Speech-Based Metadata Generation for Web Map Search

Pei-Chun Lai

2020

Thesis submitted to partially fulfill the requirement of the Master of Science in Geospatial Technologies (Erasmus Mundus).

Supervisor:
Dr. Auriol Degbelo
Institute for Geoinformatics (Ifgi),University of Münster
Münster, Germany

Co-Supervisor:
Dr. Sven Casteleyn
Institute of New Image Technologies,Universidad Jaume I (UJI)
Castellón de la Plana, Spain

Dr. Marco Painho
Nova Information Management School, Universidade Nova de Lisboa (UNL)
Lisbon, Portugal

Certificate of Originality

This is to certify that I am responsible for the work submitted in this thesis, that the original work is my own except as specified in acknowledgments or in footnotes, all the used references and sources are mentioned and that neither the thesis nor the original work contained therein has been submitted to this or any other institution for a degree.

____________________________________(Signature)

____________________________________(Date)

Acknowledgement

First and foremost, I would like to sincerely appreciate the Ministry of Education (MOE) in Taiwan for sponsoring me on the degree of Geospatial Technologies. Also, thanks to all my friends and family members in Taiwan who support me all the time.

Another gratitude to my supervisors. Dr. Auriol Degbelo, my main supervisor, always encourages my work and gives the proper guidance. I am lucky to meet Dr. Auriol, an expert in GIS ontology, and understand my research interest. Dr. Sven Casteleyn improves my JavaScript coding style, and my work can not finish efficiently without his JavaScript teaching. Dr. Marco Painho always shares constructive suggestions and criticisms and makes my work complete.

Finally, I appreciate everyone I met and every moment I had during the pursuit of the degree. These precious experiences make me transform and are the best memories!

Contents

Abstract

Metadata is indispensable for data discoverability and interoperability. Most datasets utilize automatic techniques to create metadata; nevertheless, metadata creation still requires manual interventions and editions, yet manually metadata creation is a tedious task. The study proposes a prototype that introduces speech recognition in the metadata creation process. Users can generate content by speaking. Afterward, the prototype transforms it into metadata with JSON-LD format, a popular metadata format and utilized by mainstream search engines. A user study was conducted to understand the impact of speech-based interaction on user performance and user satisfaction. The result showed no significant performance difference between speech-based and type-based by the efficiency, slip rate, and difficulty rating evaluation. In the user experience evaluation, participants consider the type-based metadata creation is pragmatic, and speech-based metadata creation is hedonic. It suggests that the mix-mode can complement mutually with the advantages of each and optimize the user experience.

Keywords: Metadata Creation, Schema.org, Speech Recognition

Chapter 1

Introduction

1.1 Motivation & Rationale

Map is an influential tool to help people understand complex information at a glance. Public and private sectors integrate map display in the management platforms for better decision making. Nowadays, it's common to see that maps are adopted as a visualization tool for multiple usages. For example, the press utilizes maps to explain the distribution of events or clusters of issues, and some blogs embed maps to illustrate the story clearly.

With the support of map software and web technologies, generating web maps becomes an undemanding task and provokes enormous web map distribution. The metadata quality and content influence the data findability [1]. Metadata is information about data [2], and it increases the data distribution and reuse. It contains predefined elements in an assigned and structured format built based on the search purposes. For instance, the book 'Harry Potter' can be searchable by including the book title, the book author, the published year, and the summary into metadata. Correspondingly, the metadata of web maps would contain map type, abstract, spatial coverage, and temporal coverage [9]. Although metadata plays an essential role in data search, most users are not aware of it. Metadata is machine-readable but mostly invisible to humans. It empowers the communication between machines by giving metadata of web pages. Through the interpretation, machines can display the summary of datasets understandably for humans, and they can examine if outcomes fit their search quickly.

Geospatial metadata can be generated by automatic, semi-automatic, and manual way [6]. Automatic metadata creation generates metadata by extraction and inference information from data. Batcheller introduces the ISO 19115 standard [3], which contains metadata elements such as title, location, then extracts information from the attribute field and turns it into the metadata element's content. Some metadata information derives from inference. For example, the spatial resolution uses the digital number of coordinates. The higher digital number represents a higher spatial resolution. The abstract can obtain from inferences by combining the existing information such as title, place name and adding conjunctions and relational words (e.g., is, part of), and finally generating a full description. [3, 4]. Some metadata extraction techniques are developed for the ad-hoc data catalog systems, and some tools support general extraction purposes. Tools such as gvSIG metadata editor, ESRI ArcCatalog, and Geonetwork supports automatic extraction and update essential information such as spatial boundary, time, and the title [6].Although the richness of metadata elements is limited, it saves lots of work and time.

Previous works successfully harvest the metadata from data by an automatic process, but it cannot replace the manual jobs. For example, CatMEDIT is a popular metadata extraction tool and offers different extraction modes to obtain metadata of a map. Nevertheless, the data quality can be influenced by the modes and input data format, and the outcome requires manual check and modification [5]. Some studies thus integrate the manual intervention to increase the usability of metadata. Kalantari [7] proposes an implicit model that records the frequently used search words for obtaining a geographical feature and provides an explicit model for users to contribute search words. The user engaging in the creation of metadata improves the discoverability of spatial data. The volunteered geographic information system is a vital resource to support Open data, but it still relies on manual work.

Moreover, some studies [11, 12, 16] scratch more information in metadata from maps since a map delivers information by visualizations, and the information in legends can represent the contents of maps. For example, the legend of the land use map reveals the type of land. The visualization can be encoded with semantics descriptions; by integrating it, people can search a map with more keyword options and receive more accurate results. The data descriptions could be done with automatic tools and but mostly is done with manual work, especially for the cases of non-English [20].

The manual work in metadata remains a necessity for creating map metadata and descriptions, while it is criticized as a tedious job and often considered obstacles for better efficiency [4]. This study attempts to introduce speech recognition into manual metadata creation and explores the user experience. Speech recognition technology allows machines to interpret a set of human speech and turn it into text. Using voice to control the device interface and fulfill the simple task have been applied in many applications. Speech recognition has been used to generate the metadata information for a live TV program, and the success rate is 82 percent in 2005 [8]. It becomes a more common tool today. So far, the speech recognition rate is accurate enough to support a smart assistant's operations such as Google Home, Amazon Alex, and Apple Siri and save daily errands' efforts. As speech recognition has been adopted in many fields, we find the potential of using speech recognition in metadata creation.

1.2 Research Questions

Given the practical use of speech recognition, this research addresses the user experience in map metadata creation process. Our research question is: What is the impact of speech-based interaction on user performance and user satisfaction in the process of map metadata creation?

1.3 Research Objectives

The study refers to metadata as a set of semantic descriptions, and the work will focus on two aspects of metadata creation. Firstly, the metadata creation for predefined elements. The elements are used to describe what the map is about, such as place name, alternatives location name, topics, and descriptions. The other focuses on creating the descriptions of geovisualization or map interpretations. The objectives of the study are listed below:

1. To implement a prototype that offers the use of speech recognition in the metadata creation process.
2. To evaluate the performance and satisfaction of speech recognition compared to the current typing approach.

1.4 Methodology and User Study

The methodology is organized into four stages.

1. Design a prototype, which can use speech recognition to create metadata and then transform metadata into semantics data. The prototype includes three features:

- Users can read a map and input descriptions for predefined elements by speaking.
- Users can annotate patterns on a map and create descriptions for annotations by speaking
- The metadata contents from mentioned features are converted into semantics data in the JSON-LD format with schema.org's vocabulary set, and semantics data lead web maps better findability.

2. Design a user study to explore the preference and performance of metadata creation. The experiment offers participants to operate the prototype and we collect operational data for the performance and the experience analyzing. The user study comprises a background survey and user interviews for exploring the background impact and causes of the preference.

3. The step collects data by recruiting people to participate in the experiment. During the experiment, participants were asked to create two types of metadata by speech or typing. The experiment applied 4 indexes to evaluate the performance of both modalities. The efficiency index is measured by the spent time. The difficulty index value is based on the Likert scale rating, the slip rate counts element's correction times, and the accuracy is derived from the speech recognition API.

4. Finally, when data collection is finished, the comparison analyst is performed by using the value of type mode to subtract the value of speech mode. Afterward, we used the bootstrap package to generate the confidence interval and examined whether the performance discrepancy between the two is significant. Additionally, the background impact was examined.

1.5 Contribution

This work introduces speech recognition into the metadata creation process and explores its satisfaction and performance. The findings of the study can be applied to the metadata application design.

1.6 Thesis Organization

The thesis comprises 7 chapters. Chapter 2 reviews the literature regarding the semantic descriptions of maps and semantic annotation. Chapter 3 elaborates on the prototype design and the architecture of experiment implementation. Chapter 4 describes how we design and carry the user experiment. Chapter

5 presents results, and chapter 6 is the conclusion. The final chapter states limitations and suggestions for the future work.

Chapter 2

Literature Review

2.1 Semantic Descriptions of Maps

Semantics technologies warp the human knowledge on webpages into a structured schema for machine automatic processing and understanding [20]. The schema is composed of the name entities, type, and its definition. When things are named similarly, semantics technologies can distinguish between them. For example, the noun 'Turkey' would refer to a bird or a country, based on the content relevant to politics or nature. With the type and definitions recorded in the schema, semantics technologies can distinguish the two. The descriptions are a set of assertions for maps [10] or statements about map contents from the user's perspective [11]. Encoding in the schema can make machines interpret information of webpages with ease. The following will introduce approaches to encode maps.

Some works focus on encoding semantic descriptions with a vocabulary set and predefined schema. Schema makes semantic descriptions present in a structured data form and can be interpreted by machines. Vocabulary set to increase the interoperability of different data for using the same terms to describe a thing. Schema.org [26] is a vocabulary set and can be used with several metadata encodings such as resource description files (RFD), Microsoft data, and JSON-LD. Mainstream search engines also adopt it. Dublin Core Metadata Initiative (DCMI) [25] is another popular standard. For geospatial data, national and international standards for geospatial metadata have developed. Gemini [27] is the national geospatial metadata standard of the United Kingdom. The standard is compliant with the geospatial stand of INSPIRE (Infrastructure for Spatial Information in Europe), the open data committee in the European Union. The United States has Federal Geographic Data Committee to formulate the metadata. Moreover, the metadata also complies with ISO standards, known as ISO 19115 [28].

However, in some cases, the information in maps is more than metadata schema can cover. For example, aside from the spatial distribution in the historical administration map, a map records the information such as crop types and road types in an ancient period, yet they are often omitted in the metadata creation [10].Those pieces of information are essential for informative questions such as 'What crops were planted in the 18th century?'. To achieve a more accurate searching result, works on semantics description have been done to enrich metadata.

Some works focus on sentence analysis. Scheider et al. [10] break descriptions into name entities (e.g., London, Land use) and relation words to describe the relationship (e.g., A part of, is a), then they utilize the name graph to represent the relationships of name entities in descriptions. A predefined vocabulary set is used to classify name entities into different objects. Objects can connect to others (e.g., buildings connect to roads) or be a part of others (e.g., a tree is in a forest). Each object in the name graph can seem as a node, and links represent relationships between objects. The named graph is searchable with some query languages such as SPARQL. Finally, they relate the semantics description with maps by adding another node. Simon et al. [14] develop YUMA, which supports collaborative semantic descriptions on historical maps and incorporates a semi-automatic approach to create linked data. The tool can identify the name entities in descriptions, search corresponding information from databases such as Geonames or DBpedia, and ask users to confirm the information's correctness.

Some works focus more on visualizations. Roula et al. [13] propose an ontology system, CartOWL, to describe map legends. Properties in legends convey information. For example, the red color refers to danger in western society but luck in China. The symbols used by map legends are meaningful to map readers. However, map legends differ from each provider. An ontology system for map legends helps organize those distinguishes and similarities. Thus, when Geodata comes from different service providers, we can mesh-up and optimize the search result and achieve divergent map services' interoperability. Gao et al. [12] propose an approach to encode map legends for a better search result. Map legend implies the content and value range of map layers. The information from map legends such as 'maps with population density larger than 1000 (people /km2)' or 'Common symbols used for highway transportation' can transform into the schema and become searchable.

Except for map legends, map elements such as map scale and the north arrow (orientation) in the map contain semantic information. Carral et al. [15] indicate that the map representation in each scale is different and proposed an ontology design pattern to describe each scale's geographical phenomenon. kadolou et al. [16] add the orientation into the ontology for historical maps. Degbelo [11] proposes a design pattern for geovisulization on maps and classifies 7 patterns for geovisualization, including observation, frequency, outlier, cluster, distribution, trend, correlation. The design pattern can incorporate the machine-readable format and make geovisualization searchable.

2.2 Semantic annotation

Annotations are added marks or selected texts of a document by users, and annotation semantics turns the selections into schema for machine understandings [20].

Annotation can process manually or automatically. Automatically semantics annotation is still a challenge today, but multiple well-developed tools support semi-automatic annotation, such as AeroDAML [18], Armadillo [17], refer [21]. The argument to support automatic annotation states that manual annotation is an expensive process. Annotators need training before annotating work [18], and manual annotating is criticized for low data quality [19]. Semi-automatic semantic annotation mainly adopts pattern-based and machine learning-based to generate annotations. The pattern-based semi-automatic annotation needs manual work for pattern formatting. It searches name entities in a corpus and generates a pattern. The pattern is modified as a new name entity appears in the corpus till it accommodates all entities through recursive modification. Machine learning based adopts induction and statistic methods to predict location and name entities in sentences [18].

