2023

Distributed Spatial Data Sharing: a new era in sharing spatial data

Majid Hojati
hoja4090@mylaurier.ca

Distributed Spatial Data Sharing: a new era in sharing spatial data

by

Majid Hojati

A thesis
presented to the Wilfrid Laurier University
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Geography and Environmental Studies

Waterloo, Ontario, Canada, 2023

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

Internal-External Member:	Marc Kilgour Adjunct Professor, Department of Systems Design Engineering The University of Waterloo and Professor of Mathematics Wilfrid Laurier University.
External Examiner:	Dr. Emmanuel Stefanakis Professor and Head of Geomatics Engineering, University of Calgary Schulich School of Engineering University of Calgary
Supervisor:	Dr. Steven Roberts Department of Geography & Environmental Studies Wilfrid Laurier University, Waterloo, Ontario, Canada
Co-Advisor:	Dr. Colin Robertson Department of Geography & Environmental Studies Wilfrid Laurier University, Waterloo, Ontario, Canada
Committee Member 3:	Dr. Rob Feick School of Planning University of Waterloo, Waterloo, Ontario, Canada
Committee Member 4:	Dr. Carson Farmer Textile.io Victoria, BC, Canada

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

The advancements in information and communications technology, including the widespread adoption of GPS-based sensors, improvements in computational data processing, and satellite imagery, have resulted in new data sources, stakeholders, and methods of producing, using, and sharing spatial data. Daily, vast amounts of data are produced by individuals interacting with digital content and through automated and semi-automated sensors deployed across the environment. A growing portion of this information contains geographic information directly or indirectly embedded within it. The widespread use of automated smart sensors and an increased variety of georeferenced media resulted in new individual data collectors. This raises a new set of social concerns around individual geopricacy and data ownership. These changes require new approaches to managing, sharing, and processing geographic data. With the appearance of distributed data-sharing technologies, some of these challenges may be addressed. This can be achieved by moving from centralized control and ownership of the data to a more distributed system. In such a system, the individuals are responsible for gathering and controlling access and storing data.

Stepping into the new area of distributed spatial data sharing needs preparations, including developing tools and algorithms to work with spatial data in this new environment efficiently. Peer-to-peer (P2P) networks have become very popular for storing and sharing information in a decentralized approach. However, these networks lack the methods to process spatio-temporal queries. During the first chapter of this research, we propose a new spatio-temporal multi-level tree structure, Distributed Spatio-Temporal Tree (DSTree), which aims to address this problem. DSTree is capable of performing a range of spatio-temporal queries. We also propose a framework that uses blockchain to share a DSTree on the distributed network, and each user can replicate, query, or update it.

Next, we proposed a dynamic k-anonymity algorithm to address geoprivacy concerns in distributed platforms. Individual dynamic control of geoprivacy is one of the primary purposes of the proposed framework introduced in this research. Sharing data within and between organizations can be enhanced by greater trust and transparency offered by distributed or decentralized technologies. Rather than depending on a central authority to manage geographic data, a decentralized framework would provide a fine-grained and transparent sharing capability. Users can also control the precision of shared spatial data with others. They are not limited to third-party algorithms to decide their privacy level and are also not limited to the binary levels of location sharing.

As mentioned earlier, individuals and communities can benefit from distributed spatial data sharing. During the last chapter of this work, we develop an image-sharing platform,

aka harvester safety application, for the Kakisa indigenous community in northern Canada. During this project, we investigate the potential of using a Distributed Spatial Data sharing (DSDS) infrastructure for small-scale data-sharing needs in indigenous communities. We explored the potential use case and challenges and proposed a DSDS architecture to allow users in small communities to share and query their data using DSDS. Looking at the current availability of distributed tools, the sustainable development of such applications needs accessible technology. We need easy-to-use tools to use distributed technologies on community-scale SDS.

In conclusion, distributed technology is in its early stages and requires easy-to-use tools/methods and algorithms to handle, share and query geographic information. Once developed, it will be possible to contrast DSDS against other data systems and thereby evaluate the practical benefit of such systems. A distributed data-sharing platform needs a standard framework to share data between different entities. Just like the first decades of the appearance of the web, these tools need regulations and standards. Such can benefit individuals and small communities in the current chaotic spatial data-sharing environment controlled by the central bodies.

Acknowledgements

There are many people who have been instrumental in my graduate student experience. I would first like to acknowledge the territories I was situated on during my time of the study. I acknowledge during my course work, I was located on the Haldimand Tract, the land promised to the Haudenosaunee of the Six Nations of the Grand River and are within the territory of the Neutral (Attawandaron), Anishinaabe, and the Haudenosaunee peoples. I am grateful to have worked with the Ka'a'gee Tu First Nation within the Dehcho region during my research time.

From the first day I met Colin, he believed in me right off the bat. I appreciate your support during all these years, during our many meetings, and asking what I need to succeed. I have leaned on you for your guidance and expertise, and I couldn't have done this without you.

Sincerest thanks to my amazing committee members, Dr. Steve Roberts, Dr. Rob Feick, and Dr. Carson Farmer, who provided valuable feedback and whose support and encouragement showed they were invested in my success. Their support, guidance, leadership, and collaboration provided me with inspiration and confidence. Thank you for your encouragement, support, and helpful feedback that helped shape my research to be relevant novel, and focused.

Thank you (Mahsi) to the Kakisa community for welcoming me with open arms and allowing me to live and learn from all of you. Thank you to the Band office members (Chief Lloyd, Ruby, and Melaine) who allowed me to hold a two-day workshop in the community.

Thanks to Dr. Alex Latta and Dr. Andrew Spring for your great help and support for the last chapter of this work and for helping me prepare and work with the Kakisa community. During my one-week amazing and memorable trip to NWT with Alex, I learned a lot about community work and social science research.

I am grateful to have such a great circle of friends across Canada. My Friends supported me during my program's confusing, stressful, and joyous times. Lastly, I am thankful to my friends and family. Thank you to my family for your support in many ways.

Dedication

This is dedicated to the one I love.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction to spatial data sharing

How geographic data and information are shared among individuals, communities, and organizations is becoming increasingly significant as geospatial data are utilized increasingly in decision-making and applications used in everyday life. Sharing in general terms means a joint use of resources. The most common use of sharing in geography might relate to sharing data between people, communities, and sectors. A sharing strategy for data can also lead to higher levels of collaboration between researchers and teams and can foster skills through having access to other disciplines' data [145]. Sharing data can also impact the quality of data and enhance data integrity; for example, providing greater transparency to government [129] and corporate [232] processes and being integrated within a comprehensive open science workflow [15, 27].

The last few decades have yielded unprecedented technological changes and advancements. The emergence of Web 2.0 and an increased capability of users to share data on the internet has changed their role from sole users to include being data producers. A transition has also occurred from data directories to catalogs and subsequently to observatories. The emergence of new data sources including the Internet of things (IOT), crowdsourcing initiatives (such as Volunteered Geographic Information (VGI), Public/Participatory Geographic Information System (PGIS/PPGIS), and GeoCollective initiatives), newly deployed Earth Observation satellite systems, and research datasets results in a big geo data movement that affects data sharing models. Emerging software and hardware architectures for the storage, manipulation, and processing of big data (such as digital earth platforms, distributed systems, and web 3.0) are elements of further technological development, which influences data sharing agendas [149].

In the recent data sharing, requirements, and expectations have been greatly influenced

by the growing awareness of data in society, exemplified in geography by the various forms of VGI [108]. While data production has shifted to a decentralized 'prosumer' model, the storage and management of such data have remained centralized. Kotsev et. al [149] in their 2020 review of the INSPIRE platform (EU 2007), an EU initiative to establish an infrastructure for spatial information across Europe marking its 10th year, defined three main routes for the EU's Spatial Data Infrastructure (SDI) post-2020; Two of their highlights are, providing solutions for heterogeneous sources of data and the standardization of a licensing framework. In most current data sharing platforms data must be shared according to some manual, semi-automated, or automated process. The specific mechanism by which data are shared can be ad hoc, or outlined in a data policy that addresses issues of ownership, data rights, and ethics [110]. In the VGI-like data-sharing model users do not have any control over collected data and in most cases, data are controlled and owned by the platform. In such platforms data storage is discrete and fragmented, making it difficult for contributors to exchange information. Such platforms often also lack transparent policies about how data may be re-used; leading to an 'economy' in the buying and selling of user-contributed data [300].

These issues get highlighted when it comes to individual, community, and indigenous level data sharing (discussed in chapter 2). At these scales, unanswered issues such as unclear data ownership, inaccessible VGI platforms due to their Terms of Service (ToS) and policies for some of the communities, lack of transparency, and limited control of the contributed data lead to a well-known power imbalance in the data economy [108].

By moving toward a new era of spatial data collection, integration, and manipulation, researchers have begun to explore a distributed approach to data sharing and processing. In distributed approaches data sharing data contributors are able to control their own contributions and the role of third-party platforms as data controllers decrease. Distributed technology can be at the storage level, such as the InterPlanetary File System (IPFS) [18], at the application level such as Decentralized Applications (DApps) [259], or at the process level, such as Apache Hadoop [71]. In a Distributed Spatial Data Sharing (DSDS) approach, resources are shared and controlled by the network entities. The recent emergence of the distributed web (aka Web 3.0) is based primarily on peer-to-peer (P2P) networks, whereby nodes can communicate directly with each other, without the need for intermediaries. This provides the capability to distribute data via distributed nodes rather than having centralized storage. Much of the emerging research into distributed data sharing is in the medical field, aiming to use distributed data sharing platforms to provide better data transparency and trust in medical research. For instance, Nugent et al. [194] used smart contracts, and distributed data storage to address data manipulation issues common in clinical trials. They showed that immutable distributed data storage

can improve the transparency of data reporting in clinical trials and, is able to capture all aspects of the data that might be subject to manipulation. Chen et al. [36] proposed a framework to store medical information on blockchain and a distributed file system, IPFS, to solve key technical challenges related to confirming rights in the medical system. In the Spatial Data Sharing (SDS) context, some recent studies have aimed to address centralized aspects of data sharing [110], exploring concerns related to privacy, transparency, trust, and data ownership. However, this research remains in its early stages, as distributed sharing platforms need further development and standardization before wider uptake within the GIScience community.

A distributed approach to sharing spatial data might be used to enhance transparency to facilitate a new model of data sharing behaviors. It is typically believed that data transparency and sharing are gained at a cost of privacy, and as data privacy has become a more prominent societal issue, the only response is to impose limitations on data transparency [175]. Such limitations have been raised because of the need for privacy protection against hacked data or providing different levels of ownership and data rights management. However, such limitations can be addressed by technological and methodological advancements in different data sharing infrastructures. There are currently no data sharing platforms that preserve privacy and promote transparency in the context of VGIs or local SDIs. More research is needed to explore the interface between distributed data storage, big geospatial data, and impacts on geospatial data sharing.

In summary, centralizing data storage of user-contributed data is no longer sufficient to meet the relevant needs of data sharing in the big data era [36]. These problems are exacerbated in small communities, where data collection and sharing data over common existing platforms (such as geoportals, web-map services, etc.) has specific legal and policy-related challenges [181]. It is necessary to look at this new era of data sharing in-depth and understand its requirements, technologies, and social aspects to be able to respond to these changes and challenges in order to solve the common issues with the spatial data sharing environment.

1.1 What is spatial data sharing?

El-Sayed Ewis [78] defines spatial data sharing as *"transactions in which individuals, organizations or parts of the organizations obtain access from other individuals, organizations or parts of the organizations to spatial data. These transactions may or may not include financial payment"*. In this definition spatial data are defined as a subset of information that represents some features, attributes, and objects of the world; typically, it includes

both physical and socio-economic phenomena. Goodchild [107] uses the term Geographic Information to mean *"a collection of tuples of the form of* $< x, z >$*, where* x *is a location in space-time and* z *is a set of properties"*. In other words, GI connects properties (attributes, characteristics, variables) to locations and times within the geographic domain. The terms geographic and geospatial are considered synonymous in Goodchild's definition. Defined in this way, GI sharing involves two main actors: a GI user and a GI provider, between which the data sharing transaction happens [108]. In the SDS terminology the term *information community* (IC) is often used, defined as *"a group that shares a common language, a common set of definitions, or a common set of data format standards such that a user within the community can understand a GI Object communicated by a provider within that community"*. So, having a common language in sharing spatial data is an essential part of such data transactions in order for interoperability to be achieved.

1.1.1 Spatial data sharing components

Spatial data sharing has many aspects to be discussed. We can identify several gaps in the field of SDS. These gaps can be identified by looking at the basic model of data sharing and current technological developments, which have potential to solve them. Spatial data sharing has several components which are outlined below and in Figure 1.1.

Technical component

Technical component includes data model and data structures [205], data interoperability, and integration of spatial data sharing platforms with other frameworks such as Digital Earth. A data model is an abstract model that organizes elements of the data. Data structure means a data organization and storage format that enables efficient access to data. Sharing data between different sources is associated with both technical and non-technical inconsistencies and heterogeneity [181]. Such inconsistencies result from the variety in the standards, data specifications, and organizations, which are used by the data provider or owner of the data. These limitations often hinder the effectiveness of data integration by users of different data sets. Data integration facilitates spatial data sharing [181], thus the development of data integration technologies may contribute to sharing. Data integration tools can include conversion tools/scripts between formats, thematic classifications, metadata, and/or new data models that serve as an interchange format (e.g., geojson).

Interoperability and open standards are core to any spatial data infrastructure, as they enable the exchange of data and the use of processing, management, and representation

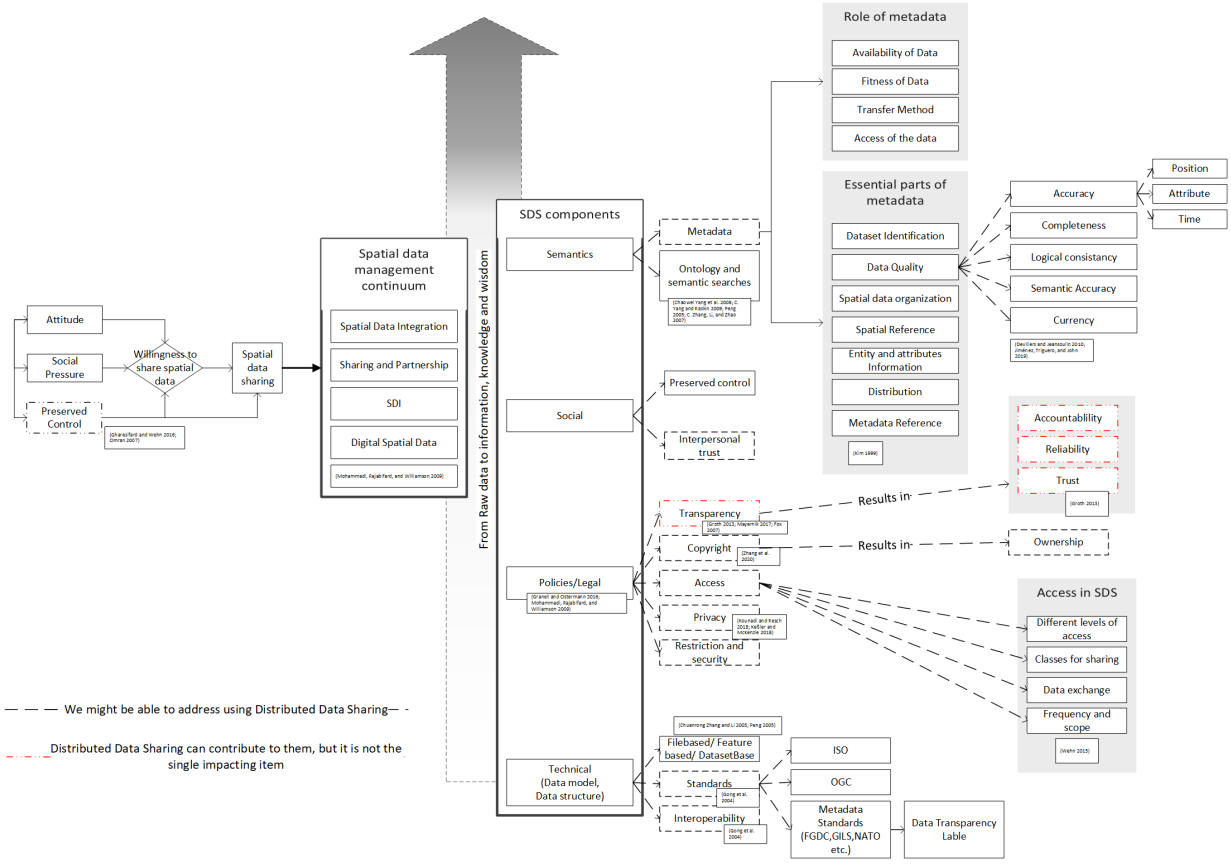


Figure 1.1: Different components of spatial data sharing

services in distributed systems efficiently. The economic benefits of building systems based on standards have been shown in many studies [195]. Looking at current sharing platforms, interoperability is being achieved through the use of three OGC specifications. The current approaches for sharing data at the feature level [297, 205] or other levels of interoperability are mainly designed for a centralized data sharing architecture [104].

Policies/ legal component

This component mainly covers the regulations and norms for using and publishing spatial data. It includes the ownership and copyright of the data [300, 203, 36], data access levels [278], data privacy, and finally data restrictions and security [272, 198, 151]. Data privacy has attracted more attention in recent years; however, this aspect has been studied in a centralized spatial data-sharing framework and needs to be formulated in potential distributed platforms. It is worth mentioning that to preserve geoprivacy a data model must address related issues. For example, a Discrete Global Grid Systems (DGGS) data model can provide some level of geoprivacy by utilizing the hierarchical relations in a grid. Data ownership and copyright data is a key political/legal dimensions of spatial data sharing [300]. Traditional copyright management frameworks are designed for a centralized approach to copyright management which has a single point of failure, and risk of illegal tempering and are not practical for many uses. Data access and ownership is one of the sub-components of data policy. It has some overlaps with the previous component since a data access level control in traditional approaches includes direct access to the database, data directories, and OGC standards [104]. Transparency is another sub-component in this category which can be a result of using the above-mentioned sub-components in a data sharing platform. Another impact of this transparency is open data movements and their direct and indirect impact on transparency, accountability and etc. [86, 175, 111].

Social component

Sharing data is as much a social activity as it is a technical one. The sub-components are 1) preserved control and 2) interpersonal trust. According to the basic model of data sharing developed by Wehn de Montalvo [278], the behavioral intention of an individual to share spatial data is based on three main constructs; attitude towards sharing the data that is linked with expectations about the result of sharing; perceived social pressure or beliefs about the expectations of others and their (dis)approval of data sharing and, finally, the perceived control or perceptions about the absence or presence of specific factors that impede or facilitate data sharing [277]. In general, a combination of more positive

and favorable attitudes, stronger positive social pressure, and greater perceived behavioral control will lead to stronger motivations and intentions to share data. Perceived control over data sharing also has a direct influence on actual data sharing behavior, which shows the importance of this factor. Individual behavioral beliefs about who is responsible for sharing data in an SDI are related to spatial scales [78]. Preserving control over data as needed by an organization can become an issue if the data to be shared is held in a central repository. This issue can also impact the attitude of the individuals willing to share spatial data [276]. For example, farmers may not want to share data with researchers if they believe the data may be shared with the government and result in additional regulation or enforcement. Interpersonal trust is another key to understanding spatial data sharing and developing appropriate technologies. Several studies on knowledge sharing in online communities have recognized the importance of trust as a determinant of the intention to participate in these communities (e.g. Chow and Chan [41], Gharesifard and Wehn [93]). Interpersonal trust in the context of knowledge sharing in online communities is defined by Chen and Hang as *"a degree of belief in good intentions, benevolence, competence, and reliability of members who share knowledge"* [35].

Semantic component

One of the challenges in sharing spatial data is the semantic (i.e., contextual meaning) aspect of sharing data [285, 193, 274, 294]. The heterogeneity of spatial data with different data collection tools, standards, and platforms is a key reason behind the lack of so-called *semantic interoperability*. An example of semantic issues in SDS is the various ways of representing and naming spatial features in different datasets, which, unlike many phenomena, have commonly recognized geographic entities (e.g., roads, rivers, forests). A different approach to overcome such limitations is standardization or database integration [205]. Metadata is one of the sub-sections of this component which is one of the requirements for providing semantic integrity between shared spatial data. In order to provide a comprehensive framework to share spatial data, data semantics at a higher level of sharing data must be considered. This component involves different methods to define geographic information and objects.

1.2 Distributed spatial data sharing

It is clear that technological change is one of the essential drivers of the reconceptualization of the nature of geographic information, from the purposes of sharing to the economic

models that make sharing rewarding to both the provider and the user [108]. These technological capabilities offer great opportunities to revolutionize the way that Geographical Information Science (GIScience) evolves by developing and using high-performance, distributed, and collaborative Geographical Information Systems (GIS) [274]. From a data collection view, technological advances, such as intelligent sensors and high-resolution satellites, are increasing both data quality and quantity [273], resulting in the need to use new approaches for data storage, management, and processing. A distributed approach is one of the possible methods to conduct data storage, management, and statistical analyses for large-scale spatial data [84]. The peer-to-peer (P2P) paradigm is popular for storing and sharing information in a totally decentralized manner[136]. Typically, a P2P file storage system is a distributed environment formed by autonomous peers that operate in an independent manner. Each peer stores a part of the available information and maintains links (indexes) to other peers. P2P systems i) provide a method to distribute the available information to peers, ii) guarantee the retrieval of any information that exists in the system, iii) achieve a reasonable index size for all peers, iv) achieve a reasonable search path for any search performed in the system, v) maintain a low cost for updating peer indexes when peers join or leave, and vi)achieve data and search load balancing, i.e., there are no peers overloaded with stored data, and there are no traffic bottlenecks in the network. Distributed systems coordinate resources and computationally intensive statistical analyses of large geographic datasets between different peers [273]. The transformation from a centralized approach of sharing to a decentralized and then distributed approach has not been fast during past decades. To some extent, they are not even today developed for practical use.

Existing GIS architectures are not able to fulfill the scalability challenges raised by the enormous number of users, data, and heterogeneous sources [159]. Thus, A new Internet GIS architecture that is able to scale up in order to accommodate these needs is essential. A Peer-to-Peer (P2P) spatial architecture, which is fully distributed, can fulfill the needs of scalability of the new architectures [159].

The use of a distributed architecture is also justified due to the most relevant features of P2P systems, such as resistance to censorship, decentralization, and security [234]. Other key advantages of P2P networks are their scalability, their fault tolerance, and their robustness due to the symmetrical nature of peers and self-organization in the face of failures. The above advantages make P2P networks suitable for content distribution and service discovery applications [**?**]. Distributing data among thousands or millions of peers provides a robust system free of restriction from a central authority [234]. Such capabilities also can provide user/creator control over data which can have an impact on the SDS behaviors of individuals as well. Another advantage of using a distributed

approach for a GIS architecture is its performance. Compared to the centralized GIS architecture, distributed GIS architecture has more servers in the system and allows new datasets to be added onto the server experiencing the lowest system load. As a result, the distributed GIS architecture balances system storage resources much better than client-server GIS, where a single server is responsible for managing all datasets in the system. In some architectures, a central directory service distributes users' requests across multiple servers. Thus, it also better balances system load than client-server architectures, where a single server is responsible for dealing with all user requests. However, in such proposed distributed architectures, when the number of users increases, the number of requests sent to the directory service also increases. For this reason, although distributed GIS architectures provide better load balance, centralized directory service is still the system load bottleneck [159].

Having the spatial data sharing components in mind, we can understand the complexities of the data sharing environment. Spatial data sharing is a multi-dimensional phenomenon and we cannot study these components separately. While, we can identify several research gaps in designing a distributed spatial data sharing framework, we don't cover all the components like economical impacts, semantics, and policies. From a technical perspective, we need to be able to perform multi-dimensional data queries in the P2P networks. The first generation of P2P systems, namely file-sharing applications such as BitTorrent, support only keyword lookups and mostly provide no load balancing. The second generation is mainly structured P2P systems supporting basic key-based routing using a distributed hash table (DHT) [159]. DHTs only support exact matching queries, and thus are not practical to support nontrivial applications like geographic or location-based services[159, **?**]. Nowadays, the need for P2P applications with multidimensional data has emerged, motivating research on P2P systems that manage spatial data. The majority of the proposed techniques are based on the distribution of centralized indexes or try to reduce multidimensional data to one dimension and share them on DHT[95].

Geoprivacy is another growing concern in the current spatial data sharing platforms. Fine-grained control over the resolution of shared spatio-temporal data is a basic geoprivacy consideration. In centralized systems, the platforms are usually responsible for obfuscating data. In distributed systems, sharing spatial data between different peers is a challenging task, because an intermediate entity does not exist and nodes in the network are responsible to control access to the data at different resolutions. Thus, we need to explore methods to preserve geoprivacy in the distributed networks too.

The initiatives, requirements, and considerations of using distributed technologies to share spatial data at the community level are one of the sub-components of the social component of spatial data sharing. The new distributed data sharing technologies promise

to give users better control over shared data and fill the ownership and control gap of the current spatial data platforms. However, we need to explore the different aspects of this new environment to find out the requirements and limitations of using these tools in small communities.

1.3 Research objectives

This research aims to identify distributed spatial data sharing as a socio-technological movement that can be integrated into a digital earth framework and contribute to a data sharing process that provides transparency, affords control over data, and can maintain the geoprivacy of users. Considering the challenges and characteristics of current spatial data sharing paradigms, the following research objectives have been identified:

1. Describe how the data sharing environment has evolved and the potential role of distributed data sharing for individual and community level spatial data sharing
2. Explore the methodology and tools for sharing and querying spatial data stored within a distributed file system environment.
3. Explore how decentralized spatial data sharing can addresses issues of control and geoprivacy.

Detailed Sub-questions related to the objectives above:

- *What is distributed spatial data sharing, why is it important and what are its main components?*

  To answer this question using a literature review we identify a new era in spatial data sharing which is heavily impacted by user contributions in data collection. We then find the limitations of the current centralized data-sharing platforms in fulfilling users' needs for more control and ownership of the data and at the end will discuss the capability of distributed data sharing as a solution.

- *How does spatial data query processing work on distributed spatial data? Which standards for sharing spatial data over a distributed data sharing platform need to be developed?*

  To answer this question any kind of existing spatial data model must be able to be shared on a distributed platform, in this case, IPFS, also shared spatial data must be

query-able. To address this capability a distributed spatial index must be developed, which is designed to work on top of IPFS platform. Such an index must be able to capture the locality of the multidimensional space and answer spatio-temporal queries.

- *How we can grant more control to users when they share spatial data using distributed data sharing approaches?*

  To answer this question we will look at social media data and develop an approach to geoprivacy for a distributed web environment. We propose *dynamic k-anonymity* as a model that provides users with fine-grained control over geographic information sharing.

- *What are the requirements of using distributed spatial data sharing for community data sharing?*

  To answer this question we went to a indigenous community and during development of a VGI application we discussed the community's needs for data sovereignty and ownership and discussed distributed data sharing benefits and it's acceptability in the community

A summary of how the research objectives will be organized into chapters and analysis is provided in Table 1.1

1.4 Overview of chapter contents

In Chapter 2 we conduct a literature review to explore the evolution in spatial data sharing. This chapter examines how spatial data-sharing practices have shifted and may be moving towards a more decentralized sharing ecosystem as technologies for a more distributed web mature. We identify this transition as more hybridized forms of data ownership and access control concerns coupled with new distributed systems (aka web 3.0). We also discuss opportunities and barriers to distributed spatial data sharing. Distributed data sharing is inevitable and it is able to benefit the new era of big geographic data collection, however, there must be protocols to share, integrate, and process data shared on the distributed networks. From the identified research gaps in this chapter, we choose three basic aspects of creating a working platform for spatial data sharing. The three aspects are 1) the ability to process queries in distributed networks, 2) the ability to cloak the location when

Chapter	Major issues addressed	Approach utilized
2	*Theoretical:* Understanding the new era of spatial data sharing and social, technological changes that resulted in more individual and community level contribution in spatial data sharing. Identifying distributed data sharing as not only a technological transition, but a transition in the social components of data sharing.	A review study of issues in spatial data sharing, defining distributed spatial data sharing and looking at technology and its attributes used to address the issue and finally identifying the GIScience and geographical research challenges in this area
3	*Methodological:* Query processing of shared spatio-temporal data on the distributed networks and addressing multi-dimensional data storage and retrieval	DSTRee is proposed as a method to address spatio-temporal query processing in the P2P networks based on IPFS and blockchain technology.
4	*Methodological:* The issue of the control and ownership of data in distributed networks.	Developing a geoprivacy framework which is illustrated with a *dynamic k-anonymity model* that links the geographic precision of shared data to social trust within in a social network to facilitate fine-grained user control over the ownership of and access to their data.
5	*Application oriented:* The acceptability of the new spatial data sharing technology in communities and social aspects of distributed spatial data sharing is explored	During meetings with the Kakisa indigenous community in North West Territories and with the purpose of the development of Harvester and safty app we discuss distributed spatial data sharing and look into the requirements to make it applicable for communities to address data sovereignty and ownership in indigenous communities

Table 1.1: How this dissertation approaches the objectives of the proposed research.

sharing spatial data on the distributed network, and 3) socio-technical consideration in using distributed data sharing to share spatial data in small or indigenous communities.

In Chapter 3 we try to answer one of the technical gaps in the distributed data sharing environment. The query processing on distributed networks is one of the fundamental technical requirements of spatial data sharing. We propose *DSTree*, a spatio-temporal indexing data structure for distributed networks, which is capable of answering common spatio-temporal topology queries. We then compare our method with the other existing models and identify the main benefits of using DSTree as a method to share log-based spatial data.

In Chapter 4 we look at one of the sub-components of the policy/legal component of spatial data sharing. Geoprivacy and control are of important concepts for sharing spatial data on different platforms. User participation in different platforms continues to require users to trust corporations to safeguard their personal data. Since these data increasingly contain geographic references that allude to individuals' locations and movements, the need for new approaches to geoprivacy and data sovereignty has grown. In this chapter, we propose a geoprivacy framework that couples two emerging technologies -decentralized data storage and discrete global grid systems- to facilitate fine-grained user control over the ownership of, access to, and map-based representation of their data. In this framework, users' spatio-temporal data are shared through a decentralized system and are represented on a discrete global grid data model at spatial resolutions that correspond to varying degrees of trust between individuals who are exchanging information. In this chapter, we discuss the concept of trust in spatial data-sharing platforms and briefly discuss some of the social aspects of this trust.

In Chapter 5 we explore the impact of distributed data sharing approaches on indigenous data sovereignty. This includes 1) the requirements for using it in small communities, 2) small communities' priorities for the data sovereignty, 3) the intersection between distributed spatial data sharing and data sovereignty

Figure 1.2 A flow chart of the approach to achieve research objectives.

