Semantics*, 58, 100516, <https://doi.org/10.1016/j.websem.2019.100516>, 2019.

Hu, Y.: Geospatial semantics, in: *Comprehensive Geographic Information Systems*, pp. 80–94, Elsevier, <https://doi.org/10.1016/B978-0-12-409548-9.09597-X>, 2018.

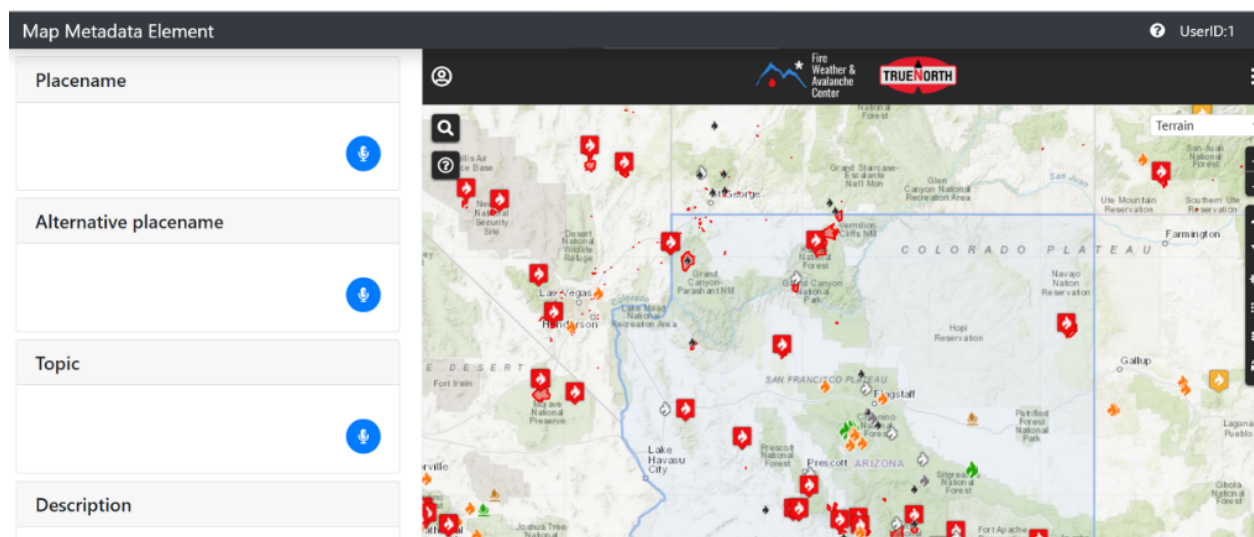


Figure 2. Screenshot of the view supporting map metadata element creation (speech-based interaction).

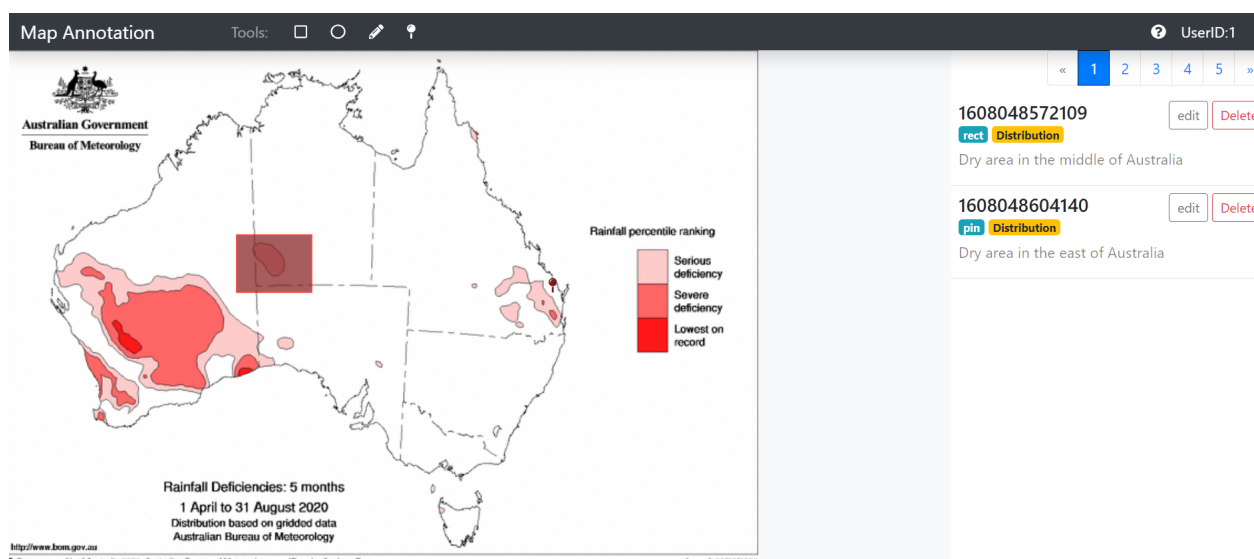


Figure 3. Screenshot of the view supporting the recording of map content observations. A variety of annotation tools (top) is available to users; the annotations and their types are shown to users who can edit them (right).