However, arguments for manual annotating articulate that the automatic annotations process is incompetent for searching and retrieving data. Data is presented differently and hard to create a general set of ontology [18]. Annotating requires domain knowledge, which demands manual intervention. Moreover, automatic annotation summarizes current materials that are not flexible enough to handle new sentences. Finally, most of the automated semantics tools are in English. Models, algorithms do not support to do automatic annotation for other languages. The manual annotating process is still essential in this view. However, manual annotation is facing the criticize of low usability. Hinze [20] proposes a prototype to improve manual annotation. The study classifies the annotation into free text, shared vocabulary (Turkey can be food or as a country name and treat as different vocabularies) and semantic identity (link the word Turkey with other background information). Later on, conducting a user study to understand their reactions toward different annotations. The result shows that all participants can select text atoms that are important to annotate in a set of texts and can identify the text atom into the corresponding class. Nevertheless, when conducting semantic identification with a dignified topic is a challenging task for non-experts.

2.3 User annotation and visualization

Some works focus on developing a better tool for manual annotation. The Senseus [23] proposes a platform that contains tools for group discussion, and users can leave comments and annotate essential data. The system arranges and visualizes those interpretations for a better group discussion experience. The user study of the platform observes the process of group data exploratory and evaluates the platform's usability. The user study task asks users using the platform to discover the cause of a large drop of bartenders around the 1930s. Users can make their comments, incorporate graphical annotation, and use the most used tool, including arrows, text, and ovals. The study collects comments, graphical annotations, and user interaction data in the online asynchronous discussion. Finally, the study found that users' comments can be further classified into different categories, such as observation, questions, and hypotheses. Besides, they found visualization plays a vital role in user data searching, and most people can extract more information when relevant data is attached with visualization elements.

Similarly, CommentSpace [22] proposes a collaborative annotation system for better group interpretation and deliberation. The solution attempts to reduce the reading burden when threads of a topic increase. It incorporates visualization hacks such as keywords tagging in a comment for quick filtering, puts comments with more views at a more visible place, and offers links to relevant comments. The user study shows that tagged comments can reduce searching time in the early stage and help keep the focus on consistent topics. The links create connections between comments and bring a better experience in an exploratory analysis. Mahyar [24] explores the use of a tabletop device in collaborative visual analysis. In the user study, participants analyzed the dignified topic in groups and can use the tabletop device that offers a visualization toolkit to present their observations. Unlike previous works, the collaborative analysis is synchronous and within a group, and more interactions differ from mentioned studies. For example, participants tend to take private notes in their papers and then share them with group members. Finally, they put the collaborative analysis on the tabletop. The study shows the necessity of taking note during the analysis process and suggests strategies when making manual annotation.

Chapter 3

Implementation

The implementation includes the work on the prototype and user experiment. The prototype allows users to use speech recognition to input the semantic information for maps and convert it into metadata with JSON-LD format. The prototype can create two types of semantics information - predefined elements, which can summarize the map with elements such as place name, alternatives location name, topics, and descriptions. The other type is semantic descriptions for geovisualization describing the contents on a map. The annotation functions will work with the second type cause it is essential in the analytic process and generation of semantic descriptions [24] . Additional functions are built according to the design of the user study, which elaborates in chapter 4. The chapter introduces the design of the prototype and the user experiment's architecture.

3.1 Prototype

3.1.1 Map Metadata Element Creation

To get an impression of the offered metadata management tools, we reviewed metadata creation tools suggested by the Federal Geographic Data Committee and Open Source Geospatial Foundation [29, 30]. In table 3.1 we summarize those who were still working (12) while writing this thesis. The tools reviewed use typing to create metadata, and some tools offer a map view to visualize the spatial extent of the resource being documented.

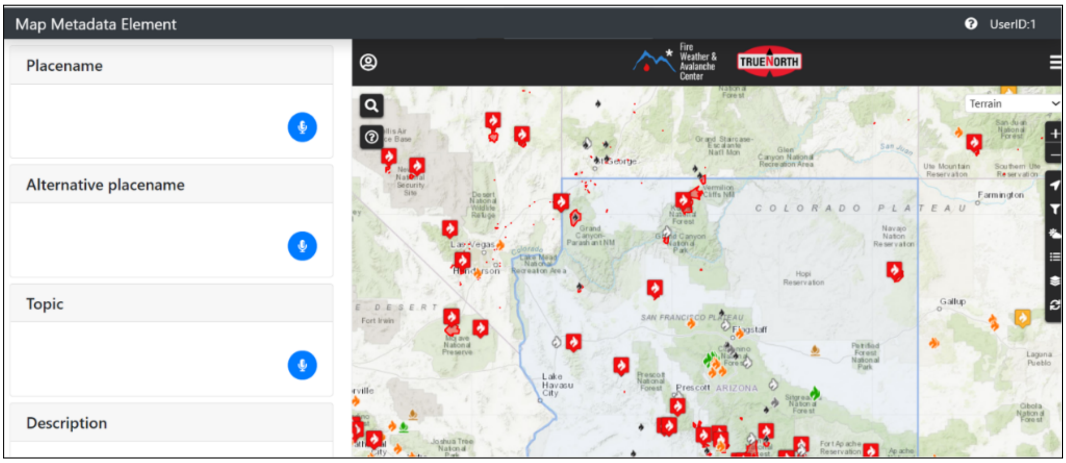
Most metadata creation tools support users to document more than 20 elements, but less than half tools provide the map view. Nevertheless, our metadata creation comprises the annotation activity, which needs a map view. As the predefined elements, we select the most common elements in metadata creation tools into our prototype. Ultimately, the interface contains two parts,

Table 3.1: Classification on Interface

Tool	Platform	OS	Metadata Creation	Map view
GeoNetwork	Web	-	Typing	V
Mapbender	Web	-	Typing	
MDweb	Web	-	Typing	
MEtadata Editor	Desktop	Windows	Typing	
CatMDEdit	Desktop	Windows	Typing	
ISO Metadata Editor (IME)	Desktop	Windows	Typing	V
EPA	Desktop	Windows	Typing	
tkme	Desktop	Windows Linux	Typing	
CoMET	Web	-	Typing	V
MetadataWizard	Desktop	Windows	Typing	V
ArcGIS	Desktop	Windows	Typing	V
QGIS	Desktop	Windows	Typing	

which are the map view and metadata creation. Users can view the map and input elements in the metadata creation area (Figure 3.1). The application support users to input metadata by voice and the voice simultaneously transforms into text and display on the interface.

Figure 3.1: Prototype of Map metadata element creation



3.1.2 Map Annotation Creation

The map annotation creation allows users to take notes on a map and generates the semantic descriptions from the visualization. The map view and data view were provided on the interface. Users can use the annotating tool to take notes on maps and input semantics descriptions afterward(Figure 3.2). The annotation tools include rectangle, circle, free-drawing pen, and the pin allow users to highlight their observations on a map (Figure 3.3).

Figure 3.2: Prototype of Map annotation creation

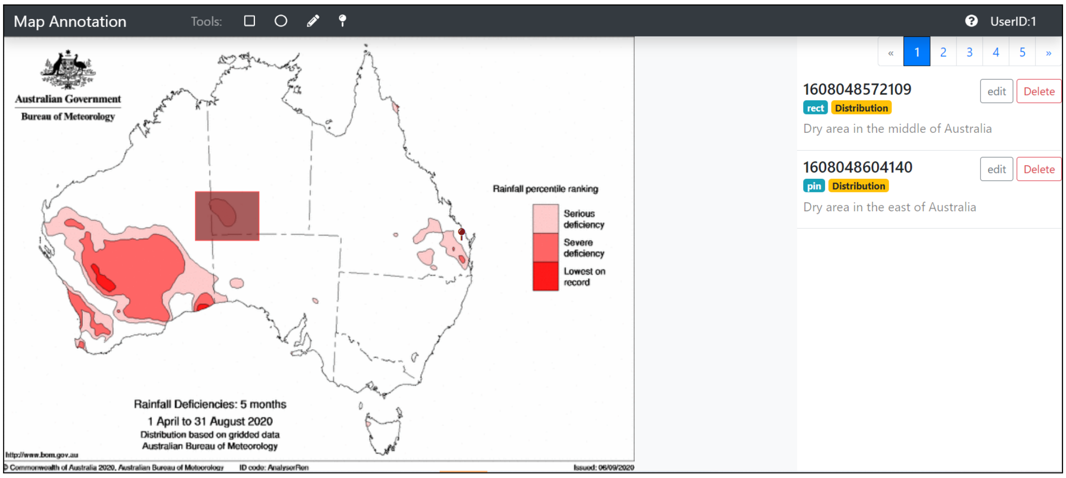
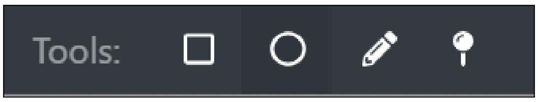
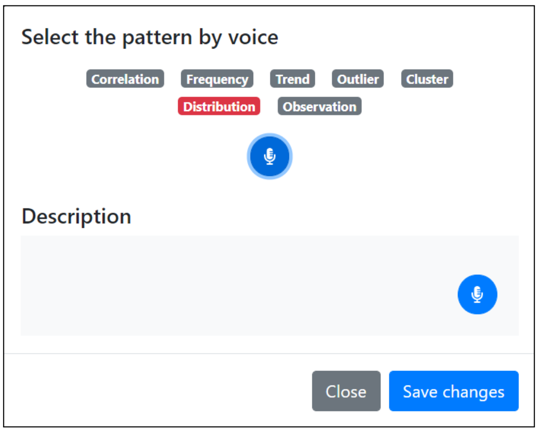


Figure 3.3: Annotation Tools



As users finish annotation drawing, a pop-up window will automatically show up, and users can select a pattern and input descriptions by speaking. Here, the pattern selection utilizes the design pattern proposed by Degbelo [11]. The design pattern includes cluster, outlier, correlation, trend, frequency, distribution, and observation. Lastly, users input descriptions for the annotation by speaking and save data.

Figure 3.4: Input values by speech



3.1.3 Speech Recognition

Speech recognition has been developed for a while, and there are several service options. The study adopts the open-source speech recognition API developed by Mozilla [31]. The API offers functions to textualize the speech and generate audio files from text synchronously. Besides, it supports multiple languages, and the default language usually depends on the language property of the webpage [32]. Since our focus is on English, the language set of language is 'en-US', and the prototype interprets speech input only into English.

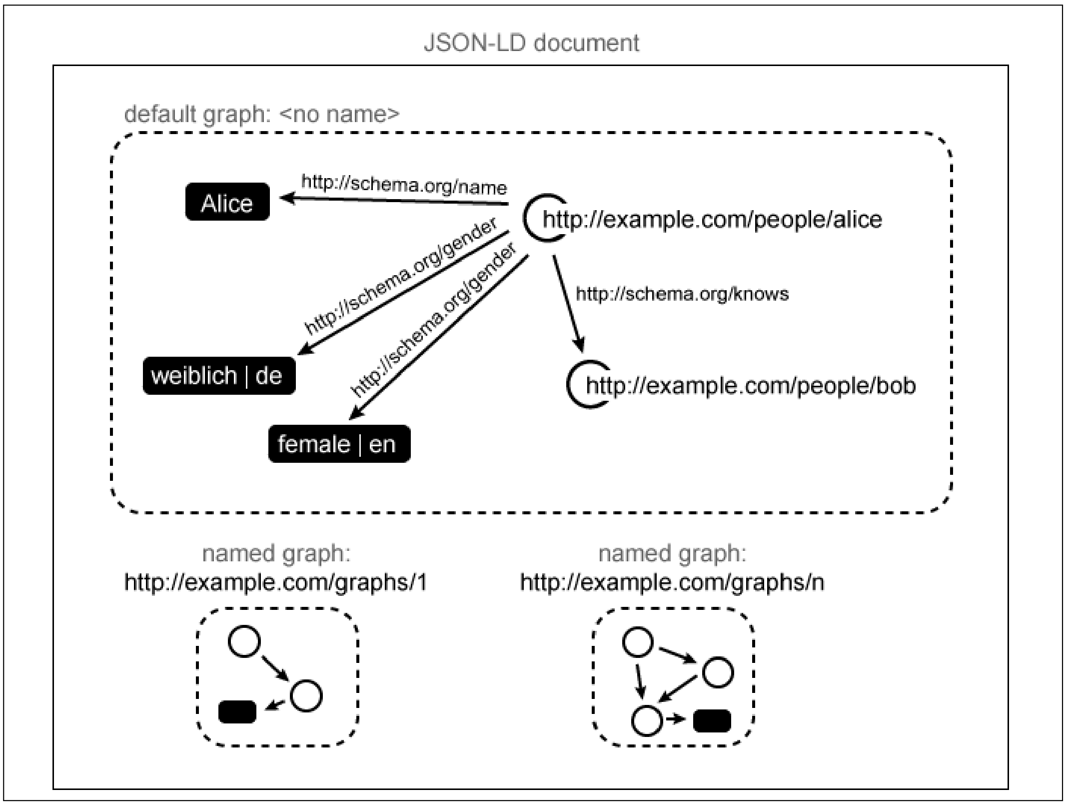
3.1.4 Metadata Conversion

Finally, when users finish the metadata creation, the application will transform metadata into the JSON-LD format with the vocabulary set of Schema.org. JSON-LD is an encoding format to save linked data. The linked data represents data with keys and values, and keys connect mutually if the relationship between data exists. The connections create a graph data structure, which allows machines to search for data from one data to another (Figure 3.5). Thus, machines can perform searches similar to human searches and increases the findability of data. Nowadays, Google uses linked data to support its knowledge graph. People can put questions such as 'how do I create metadata,' and Google can answer the question based on information grasped from linked data and list relevant resources. For compiling data in JSON-LD with standards, Schema.org is used for standardizing keys and values. Schema.org is a vocabulary set for entities, and it also defines the relationship between entities. Now the use of JSON-LD is advocated by mainstream web service providers, including Google and Facebook; hence it is selected as the study's format.