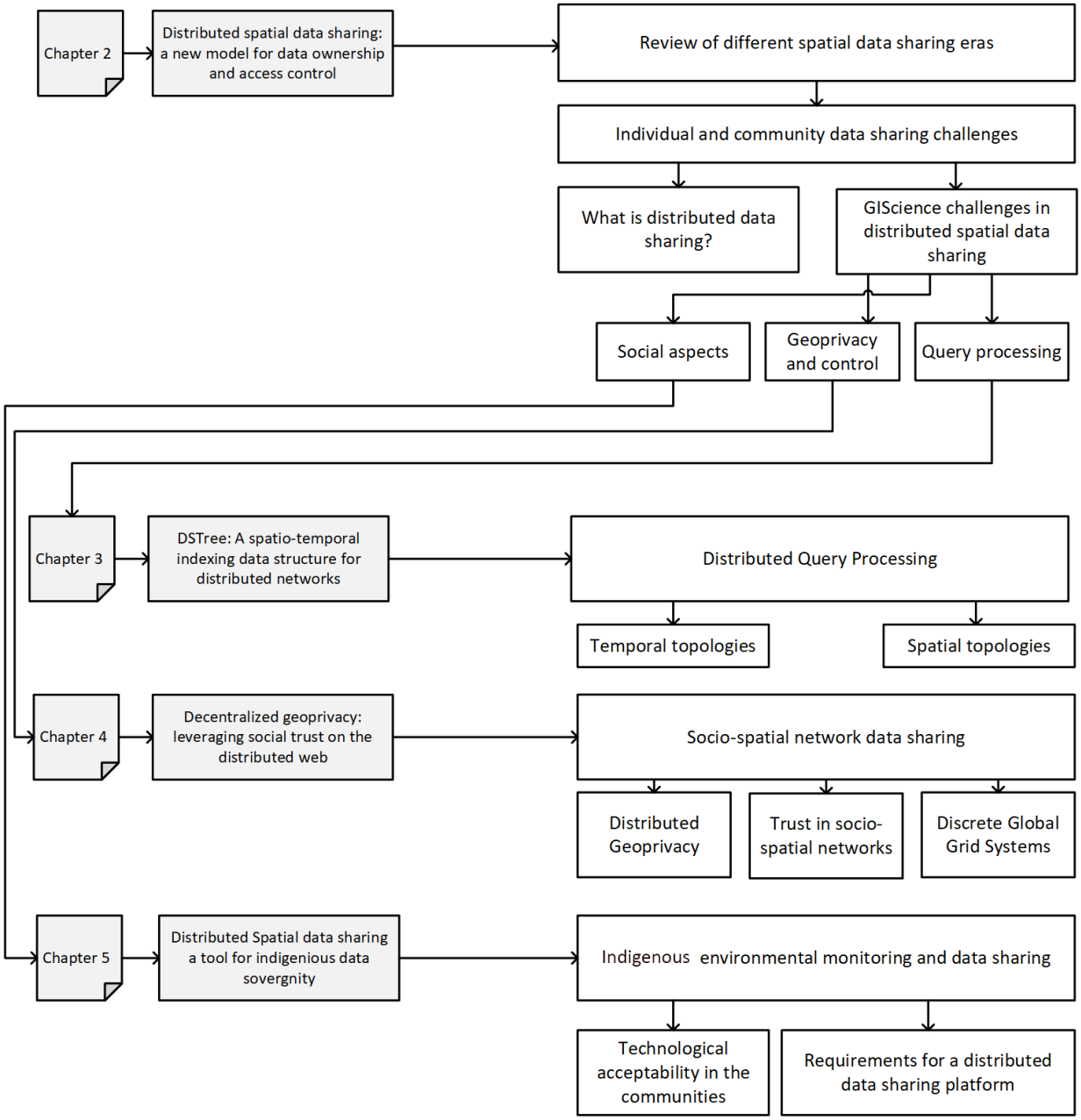


Figure 1.2: Flow chart of the approach to achieve research objectives.

Chapter 2

Distributed spatial data sharing: a new model for data ownership and access control

Abstract With new technologies and broader participation in geospatial data production, new challenges emerge for spatial data sharing. In the past decades, spatial data sharing practices have increasingly been transacted through a handful of privately controlled corporate services. Data collection has changed from that centralized one-way data collection to a more distributed assemblage of individuals sharing and interacting through centralized architectures and control regimes. These changes have mainly resulted from technological and social changes linked to the emergence of web 2.0 and widely available Internet participation tools. Distributed file storage and processing allow users to share resources independently of centralized platforms. This paper examines how spatial data sharing practices have shifted and may be moving towards a more decentralized sharing ecosystem as technologies for a more distributed web mature. We identify this transition as more hybridized forms of data ownership and access control concerns coupled with new distributed systems (e.g. web 3.0). We also discuss opportunities and barriers to distributed spatial data sharing. Distributed data sharing is inevitable and it is able to benefit the new era of big geographic data collection, however, there must be protocols to share, integrate, and process data shared on the distributed networks.

Keywords: spatial data sharing, distributed data sharing, data ownership, blockchain, IPFS, SDI

2.1 Introduction

Each day vast amounts of data are produced by individuals interacting with digital content and through automated and semi-automated sensors deployed across the environment. A growing portion of this information contains geographic information directly or indirectly embedded within it, vastly increasing the use cases for geospatial data [237]. The advancements in information and communications technology, including the widespread adoption of GPS-based sensors, improvements in computational data processing, and satellite imagery, have resulted in new data sources, stakeholders [108], and methods of producing, using, and sharing spatial data. A big portion of these changes includes the growth of user-generated geographic information through the widespread use of automated smart sensors and an increased variety of georeferenced media. It is estimated that around 2% of Tweets have explicit coordinates [225] while a greater portion (55.6%) contain locational data that can be inferred from the content of Tweets (e.g., see [68]). These changes have impacted spatial data sharing in several ways.

First, as Kotsev et al. [149] notes with respect to the post-2020 path for INSPIRE [75], national and pan-national Spatial Data Infrastructures (SDI) have been challenged to accommodate more heterogeneous data sources and a diversity of licensing frameworks. Data integration, access, and rights management have long been challenging in spatial data sharing given the need for users to conform to common standards, policies, and processes. However, the heightened role of individuals as data authors and the increasing role of private sector platforms in controlling geodata have added new dimensions to these challenges [38, 110, 166].

Second, the growing volumes of data that document an individual's transactions, movements, and other aspects of behavior at high spatial-temporal resolutions have spurred concerns over ownership, use, and control. Efforts to enable users to protect their geoprivacy, for example, while contributing personal movement data through COVID-19 exposure notification apps (e.g., see [146, 196]) are one example of this challenge; although this concern has been raised with other forms of activity and passively-generated Volunteered Geographic Information (VGI) for some time [105, 120]. Spatial data governance is less well defined within individual and group contexts and data sharing is typically one-directional from the user to the platform. Data authors thereby lose control over where their data are stored, who can view them, or what purposes they can be used for. Platform licenses and terms of agreements play a key role in limiting contributors' rights as illustrated by [224, 263]. An obvious example is Google Maps where Google owns, controls access to, and resells data that users contribute both actively (e.g., editing places on the map) and passively (e.g., cell phone movement) [307].

While data production has shifted to a decentralized 'prosumer' model, the storage, management and controlling access of such data have remained centralized; leading to a well-known power imbalance in the data economy and requiring a form of social trust between users and the platforms they interact with [108]. To address these concerns, researchers have begun to explore approaches to data sharing where control over data access and use is distributed to individual data authors and owners. From a social perspective, this *Distributed Spatial Data Sharing (DSDS)* environment recognizes existing power imbalances between users and platforms in the data economy [108, 149] and the need to operationalize more dynamic forms of social trust between data authors, other users on a network, and platforms. Technologies to enable DSDS include many of the software and hardware architectures developed for storing and processing big data (e.g., digital earth platforms, data spaces and etc. [128, 166]). Distributed technology can be at the storage level, such as the InterPlanetary File System (IPFS), at the application level such as Decentralized Applications (DApps), or at the process level, such as Apache Hadoop [71]. In a DSDS approach, resources are shared and controlled by the network entities. The recent emergence of the distributed web (aka Web 3.0) is based primarily on peer-to-peer networks, whereby nodes can communicate directly with each other, without the need for intermediaries. This provides the capability to distribute data via distributed nodes rather than having centralized storage.

In summary, the centralized data storage of user-contributed data is no longer sufficient to meet the needs of data sharing in the big data era [36]. In addition, current data storage is vulnerable to data loss when data centers are damaged or hacked. Such platforms often also lack transparent policies about how data may be re-used; leading to an 'economy' in the buying and selling of user-contributed data [300]. The trajectory from fully centralized spatial data storage, distribution, and manipulation (controlled by a government and/or corporation) to distributed individual data sharing remains technically challenging. These problems are exacerbated for smaller organizations, as the collection and sharing of data over common existing platforms (such as geoportals, web-map services, etc.) entail specific legal and policy-related challenges [181]. In this paper, we will identify what situations or use cases DSDS is suited for and identify GIScience research challenges in realizing DSDS as a solution to the current problems of SDS.

The aims of this paper are as follows:

1. A review of current models of SDS and their challenges
2. Examining how DSDS can provide potential solutions for addressing ownership and control challenges in SDS

3. Identifying GIScience research challenges in the implementation of DSDS

To move toward the above goals, we will first look at the different models of SDS from past to present and describe how SDS is transforming from a centralized data production, control, and processing model toward a more distributed model.

2.2 SDS models

One way to conceptualize the changes taking place in how we share spatial data is through the shifting roles of individuals, institutions, governments, and enterprises (private sector), acting as data producers, data controllers, and data users [138]. Data producers are the individuals and entities that generate data. Data controllers are platform owners, data intermediaries, or license holders (see Figure 2.1). These roles overlap and interact fluidly [72]. For example in the Keating et al. [138] model the role of data providers is defined as stewardship of data and data managers are responsible for quality assurance, metadata, data access control, and delivery. However, this classification is not always valid in cases such as VGI. Each SDS stakeholder, including individuals, governmental entities, enterprise entities, the community of interest, and hybrid stakeholders can be assigned to each of the above roles. We examine how large-scale shifts in the relative proportion of actors and roles have changed SDS's nature in recent decades, highlighting three main eras: centralized, user-centric, and distributed SDS eras. Classifying these eras help us to identify how SDS practices and needs are changing and what types of data are best suited for DSDS.

2.2.1 Centralized and user-centric SDS models

There are two main approaches to SDS that have occurred over the past several decades. Figure 2.2 contrasts the main types of data flows between data producers, controllers, and users. Figure 2.2 (a) depicts the centralized model, where data sharing happens primarily between governmental bodies and corporations, who act both as data producers and as data controllers and mainly covers SDS environment between the 1980s and mid-1990s [108]. In the centralized model sensors and tools are only accessible to the government and enterprise sector, controllers and users are the same entities and individuals only use a small portion of available data. In contrast, Figure 2.2 (b) portrays data flows in the current environment where individual users and communities of interest (e.g., NGOs, non-profit groups, universities with hybrid stakeholder model) play more substantial roles in authoring, controlling, and using more heterogeneous sources and forms of spatial data.

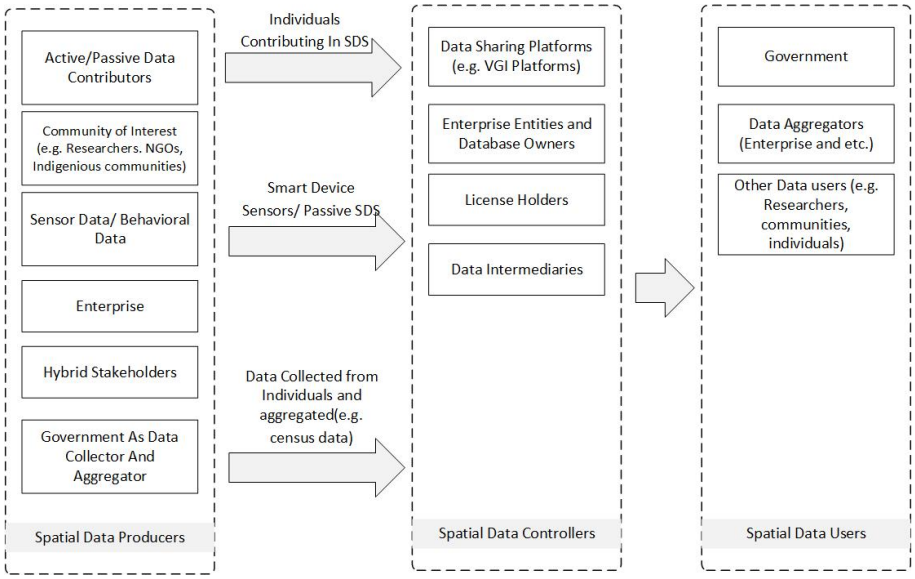
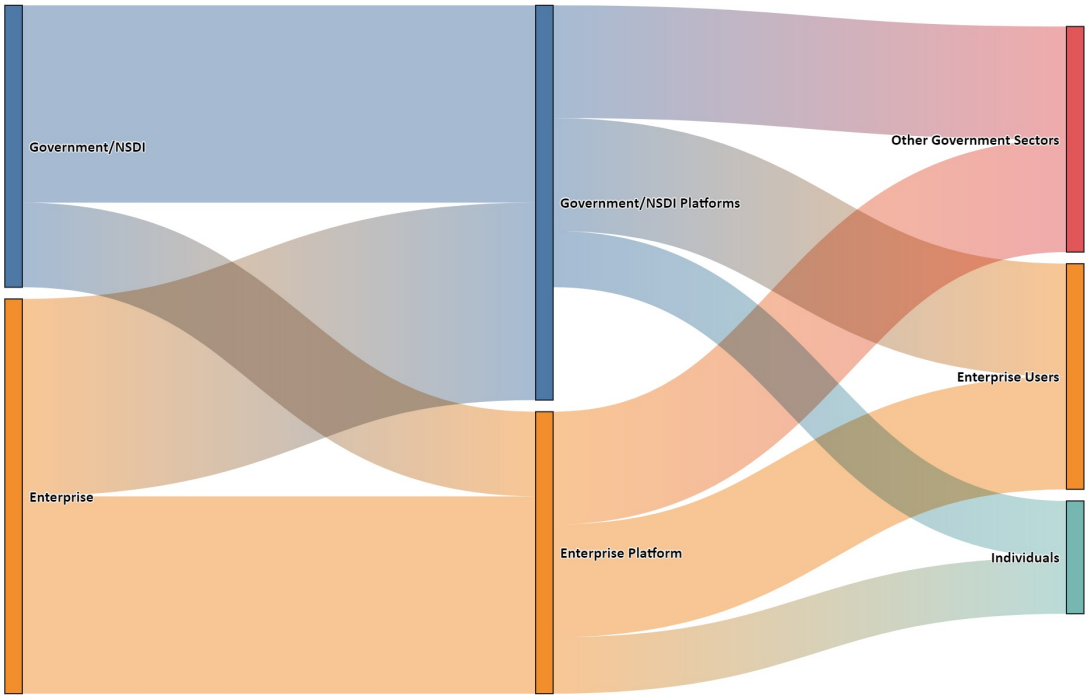
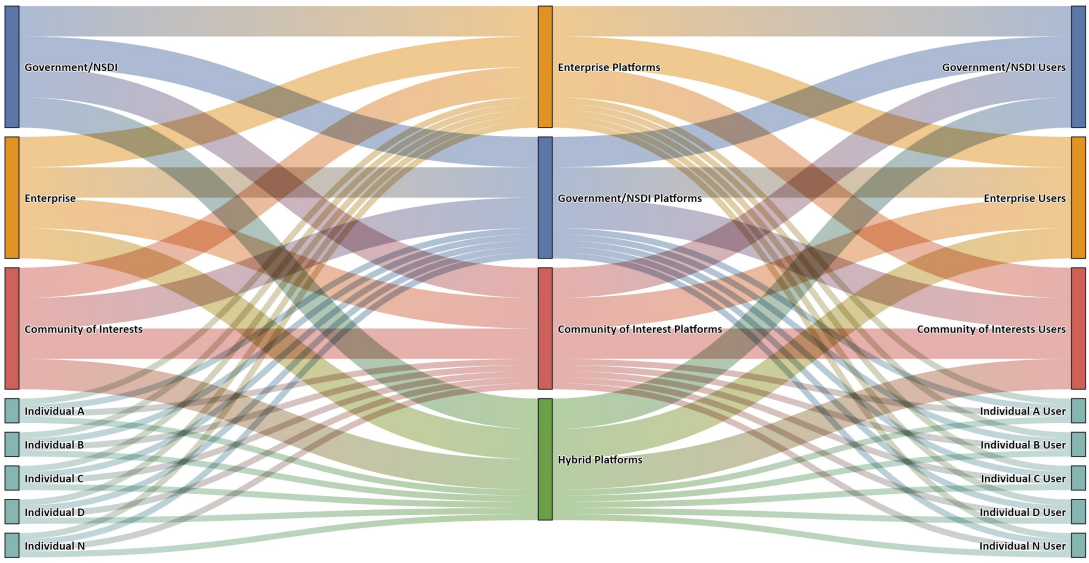


Figure 2.1: Role Players in SDS as data producers, controllers, and users

The increased use of digital data platforms over the past two decades [108] led to the currently dominant model of SDS that we refer to here as a *User-centric SDS model* [44, 118, 61]. The changing Internet environment, especially the spatial dimension and mapping capabilities, decreased the role of governments and third-party companies as data producers and individual users started to produce and use more spatial data [38]. The expanded role of citizens as data producers facilitated by corporate-controlled platforms has increased the relative power of platform owners as arbiters of SDS. In this model, new players such as communities of interest and individuals have emerged as important producers of spatial data (left of Figure 2.2 (b)). In this work grassroot groups, Indigenous peoples, as well as academic and non-profit groups, are all classified as communities of interest. Data access policies vary between open (public) and paid in such groups and are limited by different types of licenses. Data controllers remain governments, corporations, and some community of interest groups that own data sharing platforms (center of Figure 2.2 (b)), while users can be any of the mentioned groups (right of Figure 2.2 (b)). In this model, there are hybrid platforms and licenses between enterprises and individuals (e.g. google maps reviews and photos) or university-maintained repositories in which users and project teams share data and define reuse terms, and accompanying metadata includes information regarding their authorship. In the latter example, the rights are more in favor of the data producers but in the former, the users need to acknowledge enterprise TOS



(a) Data Flow in Central SDS model. The main role players in this model are enterprise and government



(b) Data Flow in user-centric SDS models. In this model, individual-level data production increases, however, their rights over contributed data are controlled by the platforms

Figure 2.2: Estimated data flow in Centralized and user-centric data models. The vertical bars correspond to the data producer, controller, and user and the data flow is from left to right.

to be able to contribute. In this model, governments are only responsible for standardizing data sharing protocols and, in conjunction with corporations, for the collection of large-scale and framework data [108]. In contrast, user-collected data can be of a higher temporal scale than the government-collected data, thereby filling gaps and capturing phenomena (e.g., public sentiment) not covered by traditional data collection practices (e.g., census data)[108]. From the perspective of data production, it is clear that the number of contributors has grown significantly [227, 190, 103]. The use of spatial data has also changed from governmental uses to more individual and community approaches (e.g., citizen science projects) and location-based services (see [201] for example). In this model, the individuals have limited rights to control their data and the relative power of platforms has increased. As such, data ownership and data control topics have become a concern among users, communities, and even the platform owners (see [38]), especially with several high-profile data breaches that commercialize user data. In the case of the community of interest, depending on the use case users are able to define the licenses and right to use, but they need to also accept the platform's TOS.

2.2.2 When centralized data sharing fails

Despite the current advancements in the centralized spatial data-sharing platform, there are unsolved challenges that are discussed in this section. Šumarada [267] has identified intellectual property rights to protect producers' right to spatial data as one of the major legal issues related to geographic data sharing. Policies and licensing issues are changing as the actors in SDS are changing [149]. Take VGI tools as an example, Scassa [224], similar to other authors like Cho [38] and Longhorn et al. [164], look at legal issues with VGI platforms and argue that data hosted on a VGI website is considered *compiled* for the purposes of copyright law. They consider data compilation as *work* in copyright law and claim that, for example, in Canada and the US only original data is protected by the law, while in the EU the protection might be broader as they have database protection laws. In addition, these legal aspects are from the viewpoint of VGI platform owners, not from a data collector's view. They suggest that VGI platform owners can use license agreements as a fundamental tool to control intellectual property rights [133], and limit the rights of users who are contributing on VGI platforms in order to avoid any future legal claims by contributors [224]. Tracing the heritage of data in VGI platforms and hence the copyright ownership can be difficult [38]. Such difficulties are raised because *confected* data might lose their copyright due to difficulties in tracing all previous owners of the data. Scassa [224] describes this problem as a *Wiki* effect, wherein multiple sources have contributed to a dataset. Granell and Ostermann [110] urge that there must be mechanisms to ensure

the privacy of gathered data and that data governance must clearly identify who owns the data and define time limits for which to use them. Michener et al. [179] mention a lack of technical implementations for acknowledging data contributions, and a dearth of easy-to-use tools for accessing data, their conversion, and analysis as some of the barriers to open data. In such situations not having a clear owner associated with the data as a result can impact user contributions and raise trust concerns when it comes to contributing to the platforms.

A lack of transparency and trust in centralized data infrastructure could be a key factor in preventing the true realization of participatory government models [230]. Trust is defined as an answer to the uncertainties in social life, which result in entities trusting others in things they cannot control. Such as trusting the government to provide certain rights. Second, trust is related to risk. By trusting an entity to provide a service for us, we accept that the entity may disappoint or harm us. Third, trust is future-oriented and is about expected actions in the future [262]. Ewis et al. [79] argue that trust is a motivational factor in data sharing behaviours and having control over one's data can increase that individual's trust and willingness to share data with an organization. Similarly, Wehn de Montalvo [276] uses the term perceived control to represent an individual's willingness to impede or facilitate data sharing. Users by using centralized platforms should trust them and with lack of transparency in their systems and how data are being stored and used adding to the distrust between contributors and platforms.

Take indigenous data sovereignty as another example. Due to historical events sovereignty communities require more control over their data [24]. Indigenous data sovereignty involves data locality [55] and access hierarchy inside and outside of the community. Indigenous knowledge needs to reside and be controlled from within the community. *Data sovereignty* has been defined as the management of information in a way that aligns with the laws, practices, and customs of the nation-state in which it is located [55]. This concept has been extended in the context of SDS to self-governing groups or nations within states (e.g., Indigenous communities). Such groups have distinct laws and practices that need to be applied to the SDS environment (see Find-able, Accessible, Inter-operable and usable (FAIR) and Collective benefits, Authority to control, Responsibility and Ethics (CARE) principles). Following these principles, in local communities, information must be available, accessible, and open to all the participants in the community (aka knowledge holders) and data access must be selective to users outside of the community [263]. In indigenous data sharing approaches, volunteered data collection need to be treated as open data for the community members and closed data for the non-community members. Inside such data sharing requirements prohibit these communities from contributing to public platforms like OSM and as a result interest is growing in tools and methods that will permit the data

sovereignty concept to extend further to individuals as well as communities.

Sharing personal information or social media contributions is another example. In the current user-centric model, a range of public data processing practice regulations are used to control how personal information is shared and represented to users. Despite such regulations, Granell et al. [110], Alessi et al. [3] and Camenisch et al. [30] argue that individuals have limited control over their data and identify the need for easy-to-use tools to control access to personal data and data governance policies that clearly delineate who owns the data and the time limits for using them. Despite the advances these platforms provide, as Camenisch et al. [30] note, they control an individual's personal information, but not data and activities which can be inferred by the platform and these platforms do not insure that user's data is not being used in their internal studies and data processing.

By looking at the above example it is clear that the challenges in the current data sharing environment come from the new role players in the SDS. The current centralized platforms either provide tools and TOS to avoid legal confrontation with contributors or simply ignore the needs of the individual or community level data sharing contributors. This issue can be solved by transferring data controller role to the data producer and developing platforms in such a way. The table 2.1 shows a summary of SDS requirements which are not all supported in the centralized platforms.

2.2.3 Distributed SDS model

With social and economic life increasingly facilitated through web-based data and services while core issues of data privacy, ownership, and access remain concentrated among data controllers, there is a growing demand for a new data sharing and governance paradigm. In the context of SDS for example the recent focus in research has changed from SDIs to the individual's geoprivacy (e.g., [141]), rights (e.g [75]), and other legal aspects of SDS (e.g [20]). We believe that we are on the cusp of a new data sharing era which may take a variety of forms. One particular form of SDS we want to highlight is that of Distributed Spatial Data Sharing (DSDS).

The critical shift in DSDS is that individuals are not only data producers and users, but they will also serve as data controllers (see Figure 2.3). Some recent initiatives related to data sharing (e.g, the right to be forgotten) emphasize the primary role of individuals. In this model, data ownership and access control are determined by the individuals, not by platforms, which provide only the infrastructure and tools. This transition is based on users' concerns regarding the behavioural data they have collected or extracted by using different platforms.

Class	note
Transparency	• Transparent decision-making procedures in PGIS • Immutable tracing the changes in the GI • Transparency in GI which is used and stored by the platforms • Transparency in the technology and its internal mechanism
Ownership	• Avoiding Wiki effect in the data contributions • Clear copyright and licensing at the feature level
Control	• Control who, when, and why contributed data are used • Benefiting the data contributors using data markets

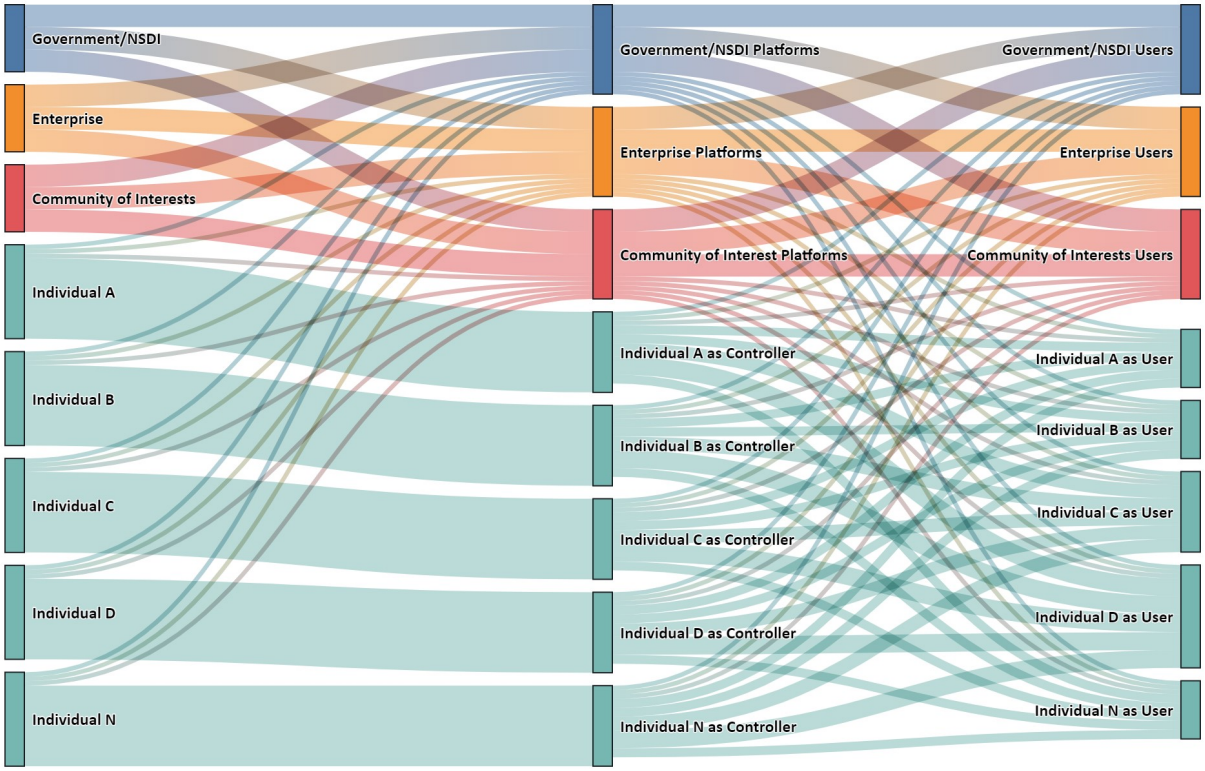
Table 2.1: SDS requirements which the current data sharing platforms are not able to fulfill.

Technical advances can be leveraged as a tool to provide users access to control their data. Two such examples are location-based services and location-based access control models [9] which can be used for data locality purposes. In a DSDS, technical factors will be more impactful, since individuals are responsible for access control, not the platforms; as such, users need access to the documentation about understanding how the data sharing mechanisms work on these networks. With the emergence of Web 3.0, including peer-to-peer protocols and distributed systems, users are not dependent on a central platform for their data sharing needs and thus have a higher level of control over data.

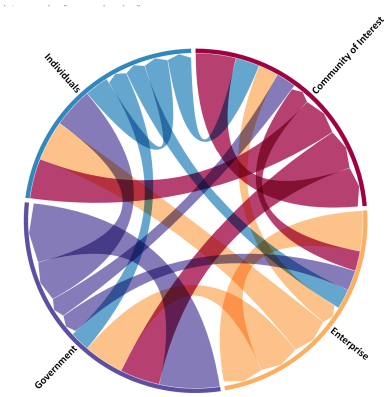
Distributed systems are defined by Tanenbaum et al. [247] as *"Transparently utilizing every kind of available resource on the network of the users and connecting the users on the network to the distributed resources transparently with support of openness and scalability."* The resources can be data [285, 163, 142], process [283, 163] or knowledge [48]. A distributed ledger is a type of distributed database in which data can only be appended or read [243]. Blockchains implement the structure and functionality of distributed ledgers in such a way that each new block of information is generated based on the previous block and appended as a new on the end of the chain. These blocks of information can even be a program or series of Smart Contract calls. Smart contracts are a set of Scenario-Response procedural rules and logic on the blockchain that run when predetermined conditions are met [275, 208]; they, therefore, allow for secure transaction issuance, without the need for third parties. For example, smart contracts can automatically move digital assets according to arbitrary pre-specified rules in a crypto-currency network [28]. The transactions that smart contracts issues are traceable and immutable [243].

In a distributed data sharing method, the data workflow would be similar to those of traditional data sharing models (Figure 2.3 (a)). The main roles as data producers, controllers, and users remain intact. However, the role of each actor will change. In such an environment, any entity within the network can be a data producer, a data controller, and/or a data user. An individual, for example, can produce data, have control over the produced data, and also gain access to the data from other entities (Figure 2.3 (b)). The economic regime of data sharing will be more complex, for example with the advent of data markets and the ability of users to control who, when, and where data is available. Individuals and community groups will be able to play a more important role in the data sharing environment.

It is worth noting that distributed systems include different technologies such as distributed file storage, blockchains, and smart contracts. Such technologies are mainly integrated together and are complementary, providing a distributed computing/sharing environment. For example, access control in distributed file-sharing systems can be done using smart contracts (e.g., [238]). In addition, because distributed ledgers are not meant to store



(a) Each actor in DSDS can be data producer, data controller and user, The complex relations in the right side of the graph indicates individual level access control



(b) Data Flow in DSDS model.

Figure 2.3: The flow of the data in DSDS model and the main role players

large data, distributed file storage can be used[189]. The peer-to-peer (P2P) paradigm is mostly popular for storing and sharing information in a totally decentralized manner [136]. Typically, a P2P file storage system is a distributed environment formed by autonomous peers that operate independently. Each peer stores a part of the available information and maintains links (indexes) to other peers. In addition, we need to distinguish between blockchain and crypto-currencies. Blockchain technology is the backbone of the cryptocurrencies such as Bitcoin. For instance, Bitcoin is a financial use case of the blockchain and its mechanism (such as its proof-of-work and energy consumption concerns) is different from other Web 3.0 applications.

Data access control on distributed platforms is often handled using public/private key management methods and smart contracts (e.g., [63, 131, 238]). In such platforms the data owner has the ability to distribute secret keys to users and encrypt shared data based on different access policies [272]. Such an ability in distributed systems enables owners to directly encrypt data and share it with users, while in centralized systems a third party is responsible for sharing collected spatial data. In centralized systems, users have to trust the cloud and the application providers – users have no choice but to rely on the security and availability of the application providers, to accept their policies, and to adhere to their licenses [228]. Such methods have been used in Distributed Personal Data Stores to provide control over sharing personal data (including location) with different platforms (e.g., see [3]). When users accept a third-party license agreement, they risk losing their intellectual property rights.

In the next section, we will look at the main features of distributed systems and explore examples of them in the context of spatial data sharing with a focus on data ownership and access control.

2.3 DSDS capabilities

Distributed systems and smart contracts have the potential to shift the way that society operates [21]. There have been studies examining the potential benefits of such systems in different sectors of society. For example, Ølnes et al. [197] have identified 16 benefits of distributed systems and smart contracts in governmental bodies in five classes: Strategic, Economic, Informational, Technological, and Organizational impacts. The benefits of such systems are attributed primarily to other technologies like encryption methods, in addition, some of these benefits are not limited to distributed technology, but to social and cultural benefits as well. However, looking at the extension of the existing technologies the aforementioned benefits have not yet been put into practice.

As distributed systems are being developed and used by different applications, it becomes possible to categorize the advantages of distributed systems as follows:

Decentralized consensus The decentralized consensus was used by Bolin [21] as one of the key features of public ledgers enabling trust-less transactions. Data can only be stored via group consensus in distributed networks, consequently becoming more transparent and verifiable [77, 21]. Such features of distributed networks prohibit data tampering - the data cannot be altered on the network since it is stored individually on the nodes on the network. The data is not exclusively maintained by a single individual or entity, but rather, is available to everyone, and the state of the data is decided upon via a consensus protocol [80].