Hu, Y., Janowicz, K., Prasad, S., and Gao, S.: Metadata topic harmonization and semantic search for linked-data-driven geoportals: A case study using ArcGIS online, *Transactions in GIS*, 19, 398–416, <https://doi.org/10.1111/tgis.12151>, 2015a.

Hu, Y., Janowicz, K., Prasad, S., and Gao, S.: Enabling semantic search and knowledge discovery for ArcGIS online: A linked-data-driven approach, in: *AGILE 2015 - Geographic Information Science as an Enabler of Smarter Cities and Communities*, edited by Bacao, F., Santos, M. Y., and Painho, M., pp. 107–124, Springer, Cham, Lisbon, Portugal, https://doi.org/10.1007/978-3-319-16787-9_7, 2015b.

Janowicz, K. and Hitzler, P.: The Digital Earth as knowledge engine, *Semantic Web*, 3, 213–221, <https://doi.org/10.3233/SW-2012-0070>, 2012.

Janowicz, K., Schade, S., Bröring, A., Keßler, C., Maué, P., and Stasch, C.: Semantic enablement for spatial data infrastructures, *Transactions in GIS*, 14, 111–129, <https://doi.org/10.1111/j.1467-9671.2010.01186.x>, 2010.

Kent, R. B. and Klosterman, R. E.: GIS and mapping, *Journal of the American Planning Association*, 66, 189–198, <https://doi.org/10.1080/01944360008976098>, 2000.

Kirby, K. N. and Gerlanc, D.: BootES: An R package for bootstrap confidence intervals on effect sizes, *Behavior Research Methods*, 45, 905–927, <https://doi.org/10.3758/s13428-013-0330-5>, 2013.

Kray, C., Pebesma, E., Konkol, M., and Nüst, D.: Reproducible research in Geoinformatics: concepts, challenges and benefits (vision paper), in: *14th International Conference on Spatial Information Theory (COSIT 2019)*, edited by Timpf, S., Schlieder, C., Kattenbeck, M., Lud-

wig, B., and Stewart, K., pp. 8:1—8:13, Schloss Dagstuhl - Leibniz-Zentrum für Informatik, Regensburg, Germany, <https://doi.org/10.4230/LIPIcs.COSIT.2019.8>, 2019.

Kuo, C.-L. and Chou, H.-C.: Metadata assessment for efficient open data retrieval, in: Accepted Short Papers and Posters from the 22nd AGILE Conference on Geo-information Science (AGILE 2019), edited by Kyriakidis, P., Hadjimitsis, D., Skarlatos, D., and Mansourian, A., Cyprus, Greece, 2019.

Lafia, S., Turner, A., and Kuhn, W.: Improving discovery of open civic data, in: 10th International Conference on Geographic Information Science (GIScience 2018), edited by Winter, S., Griffin, A., and Sester, M., vol. 114 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pp. 9:1–9:15, Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, Melbourne, Australia, <https://doi.org/10.4230/LIPIcs.GISCIENCE.2018.9>, 2018.

Lafia, S., Xiao, J., Hervey, T., and Kuhn, W.: Talk of the town: discovering open public data via voice assistants (short paper), in: 14th International Conference on Spatial Information Theory (COSIT 2019), edited by Timpf, S., Schlieder, C., Kattenbeck, M., Ludwig, B., and Stewart, K., vol. 142 of *LIPIcs*, pp. 10:1—10:7, Schloss Dagstuhl - Leibniz-Zentrum für Informatik, Regensburg, Germany, <https://doi.org/10.4230/LIPIcs.COSIT.2019.10>, 2019.

Larkin, J. H. and Simon, H. A.: Why a diagram is (sometimes) worth ten thousand words, *Cognitive Science*, 11, 65–100, <https://doi.org/10.1111/j.1551-6708.1987.tb00863.x>, 1987.

Mahyar, N., Sarvghad, A., and Tory, M.: Note-taking in co-located collaborative visual analytics: Analysis of an observational study, *Information Visualization*, 11, 190–204, <https://doi.org/10.1177/1473871611433713>, 2012.

Olfat, H., Kalantari, M., Rajabifard, A., Senot, H., and Williamson, I.: Spatial metadata automation: A key to spatially enabling platform, *International Journal of Spatial Data Infrastructures Research*, 7, 173–195, <https://doi.org/10.2902/ijdir.v7i0.279>, 2012.