According to the standard of schema.org [26], each schema contains types and properties. For example, a book includes 'bookFormat', 'illustrator', 'isbn', 'bookFormat' , 'numberOfPage', etc. The expected type for a book is a description; then, a book belongs to the entities of CreativeWork. Schema.org has defined properties, and users need to decide the type and properties for representing their data.

For the map element creation, elements including place name, alternative place name, the topic, descriptions, the start time, and the end time of a map are selected(Table 3.2). The data type is 'map,' and we have defined the corresponding schema from Schema.org and compiling it to JSON-LD (List 3.1).

Figure 3.5: The mechanism of JSON-LD [34]



In the prototype of annotation creation, users can put annotation with a selection from the designed pattern and a description. We did not find the schema for recording geometries. Therefore, we use comments to warp patterns and descriptions. Considering one map could have multiple annotations, we add dateCreated schema to distinguish annotations. The schema lookup table and conversion JSON-LD are listed below.

Listing 3.1: Example of Schema Element Metadata Creation

```
User Input:
{Place":"Western United States",
"alternateName":"southern Colorado mountains",
"Topic":"wildfire",
"Description":"places with active wildfire and
   historical records in southwestern USA."
"Time":"/2020-12-17"}
Result:
{
  "@type": "http://schema.org/Map",
  "http://schema.org/alternateName": "southern Colorado
     mountains",
```

Table 3.2: Schema for metadata element creation

Input	Schema	Definition of schema
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 3.3: Schemas of annotation creation

Input	Schema	Definition of schema
Pattern	termCode	A code that identifies a defined term within a specified term Set
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

```
  "http://schema.org/description": "places with active
     wildfire and historical records in southwestern USA
     .",
  "http://schema.org/keywords": "wildfire",
  "http://schema.org/spatialCoverage": "Western United
     States",
  "http://schema.org/temporalCoverage": "/2020-12-17"
}
```

Listing 3.2: Schema of Annotation Schema Creation

```
{ User Input:
{{"Pattern":"Cluster",
"description":"there is a bigger hotspots close to this
   cluster however they are separated and there is a
   distinct difference in the values before it becomes a
```

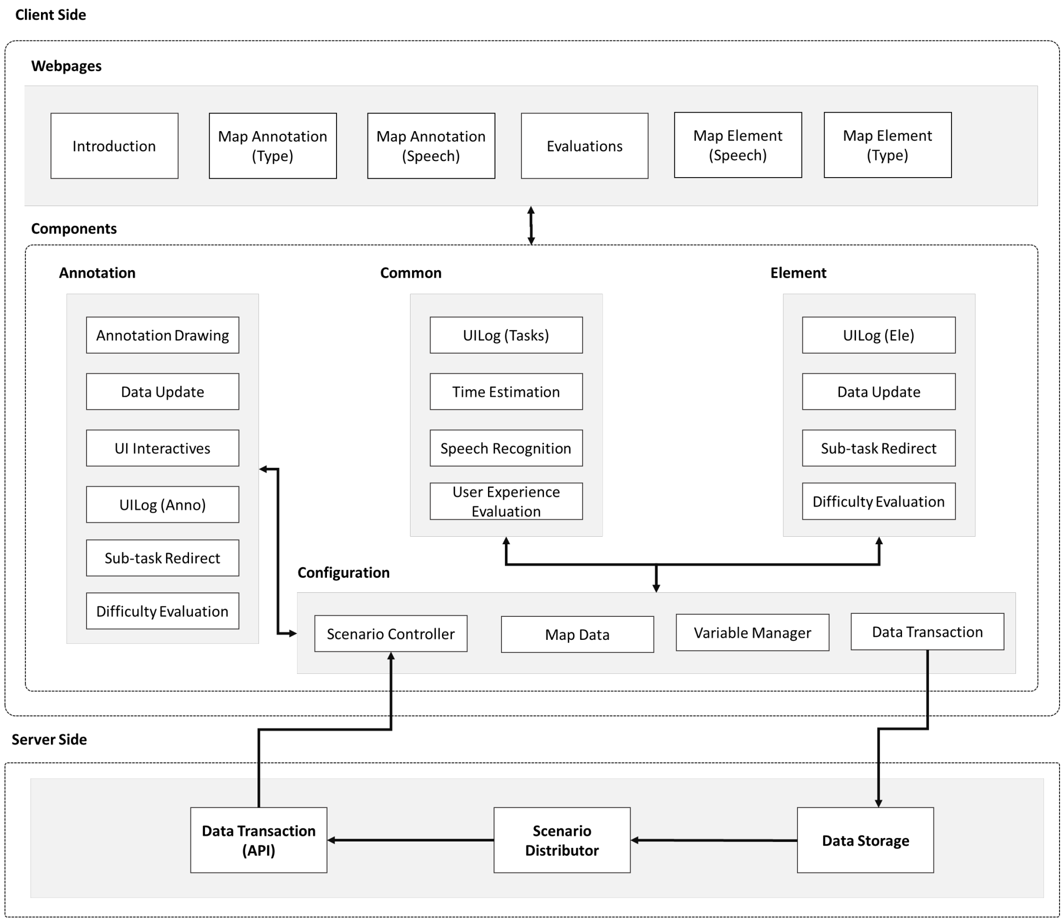
```
    cluster",},
{"Pattern":"Distribution",
"description":"there seems to be a corridor that is
   allowing for this that explains the rainfall
   deficiency however without more information about the
    elevation or other variables not sure what is the
   cause of this rainfall deficiency corridor"}

Result:
{ "@type": "http://schema.org/Map",
  "http://schema.org/comment": [
{"http://schema.org/dateCreated": {
 "@type": "http://schema.org/Date",
 "@value": "2020-12-1810:59:18"},
 "http://schema.org/description": "there is a bigger
    hotspots close to this cluster however they are
    separated and there is a distinct difference in the
    values before it becomes a cluster",
"http://schema.org/termCode": "Cluster"},
{"http://schema.org/dateCreated": {
"@type": "http://schema.org/Date",
 "@value": "2020-12-1811:0:47"},
"http://schema.org/description": "there seems to be a
   corridor that is allowing for this that explains the
   rainfall deficiency however without more information
   about the elevation or other variables not sure what
   is the cause of this rainfall deficiency corridor",
"http://schema.org/termCode": "Distribution"}
]}}
```

3.2 Architecture

The implementation consists of the prototype and the user experiment (Chpater 4) as mentioned. The final product of implication is a web application, and the architecture is illustrated in Figure 3.6. The client-side provides a user interface for operation and meanwhile recording the operation data, feedback from users. The server-side provides functions to store data. The client-side and server-side are hosted on Heroku and are accessible on the internet.

Figure 3.6: Architecture of the application



3.2.1 The Client-side

The client-side comprises the prototype and features for the user experiment. When the application initializes, it requests a scenario code from the server-side, then shows corresponding pages to participants (metadata element creation/ map annotation creation; typing/speech). The maps in scenarios are defined in source data. The element creation calls components to log user's actions and update data. The annotation creation uses components to control the annotation drawing, interface updates such as deleting and editing, user action logging, and update data. Both prototypes have a specified sub-tasks redirection, and difficulty evaluation components. Some components have universal use, such as time calculation, speech recognition for letting users create data by speech, and user experience evaluations. User experience survey and difficulty evaluation are forms that participants are aware of, but UILog components work in the background. Finally, when a participant finished the experiment, the data transaction component posts data with the restful API to the server-side.

We utilize Vue framework for client-side development. The application needs to manage many updates on interface and data on different web pages, and the Vue framework fits the demands which bind the data with the interface reactively. In other words, as data changes, the interface updates simultaneously and correspondingly. Therefore, we can reduce the scope of development and have better management of code. The full code is available on GitHub.

3.2.2 The Server-side

The work is light on the server-side. Components on the server-side manage data and assign scenarios to participants. There are 4 scenarios; each time the client-side initializes, the data transaction component requests a scenario code from the server-side. The server-side reads data from the participant file and generate a scenario code(Algorithm 1), which decides how the client-side presents the scenario. After each participant finishes the experiment, the data will send to the server-side through another restful API. Finally, the data is stored in files in CSV format.

Algorithm 1: Scenario Distributor

```
Result: Scenario
initialization;
userid = participant number + 1;
switch userid do
    case (userid mod 4 = 1): scenario1;
    case (userid mod 4 = 2): scenario2;
    case (userid mod 4 = 3): scenario3;
    case (userid mod 4 = 0): scenario4;
end
```

Chapter 4

User Study

The experiment helps understanding human and computer interactions. The chapter illustrates the experiment design, pilot study results, and the experiment's adjustments to ensure the full procedure can comply with the experimental goal.

The experiment's goal is to understand the impact of speech-based interaction on user performance and user satisfaction in map metadata creation compared to the current typing approach.

4.1 Independent and Dependent Variables

In the user experiment, independent variables are manipulated by the experimenter to fit the experiment goal with effective validity [35]. The independent variables in our study are two tasks with two interaction modalities – typing and speech. The task - map metadata element creation provides 10 wildfire and drought maps, then asks participants to interact with maps and use predefined elements such as place name, descriptions, and summary. The other task, map annotation creation, offers 5 wildfire or drought maps, then asks participants to create map annotations for each map (Appendix). Participants read a map, use the drawing tool to put annotations, and then describe it. Participants will be asked to input information by speaking or typing. We expect each participant spends 3-5 minutes on a map, and the operation of a condition lasts at least 20 minutes. If taking into account the duration of other survey events, such as the introduction and background survey, the whole experiment will still not exceed longer than 90 minutes. The design keeps participants focus on tasks with enough time and also will not be exhausted through tasks.

Dependent variables are collected from several approaches. The procedure starts with a questionnaire to gain background information of participants, and it can assist in understanding if the background relates to the preference and satisfaction. The questions are listed (Table 4.1) below, and the full questionnaire is accessible here.

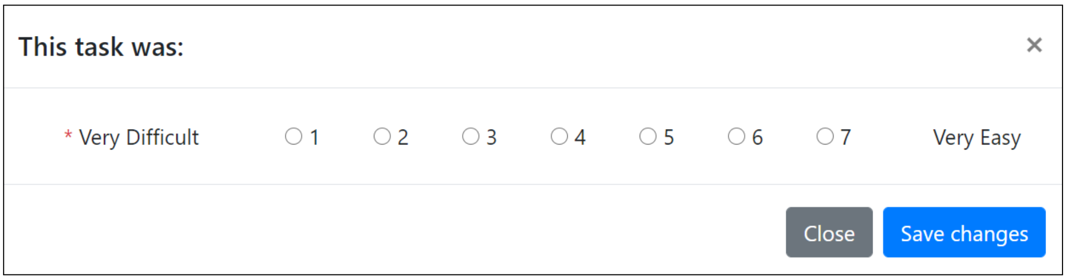
Table 4.1: Questions in Background Survey

Categories	**Questions**
Information of participants	1. Age 2. Gender 3. Where are you from (country) 4. Are you a native English speaker
Experience in using Web map	1. How often do you use web map? 2. In which platform you use web maps? 3. How did you first find out about the web map you use most often?
Experience in GIS software	1. How often do you use GIS product? 2. What kind of GIS product you've used?
Experience in metadata	1. How often do you use GIS product? 2. What kind of GIS product you've used? 3. How do you compile the metadata? 4. The task of creating metadata is (difficulty evaluation) 5. How long do you spend to create metadata for a thing on average
Experience in Speech Recognition	1. How often do you use speech-recognition technologies? 2. In which platform do you use speech-recognition 3. While using the speech recognition, how many times do you need to speak for a correct result?

The efficiency of modality is measured by time spent on given elements. The speech accuracy adopts the recognition confidence value derived from speech recognition API. The confidence value is between 0 and 1 and shows the confidence in a correct recognition [33]. Finally, the slips of value input are counted by times of modifications. Participants will not be aware of the data collection since it works in the background.

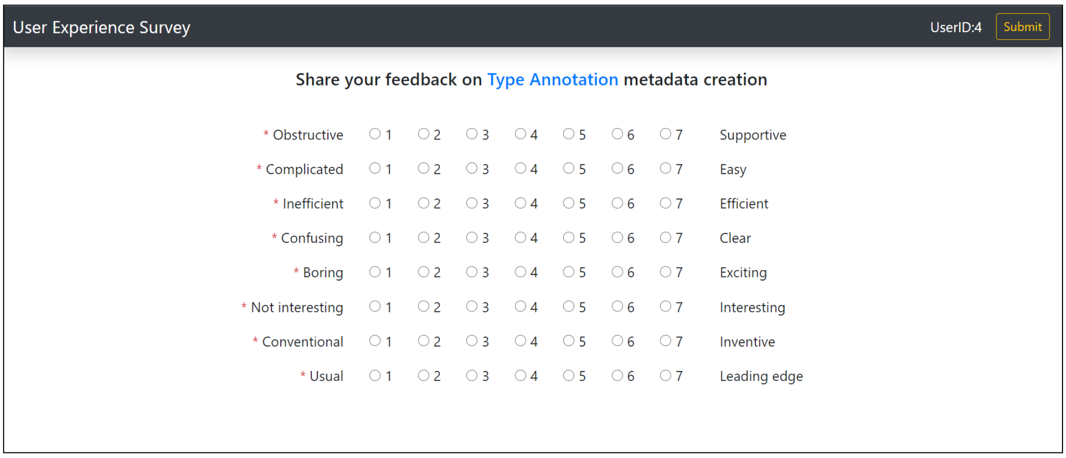
The evaluation tools are introduced to detect the difficulty level and the user experience of tasks and sub-tasks to understand user's preferences. In both tasks, as a participant finishes a map (sub-tasks), the application pops out the Likert scale to participants to evaluate the subtask's difficulty level. Likert scale (Figure 4.1) is short and easy to answer and can reduce the interruption in the experiment. The difficulty level will be evaluated before the next map comes out and ensures participants have clear memory and feeling to evaluate the map they've just finished [36].