Immutability Immutability is another feature of distributed systems. Immutability is the central reason why the participants trust distributed systems and ledgers [244, 178]. Immutability means that records and data cannot be altered and changed after they have been added to the network [119]. Distributed file-sharing systems usually preserve a unique HASH address for each unique file. Content-based addressing (in the case of the Interplanetary File System (IPFS)) provides the unique feature of tracking data in the network using its content HASH address. Content-based addressing can be used as a key component in conceptualizing a distributed network's file sharing as a database shared over nodes of the network. Immutable spatial data can be used in cadastral applications for instance to use for the rental transaction [57] or registration of land plots as spatial objects [57, 174]). The combination of content-based addressing with geohash algorithms -which encodes a geographic location into a unique short string of letters and digits [185]- or locality sensitivity hashing (see [268, 14]) might even provide a means of providing spatial data queries over a distributed network.

Append-only nature The ability to record the state of phenomena in an immutable format and to store such records as a chain of events provides a suitable means by which the state of the data at any historical block may be queried. In a distributed ledger each new record is appended to the previous record and the entire chain of records can be verified by the network nodes. This feature can help future distributed VGI platforms manage the information and life-cycle of spatial data. Such models can also be used for spatial-temporal analysis by providing the capability of querying the chain of transformations over each spatial entity.

No single point of failure From a technical perspective, decentralized storage solves the problem of the single point of failure in traditional storage systems and cloud-based systems [272]. Shafagh et al. [228] compare conventional cloud-based solutions for IoT data storage and sharing and argue that they are able to (1) provide secure data storage, and (2) provide IoT-compatible data streams, but are unable to provide decentralized access control management. Smart-contract-based applications can provide all of the above requirements for cloud-based data sharing [228].

One of the challenges to most of the current SDIs that are developed in remote areas, like the Arctic SDI, is user connectivity [188]. The accessibility of the Internet to users and user access to the central servers of SDI in remote areas is mentioned in most of the related technical documents. In the first pilot of the Arctic SDI project (www.arctic-sdi.org) challenges such as data integration, limited telecommute resource/bandwidth in the North, and end-users' concerns about the data policies are highlighted [170]. DApps and distributed data sharing methods are not dependent on central servers and data can be accessed from existing nodes on the network, potentially addressing communication issues with the 'outside' by enabling local area access within remote areas.

Scalability Distributed systems are scalable by nature. Considering the amount of geographic data which are being produced every day from many sources and at different resolutions, in different environments such as digital earth, the scalability of data sharing, management, and processing is a necessity. There have been many studies on distributed geographic processing but the scalability of geographic data sharing has not been studied before (e.g., [274]). Looking at the scalability of existing GIS architectures, it is claimed that the existing GIS architectures are not able to fulfill the scalability challenges raised by the enormous number of users, data, and heterogeneous data sources [284, 300]. Thus, a new Internet GIS architecture that can scale up to accommodate these needs is essential. A fully distributed Peer-to-Peer (P2P) spatial architecture may fulfill the scalability needs of new architectures [160]. They can scale well because it is possible to effectively partitioned access across the network. No one individual needs to store all data for that data to be available to the whole network.

2.4 DSDS as a solution

Considering spatial data sources, Cuno et al. [52] classified different data sources for an Urban Data Space. They categorize the source of data into 8 classes: (1) official

institutional, (2) enterprise, (3) research, (4) personal, (5) behavioral, (6) freely available, (7) commercially available, and (8) internally available. Looking at this classification scheme, it is obvious that some of the classes have clear data ownership, (1), (2), (4), (8), while some of the classes have arguable ownership, (3), (5), (7). For example, personal data are subject to data production regulations (especially in the EU and North America). Third parties require the data subject's consent to store and process the data. In addition, physical persons have the right to inspect the data and to initiate its removal from third parties' central servers. Freely available data can be discussed under open data sharing policies.

In order to use DSDS to share each category of the above data a different approach is required. In categories such as (1), (2), and (8) depending on the policies of the data owner they might opt to use DSDS. However, to share data that includes personal information, or collective information DSDS can be used. Take personal data for example. Users will be able to share their identity, location, and other related information in distributed social media. They are able to store and retrieve their location information with different resolutions on the P2P networks and encrypt them with their own private keys. Only contacts that are allowed to access a certain level of privacy can access the specific geo data resolution (See [120] for storage and retrieval of user location with dynamic geoprivacy from distributed networks). Figure 2.4 (c),(d) shows a simplified architecture of personal data storage using distributed systems.

Research data is another category that usually includes shared data ownership between different stakeholders under various license terms. This information can be another good example of using the DSDS model. However, the complexity of the models can increase as the licensing terms get more complicated. Figure 2.4 (a), (b) shows the architecture for storing such information. As the figure shows each stakeholder in the research data encrypts data and stores them on the IPFS network (Steps 1 and 2). The IPFS Hash will be later stored on a blockchain alongside the access level and related metadata (Step 3). A research data gateway will be responsible to index existing datasets on the blockchain and providing a method to query data from the blockchain (b). A user who wants to access the shared data needs to first find the IPFS Hash of the dataset. This can be done using a research data gateway or querying blockchain directly. In this step the user's access level can be examined in the blockchain level. Once they get a hold of the IPFS they need to decrypt the data from the dataset. The decryption can be done either by directly inquiring about data from stakeholders, requesting blockchain, or research gateway. However, more in-depth access control methods are required for such architecture.

A distributed ledger structures a series of transactions/blocks which can then be used to trace any changes made to the data. An example of such a use case is traceable logs for

health data management by [160, 194, 36]. Similar mechanisms, like [300, 58] or [213], can be used for VGI platforms to provide an immutable trace of the collected data since the rights to the collected data can be preserved for the collector. An example of copyright preservation for shared spatial data using IPFS and blockchain can be found in work done by [300]. The benefits of the DSDS for VGI platforms can be as follows: (1) Provide an immutable trace of the changes in GI and as a result, clear data ownership and avoidance of the wiki effect, (2) provide individual-level control on the data which results in a more applicable VGI platform for small communities. There are examples of VGI projects (see [305]) that use a distributed network of nodes to share sensor data. An example of such project architecture can be found in Figure 2.4 (e), (f). Such approaches use two sets of blockchains, a full blockchain to retain a history of the sensor data and a user blockchain to retain only the latest version of sensor data. Spatio-temporal data can be stored on IPFS and the query process can be done using spatial indexes. A smart contract can handle the querying data from the IPFS and returning query results to the user node.

In the above examples, smart contracts have been used as a set of functions that are responsible to perform the business logic of data sharing. They can be used as a small program to either control access or share the IPFS hashes or provide other logic to allow adding new blocks to the chain. The current versions of the popular programming languages for smart contracts such as solidity do not support native geometry objects. In order to use geographic data in the smart contract we can Discrete Global Grid Systems (DGGS) based geo hashes to perform geographic analysis inside these applications without the need to transfer/process high-resolution geographic information. In addition, it is possible to perform algebraic functions in the smart contracts which makes it possible to use regular geometry objects. Table 2.2 shows a summary of the different types of datasets and potential approaches to share them using DSDS.

2.5 GIScience research in DSDS

Distributed systems provide a transparent decision-making process by facilitating coordination and trust, and by addressing the corruption inherent to decision-making in different organizations [279]. For example, see the work done by Farnaghi et al. [81] in which they used smart contracts for a transparent and participatory site selection. They and other authors show that smart contracts can increase the openness, transparency, and accountability of participatory planning processes by democratizing data access and keeping transaction histories on every node [12, 257, 197]. Similar work has been done in [154, 172, 171] which have used distributed ledgers in PGIS projects. It is required to provide trans-

Type of the data	Sharing Approach	Addressed challenges
official institutional	Centralized Authority	
enterprise	Centralized Authority	
commercially available	Centralized Authority	
internally available	Centralized Authority	
Research including PPGIS and VGI	Store on IPFS and use blockchain for access control and licensing of different stakeholders	Access control Transparency for PPGIS Projects Monetization of the contribution for VGI
Personal and Community of Interest data	Store on IPFS and blockchain with decryption	Personal Data access control Transparency at the community level
Behavioral	Once the data product producer accesses the personal data then can generate a new product and that product can go under other types of data	Trust and access control
freely available	Can be shared on IPFS without any decryption, However, it needs data access gateways	Transparency

Table 2.2: Different types of datasets and potential approaches to share them using DSDS.

parency about the technology and the ways the technology itself ensures the traceability and immutability of the data. Once the internal mechanism of technology is transparent it will be able to provide a trustworthy backbone to the PPGIs and VGI projects.

There are also technical limitations in the distributed data sharing platforms which require more research. Spatio-temporal query processing is one of the issues which needs to be addressed. Currently, many P2P networks such as IPFS are used as storage entities. sharing geographic information at the feature level needs a method to address data retrieval, querying and transformation over P2P networks. This requires the development of methods for indexing multidimensional data on distributed networks that are capable of addressing range and k-nearest-neighbour queries, while also capturing the locality and directionality of the multidimensional space (e.g see [135, 136]). P2P multidimensional query processing refers to the execution of advanced query operators over multidimensional data stored in a distributed system [264].

Another part of the technical aspects of the GIScience research of DSDS is the need to develop standards and protocols to share geographic information. These protocols can vary from the low-level object definition to the higher level of distributed web APIs. In DSDS the shared information is in a very heterogeneous format. Geographic information representations of phenomena in DSDS need to be handled using data models that are capable of representing complex geometric, topological, and semantic elements (e.g see [43, 270] for other representations of geographic information) and be able to provide a level of the data masking to be able to use it as a privacy related concerns. At the higher level, there can be standards to look at the entire P2P network as one entity and perform spatial processing using the nodes and receive the processed data instead of just receiving raw data itself. However, distributed data processing has been studied on traditional P2P networks for many years but with the current advantages there might be need to revisit the previous approaches.

Moving toward a distributed model of spatial data sharing requires the ability to share different data from different sources and standards with DE platforms using a distributed approach. The use of emerging data models and technologies in DEgital Earth (DE) such as DGGS means the sharing of spatial data between different providers can be facilitated. Figure 2.5 shows the potential relations between DE, SDS, and SDI. SDS is the connecting bridge between DE and SDI. However, SDIs are often seen as contributors to the vision of DE [193]. Despite their common components, an SDI acts mainly as a data collection coordinator and can also provide guidelines for communications between different data owners and data customers. In addition, the role of DGGS as a data model to share and integrate data from different sources can be seen as a distributed data model too. In a DGGS each cell represents a portion of the earth's surface which can have different

identities over time.

Another challenge in DSDS is access control and data locality on distributed networks. Since data are stored on the user's devices it is required to develop protocols to provide access control based on the user's criteria (such as the user's friendship status or distance in the social network context). Distributed geoprivacy is another concept that needs more study. Once users are sharing their location with another node, they should be able to have a dynamic geoprivacy setting. Blurring geographic data in the context of distributed networks can be done by using other nodes in the network or by simply sharing data over the network with lower precision.

In order to provide a comprehensive data sharing platform, it is also required to have distributed license management systems and distributed data gateways/ marketplaces. License management systems can allow users to assign different licensing to the shared data and don't need to comply with the data sharing platform's license. These systems can store (key, value) pairs on the blockchains where the key is the IPFS hash of the stored information and the value is the licensing method attached to that particular dataset. Since IPFS hashes can be deconstructed into the feature level the licensing can also be at the feature level. Due to the limitation of the search engines on the IPFS network, it is required to have gateways to access the datasets. These gateways can also act as marketplaces where users are able to monetize their contributions on different platforms. For example, [58] has used content-based addressing methods and proposed a global data market with the goal of making contributors the stockholders in the dataset(s) they create. On their platform, the ownership of the data initially belongs to the nodes who created the data. The owner is then free to transfer the ownership of the data to others. In addition, they have also provided a mechanism whereby the data owner is rewarded when their data set is used. Another similar data marketplace platform for urban applications is also developed by [213]. Such platforms clearly define and enforce information ownership [287]. Similar initiatives also help to increase collaborative projects, leading to improved data sharing, improved data quality, and improved decision-making [109].

A first requirement to make an accessible DSDS is the technical aspect of the availability of the tools for the P2P networks. IPFS is one of the accessible P2P networks, and the community behind it, which provides many tools and SDKs to connect to the network and work with it. The recent native support of the browsers for IPFS helps to solve this gap between developers and users. Smart contracts are also lacking native support for geographic analysis, which requires users to use alternative methods such as geohash to work with geographic information inside the smart contracts. Another requirement is education and knowing how these technologies work. This can help in two ways. First, it helps with the developers and eases the process of integration and sharing geographic

Aspect	Requirements
Social	• Transparent documentation about the internals of available tools such as IPFS • Make a differentiate between cryptocurrency and distributed tools
Technical	• More multi-platform SDKs to communicate with P2P networks. • Research on query processing of spatio-temporal data on distributed networks • Developing Geo-enabled smart contracts • More research on access control methods on distributed networks

Table 2.3: The requirements to have an accessible DSDS .

data on the P2P networks and blockchains. Second, it helps with the transparency of the technology, which helps to build trust in the technology. Table 2.3 summarizes the requirements for an accessible DSDS.

2.6 Conclusions

Sharing distributed geographic information is becoming an effort visible across different scales- from global collaborations such as digital earth and global earth observation systems to small, individual-scale data sharing over time [283]. Data ownership and access control are two key aspects driving changes in the data sharing environment. Data ownership as the right to control data can be considered a reward for the data collectors in SDS [260]. Ownership of the data can be compensated to the data providers for sharing it. As a result, incentives can improve data quality by providing responsibility and liability for data collectors [263, 38]. We believe that in the current era of big geographic data collection, distributed data sharing and processing is necessary. There must be protocols in place to share, integrate, and process distributed data.

In this work, we addressed the current status of data ownership and access in different

SDS models and demonstrated how a transition to DSDS addresses some of the existing challenges. Distributed systems provide scalability, no singular point of failure, ownership, trust, and transparency. The storage, distribution, and manipulation of spatial data are changing from a fully centralized approach (i.e., controlled by the government and/or corporations) towards a distributed, individual spatial data sharing approach, though this approach remains technically challenging. Distributed technology is in its early stages, and it requires the development of tools/methods and algorithms to handle, share and query geographic information. Once developed, it will be possible to contrast DSDS against other data systems, and thereby evaluate the practical benefit of such systems. A distributed data sharing platform not only needs a standard to share data between different users but also requires a data model that can integrate different spatial data from different sources and various accuracy levels. Such standards must be in alignment with multiple sharing platforms and frameworks.

A distributed framework for spatial data sharing is not the only solution to the aforementioned challenges, but it provides a user-centered approach to address some of them. As mentioned previously, technology is only one piece of the data ownership and control puzzle, albeit a critical element. It should be noted that distributed data sharing is not a universal solution. Many types of authoritative or proprietary data sets (e.g., geodetic control, cadastral boundaries, municipal addresses) will continue to require centralized access and authoring control for social, competitive, or legal reasons. In this light, DSDS and centralized systems are complementary and together enable more fine-grained and flexible data governance architectures. For example, in the digital earth context, a mix of distributed and centralized systems can be used depending on the type of data being shared (e.g., a DEM vs a social media activity, etc.).

To sum up, we define DSDS as a new data sharing model in which individuals are active in all dimensions of data sharing: as producers, controllers, and users. The data are stored, controlled, and maintained by the data producer. The ownership and license of the data can be transferred to other users without the need for data intermediaries. The immutable spatial data are stored on a distributed file-sharing network, while the access rights, state transitions, and version history are managed by smart contracts and stored on a blockchain.

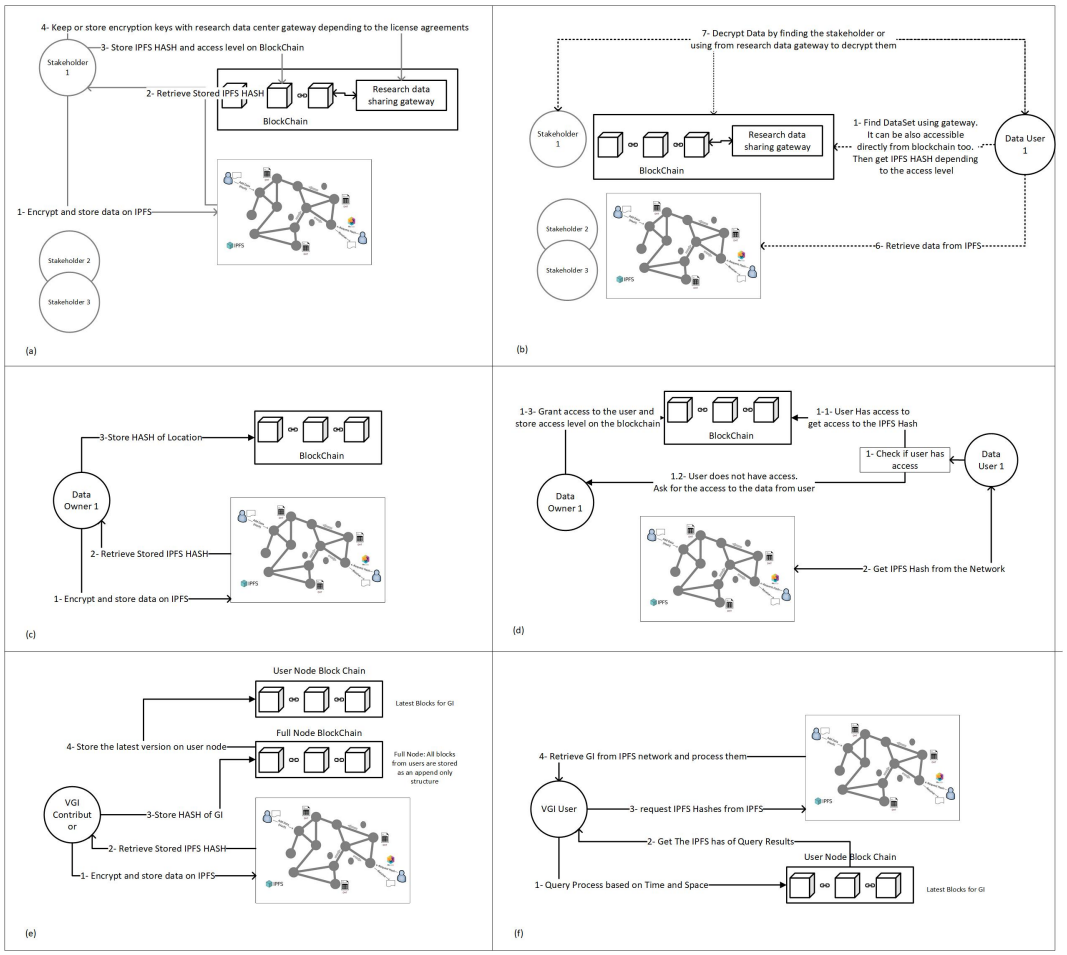


Figure 2.4: The proposed architecture of a decentralized model of data storage for (a), (b) Research Data, (c), (d) Personal Data, and (e), (f) VGI platform. The left column is the process of sharing data from the data owner's perspective and the right column is the process of sharing data from the data user's perspective

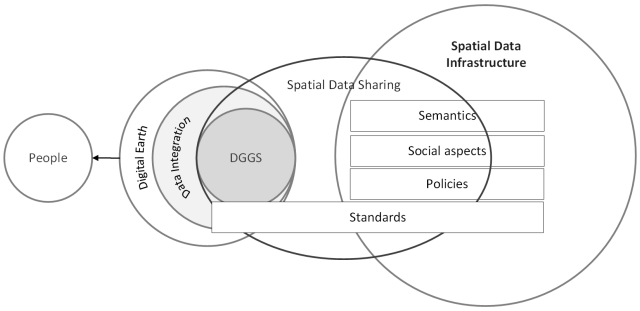


Figure 2.5: The interrelation of spatial data sharing, Digital Earth frameworks, and SDI

Chapter 3

DSTree: A spatio-temporal indexing data structure for distributed networks

Abstract The widespread availability of tools to collect and share spatial data enables us to produce a large amount of geographic information on a daily basis. This enormous production of spatial data requires scalable data management systems. Geospatial architectures have changed from clusters to cloud architectures and more parallel and distributed processing platforms to be able to tackle these challenges. Peer-to-peer (P2P) systems as a backbone of distributed systems have been established in several application areas such as web3, blockchains, and crypto-currencies. Unlike centralized systems, data storage in P2P networks is distributed across network nodes which provides scalability and no single point of failure. However, the management and processing of queries on these networks have always been a challenge. In this work, we propose a spatio-temporal indexing data structure, DSTree. DSTree does not require additional Distributed Hash Trees (DHT) to perform multi-dimensional range queries. Inserting a piece of new geographic information updates only a portion of the tree structure and does not impact the entire graph of the data. For example, for time-series data such as storing sensor data, the DSTree performs better. Despite the advantages of our proposed framework, challenges remain. We conclude that more significant research effort from GIScience and related fields in developing decentralized applications is needed. The need for the standardization of different geographic information when sharing data on the IPFS network is one of the requirements.

Keywords: spatio-temporal indexing, temporal topology, query processing, *IPFS*, distributed systems, smart contracts, blockchain

3.1 Introduction

With the advancement of Internet-based services and sensors, such as the widespread adoption of GPS-based sensors, location-based services, improvements in computational data processing, and satellite imagery, a large amount of novel spatial information is produced daily [108]. The widespread availability of tools to collect and share spatial data enables individuals and small communities to produce their own digital spatial content (e.g., Open Street Maps contributions have tripled between 2012 to 2017 [7]). This enormous production of spatial data requires scalable data management systems [66]. The data infrastructure technology supporting spatial data management and processing has changed from standalone relational database systems to spatial data warehouses that support a variety of data formats and analytical workloads [2] and from centralized infrastructures to decentralized and Peer-to-peer (P2P) systems. New data structures have been proposed such as (e.g., HDF [112], Data-Cubes [168], Geoparquet [101]) and spatial data management and analytic frameworks (e.g., Apache iceberg [124], Digital Earth [106]) have emerged to manage and perform analysis on large-scale, high temporal and spatial resolutions.

Geospatial architectures are another area of technological change developed to handle spatial data management challenges. They have changed from clusters to cloud architectures and more parallel and distributed processing platforms (e.g., Spark [291], Hadoop [71]). Peer-to-peer (P2P) systems have been established in several application areas. In the past decade, P2P systems have been used widely in web3 [16], blockchains [192] and crypto-currencies [187]. The combination of P2P file-sharing systems with blockchains provides scalability, security, immutability, and append-only attributes for sharing information amongst nodes on a network [121]. These attributes make P2P networks suitable for content distribution and service discovery applications (see [65, 286, 70, 1]). However, the main limitation of existing systems is they can only locate data on the network based on a key value using DHT [49]. While DHTs have been used as a main building block for P2P applications, they are seriously deficient in one regard: they only directly support exact match queries [212, 116, 115, 253]. The multi-dimensionality of spatio-temporal data is one of the big challenges in retrieving and querying these data in P2P networks. When querying spatio-temporal data, the ability to perform range queries is required, and it is not currently supported. The rapid increase in spatio-temporal data collection needs a new auxiliary indexing structure. These indexing structures are responsible for tracking

the behaviour of moving objects through space [280, 182]. These indexing methods allow P2P architectures to find and retrieve contents based on the user's filters and address more complex data sharing needs, facilitating data management, query processing, and delivering data to the end-user.

While much work has been done towards expediting search in file-sharing P2P systems, issues concerning spatial indexing in P2P systems are significantly more complicated due to cases such as overlaps between spatial objects, avoidance of data scattering and the complexity of spatial queries [183]. One-dimensional data mainly have been queried using DHTs in P2P networks [185, 240, 218]. DHTs are not yet designed for complex spatial queries (e.g., range query, k-nearest neighbor query), and only support the location of data items based on a key value (i.e., equality lookups) [159, 49]. For multi-dimensional data, there have been multiple approaches, including partitioning data into one-dimensional indexes using space-filling curves or kd-tree-based methods and indexing them using DHT methods (e.g. [296, 176, 134]). These methods usually work best with static data, and dynamic content relocation (locality) is dependent on the accuracy of space-filling curves [134].

The second approach is to use multiple DHT in the network (e.g. [248, 62, 29]). These methods need a higher level of network and node structure manipulation and are not commonly used in large-scale projects due to interoperability limitations of such approaches [235]. A third approach is to construct traditional indexes that have been used in centralized environments and distribute these indexes on the P2P network (e.g., [127, 49, 90, 212]). This approach can be done by constructing a tree and splitting it into parts and maintaining parts of semi-independent trees at each peer. Prefix Hash Tree (PHT) [212] and similarly, P-Tree [49] uses the same approach by storing a fraction of the overall tree on each peer. In PHT each node of the tree is labeled with a prefix which is defined recursively. Given a node with the label l, its left and right children are labeled as $l0$ and $l1$ respectively. This pattern constructs a tree structure and enables range queries on a dataset [212].

In this work, we propose a spatio-temporal indexing data structure that works in the data layer and uses a distributed InterPlanetary File System (IPFS) network. Our method is closer to the approach of Ranabhadran et. al. [212] and does not require additional DHT to perform multi-dimensional range queries. The rest of this paper is divided into two main sections. First, we introduce the Distributed Spatio-Temporal Tree (DSTree) as a data structure to perform range spatio-temporal queries on a dataset and we check its features and performance by comparing it to other existing trees. Second, we will look at the integration of DSTree with distributed networks and propose a system architecture to perform queries on the IPFS using DSTree.

3.2 Spatio-temporal data indexing methods

P2P multidimensional query processing refers to the execution of advanced query operators over multidimensional data stored in a distributed system [264]. The retrieved data from a query should be exact and complete. Exactness means the query result should not be approximate. The retrieved data should exactly belong to the query results set. This means that if we run a query on the same dataset on a centralized system the results should be exactly the same as when we run a query on the distributed system. A basic element of geospatial technology can be defined as three main components including location in space and time, and attribute of that location in space-time [56]. In this work, our focus is to retrieve a geographic object based on space or time queries.

The spatio-temporal indexing methods in central systems are mainly performed at an abstract level (Figure 3.1). They are used to improve query performance on the datasets. These indexes usually work as a separate layer on top of the data layer itself. Queries are usually performed against the indexes and then the actual reference to the geographic features are then retrieved from the index (See figure 3.1 right). In contrast, our method stores data on the nodes on the network using spatial indexing methods in both abstract and physical storage layers. For example, the purpose of indexing can also be to distribute data that are closer in space, on the nodes that are closer in the network (e.g. [49], (See figure 3.1 middle)). Through the last couple of decades, many spatio-temporal access methods have been developed. There have been several approaches for spatio-temporal indexing so far (see [91, 280, 292, 204, 250, 182, 169, 117]). Handling temporal data in GIS ranges from time-stamping GIS layers (e.g. [10]) to more object-oriented approaches such as time-stamping events and processes (e.g. [206])[292]. Another category of spatio-temporal models is trajectory-based access models which track the changes in the geographic object typologies over time (see [169]). In central systems indexing models such as OCT-tree [126] are sometimes used to process spatio-temporal queries (see [295, 169, 302]). In using OctTrees for indexing spatial data each geographic object is considered as a cuboid. The spatial dimension of the data is considered as two dimensions of each cuboid. The third dimension of a cuboid is considered a time interval. There have been some newer approaches to the query process of spatio-temporal data using blockchains. For example, [211] used a block-DAG-based index traversal algorithm, to handle spatio-temporal queries on a block-DAG. However, the main issue with the blockchain-based spatial-temporal indexes is the limited data storage capability on the blockchains [303].

A decentralized spatial indexing technique must be scalable enough to be able to handle hundreds of thousands of peers and also dynamic enough to deal with peers joining/leaving the system anytime. Another feature of such structures is their ability to preserve the lo-

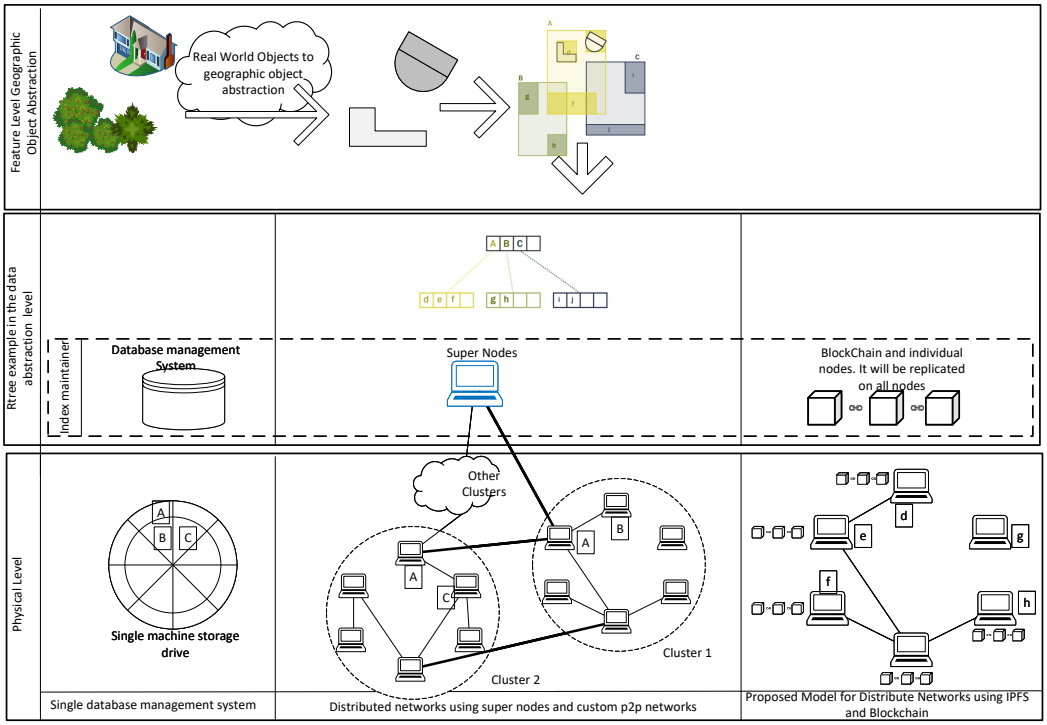


Figure 3.1: Abstraction levels of geographic information. Spatial indexes cluster units of GI at the abstract level and it is used at the storage level in the different architectures. The Spatial Index maintenance is handled by RDBMS [207] (Left) or super nodes (Middle) in some studies (e.g. [49]) or using proposed blockchain-based (Right). In the proposed model each node maintains a spatial index and the latest version of the index is always published on the blockchain. Each node only stores and serves features that they need using IPFS.

cality and the directionality of multidimensional information. Locality implies that neighboring multidimensional information is stored in neighboring nodes, while directionality implies that the index structure preserves orientation. The notions of locality and directionality are very important. If an index structure preserves these properties, then searching in the index corresponds to searching in the multidimensional space, which can highly improve query evaluation cost[183]. R-tree-based indexes can efficiently answer various types of multidimensional queries, especially range queries [162]. In addition, a spatio-temporal indexing method is required to support two types of topological relations. The first set of topological relations includes temporal typologies which are based on Allen's temporal algebra covered in [4, 89]. These relations include 7 typologies which are briefly explained in Table 3.1. A time interval for Geographic Information (GI) can be defined as the duration in which a GI feature exists with a fixed state. This interval can be as small as a few milliseconds that it takes to collect a GI feature or can be a considerably longer time period such as geological land classifications or land cover. Defining time intervals and how a GI can be attached to a newer time interval depends on the context of the study. The second type of topological relations which needs to be addressed by a spatio-temporal indexing model is spatial topology.

Each indexing method is optimized for specific types of queries. Our proposed method is more suitable for the storage, query processing, and retrieval of log-based data. These data are produced over the time that different events happen. An example of such data is data that are being collected from a sensor over time or a VGI tool to share images from different locations by users or even open data which are being shared by different government departments such as crime data that are being shared by the police. Each of these datasets can have different access levels, and geoprivacy levels and are being collected over time.

3.3 DSTree: A spatio-temporal index

A unit of geographic information (GI) in our model is defined as a spatio-temporal object which can be represented as the form of

$$Gid(GI, GI_{MBR}, GI_{Time-Interval})$$

where Gid is the identification of the object. GI is spatial location l, including longitude and latitude (GI_x, GI_y) along x and y dimensions in the interval $GI_{Time-Interval}$. Each GI also has a set of attributes, p, associated with it, and a Minimum Boundary Rectangular (MBR), GI_{MBR}, in which is constructed based on its location, l.