Roula, K., Franck, F., Robert, L., and Rima, C.: Integrating many legends through ontology for cartographic symbols, in: *Proceedings of SAGEO 2010*, Toulouse, France, 2010.

Ruan, S., Wobbrock, J. O., Liou, K., Ng, A., and Landay, J. A.: Comparing speech and keyboard text entry for short messages in two languages on touchscreen phones, *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 1, 1–23, <https://doi.org/10.1145/3161187>, 2018.

Sano, M., Sumiyoshi, H., Shibata, M., and Yagi, N.: Generating metadata from acoustic and speech data in live broadcasting, in: *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'05)*, pp. 1145–1148, IEEE, Philadelphia, Pennsylvania, USA, <https://doi.org/10.1109/ICASSP.2005.1415612>, 2005.

Sauro, J.: MeasuringU: 10 things to know about the single ease question (SEQ), <https://measuringu.com/seq10/>, last accessed: March 10, 2021, 2012.

Sauro, J. and Dumas, J. S.: Comparison of three one-question, post-task usability questionnaires, in: 27th international conference on Human factors in computing systems

- CHI 09, edited by Olsen Jr., D. R., Arthur, R. B., Hinckley, K., Morris, M. R., Hudson, S. E., and Greenberg, S., pp. 1599 – 1608, ACM Press, Boston, Massachusetts, USA, <https://doi.org/10.1145/1518701.1518946>, 2009.

Scheider, S., Degbelo, A., Kuhn, W., and Przibytzin, H.: Content and context description - How linked spatio-temporal data enables novel information services for libraries, *gis.Science*, pp. 138–149, 2014a.

Scheider, S., Jones, J., Sánchez, A., and Keßler, C.: Encoding and querying historic map content, in: *The 17th AGILE International Conference on Geographic Information Science - Connecting a Digital Europe Through Location and Place*, edited by Huerta, J., Schade, S., and Granell, C., pp. 251–273, Springer International Publishing, Castellón, Spain, https://doi.org/10.1007/978-3-319-03611-3_15, 2014b.

Schmidt, C., Grundel, B., Schumann, H., and Rosenthal, P.: Annotations in different steps of visual analytics, in: *Proceedings of the 16th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2021) - Volume 3: IVAPP*, edited by Hurter, C., Purchase, H. C., Braz, J., and Bouatouch, K., pp. 155–163, SCITEPRESS - Science and Technology Publications, <https://doi.org/10.5220/0010198001550163>, 2021.

Schrepp, M., Hinderks, A., and Thomaschewski, J.: Design and evaluation of a short version of the user experience questionnaire (UEQ-S), *International Journal of Interactive Multimedia and Artificial Intelligence*, 4, 103, <https://doi.org/10.9781/ijimai.2017.09.001>, 2017.

Simon, R., Haslhofer, B., Robitza, W., and Roochi, E. M.: Semantically augmented annotations in digitized map collections, in: *Proceedings of the 2011 Joint International Conference on Digital Libraries (JCDL 2011)*, edited by Newton, G., Wright, M. J., and Cassel, L. N., pp. 199–202, ACM, Ottawa, Ontario, Canada, <https://doi.org/10.1145/1998076.1998114>, 2011.

Trilles, S., Díaz, L., and Huerta, J.: Approach to facilitating geospatial data and metadata publication using a standard geoservice, *ISPRS International Journal of Geo-Information*, 6, 126, <https://doi.org/10.3390/ijgi6050126>, 2017.

Vanhulst, P., Évéquoz, F., Tuor, R., and Lalanne, D.: Designing a classification for user-authored annotations in data visualization, in: *Proceedings of the 13th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2018) - Volume 3: IVAPP*, edited by Telea, A., Kerren, A., and Braz, J., pp. 85–96, SCITEPRESS - Science and Technology Publications, Funchal, Portugal, <https://doi.org/10.5220/0006613700850096>, 2018.

Walny, J., Huron, S., Perin, C., Wun, T., Pusch, R., and Carpendale, S.: Active reading of visualizations, *IEEE Transactions on Visualization and Computer Graphics*, 24, 770–780, <https://doi.org/10.1109/TVCG.2017.2745958>, 2018.

Willett, W., Heer, J., Hellerstein, J., and Agrawala, M.: CommentSpace: structured support for collaborative visual analysis, in: *Proceedings of the 2011 annual conference on Human factors in computing systems - CHI '11*, edited by Tan, D. S., Amershi, S., Begole, B., Kellogg, W. A., and Tungare, M., pp. 3131–3140, ACM Press, Vancouver, British Columbia, Canada, <https://doi.org/10.1145/1978942.1979407>, 2011.