Figure 4.1: Difficulty Evaluation with the Likert scale



After participants finish a task, we ask them to evaluate their experience with the short version of the user experience questionnaire (UEQ-S, Figure 4.2). UEQ-S is easy to operate, time-saving while the evaluation is validated [37].

Figure 4.2: Short Version of the User Experience Questionnaire (UEQ-S)



4.2 Experimental Method

Considering the independent variables are two tasks bonded with two creation modes, it demands a strategy to curate the experiment. The experiment applies the within-subject approach that each user will experience all independent variables. The within-subject design can bring a more precise result than a between-subject approach due to each subject experience all independent variables and can achieve statistical validity with fewer participants [35]. However, there are some concerns in a within-subject design. Users may learn from earlier tasks and gradually master operations or realize the experiment's purpose.

Thus, we employ the tactics named the Greco–Latin squares to arrange scenarios counterbalancing (Table 4.2) and effectively prevent learning effect [38].

Table 4.2: Count of annotation per map

Scenario	**Phase1**	**Phase2**
Scenario1	Element by typing	Annotation by speech
Scenario2	Annotation by speech	Element by typing
Scenario3	Annotation by typing	Element by speech
Scenario4	Element by speech	Annotation by typing

4.3 Procedure

The procedure comprises the background survey, implementing tasks, and user interviews. At the beginning of the study, participants will receive a link to the experiment web application. On the first page of the web application, participants will be informed of the experiment process and the data collection declaration, which information of background information of participants and operational data will be collected. The experiment officially starts with approval from participants. Due to the experiment conducts online, we collect consent by video recording instead of sign. Next, participants will test their microphone device on the second page of the web application. Afterward, we send them the link to access the questionnaire created with Google forms to collect the background information. The main part of the experiment follows after the background survey.

From the third page, participants implement tasks. The whole experiment is divided into two phases, and participants do a task in each phase. The task is either annotation or element creation, and it depends on participant ID (Table 4.2). Each task contains 5 sub-tasks. Evaluations interweave with tasks and sub-tasks. The measurement of the difficulty level of sub-tasks (Figure 4.1) shows up before the following the sub-task, and assessments of user experience carry on after the task (Figure 4.2). Finally, we conduct the user interview. The experiment's duration is 45 to 60 minutes to finish.

The experiment is conducted online due to the pandemic situation, so we utilize Zoom and Google hangout meeting platforms to achieve every step of the experiment.

4.4 Participants

Participants need to have communicable speaking and writing capability and have a microphone device that can convey voice clearly. Participants are recruited by emails, Facebook, Twitter, university students, and word of mouth.

We meet the participants on the online meeting platform, and the whole procedure complete on the online platform. Before the experiment, we request approval by asking participants to video record a read-aloud consent declaration. Participants are all informed they need to share their computer screen and allow the experimenter to video record the operations, and they can withdraw their data at any time.

4.5 Pilot Study

The pilot study examines if the experimental design works as we expected [35]. There were three participants in the pilot study. Two are male, and another is female. All participants have a geography background but currently dedicate to different domains, including financial, art and performance, and environmental law. They all belong to the 30 to 34 age group, and none of them is a native English speaker. However, one participant can speak multiple languages, including French, German, Mandarin, and owns English level almost the same as a native English speaker. From the background questionnaire, we found one participant has experience in creating metadata, and two have used speech recognition, and all have experience in using GIS to generate maps.

Several adjustments were made after the first pilot study. The first-page introduction was designed to let participants understand the experiment's purpose and operations with their paces. However, the first participant skipped the introduction and had many questions during the experiment. The first participant testified the scenario that the task annotation with the typing modality in phase 1 and the task element creation with the speech in phase 2. In the second stage, speech recognition could not effectively interpret the participant's speech and resulted in multiple failed attempts in value input.

Moreover, the task element contains 10 subtasks, and participants were overwhelmed. In the end, the experiment took 90 minutes, and it was significantly longer than expected. Another feedback from the first participant is that the question of difficulty level evaluation (Figure 4.1) is ambiguous to him. Therefore, his evaluation of difficulty was answered arbitrarily due to uncertainty about the question. Changes made are listed below.

1. The introduction of the experiment is presented by slides and includes examples of task implementation.

2. The location of task instruction is highlighted during the introduction. The map number in the task element creation reduce from 10 to 5.

3. Questions of the difficulty evaluation (Figure 4.1) were rephrased. In the map element creation, participants shall see the question as: 'Overall, summarizing maps with elements was:' and 'Overall, putting annotations on the map was' in map annotation.

After the modification, each experiment's duration is about one hour and fit our expectations, and participants did not ask lots of questions during the experiment. We still slightly modify the experiment according to the feedback. The first and the second question are removed in the user interview because no practical answers are obtained from those questions. The first and second questions were to probe the pros and cons of speech recognition and typing for metadata creation. However, answers were addressed on the difficulty during making annotation and sentimental feeling on tasks. Alternately, in the third question, we asked participants to compare whether they would choose typing or speech recognition for their project, and participants clearly illustrate the pros and cons.

Table 4.3: Count of annotation per map per participant

Map ID / Participant ID	**1**	**2**	**3**
1	0	2	2
2	2	1	2
3	4	2	3
4	2	2	1
5	1	1	3

Additionally, we notice the first and second participants struggle in making lots of annotations. Participants spent a long time on the first map, then they gradually lost interest and rush to finish sub-tasks. We asked the third participant to make annotations with instincts and do not hesitate to put annotations on maps. The process went more smoothly than the previous.

However, considering all participants have difficulties in the task annotation, we count the annotation number in each sub-task (Table 4.3). Finally, we decide that all participants only make 2 annotations for each map immediately just as they observe something in map annotation creation.

Chapter 5

Results

Efficiency, accuracy, user experience, difficulty, slip rate, user background information, and feedback were collected from 12 participants. BootES, an R package [39] was utilized to bootstrap the small sample. The data was re-sampled by bootstrapping the initial sample to 2000. The analysis presents the descriptive statistical value and compares the difference between modalities (typing, speech). The comparison used value in type mode subtract the value in speech mode by the weight setting (type value * 1, speech value * -1). The sign of value indicates which mode outweighs another in the evaluation. Then, the confidence interval (CI) estimation is utilized to examine the difference is significant or not. CI also provides mean, upper bound, and lower bound value, which offers more information for explaining results.

As a confidence interval of difference includes value zero, it indicates statistical non-significance. Zero implies the possibility of no difference between a compared pair within an experimental replication. Additionally, a narrower interval indicates a greater significance since it shows that the difference in samples is consistent, not result from outliers. Finally, as the absolute difference value is larger than 0, showing a more significant discrepancy between compared variables, which implies a higher significance. The study estimated the effect size when the difference between a comparison is significant. The effect size is a measure to standardize the different level between a comparison, and the approach used in the study is Hedges' G, which is applicable for a small sample.

The influence of three background factors was compared with the following weight setting: Gender (Female = 1, Male=-1), Native Speakers (Yes=1, No=-1), Metadata Experience (Yes=1, No=-1).

5.1 Participant Background

As mentioned, participants were recruited from social media and word of months. The participants' age is between 18 to 34, and among 12 participants, there are 7 males and 5 females; only 2 are native English speakers (Figure 5.1). More than half of people use Web-Map every day, and most of them have experience in GIS usage and metadata creation (Figure 5.2). 6 out of 8 participants use the manual way to compile metadata. It shows that manual creation is still the primary alternative. However, the difficulty rating results and time spent on metadata creation experience are diverse. About the experience in using speech recognition, 2 people have never used the technology, and most of them use a speech recognition feature on mobile devices.

Figure 5.1: Background

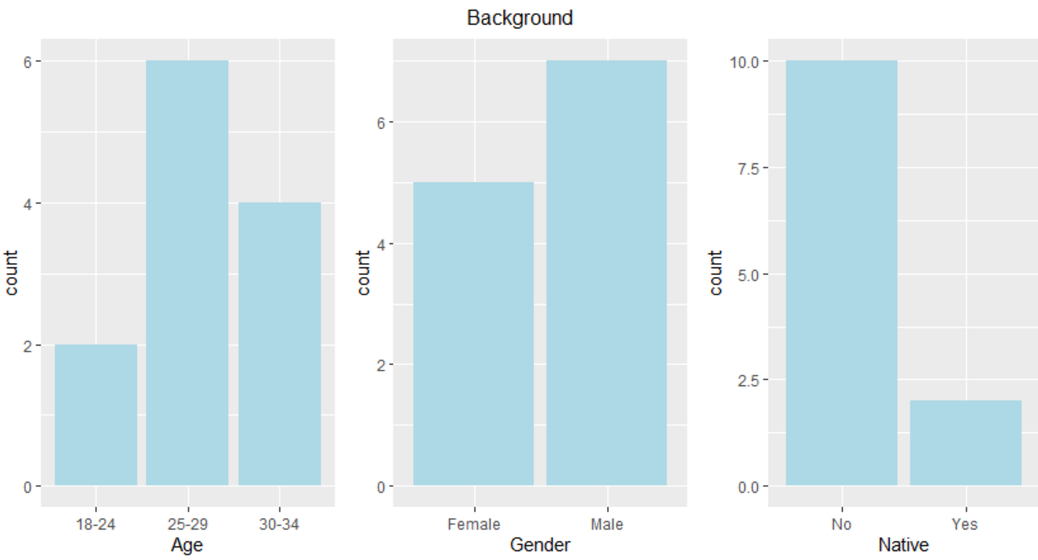


Figure 5.2: Using Frequency in web maps, GIS, Metadata, and Speech Recognition

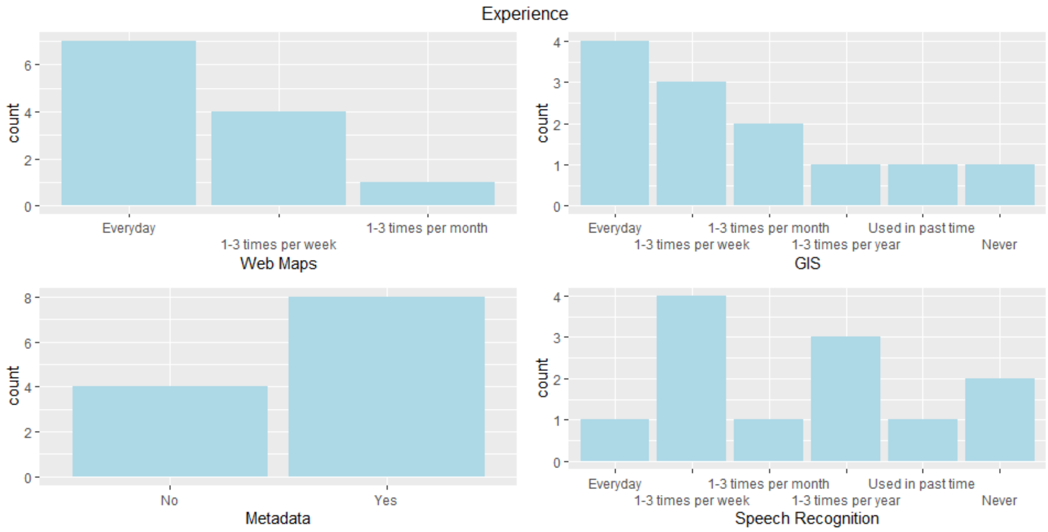


Figure 5.3: Other Background Information

5.2 Efficiency

The efficiency was measured by the time spent on each map (unit: seconds) in the task element and the annotation. All participants finish a map around 2 minutes (Table 5.1). Overall, the maximum and minimum value of the type is 119 and 68 seconds; the speech is 111 and 69 seconds.

Table 5.2 presents a statistic summary of the task element and the annotation. The maximum and minimum value of the type is 124 and 80 seconds; the speech is 125 and 84 seconds. In the task annotation, values of the type are 115 and 67 seconds; the speech is 97 and 58 seconds. Table 5.3 is the CI of the time discrepancy among the type and the speech in both tasks. The values are all negative in the task element, implying the type spent less

Table 5.1: Mean Time Spent on each map (Seconds)

Map	Modality	Overall	Task Element	Task Annotation
Map1	Type	104.42	113.50	95.33
	Speech	111.08	124.33	97.83
Map2	Type	119.75	124.00	115.50
	Speech	104.25	125.50	83.00
Map3	Type	95.17	80.50	109.83
	Speech	69.75	84.67	54.83
Map4	Type	76.67	83.50	69.83
	Speech	71.83	84.83	58.83
Map5	Type	68.33	69.17	67.50
	Speech	87.42	99.83	75.00

time than the speech. The task annotation gains negative and positive values, and the type mode spent more time than the speech in map 2,3,4. However, none of the maps is significant, indicating that the efficiency of speech and type are not different.