Table 3.1: Temporal algebra introduced by [4]. X (line border) is time interval of the first GI and Y (dashed border) is the time interval of the second GI

Time direction→	X in relation to Y	Y in relation to X	Condition
	equal (=)	equal (=)	$X_{t_1} = Y_{t_1} \& X_{t_2} = Y_{t_2}$
	meets (X m Y)	is met by (Y mi X)	$X_{t_2} = Y_{t_1}$
	overlaps (X o Y)	is overlapped by (Y oi X)	$Y_{t_1} \leq X_{t_1} \leq Y_{t_2}$
	during (X d Y)	contains (Y di X)	$Y_{t_1} \leq X_{t_1} \leq Y_{t_2} \& Y_{t_1} \leq X_{t_2} \leq Y_{t_2}$
	starts (X S Y)	is started by (Y Si X)	$X_{t_1} = Y_{t_1}$
	finishes (X F Y)	is finished by (Y Fi X)	$X_{t_2} = Y_{t_2}$
	precedes (X < Y)	is preceded by (Y > X)	$X_{t_2} < Y_{t_1}$

i stands for **i**nverse
$X = [t_1, t_2]$ and $Y = [t_1, t_2]$ are two intervals. t_1 is called the low end point and t_2 is called the high end point of the interval

DSTree is a two-level tree structure. The approach of DSTree is closer to the work done by [210, 242, 249]. They constructed a multilevel tree structure to improve the query process of trajectory data. In [210]'s method they have formed a global space-time subdivision scheme. [242] combined two trees, an R*-Tree, and a kd-Tree, to improve the query process in the centralized machines. Tao [249] also used a series of temporal quad-Trees to handle interval queries in centralized systems. In a DSTree an Interval-Tree [60, 46] is used as the top part of the tree, and a quad-Tree [82] is used at the bottom part of the tree. The interval tree is responsible for temporal queries and the quad-Tree is responsible for performing the spatial part of the queries. We define a time-interval as a pair of real numbers $[t_1, t_2]$ where $t_1 < t_2$. $[t_1, t_2]$ can be represented as $\{[t_1, t_2] | t \in \mathbb{R} : t_1 \leq t \leq t_2\}$. The different variants of the interval-trees are capable of supporting open and half-open intervals. Interval trees are optimized for querying of the intervals which overlap with a given interval, but, can also be used for point queries. Having the ability to query overlapping intervals allows us to query based on the temporal topology in Table 3.1. During each time-interval we assume that GI state does not change.

3.3.1 DSTree index

When constructing a DSTree it is possible to partition data into spatio-temporal chunks and assign a unique id to each portion of the data. The proposed DSTree indexes are

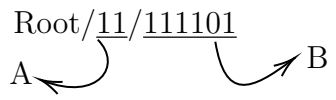


Figure 3.2: An example of DSTree Index. Part A is Interval-tree index. Its length is equal to the *Temporal level (T)*. Part B is a quad-tree index.

composed of two parts (See Figure 3.2). The first component (A) is called interval-tree index and the second component (B) is called the quad index. The interval-tree is a binary tree so each node can have only two children. By assigning 1 to the right child and 0 to the left child a series of IDs will be constructed. The length of digits in the interval-tree portion of the index equals *Temporal level (T)*. The second section of the DSTree index is a quad index. Each node in a quadtree consists of 4 children. Each child can be assigned an index from 00, 01, 10, and 11 and it can construct the quad index. Figure 3.3 shows the DSTree index corresponding quad-tree indexes. More details about the parts adhere to the specification in Figure 3.2. The top part of the DSTree is a regular Interval-tree. However, the depth of this tree is controlled by a parameter called *Temporal level (T)*. T is responsible for balancing between the top-level tree and the bottom-level tree. Once the top interval-tree is constructed the GI in the leaves (Each leaf includes n GI) is used to construct a quad-tree. As a result, we will have one quad-tree at each leaf of the interval-tree at the level of T. Regular quad-trees always have an extent equal to the minimum boundary box of the GI inserted into the tree. However, in a DSTree, all of the quad-trees should have the same extent, e.g., equal to $(-180°, -90°, 180°, 90°)$ in geographic coordinates. Having the same spatial extent bottom part of DSTree allows us to query data across all of the interval-tree leaves. Figure 3.3 shows an example of constructed DStree. Figure 3.4 shows the points which are used to construct that tree and their relative location and time-interval. Figure 3.2 details an index entry.

Insert In order to insert $Gid(GI_{CID}, GI_{MBR}, GI_{Time-Interval})$ into a DSTree, a two-step process is required. First, we need to find the proper node on the interval-tree part of the DSTree that *Gid* can to be added. Afterward, we will add the *Gid* to the proper quad-tree leaf using GI_{MBR}. To do so we first get the low value of the interval at the root of DSTree. If the root's low value is smaller than $GI_{Time-Interval}$'s low endpoint, then the new interval goes to the left sub-tree, otherwise, the new node goes to the right sub-tree. We continue the same process until the sub-tree level is equal to the *Temporal level (T)* parameter of the DSTree. Once the node is selected if there is already a quad-tree in the node (node is spatial) we insert the *Gid* to the quad-tree using its GI_{MBR} parameter. If the node

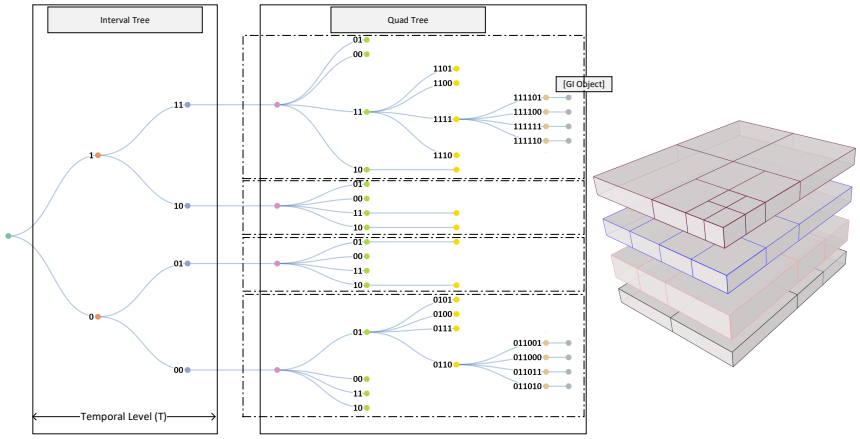


Figure 3.3: An example of DSTree Constructed from a set of sample points. Each DSTree has a temporal (Interval-tree) component and a spatial (quad-tree) component. The final graph will be a stack of quad-trees on top of each other

is empty we generate a new quad-tree and proceed with adding it. Adding *Gid* to the quad-tree is similar to the regular quad-tree insert (e.g see [82]). Once all the steps are done we update the max value of the ancestors of interval-tree portion if needed.

Delete Deleting items from DSTree is a relatively complex process due to the complexities of removing intervals from a regular interval-tree. After deleting an *Gid* from the DSTree, if the node containing that *Gid* contains no more objects, that node may be deleted from the tree. This involves promoting a node further from the leaf to the position of the node being deleted which results in the reconstruction of the top part of the DSTree (For details about deleting items from interval-trees see [46] pages 348-357).

Query A spatio-temporal range search is a query of geographic objects that intersect with a boundary box, $S = (Min_x, Min_y\ , Max_x, Max_y)$, in two-dimensional space and also is in a temporal topological relation, TP, with a time interval,$I[t_1, t_2]$ [74]. There can be three main variations of queries a DSTree which include queries with both temporal interval and spatial extents, queries with only temporal interval, and queries with only spatial extent. Here we only cover the first variation. The approach to those two variations is similar and explained in more detail in Figure 3.7. In order to perform a spatio-temporal range on a DSTree it is required to first query the interval-tree portion of the DSTree. If I is in TP

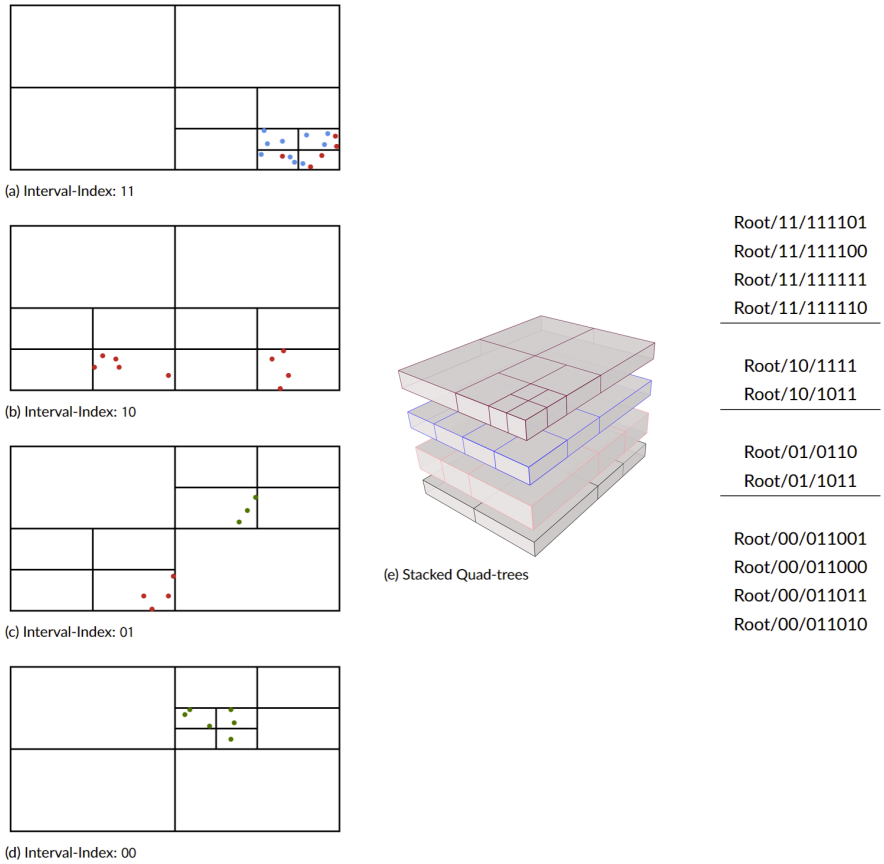


Figure 3.4: Spatial and temporal location of the points in the Figure 3.2. On the right side the DSTree-Index related to each section of the graph is listed

relation with the root's interval, we add the root's interval to the candidate nodes list. If the left child of the root is not empty and the max value of the sub-tree in the left child is greater than I's low value, recur for the left child otherwise recur for the right child. Once the candidate nodes are selected, if the selected nodes have a quad-tree as their sub-tree, we perform a quad-tree search for S, otherwise, we only check for the intersection of the selected node with the boundary box, S.

DSTree construction from bulk data Since DSTree is a two-level tree the performance of the insert, delete and construction of the tree depends on both the top level and bottom level of the trees. Construction of the DSTree from scratch for bulk data is a relatively straightforward process. First of all, an interval-tree is constructed based on the existing data. Once data are inserted into the hierarchical node structure, the algorithm traverses down from the root of the interval-tree until the tree level equals the *Temporal level (T)*. Once the appropriate level is detected in the interval-tree all of the items under sub-trees (left and right branched) of the selected node are collected into one single node and a quad-tree is constructed in that node.

3.3.2 Performance metrics

The behaviour of the proposed tree structure is measured using a number of experiments with real data. These experiments are intended to reflect the conditions of the common tasks involved in spatial and temporal queries of the data. Results of multiple models compared to the method proposed. In the first set of the experiments data on the occurrence of crime from the Waterloo Regional Police Service [1] are used. This data details all the police-reported occurrences for the calendar year. The time frame of this data is from 2017 to 2022. The data includes occurrence data and time, response time, and geographic coordinates of the occurrence. In this experiment, only the location of each event and the response time of each event is used. In order to have a comparison between DSTree and the other existing models, we have used three other methods to process spatio-temporal queries. In choosing each method, the availability of source code to perform the tests were considered. OCT-tree is one of the methods which is used in the experiments similar to the work done by [295, 169, 302] in which the third dimension of cuboid data is considered as the time-interval (source code available at [102]). The second access method used is a regular quad-tree (source code available at [99]). In this method, only a spatial query is performed and then to get the exact result set, all the results from the spatial query are

[1] https://www.wrps.on.ca/en/about-us/reports-publications-and-surveys.aspx

tested to filtered based on the temporal parameters. The third method is an Interval-tree (source code available at [97]). In this method, in contrast to the quad-tree method only a temporal query is performed, and once the results are extracted from the tree structure the spatial filter is applied to them to get the exact results. The above experiment is applied to the batches of $50k, 100k, 200k, 400k, 800k$, and $1million$ points. Each test was executed 5 times then repeated 10 times and the average time of the execution was measured. The spatio-temporal query was constant over the entire experiment. Figure 3.5 show the results of this comparison. The main query for these different models is defined as follows:

Find the events where their location has an $Intersection$ topology relation with an extent equal to

$$Spatial_{Extent} = (\frac{Max_x - Min_x}{4}, \frac{Max_y - Min_y}{4}, 3 \times \frac{Max_x - Min_x}{4}, 3 \times \frac{Max_y - Min_y}{4})$$

and their response time has a topological relation T with a temporal extent of

$$Temporal_{Extent} = [\frac{Max_t - Min_t}{4}, 3 \times \frac{Max_t - Min_t}{4}]$$

where $Max_x, Min_x, Max_y, Min_y$ are the spatial extent of the entire dataset and Max_t, Min_t are temporal extent of the data and T is the temporal topology from table 3.1

The second experiment is with the number of visited points in order to answer a constant spatio-temporal query. In the experiment only objects that the index traverses and are checked to answer each query are counted. Figure 3.6 shows the results of this experiment.

As Figure 3.5 shows DSTree shows a good performance for query processing in medium-sized datasets compared to other types of indexing methods. The metrics of the DSTree in all of the 6 temporal topologies are close to the Quadtree method. Oct tree and Interval tree also show close performance metrics. In Figure 3.6 the number of the points which are visited during the queries is shown. In this figure, DStree has fewer visited points compared to QuadTree. The reason why they have close metrics is that the Quadtree is less computationally heavy compared to DSTree.

3.4 IPFS

The first generation of P2P systems, namely file-sharing applications such as BitTorrent, support only keyword lookups and mostly provide no load balancing. The second generation is mainly structured P2P systems supporting basic key-based routing [159]. The

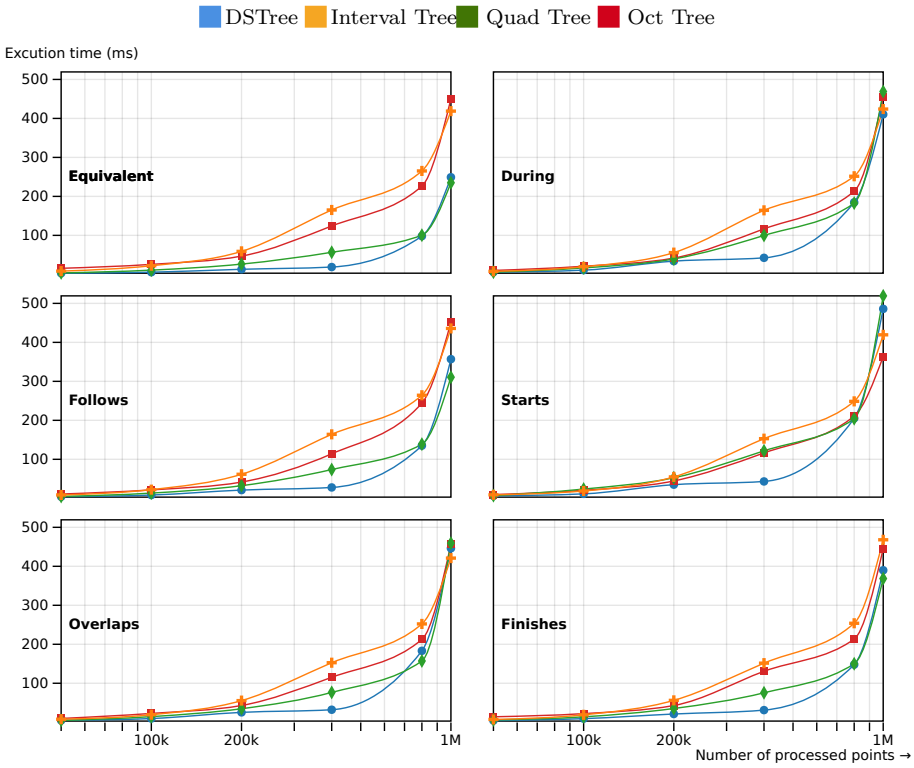


Figure 3.5: The query processing time for different spatio-temporal access methods. The spatial and temporal extent of each query remained constant. The results are measured for 6 different temporal topologies.

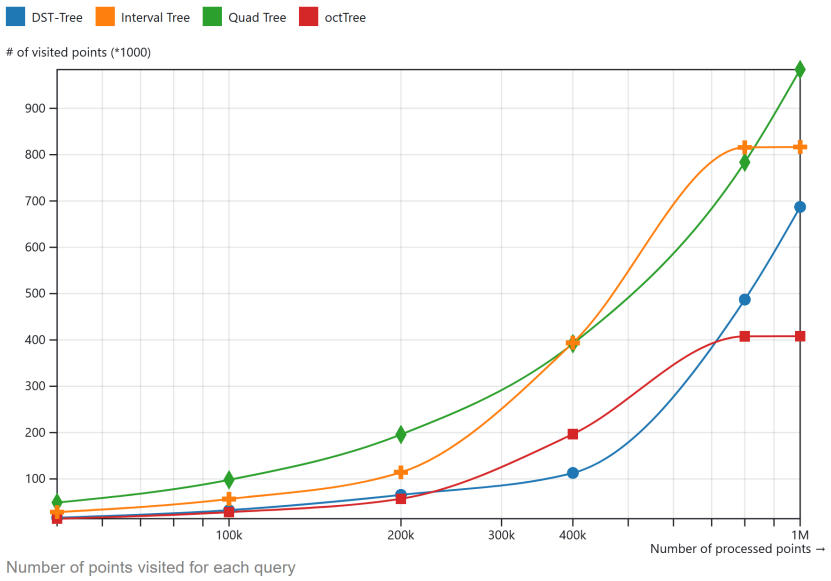


Figure 3.6: Number of the visited points in each model to answer a spatio-temporal query.

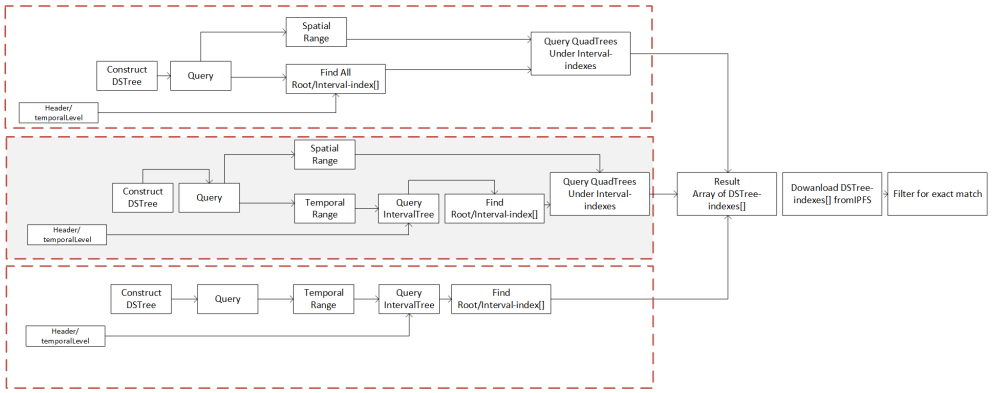


Figure 3.7: Three main scenarios to process queries using DSTree. Top: When a user requests only a spatial range in which we search all the quad-trees in the DSTree. Middle: When the user queries spatial and temporal ranges together, DSTree first queries interval-tree part of the graph and then searches quad-trees that exist at the bottom of those candidate nodes. Bottom: Cases user only provides a temporal range. As a result, we only search Interval-tree part of the DSTree and then simply query the root of the Quad-tree in each candidate node.

InterPlanetary File System (IPFS) is a protocol and a P2P network for storing and sharing data in a distributed file system. IPFS uses content-addressing to uniquely identify each file in a global namespace connecting all computing devices [18]. Each file on the IPFS network has a unique hash address which is used as a reference to request it from the network. Any user in the network can serve a file by its content address, and other peers in the network can find and request that content from any node that has it using a distributed hash table (DHT)[306].

At its core, IPFS is built on top of a data structure called The InterPlanetary Linked Data (IPLD) [18]. The IPLD model is a set of specifications in support of decentralized data structures for the content-addressable web. Content IDs (CID) are hashes generated to allow the user to interact with IPFS in a trustless manner and recover their data. IPLD deals with decoding these hashes so that users can access their data. When new content is added to the IPFS network, that content is separated into several chunks and stored in different blocks. To reconstruct the whole file, a Directed Acyclic Graph (DAG) connects each bit of content together. In a DAG, we can only move from parent nodes to child nodes as each edge is oriented. Hierarchical data in particular is very naturally represented via DAGs.

IPLD creates a series of links to data internally but also allows users to create those links themselves through simple data structures that can be stored on IPFS. This capability allows us to store a DAG graph (in our case a DSTree) on IPFS. This capability allows users to request a portion of the data from the network without the need to download the entire dataset. For example, a user is able to store a graph shown in Figure 3.3 as an IPLD object as shown in Table 3.2. IPLD's capability to store and retrieve DAG graphs allows us to store spatio-temporal data as a graph structure and as a result, we can request them based on the query parameters. DSTree only keeps a CID reference to the actual feature in each tree leaf. So depending on the query parameters we only need to retrieve a portion of the GI or specific subtree of the DSTree. For example, a DSTree leaf can be addressed as

$$DSTree_{CID}/Interval\text{-}tree_{index}/quad\text{-}tree_{index}/$$

where $DSTree_{CID}$ is the CID of the root of the DStree, $Interval\text{-}tree_{index}$ is the interval-tree portion and $quad\text{-}tree_{index}$ is the second portion of the DSTree index. Under each leaf, there will be a series of GI objects. Each GI can be stored separately on IPFS and its own CID, GI_{CID}, can be used as a reference to the object itself. So to access a single feature we can use an IPFS Address similar to $DSTree_{CID}/Interval\text{-}tree_{index}/quad\text{-}tree_{index}/GI_{CID}$. Note that GI_{CID} is generated based on the content of the GI by IPFS and to have access to it we need to query it from DSTree.

Table 3.2: An example how IPFS stores a DAG graph (based on the graph in Figure 3.3) and how to request a portion of the graph from the IPFS.

IPFS Address	DAG result	IPFS Address	DAG result
Qmb...R	Returns entire graph	Qmb...R/11	Returns Root/11/
Qmb...R/11/111111	Returns all features under Root/11/111111	Qmb...R/11/111110	Returns all features under Root/11/111110
Qmb...R/10/1011/ Qmb...d	Returns a single feature under Root/10/1011/ leaf	Qmb...R/10/1011/ Qmb...c	Returns a single feature under Root/10/1011/ leaf
Qmb...R/metadata	Returns entire metadata object	Qmb...R/metadata/ key_1	Returns key_1 under metadata

Qmb...R is $DSTree_{CID}$ which is a hash generated based on the root content of entire DAG graph from the DSTree
Qmb...d or *Qmb...c* is a GI_{CID} which is a hash generated based on the content of single GI object.

3.5 Distributed network integration

In order to process queries on distributed networks (in our case IPFS) it is required to store data in a DAG format. We use DSTree to construct a DAG graph and once the tree structure is constructed an IPLD graph is formed from the DSTree graph. Then the IPLD object is stored on the IPFS network and a CID of the uploaded contents is used as a root gateway to access the entire tree structure. DSTree in this system acts as the main index structure to perform and answer spatio-temporal queries. Since DSTree is an indexing structure it does not store the actual GI. It only stores the CID of each individual GI as a reference to the object itself. This provides a lightweight graph that can be used

by each client.

Data Management As mentioned earlier the DSTree only stores a CID reference to the actual GI. The GI can be in any format (e.g., geojson, topojson, or other feature-level standards). In constructing the DSTree each GI first is uploaded on the IPFS as a regular file or IPLD object. Then for each GI, we will then have a set of $(GI_{CID}, GI_{MBR}, GI_{Time-Interval})$. This Object is then inserted into a DSTree and the related DAG graph is then constructed. Once all the GI objects are added to the DSTree the graph structure is converted to an IPLD object and is uploaded on the IPFS and the pair of $(DSTree_{CID}, DSTree_{Metadata})$ will be shared between users.

Metadata Metadata is usually defined as data about data [47]. In order to provide interoperability between different systems it is required to include metadata objects within shared content. In the DSTree $DSTree_{Metadata}$ is used to store information related to the dataset and can include general spatial metadata objects (e.g. see [26, 144, 22]). The following metadata keys (Table 3.3) are necessary in order to provide minimum interoperability when sharing information using DSTree on IPFS.

3.5.1 Architecture and implementation

The proposed method to process queries on IPFS networks consists of four main components. Figure 3.8 shows the flow of the communication between users on a distributed network in order to query, retrieve and store GI. The start of the data sharing process on IPFS is with a user, $User1$, willing to share a GI, $(GI, GI_{MBR}, GI_{Time-Interval})$, on the network. Once the user uploads the GI content on IPFS they use its CID, MBR, and time-interval associated to it, $(GI_{CID}, GI_{MBR}, GI_{Time-Interval})$, to construct a DSTree. In this step, the user is able to keep adding as many GI objects as they want to the DSTree. Once the construction of the DSTree is finished, the necessary metadata is also added to the data structure and an IPLD object, $DSTree_{IPLD}$, is constructed from DStree's DAG graph. The $DSTree_{IPLD}$ then is uploaded on the IPFS network and the related IPFS root hash, $DSTree_{CID}$, and its metadata, $DSTree_{Metadata}$, is retrieved from IPFS. Since the $DSTree_{CID}$ is generated by IPFS based on the content of the DAG graph, we need to share this CID with other users to be able to access the index. In order to share the IPFS hash with other users a smart contract is used. This smart contract is responsible to keep a history of $DSTree_{CID}$ hashes over time and provide the latest $DSTree_{CID}$ hash to the users at any time. In our example, a simple smart contract using Solidity is developed and

Table 3.3: Necessary metadata objects required for sharing DSTree on IPFS

Metadata Key	Type	Possible/Example Values	Description
featureType	String	•geojson •topojson •dggsfeature	To identify the decoder for reading GI_{CID}
intervalKeys	Array[[]]	[[1,10],[5,15],[28,20]]	Temporal interval keys in order to decode *interval tree index* values to the exact Intervals
temporalLevel (T)	Integer	1,2,...,n	The interval-tree level before converting nodes to quad-tree root nodes
featureTypeProperties	JSON	Variable	Each *featureType* might need specific details and metadata, they can be stored as this object.
quadExtent	Array[]	$[-180, -90, 180, 90]$	The extent of the quad-trees. The default value is [-180,-90, 180,90] degrees in geographic coordinates

deployed on the Ethereum test network. This smart contract is used in a web application in order to provide access to the latest version of $DSTree_{CID}$ hash and its metadata when a user visits the web application. Once the $DSTree_{CID}$ hash is added to the smart contract it will be available to all the users who connect to this smart contract.

If a new user, $User2$, accesses the smart contract then they will be able to fetch metadata and the IPLD DAG graph structure related to the DSTree. Then they will be able to replicate a version of DSTree on their own local environment and as a result, they are able to query data from that Index. The result of the query, an array of GI_{CID}s, then are requested from the IPFS through the query process explained in Figure 3.7.

If the $User2$, wants to add a new GI, $(GI2, GI2_{MBR}, GI2_{Time-Interval})$, to the dataset, they first add it to the DSTree and then construct a new IPLD graph and upload the data on IPFS. Since the content of the new DSTree is different from the previous one, a new IPFS hash is generated and the $DSTree2_{CID}$ is returned to the user. In the next step $User2$, connects to the smart contract and adds $DSTree2_{CID}$ as a new block to the underlying blockchain. At this point all the users will be able to access the updated data, $DSTree2_{CID}$, throughout the blockchain. The older version of the DSTree, $DSTree_{CID}$, will also remain on the block history of the blockchain and will be accessible too.

3.6 Discussion

One of the reasons for moving from centralized data-sharing methods to a distributed method is the increasing amount of data which are being collected on a daily basis. Unlike centralized systems, data storage in P2P networks is distributed across network nodes which provides scalability and no single point of failure. However, the management and processing of queries on these networks have always been a challenge. The proposed method to share and query spatio-temporal data on distributed networks tackles this issue by tracking and updating a spatio-temporal index between network users. In this approach, a blockchain is responsible for keeping a history of different versions of a DSTree index. Each user is able to replicate a version of DSTree on their own node and run spatio-temporal queries on the index. Since each user performs the queries on their own side the indexing tree should check fewer items and support more topology out of the box. As shown in Figure 3.6 the number of visited items in the DSTree is generally less than other indexing structures which provides less memory consumption on the client apps. While octree also visits fewer nodes/items during its query process, its tree construction is slower (20%) compared to the other tree indexes and also it takes more time to answer the queries (see Figure 3.5) since the internal intersection functions are 3-dimensional.

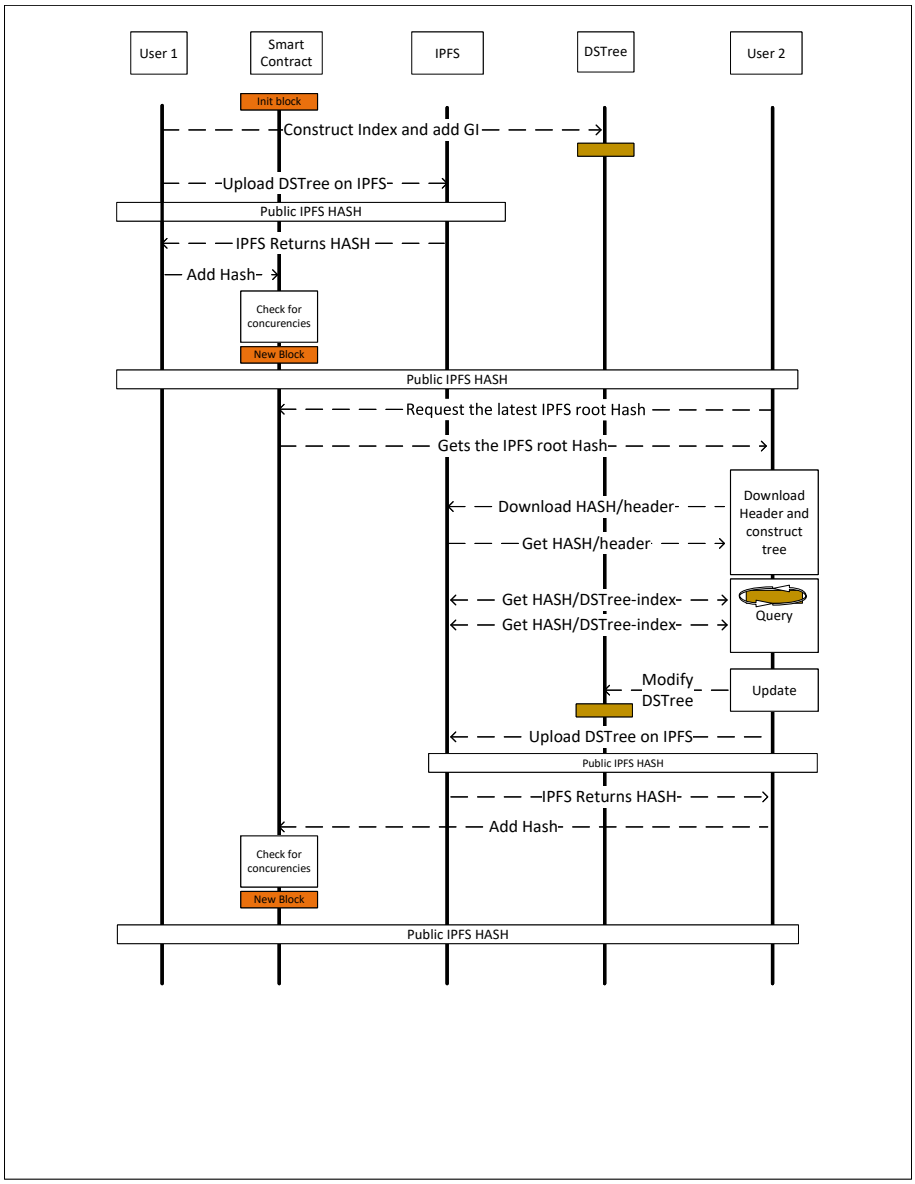


Figure 3.8: Query process and publishing spatio-temporal data on the IPFS. It shows the process of sharing the DSTree index between users using a blockchain, querying the content from the network by another user, and updating the data on the network

They can also grow very fast if the time intervals are large [169]. In a single Interval tree or quadtree approach, the results of the queries need to be checked for the spatial or temporal topologies accordingly during the post-processing stage. In addition, the DSTree can handle 6 main temporal topological relations (see Table 3.1) during the query process without the need to post-process data. Time wise the DSTree also performs well on the small to average datasets (see Figure 3.5). Update, insertion, and deletion of existing data is another requirement for the current data-sharing environments. Due to the use of an interval-tree as the top part of the tree, DSTree is not optimized for the deletion of the items. Inserting a new GI will update only a portion of the tree structure and does not impact the entire DAG graph of the data. However, adding data with large intervals in such a way that the GI temporal interval covers branches from the left to the right side of the interval-tree can cause a restructuring of the entire tree. For example, for time-series data such as sensor data or VGI, the DSTree performs better. In our police department example, the newly reported incidents can be added on top of the DStree and it does not cause restructuring of the entire tree.