Table 5.2: Type and Speech comparison between each map

Map	Type vs Speech
Map1	-6.667,CI[-91.775,27.917]
Map2	15.500,CI[-49.810,74.781]
Map3	25.417,CI[-34.468,59.372]
Map4	4.833, CI[-34.098,27.250]
Map5	-19.083,CI[-64.132,15.608]

Table 5.3: Type and Speech comparison by tasks

Map	Element	Annotation
Map1	-10.83,CI[-138.4,83.17]	-2.500,CI[-74.83,49.5]
Map2	-1.5,CI[-87.67,76.667]	32.500,CI[-42.9,140.5]
Map3	-4.167,CI[-73.1,37.810]	55.000,CI[-8.5,125]
Map4	-1.33,CI[-81.67,60.5]	11.000,CI[-17.67,58.17]
Map5	-30.67,CI[-130.5,14.33]	-7.500,CI[-72.17,42.5]

The background factors, namely gender, language capability, and metadata creation experience background, were compared. The median of 5 maps was used for the comparison. Females perform better than males in the type mode but inferior to males in speech mode. The native speaker finished tasks faster than the non-native speakers in both modalities, and people who have the metadata experience spent more time than those who do not. People with metadata experiences would consider more details and thus spent more time on creation. However, all confidence interval (Table 5.4) is large and contains zero and thus refers to insignificant discrepancies between those background groups.

Table 5.4: The efficiency under different background groups

Modality	**Gender**	**Native Speaker**	**Metadata Experience**
Type	9.43,CI[-40,52.21]	-11.4,CI[-49.45,21.13]	5.25,CI[-44.97,56.38]
Speech	-2.86,CI[-52,104.39]	-20.6,CI[-87.48,20.64]	28,CI[-8,88.09]

5.3 Difficulty Rating

The difficulty value is measured with the Likert Scale [36], participants evaluate sub-tasks by rating 1 to 7. The 1 indicates the most difficult and 7 is the easiest. Table 5.5 presents the mean value of the difficulty rating. Overall, the type gets difficulty value from 4.83 to 5.5, and speech gets values from 4.83 to 5.33. The value indicates that the difficulty is neutral. The overall comparison between the type and the speech is illustrated in Table 5.6. Most maps get positive values, implying the type is easier than the speech. Nevertheless, the discrepancy is not significant.

Table 5.5: Mean Difficulty Rating in the Modality of Speech and Type

Map	**Modality**	**Overall**	**Task Element**	**Task Annotation**
Map1	Type	5.5	5.67	5.33
	Speech	4.67	4.83	4.5
Map2	Type	4.92	5	4.83
	Speech	4.83	4.83	4.83
	Type	4.83	4.33	5.33
Map3	Speech	5	5.17	4.83
	Type	5.33	5.5	5.17
Map4	Speech	5.33	5.67	5
	Type	5.42	5.5	5.33
Map5	Speech	4.83	5	4.67

Table 5.6: Type and Speech Comparison on each Map

Map	**Type vs Speech** *
Map1	0.833,CI[-0.250,1.750]
Map2	0.083,CI[-1.167,1.250]
Map3	-0.167,CI[-1.083,0.833]
Map4	0.000,CI[-0.750,1.000]
Map5	0.583,CI[-0.167,1.500]

Taking a closer look at tasks, we can see in the task element, the ratings are from 5 to 5.67, saying that the task element tends to be easy for participants. The task annotation rating is between 4 and 5.5, suggesting a neutral difficulty (Table 5.5). The comparison in both tasks is presented in Table 5.7. Whether speech mode or type mode is easy is inconclusive in the task element, speech mode outweighs type mode randomly. In the task annotation, the type is easier than the speech in all maps. However, no significant difference was found between the type and the speech in both tasks, referring to a similar difficulty level on the two modalities.

Table 5.7: Type and Speech Difficulty Rating Comparison

Map	**Element**	**Annotation**
Map1	0.83,CI[-0.17,2.5]	0.83,CI[-0.83,2.33]
Map2	0.17,CI[-2,2.33]	0,CI[-1.67,1.5]
Map3	-0.83,CI[-2.33,0.5]	0.5,CI[-1,1.67]
Map4	-0.17,CI[-1.17,0.33]	0.17,CI[-1.33,2]
Map5	0.5,CI[-0.83,2.67]	0.67,CI[-1.17,2.5]

Finally, the comparison of different background groups is presented in Table 5.8. Participants with metadata experience's evaluation are lower than opposite groups in both modalities, indicating they tend to consider tasks are difficult. Males' rating on the type modality is higher than females, and females' difficulty rating on the speech is higher than males. However, the differences in groups are not significant.

Table 5.8: The difficulty rating under different background groups

Modality	**Gender**	**Native Speaker**	**Metadata Experience**
Type	-0.46,CI[-1.54,0.23]	0.4,CI[-0.2,0.8]	-0.88,CI[-1.88,-0.38]
Speech	0.26,CI[-1.51,1.43]	-0.3,CI[-2.8,2.2]	-1.125,CI[-2.5,0.375]

5.4 Slip Rate

The slip rate was evaluated by the correction times on each input form. In the task element, there are 6 values, including place, alternative place, topics, description, start time, and end time. Table 5.9 shows that the slip rate is from 1 to 3 times. Table 5.10 illustrates the comparison between two modalities in the task element. Values are negative or positive and present without a systematic pattern, representing the slip rate may not be relevant to the modalities. Besides, most confidence intervals are not significant. Therefore, the comparison remains inconclusive. In other words, the two modalities are similar.

Table 5.9: Mean Slip rate in the Task Element

Map	Modality	Place	Alternative Place	Topic	Description	Start Time	End Time
Map1	Type	1.33	1	1.5	2.83	1.83	1.17
	Speech	1.33	1	2.17	3.17	2.33	0.83
Map2	Type	1	1.33	1.33	2.67	2	2.33
	Speech	1.67	2.5	3.83	4.83	1.67	1.83
Map3	Type	1	2	1.33	1.5	1.17	1.5
	Speech	2.83	1.17	1	1.17	1.17	1.5
Map4	Type	1.17	1	1.5	2.67	1.83	1.33
	Speech	1.17	1	2.17	2.83	1.33	1.33
Map5	Type	1.5	1.33	1.33	1.67	1.17	1.83
	Speech	1.5	1.67	1.5	1.67	1.33	1.67

Table 5.10: Type and Speech Comparison on each Map

Map	Place	Alternative Place	Topic
Map1	0,CI[-0.67,0.33]	-	-0.67,CI[-2.83,0.33]
Map2	-0.67,CI[-1.5,-0.33]	-1.16,CI[-5.5,0.33]	-2.5,CI[-5.49,-0.58]
Map3	-1.83,CI[-9.17,0]	*0.83,CI[0.17, 2.17]	0.33,CI[0,1]
Map4	0,CI[-0.5,0.17]	-	-0.66,CI[-4,0.83]
Map5	0,CI[-1.12,0.5]	-0.333,CI[-2.5,0.17]	-0.17,CI[-1.17,0.67]
Map	**Description**	**Start Time**	**End Time**
Map1	-0.33,CI[-5.33,1.5]	-0.5,CI[-2.17,0.33]	*0.33,CI[0,0.67]
Map2	-2.17,CI[-6,0.83]	0.33,CI[-0.83,1]	0.5,CI[-1.83,2.5]
Map3	*0.33,CI[0,0.67]	-	0,CI[-0.67,0.83]
Map4	-0.167,CI[-1.33,1.31]	0.5,CI[-0.88,1.5]	0,CI[-1.17,0.5]
Map5	0,CI[-1.167,0.833]	-0.17,CI[-1.33,0.33]	0.17,CI[-1.17,1]

Table 5.11: Mean Slips Rate in the Task Annotation

Map	Modality	Pattern1	Pattern2	Description1	Description2
Map1	Type	1.67	1.5	1.5	1.17
	Speech	3.17	2.67	1.17	1.83
Map2	Type	1.5	1.167	1.33	1.17
	Speech	2.5	2.17	1.33	1.5
Map3	Type	117	1.33	1.17	1
	Speech	1.17	1.67	1.33	1.17
Map4	Type	1.17	1.33	1	1.17
	Speech	1.83	2	1.17	1
Map5	Type	1.33	1	1.17	1
	Speech	3.5	3	2.17	1.67

The task annotation recorded the slip rate of 'pattern' and 'description', and participants were asked to put 2 annotations for each map. The mean slip rate is from 1 to 4 (Table 5.11). Most comparison results are negative (Table 5.12), showing typing slips less than the speech. Especially the value pattern is verified with significance. The pattern is selected with a drop-down menu in type mode, and the design seems to reduce the slip rate significantly. More significant results are found in task annotation, suggesting that typing could be more applicable than speech in task annotation.

Table 5.12: Type and Speech Comparison on each Map

Map	Pattern1	Pattern2
Map1	-1.5,CI[-3.5,0.33]	*-1.17,CI[-2.67,-0.5]
Map2	*-1,CI[-3.17,-0.33]	-1,CI[-3.5,0]
Map3	0,CI[-0.67,0.33]	-0.33,CI[-3.66,0.5]
Map4	-0.67,CI[-1.67,0.17]	-0.667,CI[-2.7,0.167]
Map5	*-2.17,CI[-5.17,-0.5]	*-2,CI[-5.33,-0.33]
Map	**Description1**	**Description2**
Map1	0.33,CI[-0.79,0.67]	*-0.667,CI[-1.5,-0.33]
Map2	0,CI[-0.67,0.33]	-0.333,CI[-1.44,0.17]
Map3	-0.167,CI[-0.83,0.17]	-0.167,CI[-0.83,0]
Map4	-0.167,CI[-0.83,0]	0.167,CI[0,0.5]
Map5	-1,CI[-5.77,0.17]	-0.667,CI[-3.33,0]

Table 5.13 shows comparisons between different background groups. Two modalities were not compared because the initial sample is too small to bootstrap. Overall, female, non-native speakers, and people who have metadata experience slip more times on place, alternative place, topic, and description than its opposite. Participants with high slip rates may have a stricter standard on their input than those with low slip rates and thus modify their results more times.

Table 5.13: The slip rate under different background groups (Task Element)

Background	**Place**	**Alternative Place**	**Topic**
Gender	-0.14,CI[-0.86,0]	-0.14,CI[-0.71,0]	-0.51,CI[-2.43,0.2]
Native Speaker	-0.1,CI[-0.6,0]	-0.1,CI[-0.6,0]	-0.6,CI[-2.1,-0.2]
Metadata Experience	0.14,CI[0,0.43]	0.14,CI[0,0.43]	0.51,CI[-0.26,2.23]
Background	**Description**	**Start Time**	**End Time**
Gender	-0.97,CI[-3.36,0.2]	0.11,CI[-0.99,0.6]	0.06,CI[-0.43,0.4]
Native Speaker	-0.2,CI[-2.42,0.5]	0.2,CI[-0.8,0.8]	-0.2,CI[-0.7,-0.1]
Metadata Experience	0.97,CI[-0.29,3.25]	0.23,CI[-0.4,1]	-0.06,CI[-0.6,0.29]

Table 5.14 shows comparisons of different background groups. The values turn out in different signs and show an inconclusive result. Besides, all discrepancy is not significant, implying the background influence is little.

Table 5.14: The slip rate under different background groups (Task Annotation)

Background	**Pattern1**	**Pattern2**
Gender	-0.03,CI[-0.94,0.46]	0.17,CI[-1.31,1]
Native Speaker	-0.5,CI[-1.2,-0.3]	0.6,CI[-1.2,1.8]
Metadata Experience	0.37,CI[-0.26,1]	0.51,CI[-0.2,1.66]
Background	**Description1**	**Description2**
Gender	0.2,CI[0,0.4]	-
Native Speaker	-0.1,CI[-0.5,0]	-
Metadata Experience	-0.1,CI[-0.5,0]	-

5.5 Accuracy

Accuracy measure was solely applied to the speech, and the value is obtained from API. The value is the confidence of correctness and ranges from 0 to 1. Table 5.15 indicates that the task annotation's mean accuracy is from 0.17 to 0.59, and the mean accuracy in the task element is from 0.42 to 0.92. The significant discrepancy of mean value results from the mechanism of speech recognition. Inputs in the task element are short words, while the task annotation records long sentences, which increases the recognition error.

Table 5.17 shows comparisons between different background groups of the

Table 5.15: Mean Accuracy of the Task Annotation

Map	Pattern1	Pattern2	Description1	Description2
Map1	0.59	0.49	0.59	0.49
Map2	0.52	0.54	0.52	0.54
Map3	0.61	0.39	0.61	0.39
Map4	0.17	0.53	0.17	0.53
Map5	0.30	0.29	0.30	0.29

Table 5.16: Mean Accuracy of the Task Element

Map	Place	Alternative Place	Topic	Description	Start Time	End Time
Map1	0.75	0.72	0.69	0.84	0.47	0.85
Map2	0.73	0.70	0.67	0.82	0.91	0.66
Map3	0.53	0.74	0.77	0.81	0.42	0.53
Map4	0.74	0.56	0.73	0.86	0.53	0.77
Map5	0.82	0.78	0.84	0.74	0.65	0.86

task element. No native speaker in the task element, so the comparison of the group is omitted. Among background factors, we found that females get better results than males. The effect size values on those significant results are larger than 1, which implies a considerable difference between females and males. People without metadata experiences generate higher accuracy than those with experiences.

Table 5.18 shows comparisons of different background groups in the task annotation; the comparison adopts the median value from maps. The gender has a substantial influence on accuracy. The female has significantly higher accuracy than the male, and the effect size is larger than 1, which implies a considerable difference between the two groups. The native speaker groups and metadata experience groups get the non-significant result. However, we

Table 5.17: The Accuracy under different background groups (Task Element)

Background	Place	Alternative Place	Topic
Gender	0.09,CI[-0.09,0.41]*	-0.07,CI[-0.27,0.02]	0.14,CI[-0.04,0.43]
Metadata Experience	-0.17,CI[-0.43,-0.07]	-0.03,CI[-0.25,0.03]	-0.1,CI[-0.36,0]
Background	**Description**	**Start Time**	**End Time**
Gender	0.13,CI[0.05,0.19]*	0.07,CI[-0.02,0.14]	0.04,CI[-0.58,0.63]
Metadata Experience	-0.01,CI[-0.1,0.05]	0.04,CI[-0.03,0.09]	-0.28,CI[-0.82,0.07]

Table 5.18: The Accuracy under different background groups (Task Annotation)

Background	**Pattern1**	**Pattern2**
Gender	0.7,CI[0.02,0.92] *	0.61,CI[0.37,0.84]*
Native Speaker	0.03,CI[-0.68,0.73]	-0.06,CI[-0.68,0.54]
Metadata Experience	0.31,CI[-0.59,0.9]	-0.05,CI[-0.6,0.47]
Background	**Cont1**	**Cont2**
Gender	0.7,CI[0.02,0.92]*	0.61,CI[0.32,0.84]*
Native Speaker	0.03,CI[-0.88,0.72]	-0.06,CI[-0.67,0.54]
Metadata Experience	0.31,CI[-0.6,0.9]	-0.05,CI[-0.61,0.51]

* Significant

found that non-native speakers and people without metadata experience gain higher accuracy than their contrary group in the second annotation. It reflects that the accuracy could improve when people try more time.

Table 5.18 shows comparisons between different background groups. The gender has a substantial influence on accuracy. The female has higher accuracy than the male, and Hedges' G's effect size is all larger than 1, which implies a considerable difference between the two groups. The difference in native speaker and metadata experience is not significant. However, non-native speakers and people without metadata experience gain higher accuracy than their contrary group in the second annotation. In the slip rate section, we notice a similar tendency of value. It could imply the accuracy can improve when people try more time.

5.6 User Experience Evaluation

User experience was measured with the UEQ-S questionnaire and its analysis tool [37]. The 8 values in the questionnaire, such as support, interesting level, were transformed into three indexes – pragmatic quality (supportive, easy, efficient, clear), hedonic quality (exciting, interesting, inventive, leading edge), and overall. Values between -0.8 and 0.8 represent a neural attitude, values larger than 0.8 represent a positive attitude, and values less than -0.8 represent a negative attitude [37]. The speech is more hedonic than the type but less pragmatic than the type(table 5.19). The benchmark (table 5.20) illustrates value in an understandable context. The type and the speech have the strength of pragmatic or hedonic quality; therefore, this suggests that mixing both modalities can bring both pragmatic and hedonic experience to users.

Table 5.19: User Experience

Item	Speech	Type	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

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.

Table 5.20: Benchmark of user experience

Scale	**Speech**		**Type**	
	Mean	**Benchmark***	**Mean**	**Benchmark***
Pragmatic Quality	-0.125	Bad	1.56	Good
Hedonic Quality	1.42	Above Average	-0.08	Bad
Overall	0.65	Below Average	0.74	Below Average

* Benchmark from the lowest to highest is Lower Border >Bad >Below Average >Above Average >Good >Excellent

5.7 User Interview

The user interviews used 5 questions to understand participants' preferences and opinions on modalities. The first set of questions asks participants the preference for modality, and it turns out that participants prefer typing in both tasks (Figure 5.4). Another set of questions detect participants' feedback on the advantage and disadvantage of the two modalities. In the task element, 3 participants pointed out that the type is efficient (table 5.21). A participant explains that the speech demands users to switch on and off the microphone, which is not efficient for short words. Many participants value the convenience of content editing. 2 participants stated the type modality offers preparation time while the speech does not. The negative feedback on the type is not found from participants. The speech mode has both positive and negative feedback. Participants express speech is interesting, convenient, easy to use. However, many participants concern the accuracy of the speech.

In the task annotation, participants need to describe the pattern they find. Participants tend to the type because it is accurate, having time for preparation, and easy to modify (table 5.22). They do not like the speech

Figure 5.4: Modality Preference

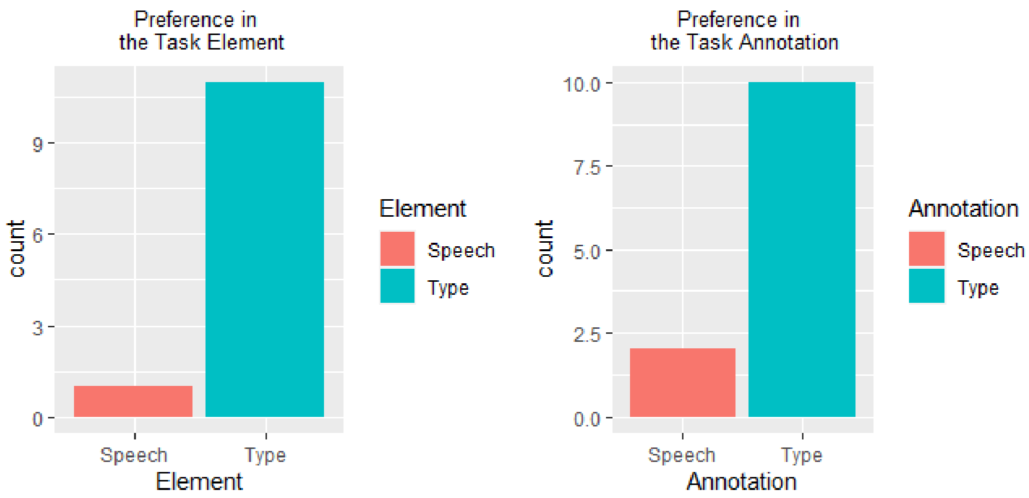


Table 5.21: Count of Feedback (Task Element)

Advantage				Disadvantage	
Type		**Speech**		**Speech**	
Accurate	2	Interesting	2	Inaccurate	4
Familiarity	1	Convenient	1	No time for preparation	1
Efficient	3	Easy to use	1	Edit	1
Edit	5	Modern	1		
Time for preparation	1	Efficient	1		

for its inaccuracy, no time for organizing ideas. Some people express that working with a native language is more comfortable than their second language. The task annotation needs to generate a summary based on observation, and the content is more complicated than the task element. Therefore, more participants emphasize time preparation and are worried about their English ability. During the experiment, the speech recognition mistook what a participant said. For example, when a participant said:' wildfire' the result became 'why do fire.' Although this is a common situation, some participants felt frustrated with their pronunciation.

Table 5.22: Count of Feedback (Task Annotation)

Advantage				Disadvantage	
Type		**Speech**		**Type**	
Accurate	3	Easy to use	1	Inaccurate	4
Familiarity	1	Efficient	2	No time for preparation	2
Edit	2			Edit	1
Time for preparation	2			Not native speakers	3
Easy to use	1			Data Privacy	1

Chapter 6

Conclusion

The results obtained from the user study show the performance of the two modalities are similar. The background factors, including gender, native speakers, and metadata experience, were compared. The result indicates backgrounds have no significant impact on the performance, except that gender influences speech accuracy, and females' speech is clearer to speech recognition than males. Based on the performance result, speech can be considered to create metadata.

The user experience result shows that participants think the type mode is practical, and the speech mode is hedonic. The result of performance and user experience evaluation suggests the potential of the hybrid modality. On one side, the two modalities have similar performance. On the other side, the type mode outweighs pragmatic quality, and the speech mode works better in hedonic quality, and the mix mode can harmonize each modality's disadvantage.

Chapter 7

Limitations and Future Works

The bootstrap tool suggests to have more than 15 in the original sample to generate a robust confidence interval. However, the user study has 12 participants due to the pandemic and the long experiment duration. Furthermore, in some cases we do not have enough English speakers to bootstrap data and can not compare. In the background analysis, we attempt to collect data about participants' using frequency in GIS, Web Map, and speech recognition and examine the influences of use frequency on the accuracy, efficiency, and slip rate. Nevertheless, the data was collected in the category form and cannot be utilized for the analysis. Thus, the difference between trained and untrained participants has remained unknown.

Besides, our choice of maps is based on conditions that no metadata exists in the map, and the topic is relevant to drought. Topics and formats of Maps are alike; maps are in interactive format in the task element and are statistic format in the task annotation. However, the result illustrates a subtle difference on each map; sometimes the type is better than the speech, while sometimes it is not. Similarly, some elements work better with the speech mode, but some don't. For instance, the speech slips less than the type on 'topic' but performs conversely on 'End Time.' Whether the result is random or influenced by other factors can't be explained at the moment. If the factors are known, it reveals more information to a suitable metadata application design.

We found our metadata creation case is easier than the real work. In the background survey, some participants showed they have handled complicated metadata, which takes around a 30 - 60 minutes. Nevertheless, in the experiment, we reduce the metadata creation loading to make participants finish a map within 5 minutes. Whether the complexity of the map can influence participants' preference is another research gap.

We discover that microphone choice and internet quality influence speech recognition accuracy. The two factors ensure the sound receiving quality. However, the pandemic made the experiment go online, and each participant attends the experiment from their home. The experimental environment is not consistent, but this technical issue in the study can be improved.

Finally, Metadata quality is essential when discussing metadata creation; however, this was not in the study's scope. The metadata quality includes completeness, trust, clarity, and levels of detail [40], which would require trained participants and distract the focus of the modality comparison. From the user interview's feedback, many participants point out their main concern toward the speech mode is its inaccurate recognition. They are willing to shift to the speech mode when the accuracy issue is solved. The opinion reflects that metadata quality is the main factor influencing the metadata creation preference. Thus, metadata quality will be a vital direction for speech-based metadata creation topics.

Bibliography

[1] Judie Attard, Fabrizio Orlandi, Simon Scerri, and SörenAuer. 2015. A systematic review of open government data initiatives. *Government Information Quarterly, 32, 4 (October 2015)*, 399–418. DOI:https://doi.org/10.1016/j.giq.2015.07.006

[2] Geospatial Metadata — Federal Geographic Data Committee. Retrieved December27, 2020 from https://www.fgdc.gov/metadata

[3] James Batcheller, Bruce Gittings, and Robert Dunfey. 2009. A Method for Automating Geospatial Dataset Metadata. *Futur. Internet* 1, 1 (2009), 28-46. DOI:https://doi.org/10.3390/fi1010028

[4] James K. Batcheller. 2008. Automating geospatial metadata generation-An integrated data management and documentation approach. *Computers 'I&' Geosciences* 34, 4 (Apr 2008), 387-398. DOI:https://doi.org/10.1016/j.cageo.2007.04.001

[5] Aneta Jadwiga Florczyk, Francisco Javier López-Pellicer, JavierNogueras-Iso, and Francisco JavierZarazaga-Soria. Automatic Generation of Geospatial Metadata for Web Resources. 2016. *Int. J. Spat. Data Infrastructures Res.* 7(2016), 151–172. DOI:https://doi.org/10.2902/1725-0463.2012.07.art8

[6] Sergio Trilles, Laura Díaz, and Joaquín Huerta. 2017. Approach to Facilitating Geospatial Data and Metadata Publication Using a Standard Geoservice. *Geo-Inf.* 6, 5 (2016) 126. DOI:https://doi.org/10.3390/ijgi6050126

[7] Mohsen Kalantari, Abbas Rajabifard, Hamed Olfat, and Ian Williamson. 2014. Geospatial Metadata 2.0 - An approach for Volunteered Geographic Information. *Comput. Environ. Urban Syst.* 48, (2014), 35–48. DOI:https://doi.org/10.1016/j.compenvurbsys.2014.06.005

[8] Masanori Sano, Hideki Sumiyoshi, Masahiro Shibata, and Nobuyuki Yagi. 2005. Generating metadata from acoustic and speech data in live

broadcasting. *ICASSP, IEEE Int. Conf. Acoust. Speech Signal Process. - Proc. II*, (2005), 1145–1148. DOI:https://doi.org/10.1109/ICASSP.2005.1415612

[9] Auriol Degbelo. 2017. Linked Data and Visualization: Two Sides of the Transparency Coin. *ACM Ref.* 8, (Nov, 2017), 1–8. DOI:https://doi.org/10.1145/3152178.3152191

[10] Simon Scheider, Jim Jones, Alber Sánchez, and Carsten Keßler. 2014. *Connecting a Digital Europe Through Location and Place, Lecture Notes in Geoinformation and Cartography*, Encoding and querying historic map content. Springer International Publishing, 251–273. DOI:https://doi.org/10.1007/978-3-319-03611-3_15

[11] Auriol Degbelo. 2020. An ontology design pattern for geovisualization content description. In *Advances in Pattern-based Ontology Engineering, 11th Workshop on Ontology Design and Patterns.*

[12] Song Gao, Krzysztof Janowicz, and Ding Wen Zhang. 2017. Designing a Map Legend Ontology For Searching Map Content. *Advances in Ontology Design and Patterns* 32, (2017), 119–130. DOI:https://doi.org/10.3233/978-1-61499-826-6-119

[13] R Karam, R K Chamoun, and R Laurini. 2010. Integrating many legends through ontology for cartographic symbols. *Proc. 6th Int. Conf.* (2010).

[14] Rainer Simon, Bernhard Haslhofer, Werner Robitza, and Elaheh Momeni. 2011. Semantically augmented annotations in digitized map collections. *Proc. ACM/IEEE Jt. Conf. Digit. Libr.* (2011), 199–202. DOI:https://doi.org/10.1145/1998076.1998114

[15] Carral D., Scheider S., Janowicz K., Vardeman C., Krisnadhi A.A., Hitzler P. ,2013 An Ontology Design Pattern for Cartographic Map Scaling. *The Semantic Web: Semantics and Big Data. ESWC 2013. Computer Science*, 77-93. Springer, Berlin, Heidelberg. DOI:https://doi.org/10.1007/978-3-642-38288-8_6

[16] Eleni G kadolou and Emmanuel Stefanakis. 2013. A formal ontology for historical maps. In *Proceedings of the 26th International Cartographic Conference. j. Cartogr. Geogr. inf.* 63, 111. Retrieved from https://www.academia.edu/2961943/A_formal_ontology_for_historical_maps

[17] A Dingli, F Ciravegna, and Y Wilks. 2003. Automatic Semantic Annotation using Unsupervised Information Extraction and Integration.