During the update process of the DSTree and publishing the latest version of DSTree by users there are cases where conflicts may appear. To resolve such conflicts there are several approaches. The conflict resolution between different versions of DSTree can be done either on the client side before pushing the latest version on the blockchain or on the smart contract before saving $DSTree_{CID}$. Both of these methods require a mechanism to detect and address conflicts. Because of the size limitations on the smart contracts, we are using a client-side conflict detection approach. In this approach, the DSTree graph is extracted from the latest version available on the blockchain and is compared with the version of the graph on the client side [147]. If the conflicts in the DSTree graph are detected (using tools like [98]), the user will need to resolve the conflict and then publish it on the blockchain. However, this approach needs a trustful user interaction with the network.

The reason for using quad-trees is that since the root boundary box of all the quad-trees is constant the quad index part of the DSTree index will always point to the same area in the geographic space over different time intervals. This provides a faster access method and the capability to exchange the quad-tree with a Discrete Global Grid System (DGGS). DGGS grid, similar to quad-tree, provides the same index value per each grid cell in the space. It also provides methods to aggregate data o multi-resolution levels [158, 215, 122] and also provides built-in data locality and directionality of space [219]. For instance, in sharing police department information over time, the reports can be censored using distributed k-anonymity methods on the P2P networks (e.g. [120]) if the DGGS system is used as feature data storage.

Data locality in the IPFS with DSTree Data locality is defined as *locality implies that neighboring multidimensional information is stored in neighboring nodes [159]*. In an IPFS network, once each node downloads some particular content it can act as a data provider. Combining data partitioning using DSTree and sharing data on IPFS at the GI level allows users to request only a small portion of the entire dataset and as a result, they can also serve small chunks of a large dataset. This can provide a data locality in the P2P network based on the user activities in the VGI platforms. In our police data sharing example once the dispatch teams share their location on the network, they have already become a data provider for that shared GI. Once the police officers visit that location, once again they download that GI, and they will become another data provider for that specific GI on the IPFS network. This approach provides a level of data locality for the nodes close to each other since the nodes in a region usually tend to explore data related to their own region. Other examples of use cases of such data locality can be for sharing geo-tagged information in small communities.

DSTree Limitations The proposed DSTree model only supports the spatial topologies that the underlying spatial indexing method supports. In this paper, we only experimented with the intersection topology relation. However, it is possible to perform other topological relations such as overlay, within, and crosses, and also perform KNN-based models. All of the temporal typologies are supported except disjoint. The focus of this paper was support for vector-based data structures. However, supporting raster data could be achieved by converting raster data into DGGS-based models or tiling the raster data and instead of generating lower level spatial index using the multi-resolution tiling structure. Table 3.4 show a summary of supported queries and future approaches to support other data models.

3.7 Conclusions

P2P has become very popular for storing and sharing information in a decentralized approach. The amount of daily spatial data which are being collected and shared in different sectors with different levels highlights the need for P2P data management, query, and processing. In this paper, a new spatio-temporal multi-level tree structure, DSTree, is proposed which aims to address this problem. DSTree is capable of performing a range of spatio-temporal queries. In order to integrate this data structure on the IPFS distributed network, a framework is proposed which uses blockchain to share the IPFS CID of the index and each user is capable of replicating DSTree and querying or updating it. However, this model is not optimized for deletion and mainly is suitable for append-only data over

Table 3.4: DSTree Capabilities for different spatio-temporal data models

Data Model	Functions	Notes
Vector	•Temporal typologies on Table 3.1 •Spatial Intersection •Spatial overlay •Spatial within •Spatial crosses •KNN	KNN is not yet implemented
Raster	•Temporal topology queries •Extent queries •Multi-band data retrieval •Multi-resolution aggregation	Can be achieved using raster tiling methods and constructing a DAG graph or converting data to other models such as DGGS and retrieving them using DGGS DAG graph
TIN	•Extent queries	Can be achieved by converting data to other models such as DGGS and retrieving using DGGS DAG graph

time. In this work, some of the challenges in sharing and querying spatio-temporal data on distributed networks are addressed. Despite the advantages of our proposed framework, challenges remain. We conclude that more significant research effort from GIScience and related fields in developing decentralized applications is needed. The need for the standardization of different feature types and feature Type Properties when sharing data on the IPFS network is one of the requirements. The possibility of using IPLD objects in sharing GI at the feature level can provide finer access to the information. In addition, it is necessary to address attribute-level query processing too, which is not covered in the current work. The use of the smart contract to control access of the users to read and write data to the main chain can also be studied. This access control can even be at the feature level.

3.8 Data and code availability statement

The test dataset for the occurrence of the crime from the Waterloo Regional Police Service is available at (https://www.wrps.on.ca/en/about-us/reports-publications-and-surveys.aspx). Other data used in this paper are available upon request from the corresponding author. The prototype application, screenshots, and the source code for the app and its online version are available on GitHub at https://github.com/am2222/dstree.

Chapter 4

Decentralized geoprivacy: leveraging social trust on the distributed web

Abstract Despite several high-profile data breaches and business models that commercialize user data, participation in social media networks continues to require users to trust corporations to safeguard their personal data. Since these data increasingly contain geographic references that allude to individuals' locations and movements, the need for new approaches to geoprivacy and data sovereignty has grown. We develop a geoprivacy framework that couples two emerging technologies -decentralized data storage and discrete global grid systems- to facilitate fine-grained user control over the ownership of, access to, and map-based representation of their data. The framework is illustrated with a dynamic k-anonymity model that links the geographic precision of shared data to social trust within in a social network. In this framework, users' spatio-temporal data are shared through a decentralized system and are represented on a discrete global grid data model at spatial resolutions that correspond to varying degrees of trust between individuals who are exchanging information. Our framework has several advantages over centralized geoprivacy approaches, namely, trust in a third-party entity is not required, and geoprivacy is dynamic and context-dependent, with users maintaining autonomy. As the distributed web emerges, so too can the next generation of geographic information-sharing tools.

Keywords: socio-spatial networks; decentralized geoprivacy; distributed spatial data sharing; discrete global grid systems;

4.1 Introduction

Socio-Spatial Networks (SSN) have gained considerable attention in recent years due to their ability to reveal underlying relationships between social and spatial entities in the network. Socio-Spatial Networks have been used to explore the connectedness and social interactions between entities in a network including people and their relation to their surroundings [223]. In this chapter, we introduce a new algorithm (*dynamic k-anonymity*) that encapsulates these properties of geographic information sharing within a new computational framework for geoprivacy (*decentralized geoprivacy*).

In a datacentric sense, a social network consists of a virtual social graph where users (nodes) are connected through relationships (edges) [132]. In a socio-spatial network, nodes (and sometimes edges) on a graph are encoded with geographic information that offers the possibility of linking social interactions with the geographic context(s) in which they occur [67, 223]. Many studies have demonstrated the importance of online SSN data from platforms such as Facebook, Twitter, and Instagram for analysis of activity spaces [155, 269], health and places [50, 251], and crisis/disaster response, among others, [45, 59]. A core critique of using online SSN data for research or commercial purposes is that they often fail to adhere to the principles of informed consent [123]. Few users read or understand the implications of Terms of Service agreements and most are not aware of potential repurposing of data by platforms or third-parties. Privacy settings in social media apps that are limited, obscured, or favour platform providers accentuate this problem [254]. New tools are required to specifically address geoprivacy concerns and empower users to have greater control over their data [141]. To clarify our use of terminology, we denote socio-spatial networks in general as *SSN*s and online variants as *online SSNs*. Specific graph relations derived from an SSN we term a *SSN Graph*.

Sharing geographic data in online SSNs can take many forms. In addition to explicit sharing of spatio-temporal data / coordinates, more nuanced inferences can be derived from spatio-temporal data that violate users' geoprivacy, such as extracting daily routes from content of historical text messages. A recent study investigating location privacy found that 44% of participants did not want to share their location with online SSNs (e.g., Twitter, Facebook, Instagram), 66% did not want to have their location stored, and approximately 88% of participants were opposed to third parties deriving inferences about them based on their activities [173]. Further, Urban et al. [258] found that 92% of participants did not want their location to be used for ads, and 70% did not want spatio-temporal data to be saved on their own devices. Despite these geoprivacy concerns, the majority of users do not have transparent access to the data that flows between their devices' operating system, their apps, and online SSN platforms [173]. Even when privacy-preserving options

are available, it is unclear how effective they are at preserving geoprivacy. For example, some studies have found users were generally unaware of sharing settings [88, 5, 161], while other studies [19, 139] found that many users did understand when and where they were sharing their geographic information. Fine-grained control over the resolution of spatio-temporal data being shared and with whom they are sharing is not provided by most geoprivacy approaches, which tend to rely on simple binary classifications (e.g., in my friend network or outside of it)[161]. We suggest that the social and spatial properties of online SSNs provide an opportunity to develop personalized geoprivacy configurations that may enhance individuals' ability to reassert sovereignty over their data and geoprivacy. In this chapter, we seek to explore these issues by addressing the following research objectives:

1. Developing an approach to geoprivacy for a distributed web environment,
2. Proposing *dynamic k-anonymity* as a model that provides users with fine-grained control over geographic information sharing,
3. Proposing a geographical trust model for dynamic geoprivacy, and
4. Examining the challenges and opportunities of handling geoprivacy in the distributed systems.

4.1.1 Geoprivacy as a concern in socio-spatial networks

Information about people's location is very different from other types of data that can be collected from online SSNs. Geographic data, whether explicit coordinates or less precise place- or activity-based references, allow SSN content to be linked to the spatio-temporal context it was produced within as well as to ancillary data external to the online SSN [245]. According to Duckham et al. [69] *geoprivacy* is defined as "a special type of information privacy which concerns the claim of individuals to determine for themselves when, how, and to what extent location information about them is communicated to others. In short, control of location information is the central issue in location privacy".

Kebler and McKenzie [141] recently reviewed how spatio-temporal data can be used to infer sensitive information about the social, economic, and political behaviour of users. For online social network providers, combining users' spatio-temporal data with social network data greatly increases the value of their data and the quality and relevance of location-based services [13]. As a result, most online social networks tend to store user location frequently. *Twitter Trends* for example, incorporates user location as an input to their algorithm to ensure detected trends are relevant to users' spatial context [255]. Additional services such

as routing and traffic monitoring are also highly dependent on aggregating users' near real-time locational data over time, often without their knowledge [258]. Spatio-temporal data can also be revealed through more malicious tactics such as de-anonymization methods, which employ location-based inferences that reveal information from users' behaviour and movement patterns [141, 216]. Such approaches can be counteracted as Zakhary et al. [293] demonstrate, with a service that shows Twitter users random topics linked to different locations so that they can obfuscate their location and avoid location inference using their tweets.

An alternative to the current paradigm of opaque user agreements as the basis for privacy preservation (i.e., opting out of services) would see users sharing their location through explicit release or enablement of their spatial location via GPS sensors or more generic place-references via granular application and user-specific mechanisms [216, 6]. To realize this, two changes are required to the current status quo of online SSN usage. First, new methods for preserving geoprivacy are required that allow users to control the degree to which they share location data and with whom on online SSN. Second, and more fundamentally, online SSN providers must return control of personal data to their platform users to truly limit data repurposing and to reduce the impact and scope of data breaches. We contend that the majority of geoprivacy research in the GIScience community has centered on the first issue and the tangible actions that individuals or responsible agencies can take to protect personal geoprivacy [8, 42, 301, 245]. Examples include work by Kounadi and Resch [151] who provide guidelines for sharing geographic data and preserving geoprivacy or Kounadi and Leitner [150] who identify research papers that display confidential information as a way to draw attention to geoprivacy in academic research. Comparatively few solutions address the issue of user control on a fundamental network-wide and privacy-by-design basis [33]. This is particularly important as critical big data studies have highlighted how ownership of data and the power dynamics embedded in our relationships with technologies and SSN service providers can act to reproduce patterns of inequality and divergence [229, 54].

4.1.2 Models of geoprivacy

Different models for geoprivacy preservation during spatial data sharing have been proposed over the last two decades. Location obfuscation mechanisms are models that aim to protect privacy by deliberately degrading the precision of location information in a way that the service can still be carried out to some acceptable extent without revealing an individual's true location [202]. An overview of the key approaches and their characteristics is provided in Table 4.1.

Method	Details	
	Key Characteristics	Examples
Dummy locations	A user reports false locations.	Fabricating locations [289]
Jittering locations	Adding noise to location data. It includes adding random noise to the entire dataset or a portion of it.	Adding noise to trajectories [143, 290], Location jittering for health-related data [87, 17]
Space transformations	Transforming to another space while preserving the statistical and spatial relations inherent in the source data [11, 156, 288]	One-way transformation [94]
Spatial cloaking	Reducing the spatial and/or temporal precision of individuals' data. This can involve point aggregation (e.g., replacing a set of points by one individual polygon or a point such as mean center) or representing a point location with a false point location that lies within an area-based geo-mask.	Geo-masking [8, 298], Adaptive geo-masking, Voronoi-based aggregation system [51], Triangular displacement [186] K-anonymity, [221, 246]

Table 4.1: Summary of Geoprivacy models.

Some of the models in Table 4.1 are mainly developed for privacy preservation for the purpose of sharing research data, and not all are suitable for spatial data sharing in online platforms. For example, reporting false locations cannot be used for a location-based service and as a result spatial cloaking has received more attention than other methods in preserving geoprivacy in online platforms [95, 150].

Privacy models in online SSNs have been studied intensively in recent years. These models have proposed different privacy models such as access control [200, 157], relationship-based models [37], and trust-based models [271]. These models mainly use social network metrics and user network qualities to determine the level of access others have to certain information in the online network. For instance, Voloch et al. [266] present a dynamic access control system for SSNs based on trust-based metrics extracted from user attributes. The attributes used to derive such metrics for trust differ depending on the context of the network. The more common attributes which are used in such models are the total number of friends, age, followers/followers rate, mutual friends, friendship duration, similarity indexes, user interactions such as comments, private messages, and so on (i.e., see works by Misra et al. [180] or Gudes and Voloch [113]). For online SSN data, there may be potential to adapt trust-based methods to geoprivacy protection in spatial data sharing.

Another key drawback of most of the methods outlined in Table 4.1 is that they are mainly developed within the context of privacy preservation in centralized frameworks where data sharing is controlled by the platform provider and/or associated third-party entities. The requirement for a centralized body to control access to users' personal geographic data has inherent limitations. Given eroding public trust in corporate social media platforms [199], there is an urgent need to develop alternative socio-technological arrangements that provide user-control and ownership over their data. Other concerns stemming from centralized architectures inherent to current SSNs include the transparency of the central platform and providing a single point of failure [159]. Moving from online SSNs on centralized architectures to a more distributed architecture may help to resolve some of these core issues underlying current challenges with realizing geoprivacy [3, 191].

4.1.3 From centralized to distributed online SSN

The issues of ownership, transparency, and trust discussed above are not confined to personal geographic data in online SSNs, but rather underpin almost all services and activities that take place on the Internet. There is renewed interest in the idea of decentralized or distributed web (i.e., *the dWeb*) models where decentralized apps, platforms, and services interact through peer-to-peer networks [28]. Decentralized social network platforms such

as federated online social networks or peer-to-peer (P2P) networks in which users can be their own data storage provider or choose a data storage provider to store their data [226]. Peer-to-peer networks have shown a capability to provide scalability and transparency [281]. Some examples of existing P2P online social networks include Arbore, Peergos and Peepeth. FireEagle was an early example of a centralized location access management platform that Yahoo offered from 2008 to 2013 to store a user's location and share it with other authorized services [64]. *FireEagle* allowed users to choose who to share their location with and to decide what scale their location was shared at.

Studies examining issues of geoprivacy in non-centralized environments such as P2P networks are limited. Chow et al. [39] proposed a framework that masks user locations by grouping it with nearby P2P network users, ensuring an individual user cannot be identified among their k-neighbours in the group. Later the same authors enhanced their model for mobile P2P networks to preserve geoprivacy without the need for any central servers [40]. In this model, an information-sharing scheme for such networks to share spatio-temporal data between socially close nodes and a historical location caching scheme is developed. Using these two schemes, a distributed P2P algorithm cloaks a user's location when other network users want to query the user's data from the network. X-regions is another model developed to enable mobile users to share an aggregate location reference which can anonymize a user's location among other anonymous users [34]. This model is based on the k-anonymity model that considers the size of the x-region and the density of the users as parameters. These models are mainly static and do not allow individuals to safeguard their geoprivacy dynamically in response to context (e.g., emergency response versus online survey) or to differentials in trust between recipients of their data (e.g., communication with friends versus commercial firms).

4.2 A new framework for decentralized geoprivacy

We propose that a *geoprivacy-by-design* approach for SSNs is underpinned by three framing requirements. First, users should be able to define local control of how their spatio-temporal data are shared with others in an online SSN. Second, users' decisions to share geographic or other personal information should not be construed as a one-time decision or app setting after which control is forfeited. Instead, users should retain autonomy to define the social, spatial and thematic contexts that personal (geo)information is shared and be able to alter sharing parameters in response to changing needs or trust relationships. Third, variable and dynamic geoprivacy within an online SSN requires a data model that has native support for a multi-resolution representation of users' spatio-temporal data anywhere on the globe.

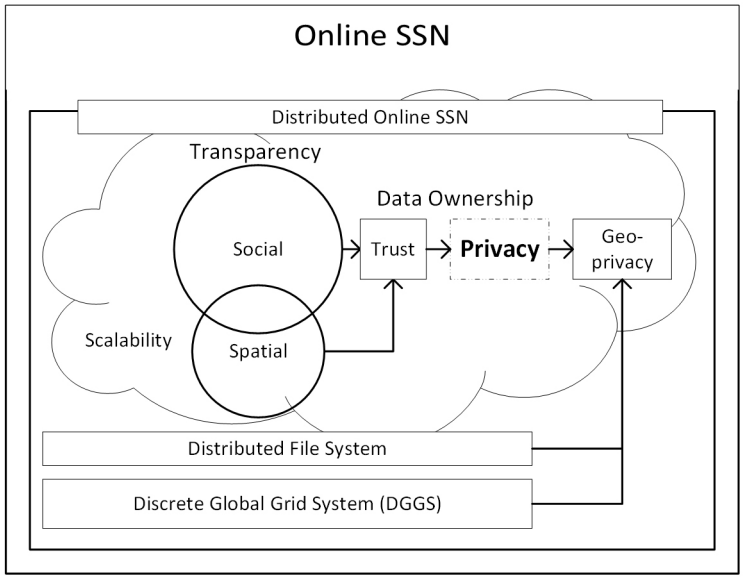


Figure 4.1: Components of a dynamic geoprivacy model based on trust in a distributed SSN.

We suggest that one way that these requirements can be met is through a decentralized geoprivacy framework as illustrated in Figure 4.1. The first component is a discrete global grid system (DGGS) which is used as the geographic data model. A DGGS is an emerging data model for big data GIS capable of embedding uncertainty into the grid resolution and thereby providing a robust mechanism for cloaking spatial locations. The second component of this framework is a distributed file system approach to storing SSN data. A distributed approach for online SSN eliminates the need for a centralized entity to store and control access to the user's data. We will briefly review the two core technologies making up the framework in Figure 4.1, DGGS and distributed file systems.

4.2.1 Discrete global grid systems

A discrete global grid system (DGGS) represents locations on the Earth's surface through a series of multiresolution and hierarchical equal-area gridcells. In a DGGS, each cell has a unique ID, which we term *dggid*, and a specific resolution based on the cell size that represent spatial measurement uncertainty [220, 166, 107]. The main characteristics of a DGGS are: i) tessellation geometry, ii) aperture iii) method of cell indexing, and iv)

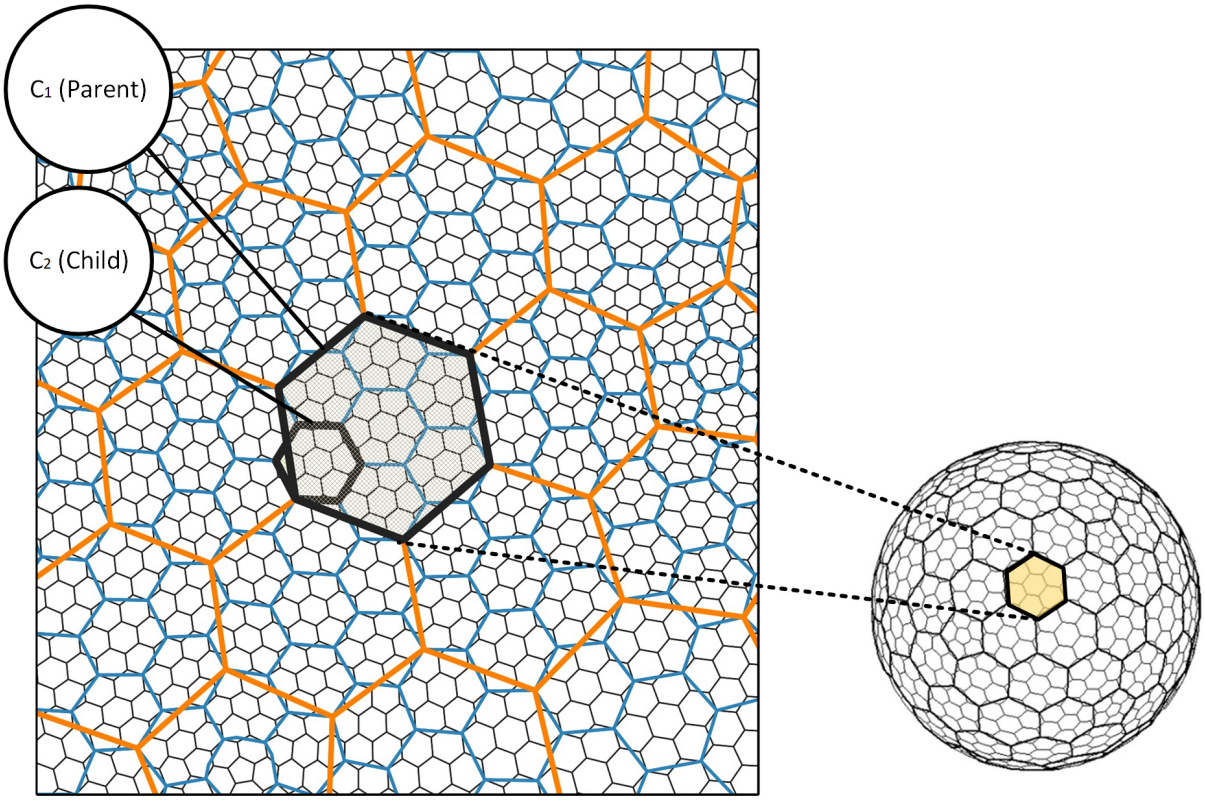


Figure 4.2: A hierarchical structure of grid and parent and child relation of the cells in an aperture 7 DGGS grid.

quantization strategy and associated mathematical functions [209]. We utilize a DGGS as a multi-resolution data model, consisting of sets of parent and child cells. Each parent can be constructed from a set of child cells, and the ratio of children to a parent is known as the system's aperture [167]. A cell C_1 is a parent of another cell C_2 if the resolution of C_1 is lower than the resolution of C_2 , and if C_2 is a member of a set of sub-cells (children) linked to C_1 (see highlighted cells in Figure 4.2). The relation between parent and children cells can be one-to-one (for aperture 7) meaning that each child is only covered by one parent or one-to-many (for aperture 3) meaning that a child can have more than one parent in a lower resolution.

We use a hexagonal-based DGGS with aperture 7 (using H3-js library developed by Uber [256]) as the spatial data model. In our proposed decentralized (online) SSN model,

a DGGS provides two levels of geoprivacy protection. First, geographic coordinates are replaced with *dggid*s as the base unit of geographic information which provides a base level of geoprivacy that lends itself to a wide range of encryption methods. Second, due to parent-child relationships embedded within the DGGS, aggregation to higher levels (i.e., parent cells) of the data model provides the ability to develop spatial cloaking algorithms within the system.

4.2.2 Interplanetary file system for decentralized SSNs

The InterPlanetary File System (IPFS) is a protocol based on P2P networks that is used for distributed data storage and sharing. IPFS utilises content-based addressing to identify files in the global P2P network [18]. In the IPFS network, each user keeps a part of the overall data and shares the content address of the files that they have. Other peers in the network can find data owners using the content's addresses with a distributed hash table (DHT). Figure 4.3 shows, from the top left, a user providing content on the network which is subsequently hashed and stored in a DHT under the given hash. Each node in the network is responsible for updating the DHT. A user then requests content (Figure 4.3, bottom right). Nodes in the network query the DHT and return (portions of) the file if they have it. If they do not have the file, they request it by communicating with peers that are *closer* to the requested content hash. In P2P applications and platforms, DHTs are a key technology for facilitating information exchange due to their robustness and scalability [176, 140]. Several P2P projects utilize DHTs to distribute the data between peers to help storage in the users' community [306].

IPFS is designed to share information on the network without built-in support for data access control. Access control mechanisms in P2P systems can be easily implemented via encryption keys, however, some limitations such as effective encryption key storage and management or the need for availability of encryption key owner's node for decryption remain in practical applications. Despite this, many new and emerging data sharing applications are turning to distributed networks and data storage (e.g., social networks by Kapoor et al. [137], geo-collaboration systems by The Gordon Foundation [85], professional communities medical by Dagher et al. [53]).

4.2.3 Dynamic K-anonymity based on decentralized geoprivacy

K-anonymity was first proposed by Sweeney [246] and has been widely studied for privacy purposes in many sectors including geoprivacy (see Table 4.1). The primary purpose of

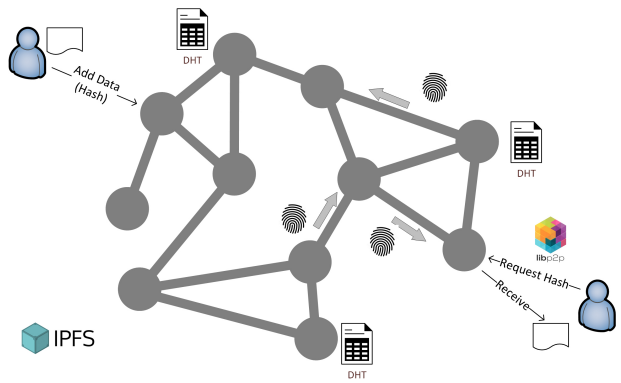


Figure 4.3: A simplified IPFS network and its components. A user adds data to the network and its HASH will be stored in a DHT. Another user request data and nodes search the DHT and will return the requested file. (Fingerprint icons means a secure P2P connection).

spatial k-anonymity is to anonymize a location among k other locations in such a way that the true location cannot be distinguished [150, 92].

Anonymizing the location before sharing:

In our model, the DGGS grid is used to anonymize a user's location in a social network based on k peer locations. Each user defines two parameters, k_{min} and k_{max}. The former is the k for the model to share location with a user with the least level of trust, and the latter is the k value for k-anonymity algorithm to share location with the user with the highest level of trust. Other users will have access to a varying value of k between these two values based on the level of trust assigned by the data owner (O). When sharing p data with locational content, (O) requests at least $k-1$ locations (as $L = \{L_i | i \in [k_{min}, k_{max}]\}$) from network peers. In this step, each peer shares their location, as DGGS cell IDs, with the level of privacy they define individually in the network. L_o (owner's location) and L_i are DGGS cell IDs each with a resolution (r) assigned to them. The function $f(x)$ returns the resolution of cell ID x, and $p(x, r)$ returns the parent cell of x in resolution r (i,e., $r \leq f(x)$). These functions are defined based on DGGS grid definition and vary per each DGGS. Once (O) retrieves k_{max}–1 other locations from the network, the following steps are applied to construct a k-anonymity algorithm,

1. Sort all the members of L by the distance from the data owner's location (L_o);

2. Compare each pair of (L_o, L_i) and assign a resolution value (r_s) to each pair. To do so we first need to define a set of resolutions of the common parents called *hierarchical distance*, D, as:

$$D = \{r | r \in S \wedge p(L_o, r) = p(L_i, r)\} \tag{4.1}$$

then

$$r_s = max(D) \tag{4.2}$$

where r_s is the resolution of the smallest common parent of two DGGS cells since in a DGGS a larger resolution grid contains grid cells with a smaller cell size. The parent of L_o at resolution r_s is defined as the cloaking region for L_o and L_i. S is an integer set of all possible resolutions in DGGS (e.g., a range of $[0, 15]$ for Uber H3 library). Note that L_o and L_i may not be cell references in the same resolution. In some cases finding a common parent may yield a very large cloaking region, and we may want finer control and lower information loss while still providing guarantee of locational privacy. In these cases we can define the cloaking region based on a set of connected neighbours of cells at a resolution higher than r_s. To do this we introduce a new parameter b termed *level jump*, which controls the information loss. If the parent cell in resolution r_s, which covers both L_o and L_i has a larger resolution difference with the minimum resolution of L_o and L_i than b, then we use the neighbouring cells in the resolution $r_s - 1$ as the cloaking region, otherwise it uses the common cell in r_s. So r_d is defined as:

$$r_d = |min(f(L_o), f(L_i)) - r_s| \tag{4.3}$$

where r_d is the minimum of the resolution of the origin and the resolution of the candidate location minus the resolution of the shared parent cell. This measure will give us an estimate of information loss. A higher r_d value is associated with a larger DGGS cell size and higher levels of location cloaking and information loss. If the result of equation 4.3 is higher than b we use cells in $r_s - 1$ resolution as follows:

$$n(L_o, L_i) = \begin{cases} p(r_s, L_o) & \text{if } r_d < b \\ p(r_s - 1, L_i) \cup p(r_s - 1, L_o), & \text{otherwise} \end{cases} \tag{4.4}$$

where $n(L_o, L_i)$ is the cell IDs of the cloaking region. Equation 4.3 can be ignored here (i.e., $n(L_i, L_o) = p(r_s, L_o)$) because it is only used to control information loss, but will result in higher information loss (See section4.5). Equations 4.1-4.4 pertain to finding the cloaking region for a single location/user origin (finding up to k_{max} neighbours). By running $n(L_o, L_i)$ per each L_i we will end up a set with an equal number of members as L.

3. Grouping L_i locations based on the assigned $n(L_o, L_i)$ output, and assign each category a unique ID called the priority index (P_i). This index is used to determine different levels of anonymity, which can be achieved based on the locations of peers. For example, in Figure 4.4c the k value between 4 and 6 will result in the same (P_3). So, users who have access to a k value in this range will all have access to the same geographic extent. This index is also used for caching spatio-temporal data between users in the IPFS network. As a result, users don't have access to peer locations (L_i) directly and thus decreases the chance of inferring the exact L_o from the cloaked locations.

Figure 4.4 shows different steps for applying k-anonymity algorithms on data converted into a DGGS model. In this example, we suppose that the source user (highlighted hexagon) wants to share their location in a P2P network and has set their privacy settings as $k_{min} = 4$ and $k_{max} = 12$. In the first step, the user requests location of its peers (Figure 4.4a). The user needs to receive at least 11 ($k_{max} - 1$) locations. After getting the peer locations, which may also be obfuscated locations, the algorithm finds a lower resolution DGGS cell that covers all 12 ($k_{max} = 12$) received locations (Figure 4.4b). The user's information is published with this level of geoprivacy. In the next step, another user (called target user), asks for a finer location from the source user, and based on the trust value assigned to the connection between them, a $k = 6$ is selected. Then the first 6 closest cells, $\{L_1, L_2, L_3, L_4, L_5, L_6\}$, are used to find a DGGS cell (P_4) which covers them. It is worth noting that if $k = \{5, 6, 7\}$ is assigned to a user based on their trust level, all of them will have access to the same geographic extent as P_3 (Figure 4.4c). For another target user with lower trust value than the target user $k = 10$ will be used and 9 closest locations, $\{L_1, L_2, L_3, L_4, L_5, L_6, L_7, L_8, L_9\}$, will be considered for obfuscating location (Figure 4.4c). For any user with trust value equal to 0 (no trust) $k = 12$, which is the default k_{max} value based on a user's setting will be used (Figure 4.4e). It worth noting that after running the model we only need to store a list of (P_i, $dggid_i$, k_i) values. In this way, the algorithm selects the proper k based on the trust score and returns its corresponding set of *dggid*s.

4.3 Architecture and implementation

4.3.1 Sharing information with geographic location

The proposed framework consists of two layers of geoprivacy. At the data level, the framework uses the DGGS model to apply an uncertainty boundary for each user's location. At

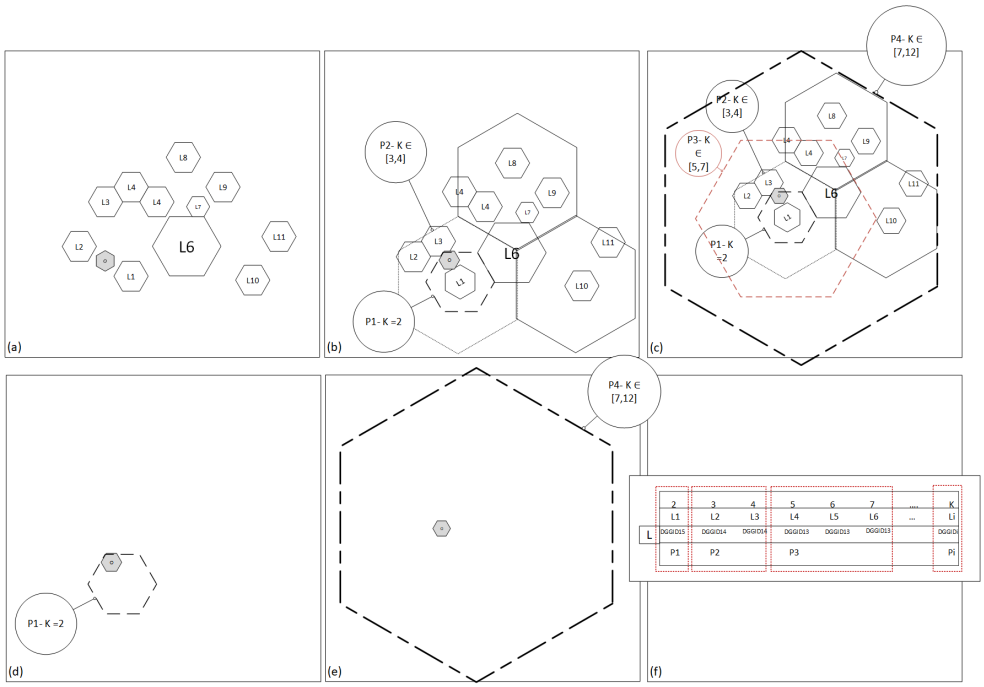


Figure 4.4: K-anonymity model based on DGGS. (a) a user (shaded hexagon) with a $Kmax = 12$ requests peers' locations, 11 *dggid*s is required for it. (b) a k-anonymity model with $k = 2$ and a $k \in [3, 4]$. (c) a k-anonymity for $k \in [5, 7]$ and $k \in [7, 12]$. (d) User's location (L_o) and its cloaking hexagon as it would appear to others with maximum trust (minimum k). (e) User's location (L_o) its cloaking hexagon as it would appear to others with minimum trust (maximum k), and (f) K-anonymity model calculation based on DGGS model. $L_o$is the user's location and L_i are the location of its peers. $dggid_c$ is the cloaking cell for Lu and L_i and P_i is priority index. Note that this image is just for demonstration and in actual DGGS we will not have overlapping cells in the same resolution, and the parent and child coverage will be different depending on the aperture.

the user level, each user with a different ranking and trust level has access to a different level of obfuscated location. Figure 4.5 shows the process of publishing information on the platform and storing spatio-temporal data. The first step in this process is to store the trust values on the network. This step happens when a user initializes a connection with another user in the SSN network for the first time. The trust matrix will be calculated for the user and the order of all the contacts in the user's contacts list will also be updated.

4.3.2 Geographic and social trust for geographic information sharing

The dynamic k-anonymity algorithm defines different values of k per each user based on users' trust relationships within the online SSN. In this framework, a trust score between pairs of users is calculated based on their social and spatial relationships. In some studies, the closeness of a pair of network users is calculated based on parameters such as friendship duration, mutual friends, etc. [265]. Due to the nature of SSN networks, such models must also consider the spatial aspect of an SSN trust score. In this proposed model, users and their relationships are defined as a directed graph, which defines users as nodes and their connections as edges of the graph. An SSN-directed graph can be defined as $G = (V, E)$ where $E \subseteq \{(i, j)|i, j \in N \wedge i \neq j\}$, V is a set of nodes, E is a set of edges and N is natural numbers set. In our case, a node is a user in an online SSN and the edge in the graph is the friendship relation between two users. In a directed SSN graph $e_{ij} \neq e_{ji}$ (e.g., Twitter's following connection), while in an undirected graph friendship connection $e_{ij} = e_{ji}$ (e.g, Facebook friendship connection). After constructing the graph, a trust score matrix, (T), is constructed with elements defined as:

$$t_{ij} = (\alpha \times s_{ij} + (1 - \alpha) \times d_{ij}) \quad \rightarrow \quad 1 \leq \text{i,j} \leq n \quad \wedge \quad 0 \leq \alpha \leq 1 \tag{4.5}$$

where t_{ij} represents the trust score between users i and j. The nodes in this graph don't need a geographic location assigned. Similar to Sarkar et al. [222], a balancing parameter α is defined in the model to balance the weight between social or spatial components. The term s_{ij} is the inverse social distance between two nodes scaled between 0 and 1. This can be calculated using different existing metrics like models developed by Moreno et al. [184] . Trust is a complex multidimensional construct that cannot be defined by social and spatial metrics alone. However, in this section, we aim to emphasize the potential role of location in estimating a trust score and how that can be incorporated into a dynamic geoprivacy method. More complex methods of estimating trust could be implemented here within this framework. Here, a standardized friendship duration is considered the social aspect

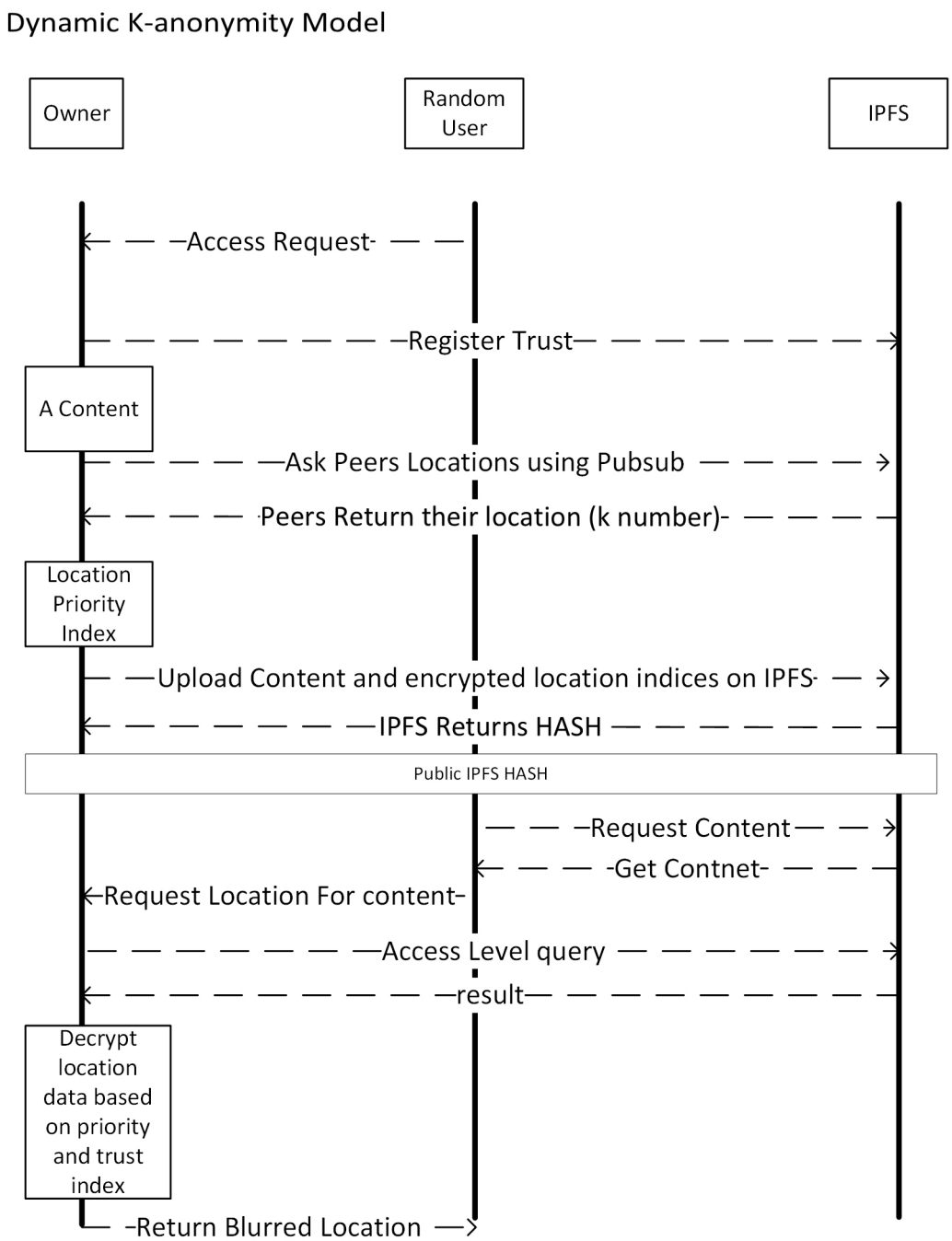


Figure 4.5: Publishing and storing spatio-temporal data on the platform. It shows the process of sharing social network content on the IPFS network, requesting the content from the network by another user, applying a dynamic k-anonymity algorithm on the data, and accessing a finer location by another user.

of the relationship between two users. The longer two users are in contact, the higher their social trust is. The term d_{ij} is defined as the normalized inverse spatial distance. To calculate the spatial connection strength within the variable resolution DGGS data model, a vertical distance is defined. The main assumption for defining a vertical distance is that the spatial component of trust between two users depends on both the Euclidean distance between source and target nodes and also the spatial scale at which they are sharing data. Here, we encode spatial granularity as a key dimension of geographic trust in that users sharing information at higher spatial resolutions have a greater distance to a fixed external node than if they were sharing information at a more generic level. This conceptualization ensures greater trust for more precise spatio-temporal data.

In order to calculate vertical distance in a DGGS grid, first we find a common parent of the locations of the source and target nodes. If there is a common parent between two cells, the vertical distance is defined as difference between source node's resolutions on the grid and a common parent's resolution. If both cells are in the same resolution, the vertical distance would be less than when both cells are at different resolutions. Figure 4.6 shows all possible cases for different vertical distances.

There are cases when two nodes are too far from each other and do not share any parents. In this case, distance is defined as the product of the vertical and horizontal distance. The geographic distance, in this case, is equal to the resolution of the source node multiplied by the horizontal distance between parents of both source and target cells in resolution 0.

After calculating the distance between two nodes for each source, the inverse spatial distance is normalized (between 0 and 1), and this is used to calculate the trust value using equation 4.5. The resulting rank is used as an indicator of the trust level of the users. The first time a user subscribes to another user's feed in a distributed social network, their trust value will be stored in each user's database. However, the trust level must be updated regularly since user interaction is dynamic. As mentioned earlier, the proposed equations can be modified to provide more complex and customized trust models. The tuning parameters such as β or γ provide more flexibility for the users, or while calculating d_{ij} we can add other parameters to modify the trust level by emphasizing vertical or horizontal distances (e.g., $d_{ij} = \beta \times D_v \times \gamma \ \times D_h$).

4.3.3 Geographic data sharing scenarios in a distributed SSN

In distributed systems, since the P2P network is fully dependent on the nodes themselves (not a centralized service), it is possible that communication between peers can break

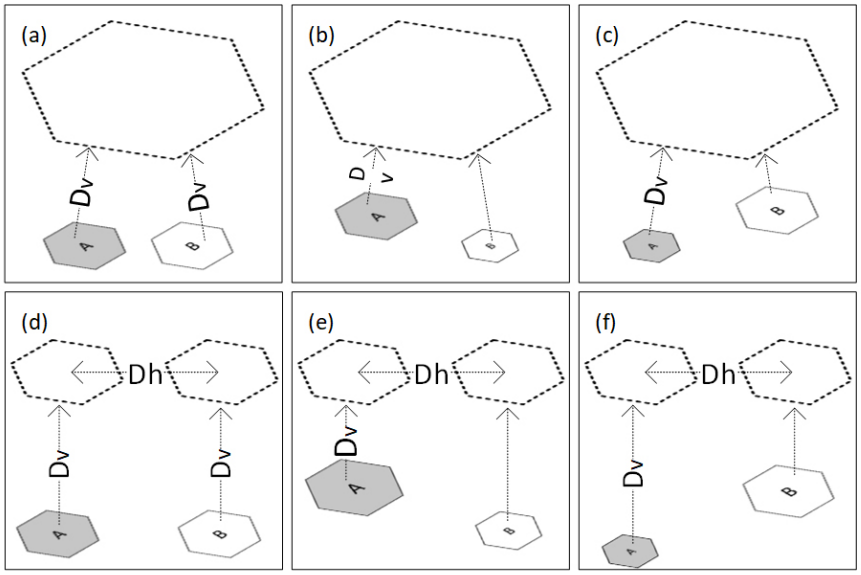


Figure 4.6: Different cases for the distance between two nodes in the DGGS grid. A is the source node, and B is the target node. (a) both source and target nodes are both in the same resolution. (b) The source node has a coarser resolution. (c) The source node has a finer resolution. (d) when source and target nodes don't have a common parent, the final distance equals to $D_v \times D_h$.(e) when nodes don't have a common parent, and the source's resolution is coarser, the final distance equals to $D_v \times D_h$. (f) nodes don't have a common parent. Final distance equals to $D_v \times D_h$.

down to some degree. In the following, we discuss these possible scenarios and propose some reasonable solutions for addressing them.

Scenario 1: When the user (O) as the owner is offline

The proposed framework depends on the owner's availability to decrypt and retrieve geographic data at a fine resolution to share it with the requester. To solve this issue a caching system is defined such that, when the owner of a given piece of data is not available, other users who have already retrieved the same data can respond to the requests. When a user asks for finer geographic data on the network, the request is sent first to the owner. If the owner of the content is online it will return the priority index and higher resolution geographic data based on the given trust level (t in equation 4.5). The content requester will then cache the response from the owner for that content. If in the future, another user asks for the same content and fails to get a response from the owner, the request will be redirected to the entire network. In this case, all network users get the content hash and, if they have it cached, they will query the trust level between the owner and requester. If the priority index (P_i) is the same for the caching user and requester (i.e., the trust level between them is equal or higher), they will respond to the request with the cached geographic content. Figure 4.7 illustrates the steps for caching more precise locations by the users when Scenario 1 happens. This solution requires that peers respond honestly to queries for other peers.

Looking at Figure 4.5 and Figure 4.7, shows that the data owner firstly encrypts data and hosts their data on their own node. When another user requests their location information, they will access the encrypted data and they also become another host of the location data of the data owner. The second user only shares data with the other users with the same or higher trust level.

Scenario 2: When user (O) cannot fetch k_{max} number of locations to share contents

A k-anonymity algorithm needs at least $k-1$ records. In the P2P networks the number of the available spatial data depends on the number of online, so there can be moments that there are not enough available locations around the user. This can be either when the k_{max} value is very high or the number of online users is very low. In both situations, the user cannot publish content with geographic data. To solve this issue, the history of previously shared locations based on time is stored by each user. Every time a user requests spatio-temporal data from the network, each peer shares their location with a timestamp. These

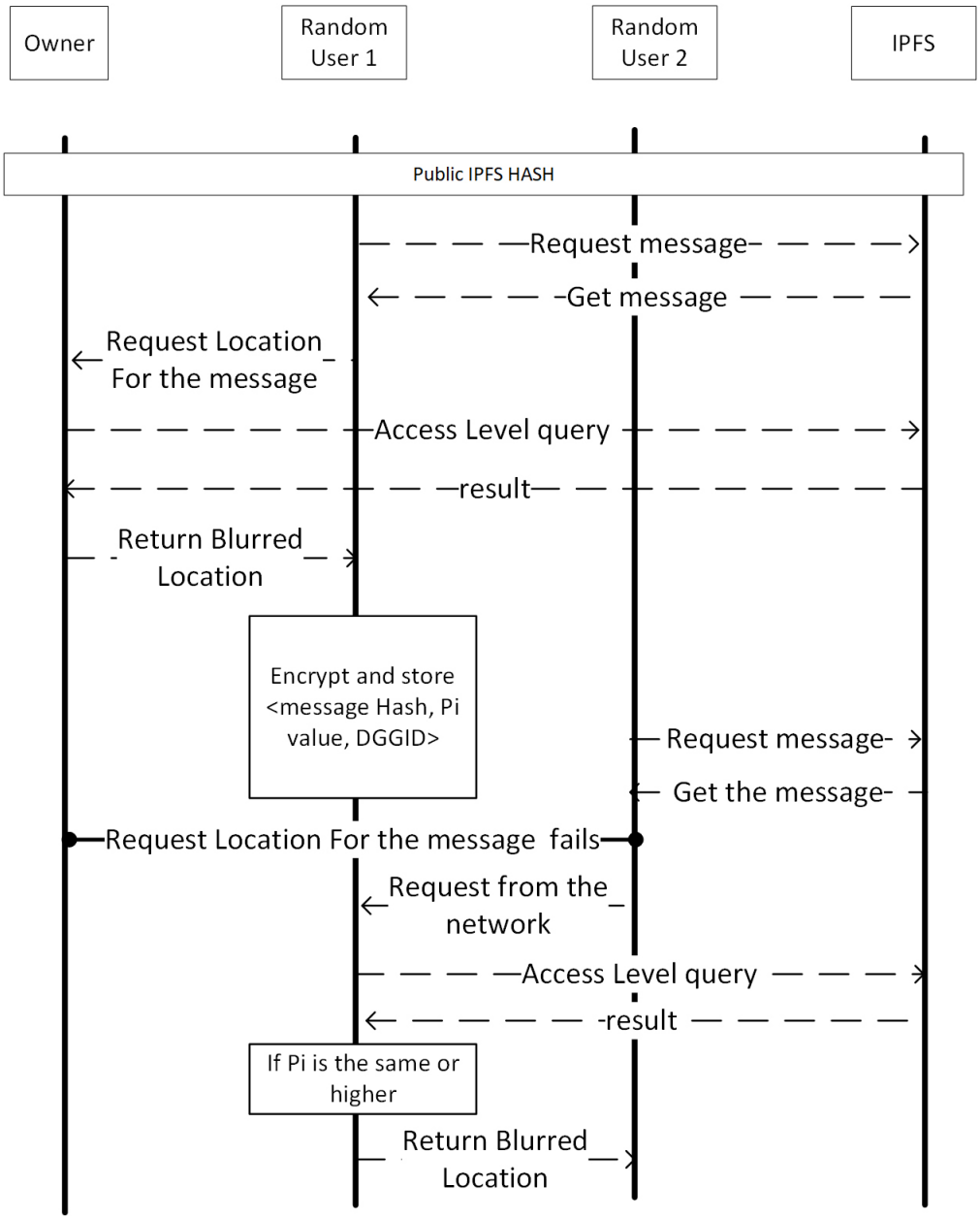


Figure 4.7: The process of caching more precise locations by the peers on the distributed network when data owner is offline.

location data and timestamps are not connected to the peers so other peers also can listen to them and store them. When a user wants to share content, and if they cannot retrieve enough shared locations, it uses its cached results from previous requests made by other users based on their timestamps and distances.

4.4 Case study analysis

To demonstrate how the proposed model could enhance individuals' geoprivacy, freely available Twitter data that collected by Lamsal [153] was used. This dataset includes IDs of geo-tagged tweets created between March 20, 2020, and April 28, 2020. The Hydrator application developed by DocNow [100] was used to retrieve the spatio-temporal data and tweet message text. The Twitter API was used to extract the social network's graph, resulting in an SSN dataset of 4000 records to illustrate our geoprivacy framework and the dynamic k-anonymity model.

Approximately 20% (4000 records) of the entire downloaded dataset contained precise latitude and longitude, which made it difficult to extract enough peer locations for each tweet retrospectively. To simulate more complete data for the k-anonymity algorithm, a time-domain of one day was used to assign locations from geotagged tweets to tweets that were published during the same day and time period. Default values of $k_{min} = 5$ and $k_{max} = 30$ were used in this case study. After collecting the tweet data, for each user, an IPFS object is generated, and the tweets are stored on the IPFS. Each user's IPFS object includes user info, tweet data, and a list of the peers who are connected to the user with their associated trust value. To simulate peer interactions, an IPFS node is generated based on each user's ID. In each user's profile, the list of peers who are following the user is shown, and by selecting each peer, it shows the owner's contents with the masked locations based on the trust between those two users. The prototype application, screenshots, and the source code for the app and its online version are available on Github.

Figure 4.8 shows the SSN graph with spatial distance and trust values calculated using equation 4.5. The figure illustrates low, medium, and high user trust values ($\alpha \in \{0, 0.5, 1\}$). In this figure two users have been selected and their friends on Twitter are extracted. To calculate the trust value we need a social and spatial trust. The social trust is the friendship length extracted from Twitter profile and for the spatial trust the location of each user is used. For the location, each user's profile information is used. To determine the proper DGGS resolution for each node in the SSN graph, a bounding box is created from the reported location in each user's Twitter profile. Based on the area of each boundary box, the appropriate DGGS grid resolution is assigned to each node. Based

on equation 4.5 the final trust value between users are calculated. These trust values later can be mapped to different k values. In the current version of the demo application, this trust score is calculated outside of the IPFS network and only the values are stored as IPFS objects. Different values of α demonstrate how this can be used to balance trust values between social and spatial components of the trust metric.

A sample of results is provided in Figure 4.9. In Figure 4.9a, a user with the lowest trust level ($t = 0, b = 0$) is presented, where the output of the k-anonymity algorithm with k_{max} and the lowest level of trust is shown. Since the Level Jump in this configuration is 0, the algorithm does not consider any Level Jump limitations and reaches resolution 2. Figure 4.9b shows a user with a medium trust level ($t = 0.5, b = 0$). Here, the same scenario with a trust level of 0.5 results in a k value equal to 13, and the output cell is of resolution 3 of the DGGS hierarchy. In Figure 4.9c, a user with the highest trust level ($t = 1, b = 0$) requests a location. In this scenario, a much higher resolution result is provided representing higher trust and thus more precise geographic information sharing.

4.5 Information loss metric

To quantify the performance of the proposed k-anonymity algorithm and compare it with other methods, an information loss measure (IL1s) was used [282]. The IL1s measure is defined as:

$$IL1s = \frac{1}{pn} \sum_{i=1}^{p} \sum_{i=1}^{n} \frac{|X_{ij} - Z_{ij}|}{\sqrt{2} S_j} \tag{4.6}$$

where p is the number of continuous variables, n is the number of records in the dataset, X_{ij} and Z_{ij} are the values before and after anonymization for variable j and individual i. S_j is the standard deviation of variable j in the original data. In this section, the IL1s metric of a regular k-anonymity model is based on the bounding box of a k number of points, A DGGS-based k-anonymity with $b = 0$, and a DGGS-based k-anonymity with $b = 2$ for $n \in \{1000, 2000\}$ and $k \in \{10, 20, 50, 100, 200, 400, 800\}$ are compared. At each step, the area of the resulting obfuscated location is compared with the input DGGS cell. For each combination of k and n the model is run 100 times, and then S_j of the iteration is calculated. Figure 4.10 illustrates the results of this metric.

As Figure 4.10 shows, the DGGS-based models of k-anonymity have a higher information loss when the value of k is low, while they tend to have a flatter slope compared to a regular k-anonymity method. This is mainly because of the aperture of the DGGS

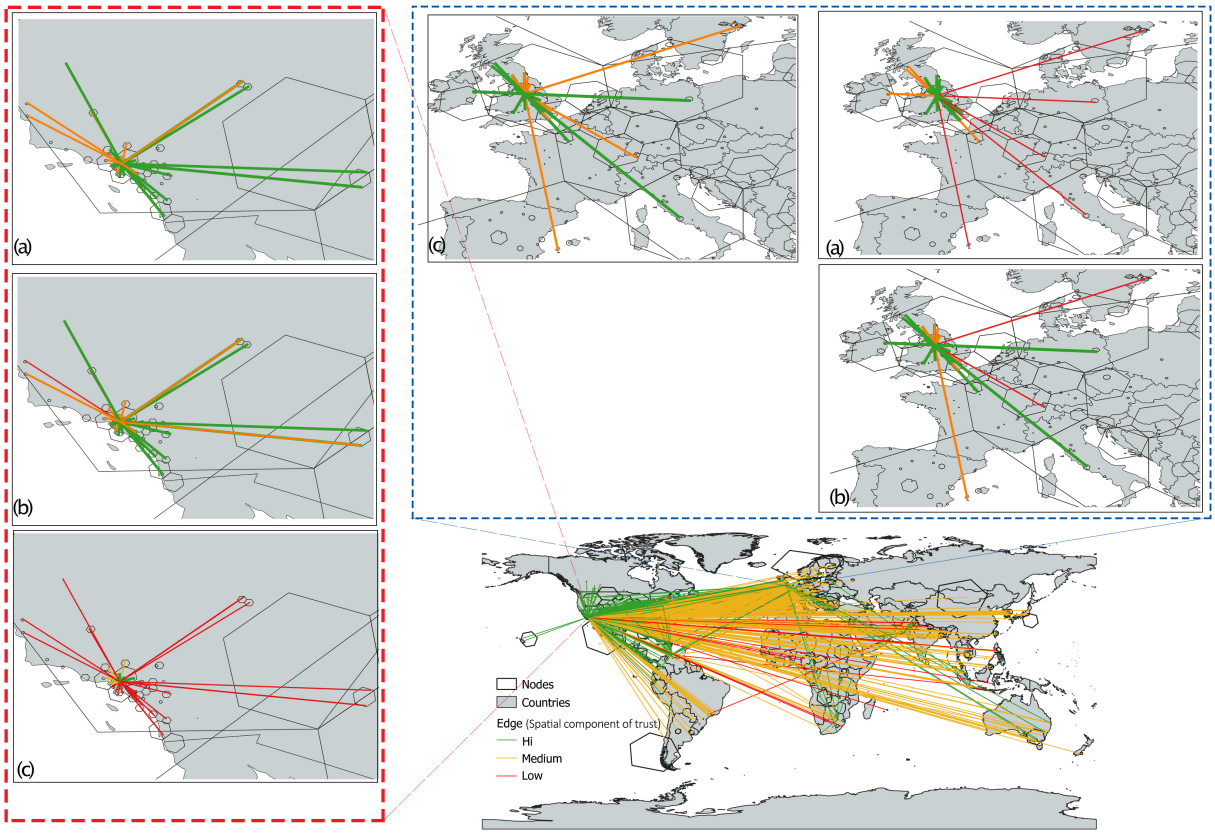


Figure 4.8: SSN graph of the distance metric for the 2 source users and their connected peers. Each edge represents a connection between users and the color of each edge shows the level. Trust values are calculated using equation 4.5 (a) The trust value with $\alpha = 1$. Only the social component of trust is effecting the final trust (b) the trust between nodes with $\alpha = 0.5$ representing balanced between social and spatial components of trust (c) the trust with $\alpha = 0$ representing the entire effect of the spatial component of trust

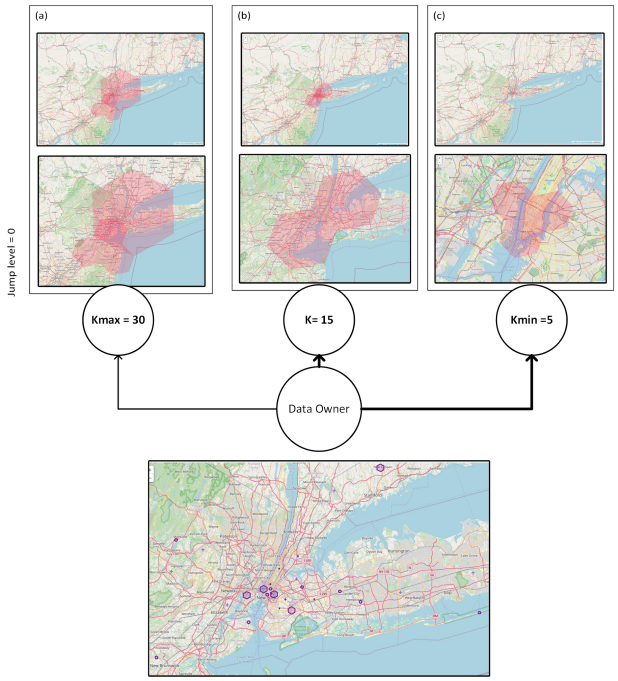
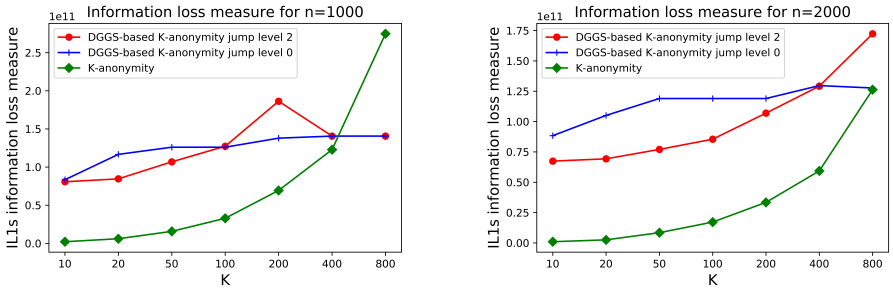


Figure 4.9: Different scenarios for a tweet and its network peers are shown. The Level Jump in this case is 0. Top row: All the maps in this row are in the same scale/zoom level as (a), Bottom row: maps are zoomed in to the objects' extent. (a) Scenario 1: A user with the lowest trust level ($t = 0$), (b) Scenario 2: A user with a medium trust level ($t = 0.5$), (c) Scenario 3: A user with the highest trust level ($t = 1$).



(a) 1000 random points.

(b) 2000 random points.

Figure 4.10: ILS1 information loss metric. Three models of regular K-anonymity, dynamic k-anonymity with $b \in \{0, 2\}$ are compared.

grid, which directly effects the slope of the graph. A smaller aperture provides a smoother transition between resolutions, resulting in a smaller change of cell area. Since DGGS is based on discrete equal-area cells that are dependent on the aperture of the grid, there will be a different ratio of the area between cells in two successive resolutions. This ratio results in higher information loss when geographic data are aggregated to a higher resolution. For example, in an aperture 7 DGGS the parent cell has an area almost 7 times that of its children, while in an aperture 3 this ratio is about 3 which results in lower information loss. Defining a parameter (i.e., the level jump parameter or b) for the k-anonymity algorithm provides control over information loss in the data aggregation, when it comes to using a higher aperture. In contrast, the regular k-anonymity provides a higher information loss with a larger value of k. The use of DGGS-based k-anonymity can provide a higher level of anonymity when the number of peers who are sharing their location is less. A regular k-anonymity based on the bounding box first needs to have an extra parameter of the minimum area to guarantee the anonymity of location, but it is also a suitable option for anonymization of the geographic data when there are many data point.

4.6 Discussion

One of the main reasons for moving from a centralized online social network to a distributed social network is to enhance user control and ownership over data. Unlike centralized systems, data storage in P2P networks is distributed across network nodes. P2P networks also provide scalability, no single point of failure, and transparency. However, privacy and access control have always been a challenge in P2P networks. Recent studies have used blockchain for access control over P2P networks (e.g., Steichen et al. [238]) but these systems are not designed to store large amounts of data. Others such as Voloch et al. [266], have proposed a binary access control over k-anonymized spatio-temporal data in a centralized platform. In contrast, the approach presented here is based on each user storing their own geographic information and this information is not available to other users unless they ask for access from the content owner. In addition, a location priority index based on DGGS is proposed, which not only provides an ability to have different levels of anonymity in an online social network, but also it provides the ability to store only k_{min} anonymized source location and the index of the data. Using this method users do not need to store other peers' geographic data.