In *Proceedings of the Knowledge Markup and Semantic Annotation.* Retrieved from https://www.researchgate.net/publication/243602775_Automatic_Semantic_Annotation_using_Unsupervised_Information_Extraction_and_Integration

[18] Lawrence Reeve and Hyoil Han. 2005. Survey of Semantic Annotation Platforms. In *Proceedings of the 2005 ACM symposium on Applied computing* Association for Computing Machinery, New York, NY 1634–1638. https://doi.org/10.1145/1066677.1067049

[19] Siegfried Handschuh Tag, Rudi Studer Korreferent, and Christof Weinhardt. 2005. Creating Ontology-based Metadata by Annotation for the Semantic Web. *Integr. Vlsi J.* (2005). https://doi.org/10.5445/IR/1000003776

[20] Annika Hinze, Ralf Heese, Alexa Schlegel, Adrian Paschke, 2019. Manual semantic annotations: User evaluation of interface and interaction designs. Journal of Web Semantics, 58, (2019), 100516. DOI:https://doi.org/10.1016/j.websem.2019.100516

[21] Tabea Tietz, Joscha Jäger, Jörg Waitelonis, and Harald Sack. 2016. Semantic annotation and information visualization for blogposts with refer. In *Proceeding of the 2th Int. Workshop on Visualization and Interaction for Ontologies and Linked Data 2016, co-located with ISWC 2016* 1704, (2016), 28–40.

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

[23] JeffreyHeer, Fernanda B.Viégas, and Martin Wattenberg. 2009. Voyagers and voyeurs: Supporting asynchronous collaborative visualization. *Commun.* ACM 52, 1(2009), 87–97. DOI:https://doi.org/10.1145/1435417.1435439

[24] Narges Mahyar, Ali Sarvghad, and MelanieTory. 2012. Note-taking in co-located collaborative visual analytics: Analysis of an observational study. *Information Visualization.* 11, 3(2012), 190–204. DOI:https://doi.org/10.1177/1473871611433713

[25] DCMI: DCMI Schemas. Retrieved December 13, 2020 from https://www.dublincore.org/schemas/

[26] Getting Started - schema.org. Retrieved December 13, 2020 from https://schema.org/docs/gs.html

[27] UK GEMINI. Retrieved December 13, 2020 from https://www.agi.org.uk/agi-groups/standards-committee/uk-gemini

[28] Jean Brodeur, Serena Coetzee, David Danko, Stéphane Garcia, and Jan Hjelmager. 2019. Geographic information metadata—an outlook from the international standardization perspective. ISPRS Int. J. Geo-Information 8, 6(2019).DOI: https://doi.org/10.3390/ijgi8060280

[29] Federal Geographic Data Committee, Retrieved November 14, 2020 from https://www.fgdc.gov/metadata/geospatial-metadata-tools

[30] Open Source Geospatial, Retrieved November 14,2020 fromhttps://wiki.osgeo.org/wiki/Metadata_software

[31] SpeechRecognition. Retrieved December 26, 2020 from https://developer.mozilla.org/en-US/docs/Web/API/SpeechRecognition

[32] SpeechRecognition.lang. Retrieved December 26, 2020 from https://developer.mozilla.org/en-US/docs/Web/API/SpeechRecognition/lang

[33] Web Speech API. Retrieved December 29, 2020 from https://wicg.github.io/speech-api/#speechreco-section

[34] JSON-LD 1.1. Retrieved December 28, 2020 from https://json-ld.org/spec/latest/json-ld/

[35] Kasper Hornbæk. 2011. Some whys and hows of experiments in human-computer interaction. *Foundations and Trends® in Human–Computer Interaction* 5, 4(2011), 299–373. DOI:https://doi.org/10.1561/1100000043

[36] Jeff Sauro and Joseph S.Dumas. 2009. Comparison of three one-question, post-task usability questionnaires.In *Proceeding of the CHI '09: CHI Conference on Human Factors in Computing Systems* ACM Press, New York, NY, 1599–1608. DOI:https://doi.org/10.1145/1518701.1518946

[37] Martin Schrepp, Andreas Hinderks, and Jörg Thomaschewski3. 2017. Design and Evaluation of a Short Version of the User Experience *International Journal of Interactive Multimedia and Artificial Intelligence* 4, (2017), 103–108. DOI:https://doi.org/10.9781/ijimai.2017.09.001

[38] Jonathan Lazar, Jinjuan Heidi Feng, and Harry Hochheiser. *Research methods in human-computer interaction*(2nd. ed), Morgan Kaufmann,Burlington, Massachusetts.

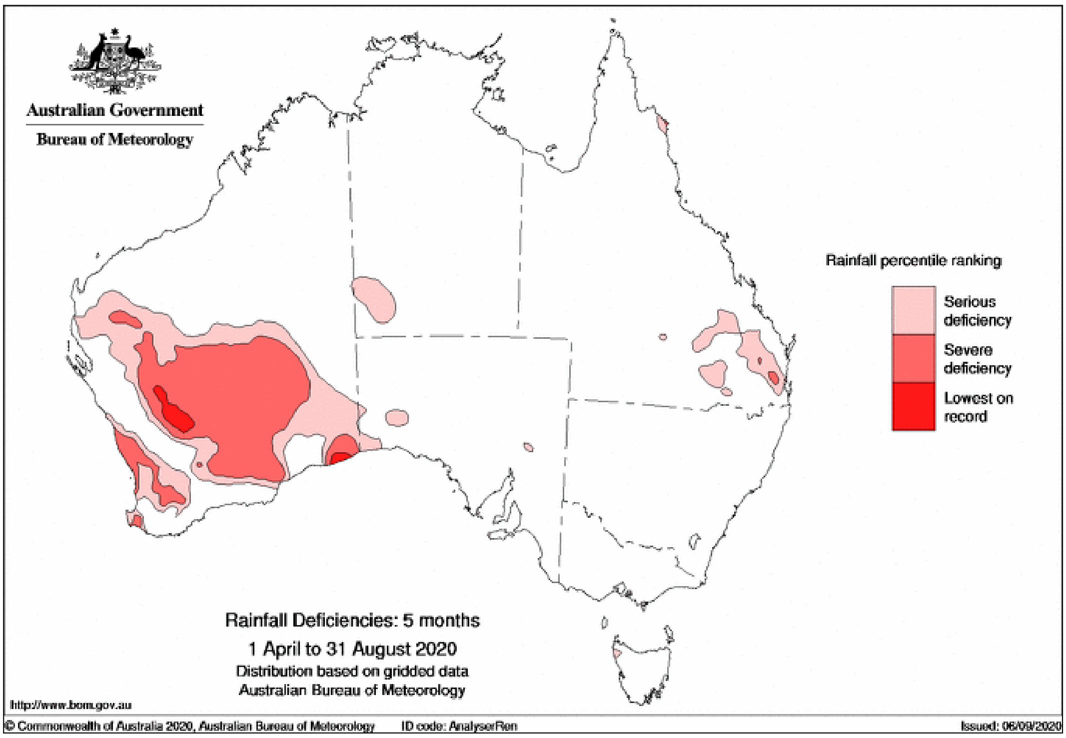
[39] Kris N.Kirby and Danie lGerlanc. 2013. BootES: An R package for bootstrap confidence intervals on effect sizes. *Behav. Res. Methods* 45, 4 (Mar, 2013), 905–927. DOI:https://doi.org/10.3758/s13428-013-0330-5

[40] Tarek Sboui, Mehrdad Salehi, andYvanBedard. 2009. Towards a Quantitative Evaluation of Geospatial Metadata Quality in the Context of Semantic Interoperability. *Spat. Data Qual.* (2009), 31–39. DOI:https://doi.org/10.1201/b10305-5

Appendix A

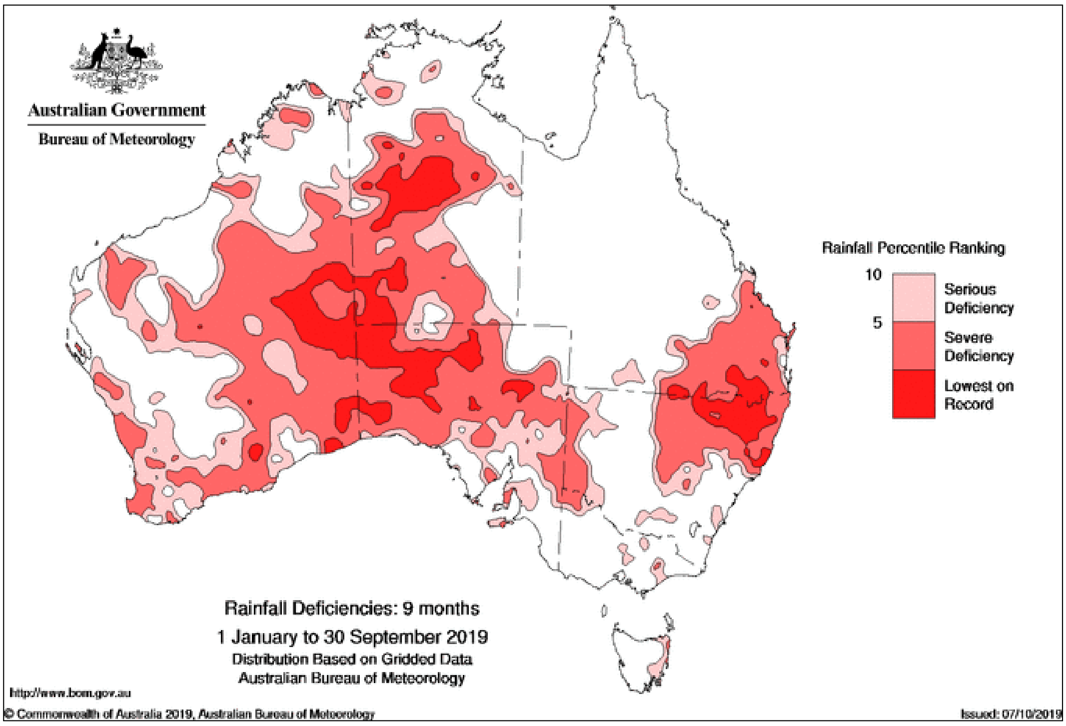
Map for the task annotation

Figure A.1: Map1



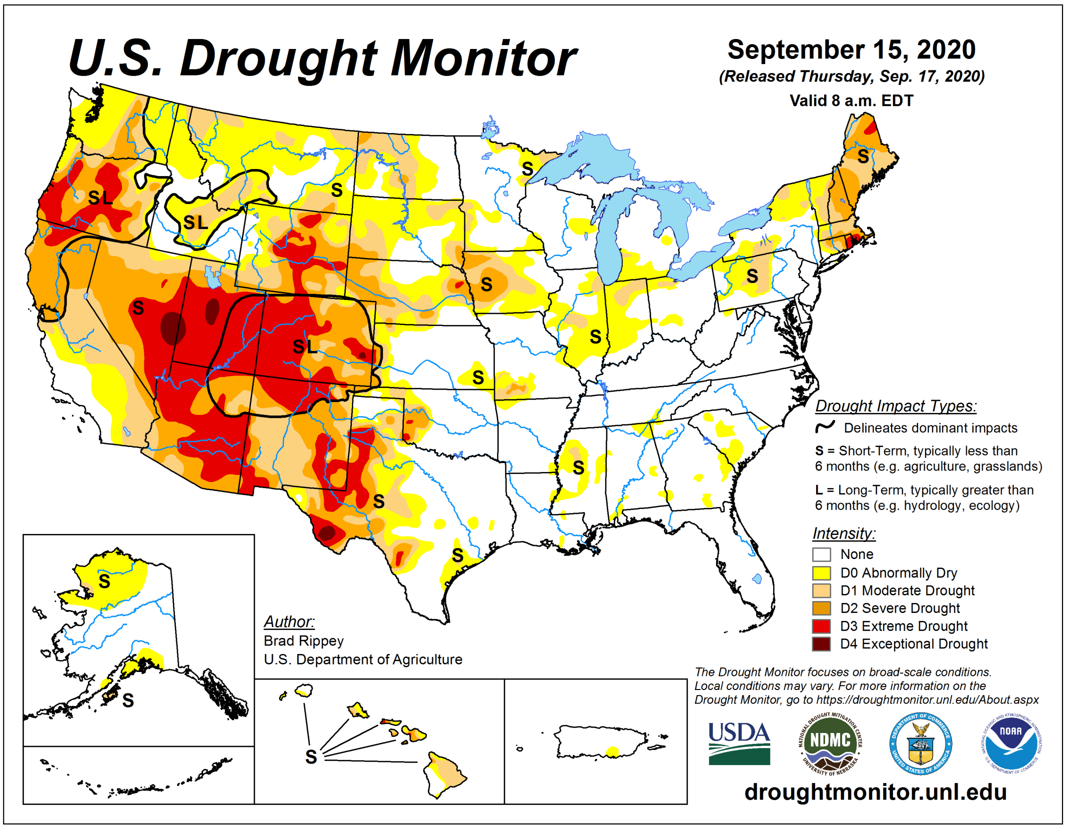
Source: `http://www.bom.gov.au/climate/drought/archive/20200908.drought1.lr.col.gif`