The use of a DGGS data model in our framework provides an explicit level of uncertainty and geo-masking embedded into each cell. Every cell in the DGGS grid has a certain area coverage which is used here to provide an extent for privacy preservation. Parent and

child relationships between cells in the DGGS grid are used to prevent storing actual coordinates of the peer locations. Such relations provide the ability to share geographic information with different levels of masking or uncertainty without revealing coordinates. In the traditional k-anonymity algorithms, the k number of records has always been used for obfuscating geographic data and a geographic extent is not considered in the obfuscated output. As a result, when the k number of points are dense in a small area, they cannot guarantee that the obfuscated location has a certain level of uncertainty. Some studies (e.g., [96, 40]) have addressed this by employing a minimum area in addition to the k value. A DGGS based k-anonymity embeds the minimum area based on the resolution of the grid and will not face the issue of contracted input points.

The dynamic geoprivacy method proposed in this paper utilities a novel trust level between users in an SSN network. As Martin and Nissenbaum [173] have discussed, individuals have specific privacy expectations about how, when, and where spatio-temporal data should be shared. A dynamic location sharing and geoprivacy preservation framework builds on this concept of geoprivacy. To determine the accuracy of geographic data sharing in SSN network, a trust metric can be developed based on social, spatial or a combination of these two components. Dynamic geoprivacy permission provides a robust and flexible option for end users to share their location and use services. While some services need accurate location information (e.g., routing services), many other services such as weather apps, loyalty apps and etc. can also work without accessing precise spatio-temporal data.

4.7 Conclusion

Individual dynamic control of geoprivacy is the main purpose of the framework introduced in this paper. However, the proposed algorithm is not limited to online social networks or distributed instant messaging applications and may have uses in other spatial data sharing contexts. Sharing data within and between organizations can be enhanced by greater trust and transparency offered by distributed or decentralized technologies. This is especially true when there are power imbalances or historical reasons underlying a lack of trust among participants in the network. For example, Indigenous governments responsible for stewardship of traditional territories may want fine control over when, what and with whom they are willing to share their geographic data pertaining to traditional use, activities, ceremonial sites, etc. [55]. Rather than depending on a central authority to manage access, a decentralized framework would provide such fine-grained and transparent sharing capability. VGI applications and PGIS projects can also benefit from this framework by providing decentralized control and ownership of the collected data by the users; thus

participation in the project does not require all participants to fully trust initiators of the project. Where place-based conflicts arise, this can be a critical project feature.

The proposed dynamic k-anonymity algorithm has similar limitations to other k-anonymity models. For example, there is low information loss when the density of points in one area is high, so some controlling parameters are used to provide a control on information loss and address such limitations. However, combining k-anonymity with DGGS provides ability of adding geoprivacy in the peer's spatio-temporal data which are used to run k-anonymity. This can add one extra geoprivacy level for the peers. Users are also able to control the precision of shared spatial data with others and control access to their data. They are not limited to third party algorithms to decide their privacy level and also not limited to the binary levels of location sharing.

Despite the advantages noted above, moving to a decentralized architecture also has some major limitations and costs. A key drawback is the requirement for data owners to be online in order to share data. The current version of proposed framework uses a caching method to overcome this issue, however this approach is limited by the number of online users and the variety of trust levels. Another technical limitation is that networks like that proposed here are not easily integrated into existing infrastructures such as existing centralized SSNs [304]. For example, they fail to take advantage of existing location information that an operator has about its customers, or when it comes to accessing and measuring trust metrics from a centralized platform, they fail to integrate this existing information. In other words, decentralized architectures often have to reinvent things already solved in centralized systems (e.g., access control). Despite providing better transparency, technologies like distributed networks should not be viewed on their own as a solution for more complex issues at the interface of society and technology, including the right to be forgotten [76]. In P2P networks such as IPFS the data remains on the peer's devices unless no one requests them anymore. In a centralized architecture context, users can ask for data to be explicitly removed, while in distributed networks such as IPFS, it is not possible to force peers to remove data, if they have already store them. Requiring data owners to decrypt data using their keys is another common issue, which here is addressed through a proposed caching system.

In this work some of the challenges in preserving geoprivacy distributed networks are addressed and the importance of having control on shared geographic information in different levels of sharing platforms is considered. Despite the advantages of our proposed framework, challenges remain. We conclude that greater research effort from GIScience and related fields in developing decentralized geoprivacy applications and tools may lead to a more secure, yet location-aware services. In particular, more precise models for trust between nodes in distributed SSN is required to address issues such as the need for having

trustable nodes in the network to use the k-anonymity algorithm.

4.8 Data and codes availability statement

The source code of the developed web application and processed data are available at https://doi.org/10.6084/m9.figshare.12816164.v1 . Please note that the developed application is using these data only to demonstrate the model and application flow. The github repository is available at https://github.com/am2222/dygeoprivacyp2p.

Chapter 5

Harvester safety in the northern indigenous communities: a distributed approach to share geographic information in small communities

Abstract Long-term research with Ka'a'gee Tu First Nation has involved efforts to identify and map traditional trails and related hazard areas during the Kakisa Atlas project. But the dynamic character of climate-related risks calls for further tools. During 2021-22, researchers developed a prototype harvester safety app that would allow real-time, geo-located hazard identification and risk assessments to be shared between community members on their smartphones. Community consultation with the app's prototype to be instrumental in identifying some key hurdles that must be overcome if the app is accessible and attractive to harvesters. Meanwhile, on the app's back end, considerations of confidentiality and data sovereignty pose challenges that lay below the surface of the more apparent considerations around user interface and overall efficacy. A distributed data storage and management approach holds the promise of community control but poses some challenges that may prove difficult to overcome on the scale of a single community.

Keywords: IPFS, Blockchain, Participatory Project, Harvester Safety, Kakisa

5.1 Introduction

Indigenous communities have been active in their lands for generations by sharing their knowledge, observation, monitoring, and actively participating in research and decision-making [239]. Significant differences between Indigenous Knowledge (IK) and non-Indigenous (Western/Eurocentric) knowledge should be considered in participatory projects and indigenous data governance. In a broad context, IK is considered qualitative, holistic, experiential, and oral, in contrast to Eurocentric knowledge characterized as qualitative, objective, and written [217, 24]. However, with an increased rate of environmental change, new challenges are presented to northern Indigenous communities to adapt [125, 130].

The spatial data-sharing needs of Indigenous organizations vary widely both geographically and socially. The Northern organizations have different levels of spatial data management and sharing requirements. However, the barriers and needs are similar in producing, using, managing, and manipulating spatial data [165]. Veland et al. (2014) [261] identify five main concerns regarding digital data sharing for the indigenous communities: (1) consulting community members at every stage, (2) maintaining ownership and copyright, (3) maintaining privacy rights,(4) providing continuous access and review of databases, (5) preserving intellectual property and having the right to control future uses of the data. Briggs et al. (2020) [24] also identify such concerns as the *place-space gap*, *control gap* and *relational gap* and argue that the geospatial community's aim should be to leverage different technologies better to serve Indigenous communities and Indigenous Knowledge.

The existing platforms indigenous communities use for sharing spatial data lack transparency about the stored data. Also, communities cannot use open data platforms due to indigenous data sovereignty concerns. The remaining options are either to agree with the third-party cloud services' terms of service (TOS) or to set up their infrastructure, which usually impacts the project's sustainability and is limited economically. In addition, The individual and community level of geoprivacy cannot be reliably assessed because it is impossible to know what auxiliary information a third party may have access to [141]. One of the other challenges to most of the current Spatial Data Infrastructures (SDIs) developed in remote areas, like the Arctic SDI, is the connectivity of users [214]. The accessibility of users to the Internet and their access to the central servers of SDI in remote areas is mentioned in most of the related technical documents. The current SDIs for remote areas aim to work with local organizations to make their data available, help manage the information and lifecycle of geospatial data, and support spatial data sharing on top of standard open data platforms. In the first pilot of the Arctic SDI project [1], challenges such as data inte-

[1] www.arctic-sdi.org

gration, limited telecom resource/bandwidth in the North, and end-users' concerns about the data policies are mentioned [170]. Using distributed spatial data sharing (DSDS), the remote communities don't need access to a central server. A local network that can run distributed geo apps provides the infrastructure to share, manipulate, and store spatial data.

In this chapter, in the form of an image-sharing platform, aka the harvester safety application, we investigate the potential of using a DSDS infrastructure for small-scale data-sharing needs in indigenous communities. The main goals of this chapter are as follows:

- Using a participatory design to develop a safety on-the-land application for small communities.
- Designing and proposing a DSDS architecture to allow users in a small community to share photos and safety issues on the land.
- Exploring the potential use case and challenges of using DSDS in small communities.

Participatory Geographic Information Systems (PGIS) principles and experiences can help to empower indigenous communities. In the harvester safety application as a PGIS, we must consider IK and its perceptions, such as the definition of place, feelings, and cultural and social values. The role of indigenous people as stakeholders should be considered within the project. Indigenous people's sovereignty rights make them different than other stakeholders, such as non-profit groups. Like grassroots groups, indigenous people face technical, legal, and economic barriers to data access, such as limited funding for software and hardware, connectivity, infrastructure creation, and capacity-building [231, 72]. Distributed platforms can reduce the dependency of such groups on extensive infrastructure and centralized entities. Their scalability can increase data processing capabilities, and they can use the community itself to develop sustainable data-sharing infrastructure.

5.1.1 Indigenous data sovereignty and distributed data sharing

Data sovereignty is one of the essential concepts of spatial data sharing with indigenous people. Regarding the maintenance of Indigenous ownership and control of digital products, researchers have identified critical issues which have been discussed under the term: "Data sovereignty"(DS) [24]. Taylor et al. [152] define DS as *"managing information in a way that is consistent with the laws, practices, and customs of the nation-state in which it is*

located.". Governmental bodies define data governance in western communities, and in Indigenous communities, it may manifest in the Indigenous community hierarchy [55]. Data sovereignty is mainly based on the right of Indigenous communities and nations to control the collection, governance, ownership, and application of data about social, geographical, and natural resources. In Canada, the First Nations Information Governance Centre [83] defines the OCAP framework to protect the data sovereignty of Indigenous communities as the following: Ownership of the knowledge/data and information, right to Control of all aspects of their life and institutions extended to information and data from collection, usage, disclosure, and destruction. Access to their information regardless of their physical location and the decision of who can access it and Possession reflects the state of data stewardship. Other organizations such as the Indigenous Data Sovereignty Interest Group (International Indigenous Data Sovereignty IG [31]), the Maori Data Sovereignty Network [152], and the US Indigenous Data Sovereignty Network [152] also emphasize data sovereignty in Asia-Pacific and the US. A similar term to Indigenous Data Sovereignty (IDS) is *"Data Locality (DL)"*, which is an obligatory legal requirement to store and share data in a specific territory whose main aims are to protect privacy, improve safety internationally, and raise national security issues [55]. While data locality mainly focuses on the physical storage of data and controlled access by users inside political boundaries, sovereignty includes the right to control data usage. All three aspects: data localization, sovereignty, and privacy, are interdependent [55]. Privacy and data sovereignty both cover personal and state/community scales.

5.1.2 Engagement and Participatory Design

The main goal of participatory design practices is to extend the object of the research beyond the project itself and turn it into a long-term process that impacts the community. In this work, we developed our project according to participatory design dimensions. Smith and Iversen [233] introduce three dimensions of engagement of *scoping*, *developing*, and *scaling*, essential for participatory design, which results in sustainable social practice. The scoping dimension is meant to create a space where diverse participants can explore the potential futures together and develop directions to suit (evolving) goals. In the developing dimension, the focus changes from the project's outcome towards working with technologies at different levels of abstraction and the stakeholders, creating frameworks and processes for developing understandings and digital design literacy among community members. This allows the stakeholders to experiment and co-develop their own platform needs. The scaling dimension focuses on creating multiple opportunities for sustaining and scaling projects beyond the agency of individual actors at various levels.

5.2 A Background on the Case Study

Ka'a'gee Tu (kah-gee-too) First Nation (KTFN) community is situated right at the mouth of Kakisa Lake and river (See figure 5.1) and was established in the 1960s. Kakisa community is situated in the administrative South Slave Region of the North-West-territories (NWT), but within the Dehcho Region based on their close relationships and association with Dehcho First Nation [130]. The area is mainly covered with cold-tolerant tree species, lakes, rivers, and wetlands [23, 148]. The community has a close cultural and spiritual relationship with the boreal forest's lands, waters, and animals. Their traditional ways of life include harvesting, hunting, fishing, and processing natural resources [148].

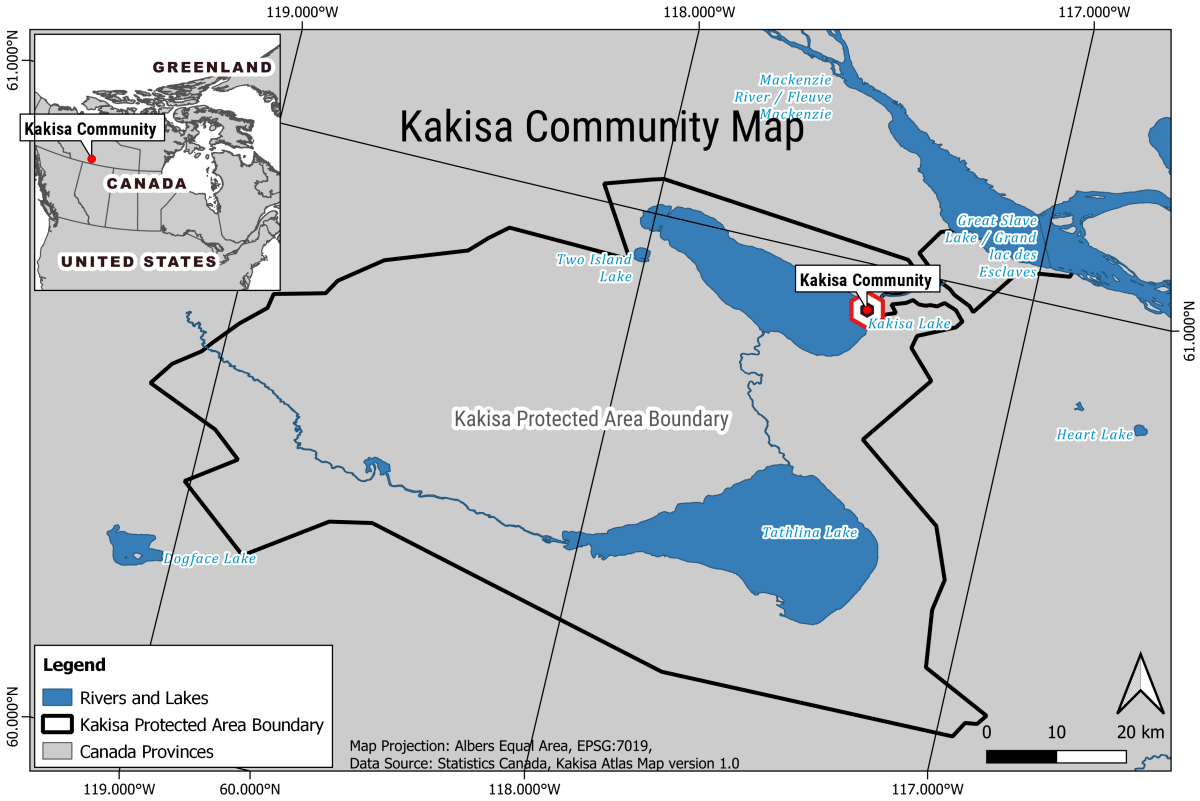


Figure 5.1: Kakisa Community Map

Over the past few years, multiple research projects emerged in response to the community concerns surrounding climate change to identify some of the physical changes Kakisa

community members are observing in the surrounding land, which has been found to affect Indigenous food systems in multiple ways [130, 148]. Spring (2018) [236] discusses the photographs community members were sharing and capturing the change they have noticed over the years. The shared photos may include conditions and hazards, animal sightings, or other interesting occurrences. The community is already engaged in informal monitoring of change by taking pictures and sharing information from the land. Later, Kok (2020) [148] developed an understanding of how environmental changes impact the community and utilized the shared photographs. They used Participatory Action Research (PAR) and a modified photovoice method that was used to identify physical landscape changes at a local level impacting local food security with the aid of traditional knowledge and individual relationships with the land. In 2021 Jayaratne [130] focused on climate change adaptation in a northern Indigenous community context through participatory geographic information systems (PGIS). Kakisa's interest in monitoring changes around them has increased with their observation of changing environmental patterns.

Harvesters are the primary source for accessing traditional food for the Kakisa community [130]. Climate-driven change due to warming temperatures is impacting the availability and access to traditional food, as well as the safety of harvesters. There is a strong association between access to food and the safety of harvesters [25, 130]. One of the impacts often mentioned by harvesters is the increased risk they face due to unpredictable weather, ice, water levels, and other factors. Improving harvester safety can help ensure that traditional foods remain part of Indigenous diets, with their dual nutritional and cultural value. The unusual changes caused by climate change affect various land activities directly or indirectly, such as more disconnection between elders and youth teachings and unusual animal activity [236].

During the above research projects, an online web-based Atlas version 1.0 was developed (see Figure 5.2) on the ESRI infrastructure for the Kakisa community to share and visualize GIS data. However, the initial version of the atlas was mainly developed by experts. It lacked a sustainable approach to updating the shared information and providing an easy-to-use gateway for the community members to use and update its data which does not address the need for constant monitoring of the changes on the land. In contrast, the project presented in this thesis presents, a multi-platform application to serve as an easy-to-use tool to share and communicate the safety issues on the land in the community. Due to the concerns with indigenous knowledge sovereignty, we proposed a distributed data-sharing platform to share and retrieve geo-tagged photographs shared by the community members.

Figure 5.2: User interface of the version 1.0 of the atlas developed by [130]

5.3 Design and results

In order to develop the Kakisa Atlas version 2.0, the team decided to follow the participatory design principles while developing the harvester and safety application. The project's scoping step aims to identify the applications, goals, and processes of the project with collaboration between community members and the project team. With the experiences we gathered from the Kakisa Atlas application [130] we identified a set of requirements for a VGI application. Then we formed the following steps to tackle the scoping stage of the project.

- Developing the prototype of the harvester and safety app
- A two-day workshop in the Kakisa community and demoing and designing the tool and processes

Developing the prototype of the harvester and safety app: In this step, in order to have a demo of the application and be able to practically get the views of the community members, a Progressive Web App (PWA), is developed. The table 5.1 and figure 5.4 shows the capabilities of the developed application.

Table 5.1: A summary of the developed prototype application

Screen	Features
Main Screen	It includes a map of the recent events View event locations, type, and severity View current user's location
Add New Event	Each user can add a new event and choose from the list of event types.
Add New Event	Each user can set the event's severity. High-severity events will send a push notification to the other servers
Location History	By Clicking on each location on the map. A list of the events that happened over the time around this location will be shown
Notifications	The events around the current user will be shown in chronological order

A two-day workshop in the Kakisa community and demoing and designing the tool and its processes: In August 2022, a two-day workshop was held in the Kakisa community to demo the application (Figure 5.3). On the first day of the workshop, community members used the application and shared their feedback. On the second day, the community elders discussed the importance of the harvester and safety and the opinions of the indigenous community about safety on the land. The following items were discussed during the workshop:

- Application UI and UX
- What kind of Information should be shared with the world? (Data Privacy concepts)
- Where to save the shared information? (Data Sovereignty and access control)

During the workshop, the comments of the community members were recorded. The concerns about the project and the needs of the community members are identified as follows:

Social considerations This category of community considerations covers *Data sovereignty*, *Data access control*, *Geo-privacy*, and ability to use it for data markets. During the workshop, the discussion was shaped around data sovereignty's importance and traditional knowledge's copyright. The following quote explains more about this concern:

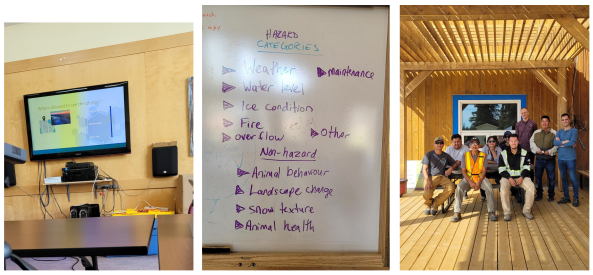


Figure 5.3: Discussion of the application and setting goals during workshop

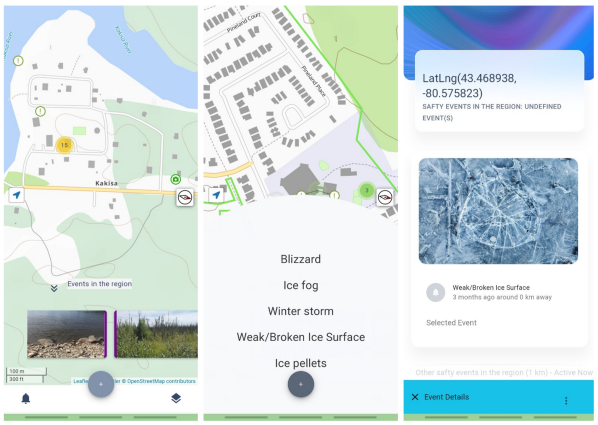


Figure 5.4: A few screenshots of the developed prototype app

> "TK and copyright are really important and nobody is speaking about that. We should [add] that all in place[...]. I look at my community here, everyone wants to share that data with you and that is beautiful[...]. It took us a lot of time to trust people with the information."

From the community member's perspective, it is important that during the project, we provide mechanisms to hold the copyright of the digital contents and also respect the data sovereignty of the indigenous community. We tried to answer the question, "Where does the collected information by users be saved?". This question is asked to raise concerns about the impact of centralized systems and their potential to dismiss data sovereignty concerns. *Geoprivacy* is another concept that was discussed during the workshop. We aimed to answer the question, "What is being shared with each geo-tagged image?". We discussed the potential ways a user's location can be inferred from the shared information

on the new platform. Workshop attendees were asked if they would like to add their name when they share an event on the application, and they agreed to only allow users who have registered in the app and are allowed by the community to see the details of the photographer.

> "Access level to the photo-related information. People who login into the application are able to see the photographer's name and information[...]"

Another aspect of geoprivacy that community members mentioned is the impact of sharing the location of the harvesters when they go on the land. Users like tourists and even community members can enable location sharing to let others know their whereabouts and this can help in emergency cases.

> "Knowing where the people are can be useful safety-wise."

During the discussion, it was suggested that users could opt to share their locations with different precision or add it to the geo-tagged images. In emergency cases, other community members can identify the latest active location of the user on the land. At this step, the other concern, which is *Data access control* came up in the discussion. The members asked for different levels of access control for the non-community member users. Users such as tourists can only access the application when they are in the community. We refer to the *data locality* which enables small communities to control outsiders' access to their sensitive data.

> "When tourists come to the community they can access the atlas data."

The concept of trust between the indigenous communities and people outside of the community was also discussed. The community members note that it took a lot of time for the indigenous people to trust and there should be protocols in place to allow others to access the shared data on the digital platforms.

> "we need to work on protocols for community work."

These sorts of protocols can be used to structure the workflow of the application to control access levels.

Table 5.2: Summary of the community concerns around the safety on the land application.

Class	Concern
Social	Data sovereignty
	Data access control
	Geoprivacy
	Data Locality
Technical	developing easy to use tools
	network connectivity

Technical Considerations The second category of needs of the Kakisa community members is the technical difficulties that exist in the community. These concerns include *developing easy-to-use tools* and *network connectivity*. Many of the community members were concerned about the development of tools that are easy to use. The main concern was that in the winter they need an easy way to share their location and other information with the other members.

> "Keep it [the processes] very simple."

This can include both user interface and user experience. The current version of the Kakisa Atlas which is developed using the ArcGIS online platform can not serve the community members in a practical and daily basis since it is not a mobile-friendly application. Also, it needs experienced users to work with the platform. During the workshop, we discussed the categories of events that users can select in the application and the process of submitting new events. The community members agreed they need a platform-specific application (IOS and Android) to monitor and submit new events.

Network connectivity is another important challenge that users in the Kakisa community mentioned. Due to limited network coverage, most users who go on the land face connectivity issues. This issue caused a couple of safety events during the past few years. The use of two-way communication GPS tools and being able to connect these tools with the atlas application also came up during the workshop. Community members need an easy-to-use method to track the location of the GPS devices in the community to use in emergency events.

Table 5.2 shows a summary of the community concerns around the safety on the land application.

5.4 Discussion

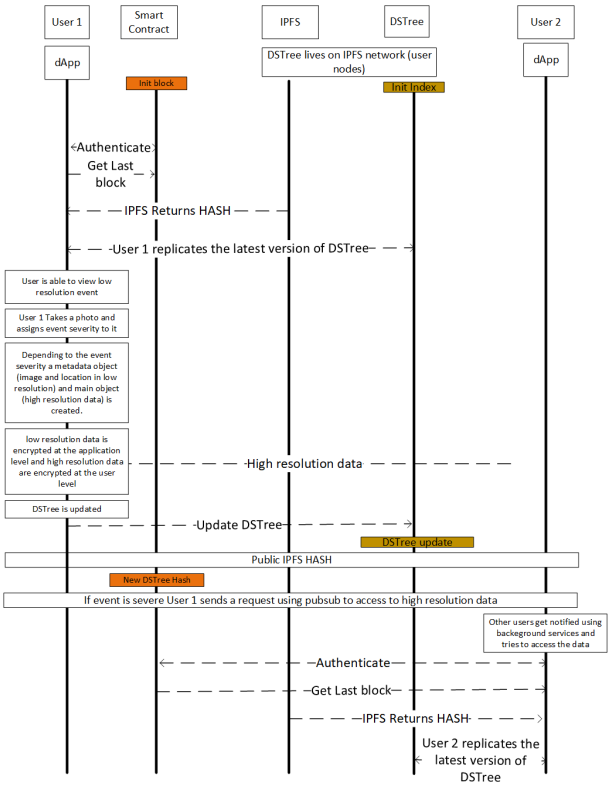
Considering the Kakisa community's concerns related to the safety on the land we proposed a software architecture for the application. We aimed to use the existing distributed system components in this proposed architecture. Figure 5.5 shows the proposed architecture.

The system architecture uses a DSTRee (See Chapter 3), Blockchain, IPFS, and Pub-Sub as its principal components. Figure 5.6 shows these components. A DSTree is the primary storage backend. Each user uses a dApp responsible for accessing the blockchain and accessing the IPFS hash of the latest version of the DSTree index. Each user stores two versions of the data on DStree. A spatially low-resolution version is only encrypted at the application level, and every registered user can visit it, and users' private keys encrypt a high-resolution version (See figure **??** (a)). Once a new user replicates the latest version of DSTRee on their device, they can only access the low-resolution information. To access the high-resolution data, they send a request using pub-sub to the entire network with the following details:

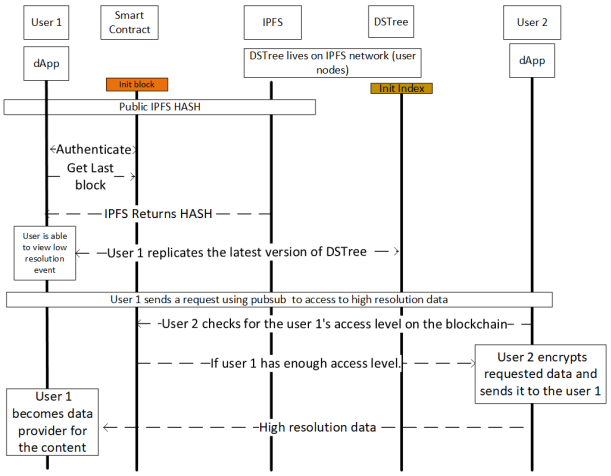
```
{
"messageId": "Requested IPFS Hash",
"userId": "The User Id of the user who is asking for higher-resolution data",
"publicKey": "The User's public key"
}
```

The network checks the user id's access level stored on the blockchain and if any of the network nodes have data associated with the *messageId* they will check the *userId* to see if the requesting user has access to the high-resolution data. If the requesting user has access, the node which hosts the data encrypts the data with the user's public key and returns it. The access level can be added to the blockchain by using dApp's admin or there can be mechanisms in place so the person who creates data can choose different access levels and store it on the blockchain (See figure **??** (b)).

The Data Sovereignty of TK is one of the essential concerns of the community. In distributed systems, a few mechanisms provide different levels of data sovereignty. The main benefit of distributed systems is the lack of a central entity that stores the data (usually the cloud services controlled by different entities). In the proposed architecture, users encrypt the data using their private key. Only metadata will be shared with others (using a DSTree). The shared metadata includes the event type, obfuscated location, and



(a) The process of sharing image data and location in multiple resolutions on IPFS network



(b) The process of accessing the high-resolution data by one user

Figure 5.5: Proposed DSDS platform architecture.

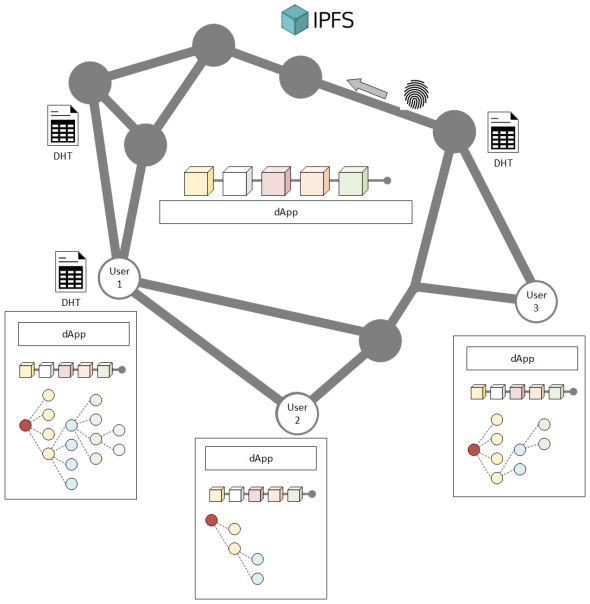


Figure 5.6: The different components of proposed DSDS to share spatial data on IPFS using DSTree

image. Depending on the severity of the events, some of the events need to be shared publicly (e.g., flooding, forest fire, or animal attacks). Still, for the rest of the events, users can choose the visibility level of the data. As a result, users can control who, when, and at which level other users can access the shared data. Table 5.3 shows different data stored in each object in a DSTree.

Geoprivacy is one of the other concerns which needs to be addressed in the application. Due to the low number of community users, it is impossible to use methods such as *Distributed Geoprivacy (See Chapter 4)*. However, we will use a slightly different method to apply geoprivacy to the data. When storing data in the DSTree we use a different resolution of a DGGS (same as what we did in the geoprivacy) and only store an obfuscated cell. Once users allow another user to access the content, they can also choose a different resolution to be shared with the third user. Once again, depending on the category of the submitted event, sometimes a more accurate location needs to be shared with the public for safety reasons.

An example of the image object is as follows:

```
{
```

Table 5.3: Each Stored object on IPFS includes the following details

Class	encryption	Details
Low-resolution DGGS cell	No	used in the query process
Photographer's username and picture	Yes, application level	only registered users in the app will be able to see the username and avatar of the user
Photographer email	Yes, data owner level	To access each user's email address, they need to get permission from the data owner
Event Type	No	
Event Photo	Yes, data owner level	Severe events will not be encrypted
Event Description	Yes, data owner level	Severe events will not be encrypted

```
        "lowResDggs [Encrypted by dApp]": "GeoHash of the location in low resolution",
        "eventType [Encrypted by dApp]": "Integer Value describing the event type",
        "eventDate [Encrypted by dApp]": "Date and time of the event",
        "user: [Encrypted by user]": {
                "userEmail": "Email Address of the user",
                "userName": "Unique user's wallet id",
                "userAvatar": "IPFS Hash of the user Avatar photo",
        },
        "description: [Encrypted by user]": "The User's description of event",
        "photo: [Encrypted by user]": "IPFS Hash of the shared photo.",
        "highResDggs: [Encrypted by user]": "GeoHash of the location in high resolution",
}
```

Sending notifications to the community members is another requirement of safety on the land application. In the current P2P networks, the pub-sub technology is well developed, and each user, upon installing the application, will be able to register to the provided channel to send and receive messages using pub-sub technology. Since it is important to send notifications to the users for severe events, the messages will be encrypted at the

application level only, and all of the application users will be able to access these safety notifications.