Figure A.2: Map2



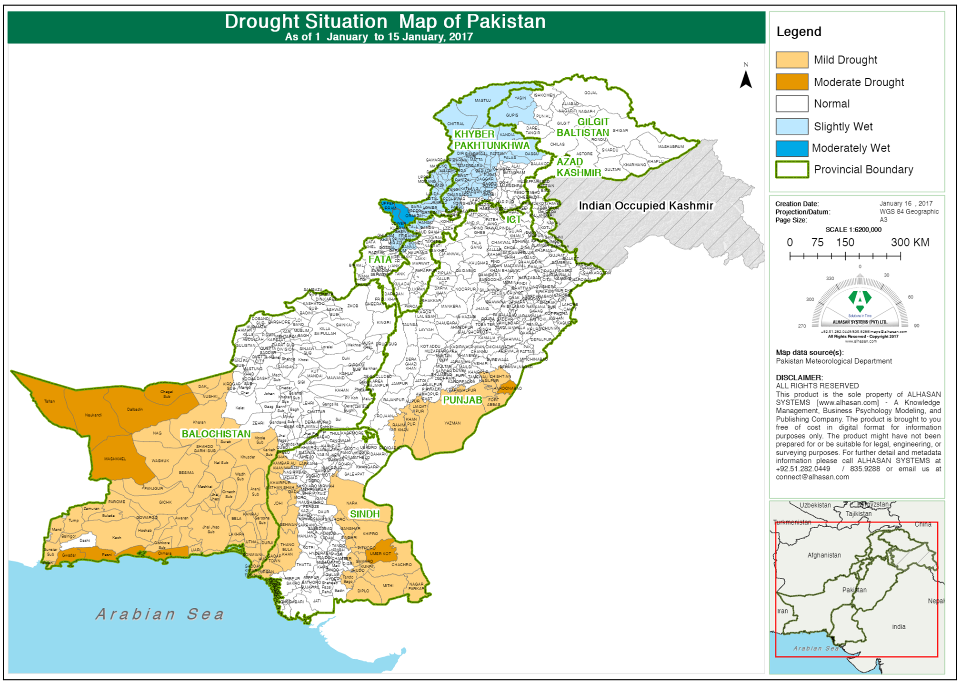
Source: `http://www.bom.gov.au/climate/drought/archive/20191008.drought1.lr.col.gif`

Figure A.3: Map3



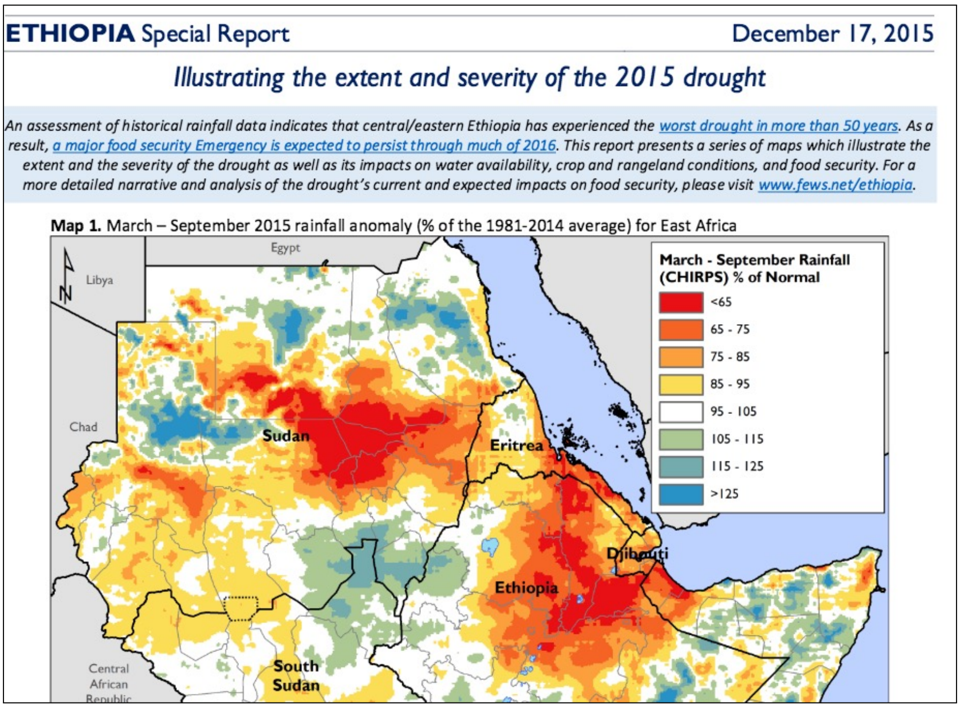
`https://droughtmonitor.unl.edu/data/png/current/current_usdm.png`

Figure A.4: Map4



Source: `https://reliefweb.int/map/pakistan/pakistan-drought-situation-map-pakistan-1-january-15-january-2017.png`

Figure A.5: Map5

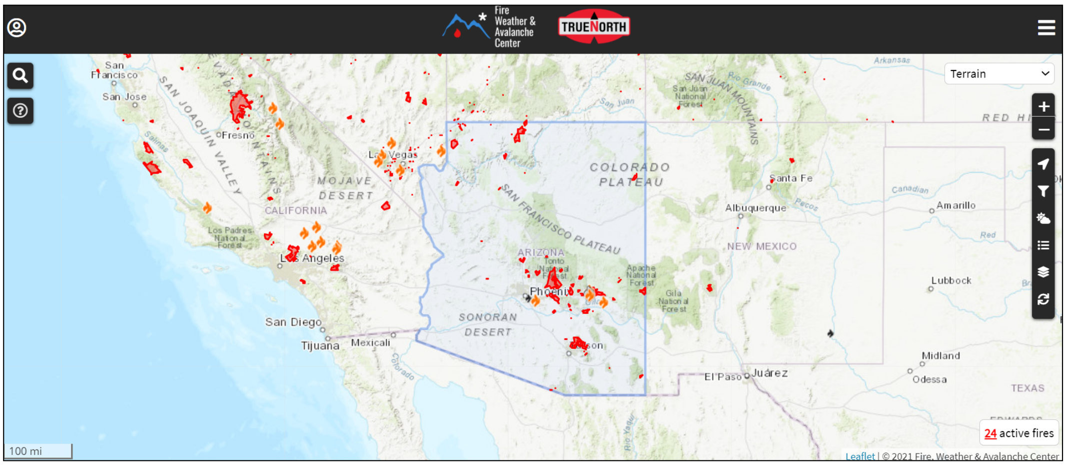


`https://martinplaut.files.wordpress.com/2015/12/drought-horn-of-africa.jpg`

Appendix B

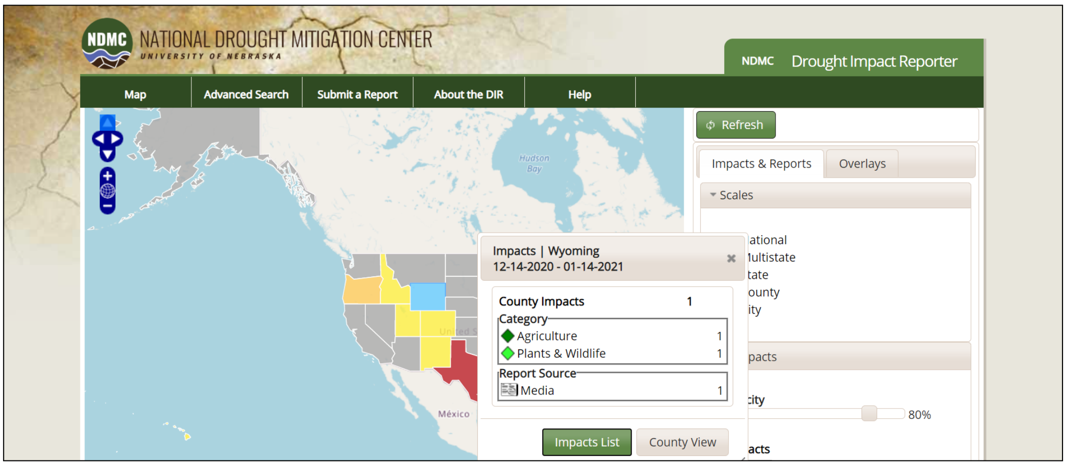
Map for the task element

Figure B.1: Map1



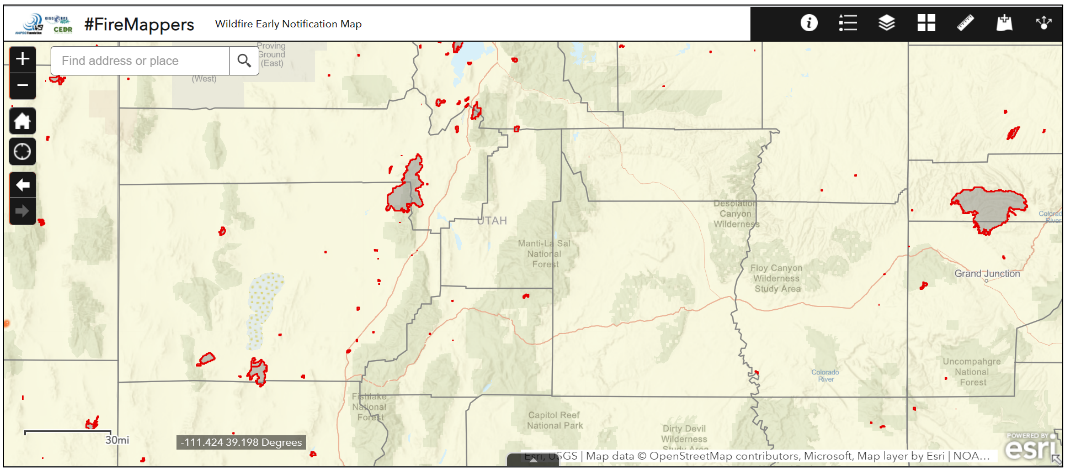
Source: `https://www.fireweatheravalanche.org/fire/state/arizona`

Figure B.2: Map2



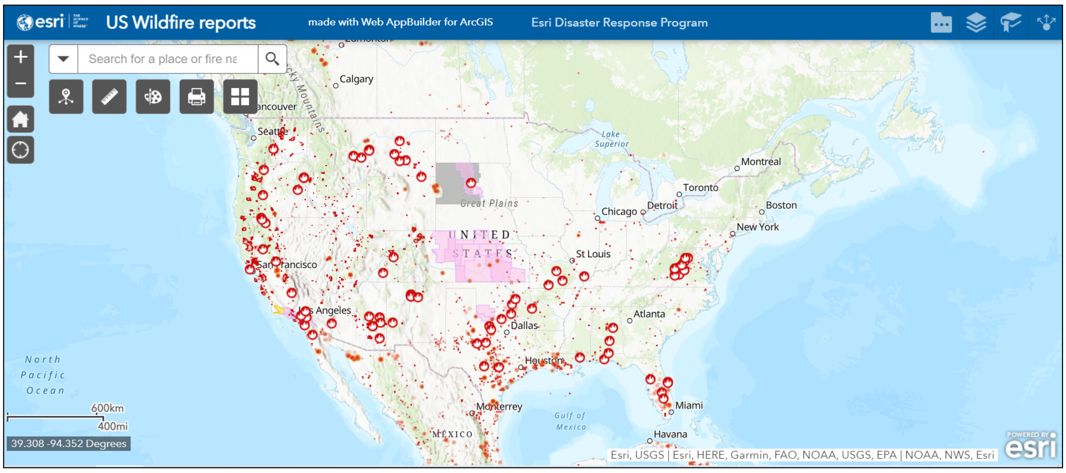
Source: https://droughtreporter.unl.edu/map/

Figure B.3: Map3



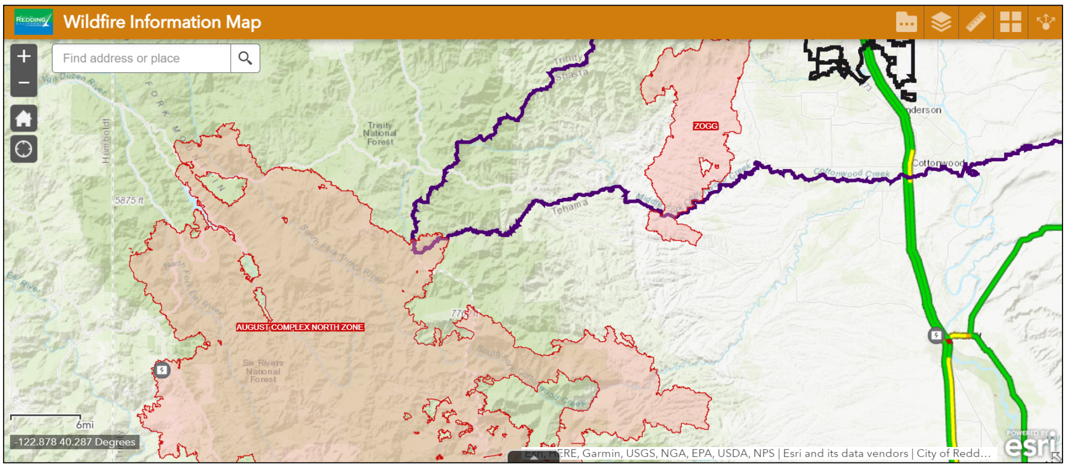
Source: https://napsg.maps.arcgis.com/apps/webappviewer/index.html?id=6dc469279760492d802c7ba6db45ff0e

Figure B.4: Map4



Source: https://disasterresponse.maps.arcgis.com/apps/webappviewer/index.html?id=2ff1677111ae4018ac705fcce7c3312f

Figure B.5: Map5



Source: https://www.arcgis.com/apps/webappviewer/index.html?id=94b379a91e0f47cb91712da22f603d39&extent=-13632561.8405%2C4954358.0786%2C-13623867.1285%2C4960664.1334%2C102100