During the workshop's two days, the community members' emphasis on the easy-to-use application was obvious. This can be discussed at different levels. In the scale of the application, the developed application and tools need to be easy to use, and the UI and UX should be clear and without difficulties. Such a need can be solved during more workshops and by getting more feedback from the community members. Developing platform-specific tools is another limitation. Version 1.0 of the atlas is developed with the ArcGIS portal platform, which does not provide a user-friendly interface for exploring the map on different platforms. Since the current Atlas version 1.0 is unavailable on mobile phones, community members don't use it often. This issue gets highlighted for the distributed network tools and libraries since some of them are not compatible with all the mobile platforms and it can be a limiting factor to making it usable for small communities. Most current distributed technologies depend on private/public keys, data encryption at the user level, and user identification using crypto wallets. Such technologies are new, and the developed tools are not accessible and easy to use. This issue can also impact the participatory design steps too. In the second and third steps of a participatory design, we need to integrate the project development and its related steps into the community and empower the community to transfer/keep the knowledge in the community to keep the projects sustainable. However, this requires a technical background in blockchains, data encryptions, etc.

The considerations mentioned above raise the importance of the different user groups in using the new technologies, such as distributed data sharing in small communities. The users can be categorized as end users and intermediate users. In community-based data sharing, the end users are the community users, tourists, and even scientists who come to the community. Such users have different backgrounds, and the impact of their background on the data sharing behavior and using the tools is clear. The developed platform needs to match users' needs and have the flexibility to allow the users to share data from their own perspectives. Limiting indigenous community members to the existing geographic data models can cause knowledge loss during data sharing. The ability of the users to define the geographic objects with more flexibility (e.g., DGGS based) can satisfy this need. However, it can also increase the complexities of the developed tools. The heterogeneous data can be shared on the distributed networks (for example, the objects stored on a DSTree, Covered in chapter 3, can be different per user), and this can cause difficulty in querying and indexing these data semantically. It also impacts the quality of the collected data. Another group of users is intermediate users, who are responsible for keeping the technical aspects of the project sustainable. Such users need more technical knowledge about the software and tools. In the case of distributed networks, it can be a challenge

since the current technology is new, and there is not enough documentation. This can directly impact the participatory design flow.

5.5 Conclusion

With the background of the collaborations between the Kakisa community members during the past few years, the harvester safety on the land is identified as one of the main challenges of the community. This challenge was partially addressed during the Atlas version 1.0 project by providing a digital version of the collected geographic information and sharing it with the community members. However, the lack of a community-based monitoring tool is highlighted more than before due to the rapid environmental change. In this project, a prototype application to share geo-tagged photographs is developed, and a three-stage participatory design framework is used to manage this project. During two days of the workshop, we collected community feedback, and based on this feedback, we proposed an architecture for a DSDS application.

There are several limitations to using DSDS for small communities, which are shown during our initial assessment. Looking at the current availability of the tools, it is clear that the sustainable development of such tools needs technical expertise. It is one of the challenges in the scaling stage of participatory projects. In each project, the participants can be from different groups. Some of them have a more technical background which impacts the scaling of the project. However, some advocate members care about the importance of data sovereignty and geo-privacy and can be more progressive in using new technology for their needs. But most participants are willing to be involved only in less technologically complex processes. This impacts the acceptability of the technologies in smaller communities. We need more education about the technologies and easy-to-use tools to interact with distributed technologies.

The security of the shared data on distributed networks is highly dependent on public/private key encryption. As such, the security of these keys is in the user's hands. This can be a challenge in two ways. First, it adds difficulties for the end users and can discourage them from using DSDS technologies. Second, user's private keys might get exposed, and others can decipher their data. However, upon private key exposures, only one user will be impacted.

Indigenous Knowledge has different levels, including public areas, accessed after negotiations, or highly restricted areas like sacred knowledge sites, practices, etc. [114]. Culture and initiatives oppose knowledge transfer, which technology was designed to support [260].

It is clear that technological movements are facilitating the transfer, processing, and storage of data, but there are socio-economical aspects that also control data sharing between different entities. While working with local communities, the cultural and social aspects of data sharing must be considered in every design step. An important feature of technology for sharing IK of places while protecting it is the ability to support restrictions on who shares knowledge, with whom, and where [24]. In small communities such as NWT communities, access control can be at the community level as knowledge sharing is at the community level. However, in DSDS, the access control is at the individual level. This can be a challenge for sharing traditional knowledge. However, this can be solved by developing a collaborative distributed access control such as [241] or [299].

Looking at some successful community-based projects, such as the Arctic Borderlands Ecological Knowledge Society, it is clear that community empowerment and sustainability are other aspects of such projects [32]. There must be initiatives for community members to participate independently in such projects. As mentioned earlier, data ownership can be one of these motivations for sharing data. In addition, it is possible to use business models such as data markets to provide economic initiatives. The ability to be independent of outside communities in terms of analysis and the avoidance of complexities in applications and tools can lead to sustainable projects for Indigenous communities.

Along with the challenges, community members need social capital, such as having land experts, people with digital and computer literacy skills, and the human capacity and time to participate in data gathering [177]. A higher level of user education about DSDS is needed to allow users to make better decisions about the tools and services they are using. More educated users can push for more restrictions to force services to provide more transparency [141]. This can increase the potential of moving from centralized to more distributed and open systems. In addition, with more information about DSDS, we can increase the acceptability of the technology by the users and their willingness to use the DSDS for their needs.

Collecting sensitive information and integrating it into an online platform creates concerns for many Indigenous communities. This is because of the possible exploitation and misuse of the information in light of Canada's history towards Indigenous people and ongoing issues for land rights [73]. By restricting access to mapped content and authorizing the community to control the project, it provides the community with entitlement, strengthens protection and sovereignty over the information [73]. Indigenous community's right to keep the information and knowledge inside the community and control their access is defined in many human rights acts [252]. While traditional methods for geoprivacy preservation are based on data aggregations done by central entities, the new technologies in GIScience, such as digital earth platforms, can provide tools to incorporate individual privacy while

sharing spatial data. Having control over the level of privacy based on trust and dynamic data anonymization [120] are examples of methods that can help share Indigenous data. Inferring the location of the people using the shared information is one of the challenges. As discussed earlier, users in the community suggested only showing the photographer's details once a user logged in and they had access rights. However, this does not fully address the problem due to the possibility of using photographs to detect sensitive places and information.

Acknowledgements

Thank you (Mahsi) to the Kakisa community for welcoming me with open arms and allowing me to live and learn from all of you. Thank you to the Band office members (Chief Lloyd and Melaine) who allowed me to hold a two-day workshop in the community. Also, a great thank you to Dr. Alex Latta and Dr. Andrew Spring for their support during this project and during my trip to the Kakisa community.

Chapter 6

Conclusion

6.1 Overview

In this chapter, we will cover the main contributions of the research, followed by Challenges and future research directions.

6.2 Contributions of this research

Considering the novelty of the distributed data-sharing concepts, this research had the following contributions in this area:

Literature During this work, we investigated a change in spatial data-sharing behavior and the highlighted role of individuals in sharing spatial data. We defined three main spatial data-sharing eras, addressed the current status of data ownership and access in different SDS models, and demonstrated how a transition to DSDS addresses some of the challenges caused by increased individual contributions. In chapter 2, we defined DSDS as *"A new data sharing model in which individuals are active in all dimensions of data sharing: as producers, controllers, and users. The data are stored, controlled, and maintained by the data producer. The ownership and license of the data can be transferred to other users without the need for data intermediaries. The immutable spatial data are stored on distributed file sharing networks, while the access rights, state transitions, and version history are managed by smart contracts and stored on a blockchain."*. After defining

the meaning of DSDS we identified some of the study gaps in and aimed to address some of them.

Social: Social component of spatial data sharing impacts the sharing behavior of the users and their willingness to contribute to the sharing data. This work tried to contribute to this component by first discussing social trust and its influence on geoprivacy. in chapter 3, we covered a basic concept of *social trust* and how it can be used to share user locations. The same concept can be used to share traditional knowledge within and outside the indigenous communities. Another aspect of the social component of DSDS is the user's willingness to share spatial data. In chapter 5, during a project to develop a harvester safety application, we discussed different concepts which can impact the willingness of the indigenous community members to share geotagged images in an application. We also explored the potential of using DSDS for such VGI projects.

Technical: In the technical part of distributed spatial data sharing, we covered two main research gaps. First is addressing the challenge in processing Spatio-temporal queries on the distributed networks by proposing DSTree (covered in chapter 3). The traditional P2P indexes usually need multiple DHTs or use federated nodes to process multidimensional queries. DSTree, works on top of IPFS and does not require additional DHT to perform multi-dimensional range queries. Inserting a new geographic object updates a portion of the tree structure and does not impact the entire data graph. DSTree's performance is also measured and compared with the existing spatio-temporal indexes, showing that it can outperform some existing models. Considering the structure of DSTree, it is a good option for storing and querying time-series data, such as sensor data which is being stored regularly over time.

The second technical contribution of this work is to propose geoprivacy (covered in chapter 4) models for distributed networks. In this chapter, we develop a geoprivacy framework that couples two emerging technologies -decentralized data storage and discrete global grid systems- to facilitate fine-grained user control over the ownership of, access to, and map-based representation of their data. We proposed a framework to link the geographic precision of shared data to social trust within a social network. This framework shares users' spatio-temporal data through a decentralized system. They are represented on a discrete global grid data model at spatial resolutions corresponding to varying degrees of trust between individuals exchanging information. We compared the dynamic k-anonymity model based on DGGS with the current models and discussed the different scenarios for sharing a user's location in a distributed network. Our framework has several advantages

over centralized geoprivacy approaches, namely, trust in a third-party entity is not required, and geoprivacy is dynamic and context-dependent, with users maintaining autonomy.

6.3 DSDS: challenges and opportunities

In chapter 2, we addressed the current status of data ownership and access in different SDS models and demonstrated how a transition to DSDS addresses some of the existing challenges. The storage, distribution, and manipulation of spatial data are changing from a fully centralized approach (i.e., controlled by the government and/or corporations) towards a distributed, individual spatial data-sharing approach, though this approach remains technically challenging. Despite providing better transparency, technologies like distributed networks should not be viewed as a solution for more complex issues at the interface of society and technology, including the right to be forgotten. As mentioned in past chapters, in P2P networks such as IPFS, the data remains on the peer's devices unless no one requests them. In a centralized architecture context, users can ask for data to be explicitly removed. At the same time, forcing peers to remove data in distributed networks such as IPFS is impossible if they have already stored them.

A distributed framework for spatial data sharing provides a user-centered approach to address ownership issues of individually collected data. As mentioned previously, technology is only one piece of the data ownership and control puzzle, albeit a critical element. It should be noted that distributed data sharing is not a universal solution. Many authoritative or proprietary data sets (e.g., cadastral boundaries) will continue to require centralized access and authoring control for social, competitive, or legal reasons. In this light, DSDS and centralized systems complement each other and enable more fine-grained and flexible data governance architectures.

One of the reasons for moving from centralized data-sharing methods to distributed strategy is the increasing amount of data being collected daily. Unlike centralized systems, data storage in P2P networks is distributed across network nodes, providing scalability and no single point of failure. However, managing and processing queries on these networks have always been challenging. The proposed method in chapter 4 tried to tackle this issue by tracking and updating a spatio-temporal index between network users. In this approach, each user can replicate a version of DSTree on their node and run spatio-temporal queries on the index. Since each user performs the queries on their side, the indexing tree should check fewer items and support more topology out of the box. However, this indexing method is not suitable for content updates of the data.

Distributed technology is in its early stages and requires the development of tools/methods and algorithms to handle, share and query geographic information. Once developed, it will be possible to contrast DSDS against other data systems and thereby evaluate the practical benefit of such systems. A distributed data-sharing platform needs a standard to share data between different users. It requires a data model to integrate spatial data from other sources and various accuracy levels. Such standards must be in alignment with multiple sharing platforms and frameworks.

6.4 Future research directions

As the distributed web emerges, so too can the next generation of geographic information-sharing tools. Despite the advantages of our proposed framework, challenges remain. We conclude that more significant research effort from GIScience and related fields in developing decentralized applications is needed. Distributed technology is in its early stages and requires developing the tools/methods and algorithms to handle, share and query geographic information. Once developed, it will be possible to contrast DSDS against other data systems, thereby evaluating the practical benefit of such systems. These standards must be on different levels, from the data model and storage level to the API and processing levels. The ability to be integrated with the existing standards is a main interoperability requirement in sharing spatial data.

In chapter 2 some challenges in preserving geoprivacy in distributed networks, and the importance of controlling shared geographic information in different is discussed. In this chapter, we conclude that, it is still required to develop more fine-grained decentralized geoprivacy applications and tools. This leads to more secure yet location-aware services. In particular, more precise models for trust between nodes in distributed networks are required. This helps with one of the biggest assumptions of this research which is the necessity of having trustful actors in the network.

This work also addresses some of the challenges in sharing and querying spatio-temporal data on distributed networks. Developing a standard framework for sharing data on the IPFS network is one of the big requirements. The possibility of using IPLD objects in sharing GI at the feature level can provide finer access to the information. In addition, it is necessary to address attribute-level query processing too, which is not covered in the current work. The use of smart contracts to control read and write access to data to the main chain can also be studied.

Along with the social considerations, social capital for the communities needs to be developed. It is required to have a higher level of user education about DSDS to allow the

users to make better decisions about the tools and services they are using. Such education processes need to follow certain frameworks to increase technological acceptability and trust in the different generations. In the smaller communities, this framework should be in place in order to provide trust between community members and the researchers with the shared geographic data. The challenges and difficulties of using DSDS such as protecting encryption keys in the communities should be addressed in such a framework. Depiction of a clear image of distributed tools and how they protect users' data ownership can help with technological acceptability at the community level.

Distributed technologies are also capable of addressing the scalability of sharing and processing geographic data. During this research, we did not explore the distributed processing of spatial data. A blockchain can be used to track distributed and parallel processes at the node level and can improve spatial analysis capacity on P2P networks. However, such approaches to compute over data need more in-depth studies.

The proposed model for querying spatial data supports the spatial topology that the underlying spatial indexing method supports. In this study, we only experimented with the intersection topology relation. However, it is possible to perform other topological relations such as overlay, within, and also perform KNN-based models. From temporal typologies, all of them except disjoint are supported. The focus of this paper was support for vector-based data structures. However, supporting raster data could be achieved by converting raster data into DGGS-based models or tiling the raster data instead multi-resolution tiling structure.

6.5 Key limitations in this research

During this research, several challenges have limited us to some extent. The first set of challenges is the availability of tools and libraries to work with distributed networks. Distributed technology is still in its early ages, and the problems already solved in the current centralized platforms, such as user management and access control, need to be handled in these networks. In other words, decentralized architectures often have to *reinvent* things already solved in centralized systems. The existing libraries and tools are also limited to web platforms, which limits the usability of distributed tools on mobile devices. Let alone that most individual data producers use mobile devices to collect and share geographic data. This limitation increases the difficulties of integrating such tools into different operation systems [304] (see chapter 4 for a more in-depth discussion). Also, it increases the difficulties of using them, discouraging individuals from being able to utilize them. Another key drawback is the requirement for data owners to be online to share data. In

chapter 4, we used a caching method to overcome this issue. However, as mentioned in that section, it is limited by the number of online users and the variety of trust levels.

Another limitation of this research is the duration of the study. Due to the technical requirements of such systems the development of a practical application for the use cases such as community-based data sharing (see chapter 5). Due to the sensitivity of such projects and the importance of data sovereignty, the time required to develop, test, and deploy a distributed application will be significantly longer. As a result, during the last chapter of this work, we only focused on proposing the system architecture and how different pieces of distributed technology can address the needs of community-based projects. Future research must develop the practical application and understand the impact of individual-level data sharing, process, and storage on the data-sharing environment.

Collecting sensitive information and integrating it into an online platform creates concerns for many Indigenous communities. While traditional methods for geoprivacy preservation are based on data aggregations done by central entities, the new technologies in GIScience, such as digital earth platforms, can provide tools to incorporate individual privacy while sharing spatial data. Inferring the location of the people using the shared information is one of the challenges. As discussed earlier, users in the community suggested only showing the photographer's details once a user logged in and they had access rights. However, this does not fully address the problem due to the possibility of using photographs to detect sensitive places and information. Due to this issue and the limited access control methods in distributed networks, developing such tools for the communities needs more attention and concern.

Several limitations to using DSDS for small communities are discussed during our initial assessment in chapter 5. Looking at the current availability of the tools and such, it is clear that the sustainable development of such tools needs technical expertise. It is one of the challenges in the scaling stage of participatory projects. In each project, the participators can be from different groups. Some of them can be a more technical background which impacts the scaling of the project. Most participators are willing to be involved in less technologically complex processes. This impacts the acceptability of the technologies in smaller communities. We need more education about the technologies and easy-to-use tools to interact with distributed technologies. For example, the security of shared data on distributed networks is highly dependent on public/private key encryption. As such, the safety of these keys is in the user's hands. It adds difficulties to users' end and can discourage them from using DSDS technologies. It can also add to the risk that a user's private keys get exposed, and others can decipher their data. Such challenges are mainly caused due to the lack of enough technical tools in this research area.

To fully understand the complex nature of spatial data sharing environments, it is essential to take a comprehensive approach. In our research, we focused on the socio-technological aspects of Distributed Spatial Data Sharing (DSDS), but we acknowledge that there are other important dimensions to consider, such as the economic, policy, and semantic aspects. One significant benefit of DSDS is its potential to have positive social and economic impacts for individuals. Access control policies play a crucial role in enabling these benefits by regulating data access and usage. Additionally, the utilization of data markets can be advantageous for individuals and smaller communities within the DSDS environment. On the contrary, enterprise entities may have less interest in DSDS due to various factors. Furthermore, it is important to explore the economic consequences of distributing data storage on user devices within the DSDS framework.In summary, our work focused on the socio-technological aspects of DSDS, while recognizing the need for further investigation into the economic, policy, and semantic dimensions. We highlighted the potential social and economic benefits for individuals, emphasized the role of data markets, and acknowledged the comparatively lower economic interest from enterprise entities. Furthermore, we emphasized the need for additional research on the economic impact of distributing data storage on user devices

References

[1] Meriem Achir, Abdelkrim Abdelli, Lynda Mokdad, and Jalel Benothman. Service discovery and selection in iot: A survey and a taxonomy. *Journal of Network and Computer Applications*, 200:103331, 2022.

[2] Sohaib Al-Yadumi, Tan Ee Xion, Sharon Goh Wei Wei, and Patrice Boursier. Review on integrating geospatial big datasets and open research issues. *IEEE Access*, 9:10604–10620, 2021.

[3] M. Alessi, A. Camillo, E. Giangreco, M. Matera, S. Pino, and D. Storelli. Make users own their data: A decentralized personal data store prototype based on ethereum and IPFS. In *2018 3rd International Conference on Smart and Sustainable Technologies (SpliTech)*, pages 1–7, 2018.

[4] James F Allen. Maintaining knowledge about temporal intervals. In *Readings in Qualitative Reasoning About Physical Systems*, pages 361–372. Elsevier, 1990.

[5] Hazim Almuhimedi, Florian Schaub, Norman Sadeh, Idris Adjerid, Alessandro Acquisti, Joshua Gluck, Lorrie Faith Cranor, and Yuvraj Agarwal. Your location has been shared 5,398 times! a field study on mobile app privacy nudging. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, Chi '15, pages 787–796. Association for Computing Machinery, 2015.

[6] Fatma Alrayes and Alia Abdelmoty. No place to hide: a study of privacy concerns due to location sharing on geo-social networks. *International Journal On Advances in Security*, 7(3):62–75, 2014.

[7] Jennings Anderson. Openstreetmap contributor lifespans - revisiting and expanding on 2018 research paper, 2021.

[8] Natalia Andrienko, Gennady Andrienko, Georg Fuchs, and Piotr Jankowski. Scalable and privacy-respectful interactive discovery of place semantics from human mobility traces. *Information Visualization*, 15(2):117–153, 2016.

[9] Claudio A. Ardagna, Marco Cremonini, Ernesto Damiani, Sabrina De Capitani di Vimercati, and Pierangela Samarati. Supporting location-based conditions in access control policies. In *Proceedings of the 2006 ACM Symposium on Information, computer and communications security*, Asiaccs '06, pages 212–222. Association for Computing Machinery, 2006.

[10] Marc P Armstrong. Temporality in spatial databases. *GIS/LIS 88 Proceedings: Accessing the world*, pages 880–889, 1988.

[11] Marc P. Armstrong, Gerard Rushton, and Dale L. Zimmerman. Geographically masking health data to preserve confidentiality. *Statistics in Medicine*, 18(5):497–525, 1999.

[12] Marcella Atzori. Blockchain technology and decentralized governance: Is the state still necessary? *Social Science Research Network*, 1(Id 2709713), 2015.

[13] Brooke Auxer, Rainie Lee, Monica Anderson, Andrew Perrin, Madhu Kumar, and Erica Turner. Americans and privacy: Concerned, confused and feeling lack of control over their personal information, 2019.

[14] Bahman Bahmani, Ashish Goel, and Rajendra Shinde. Efficient distributed locality sensitive hashing. In *Proceedings of the 21st ACM international conference on Information and knowledge management*, pages 2174–2178, 2012.

[15] Monya Baker. 1,500 scientists lift the lid on reproducibility. *Nature*, 533(7604):452–454, May 2016.

[16] Joost Bambacht and Johan Pouwelse. Web3: A decentralized societal infrastructure for identity, trust, money, and data. *arXiv preprint arXiv:2203.00398*, 2022.

[17] B Sue Bell. Spatial analysis of disease–applications. *Biostatistical Applications in Cancer Research*, pages 151–182, 2002.

[18] Juan Benet. IPFS - content addressed, versioned, P2P file system. 2014.

[19] Michael Benisch, Patrick Gage Kelley, Norman Sadeh, and Lorrie Faith Cranor. Capturing location-privacy preferences: quantifying accuracy and user-burden tradeoffs. *Personal and Ubiquitous Computing*, 15(7):679–694, 2011.

[20] Gopal Bhushan and Madhusudhan Margam. Access versus ownership of information and DESIDOC's balancing act. *DESIDOC Journal of Library & Information Technology*, 36:320–324, 2016.

[21] Kelsey Bolin. Decentralized public ledger systems and securities law: New applications of blockchain technology and the revitalization of sections 11 and 12(a)(2) of the securities act of 1933. *Washington University Law Review*, 95(4):955–980, 2018.

[22] Terry Bossomaier and Brian A Hope. *Online GIS and Spatial Metadata, Second Edition*. CRC Press, London, England, 2 edition, December 2015.

[23] J P Brandt, M D Flannigan, D G Maynard, I D Thompson, and W J A Volney. An introduction to canada's boreal zone: ecosystem processes, health, sustainability, and environmental issues. *Environ. Rev.*, 21(4):207–226, December 2013.

[24] Carolyn Briggs, Ingrid Burfurd, Matt Duckham, Olivia Guntarik, Di Kerr, Mark McMillan, and Daisy San Martin Saldias. Bridging the geospatial gap: Data about space and indigenous knowledge of place. *Geography Compass*, 4:e12542, 2020.

[25] Todd Brinkman, Karonhiakta'tie B Maracle, James Kelly, Michelle Vandyke, Andrew Firmin, and Anna Springsteen. Impact of fuel costs on high-latitude subsistence activities. *Ecol. Soc.*, 19(4), 2014.

[26] Brodeur, Coetzee, Danko, Garcia, and Hjelmager. Geographic information metadata—an outlook from the international standardization perspective. *ISPRS Int. J. Geoinf.*, 8(6):280, June 2019.

[27] Chris Brunsdon. Quantitative methods I. *Prog. Hum. Geogr.*, 40(5):687–696, October 2016.

[28] Vitalik Buterin. Ethereum white paper: A next generation smart contract & decentralized application platform. 2013.

[29] Wenyuan Cai, Shuigeng Zhou, Weining Qian, Linhao Xu, Kian-Lee Tan, and Aoying Zhou. C2: a new overlay network based on can and chord. *International Journal of High Performance Computing and Networking*, 3(4):248–261, 2005.

[30] Jan Camenisch, Ronald Leenes, and Dieter Sommer. *Digital Privacy: PRIME - Privacy and Identity Management for Europe*. Springer, 2011.

[31] Stephanie Carroll, Maui Hudson, Maggie Walter, and Per Axelsson. International indigenous data sovereignty IG, 2017.

[32] Stephanie Russo Carroll, Desi Rodriguez-Lonebear, and Andrew Martinez. Indigenous data governance: Strategies from united states native nations. *Data Science Journal*, 2019.

[33] Ann Cavoukian. Privacy by design - the 7 foundational principles. *Information & Privacy Commissioner Ontario, Canada*, page 12, 2009.

[34] Yanzhe Che, Qinming He, Xiaoyan Hong, and Kevin Chiew. X-region: A framework for location privacy preservation in mobile peer-to-peer networks. *International Journal of Communication Systems*, 28(1):167–186, 2015.

[35] Chih-Jou Chen and Shiu-Wan Hung. To give or to receive? factors influencing members' knowledge sharing and community promotion in professional virtual communities. *Inf. manag.*, 47(4):226–236, May 2010.

[36] Wanghu Chen, Yuxiang Mu, Xiaoyan Liang, and Yaqiong Gao. Medical data sharing model based on blockchain. *Journal of Physics: Conference Series*, 1267:012014, 2019.

[37] Yuan Cheng, Jaehong Park, and Ravi Sandhu. Relationship-based access control for online social networks: Beyond user-to-user relationships. In *2012 International Conference on Privacy, Security, Risk and Trust and 2012 International Confernece on Social Computing*, pages 646–655, 2012.

[38] George Cho. Some legal concerns with the use of crowd-sourced geospatial information. *IOP Conference Series: Earth and Environmental Science*, 20:012040, 2014.

[39] Chi-Yin Chow, Mohamed F. Mokbel, and Xuan Liu. A peer-to-peer spatial cloaking algorithm for anonymous location-based service. In *Proceedings of the 14th annual ACM international symposium on Advances in geographic information systems - GIS '06*, page 171. ACM Press, 2006.

[40] Chi-Yin Chow, Mohamed F. Mokbel, and Xuan Liu. Spatial cloaking for anonymous location-based services in mobile peer-to-peer environments. *GeoInformatica*, 15(2):351–380, 2011.

[41] Wing S Chow and Lai Sheung Chan. Social network, social trust and shared goals in organizational knowledge sharing. *Inf. manag.*, 45(7):458–465, November 2008.

[42] Keith C. Clarke. A multiscale masking method for point geographic data. *International Journal of Geographical Information Science*, 30(2):300–315, 2016.

[43] A. G. Cohn. A hierarchical representation of qualitative shape based on connection and convexity. In Andrew U. Frank and Werner Kuhn, editors, *Spatial information theory a theoretical basis for GIS*, pages 311–326. Springer Berlin Heidelberg, 1995.

[44] David J. Coleman, Abbas Rajabifard, and Kris W. Kolodziej. Expanding the sdi environment: comparing current spatial data infrastructure with emerging indoor location-based services. *International Journal of Digital Earth*, 9(6):629–647, 2016.

[45] Louise K. Comfort, Kilkon Ko, and Adam Zagorecki. Coordination in rapidly evolving disaster response systems: The role of information. *American Behavioral Scientist*, 48(3):295–313, 2004.

[46] Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest, and Clifford Stein. *Introduction to Algorithms, Third Edition*. The MIT Press, 3rd edition, 2009.

[47] Stéphane Coulondre, Thérèse Libourel, and Laurent Spéry. Metadata and gis: A classification of metadata for gis. 1998.

[48] Max Craglia, Kees de Bie, Davina Jackson, Martino Pesaresi, Gábor Remetey-Fülöpp, Changlin Wang, Alessandro Annoni, Ling Bian, Fred Campbell, Manfred Ehlers, John van Genderen, Michael Goodchild, Huadong Guo, Anthony Lewis, Richard Simpson, Andrew Skidmore, and Peter Woodgate. Digital earth 2020: towards the vision for the next decade. *International Journal of Digital Earth*, 5(1):4–21, 2012.

[49] Adina Crainiceanu, Prakash Linga, Johannes Gehrke, and Jayavel Shanmugasundaram. Querying peer-to-peer networks using p-trees. In *Proceedings of the 7th International Workshop on the Web and Databases: Colocated with ACM SIGMOD/PODS 2004*, WebDB '04, page 25–30, New York, NY, USA, 2004. Association for Computing Machinery.

[50] Altha J Cravey, Sarah A Washburn, Wilbert M Gesler, Thomas A Arcury, and Anne H Skelly. Developing socio-spatial knowledge networks:. *Social Science & Medicine*, 52(12):1763–1775, 2001.

[51] William Lee Croft, Wei Shi, Jörg-Rüdiger Sack, and Jean-Pierre Corriveau. Location-based anonymization: comparison and evaluation of the voronoi-based aggregation system. *International Journal of Geographical Information Science*, 30(11):2253–2275, 2016.

[52] Silke Cuno, Lina Bruns, Nikolay Tcholtchev, Philipp Lämmel, and Ina Schieferdecker. Data governance and sovereignty in urban data spaces based on standardized ICT reference architectures. *Data*, 4(1):16, 2019.

[53] Gaby G. Dagher, Jordan Mohler, Matea Milojkovic, and Praneeth Babu Marella. Ancile: Privacy-preserving framework for access control and interoperability of electronic health records using blockchain technology. *Sustainable Cities and Society*, 39:283–297, 2018.

[54] Craig M Dalton, Linnet Taylor, and Jim Thatcher (alphabetical). Critical data studies: A dialog on data and space. *Big Data & Society*, 3(1):2053951716648346, 2016.

[55] Angela Daly, Monique Mann, and S. Kate Devitt. *Good Data*. Lulu.com, 2019.

[56] Jack Dangermond and Michael F. Goodchild. Building geospatial infrastructure. *Geo-spatial Information Science*, 23(1):1–9, December 2019.

[57] Desiree Daniel and Chinwe Ifejika Speranza. The role of blockchain in documenting land users' rights: The canonical case of farmers in the vernacular land market. *Frontiers in Blockchain*, 3, 2020.

[58] David Dao, Dan Alistarh, Claudiu Musat, and Ce Zhang. Databright: Towards a global exchange for decentralized data ownership and trusted computation. *arXiv:1802.04780 [cs]*, 2018.

[59] João Porto De Albuquerque, Benjamin Herfort, Alexander Brenning, and Alexander Zipf. A geographic approach for combining social media and authoritative data towards identifying useful information for disaster management. *International Journal of Geographical Information Science*, 29(4):667–689, 2015.

[60] Mark de Berg, Otfried Cheong, Marc van Kreveld, and Mark Overmars. *Computational geometry*. Springer, Berlin, Germany, 3 edition, March 2008.

[61] Maurice de Kleijn, Niels van Manen, Jan Kolen, and Henk Jan Scholten. Towards a user-centric sdi framework for historical and heritage european landscape research. *Int. J. Spatial Data Infrastructures Res.*, 9:1–35, 2014.

[62] Murat Demirbas and Hakan Ferhatosmanoglu. Peer-to-peer spatial queries in sensor networks. In *Proceedings Third International Conference on Peer-to-Peer Computing (P2P2003)*, pages 32–39. IEEE, 2003.

[63] Damiano Di Francesco Maesa, Paolo Mori, and Laura Ricci. Blockchain based access control. In Lydia Y. Chen and Hans P. Reiser, editors, *Distributed Applications and Interoperable Systems*, Lecture Notes in Computer Science, pages 206–220. Springer International Publishing, 2017.

[64] Dejan Dimitrijević, Ivan Luković, Vladimir Dimitrieski, and Ivan Vasiljević. Orchestrating yahoo! fireeagle location based service for carpooling. In *3rd International Conference on Information Society Technology and Management, Kopaonik, Serbia*, 2013.

[65] Brahim Djellabi, Mourad Amad, and Abderrahmene Baadache. Handfan: A flexible peer-to-peer service discovery system for internet of things applications. *Journal of King Saud University-Computer and Information Sciences*, 2022.

[66] Christos Doulkeridis, Akrivi Vlachou, Kjetil Nørvåg, Yannis Kotidis, and Michalis Vazirgiannis. Efficient search based on content similarity over self-organizing p2p networks. *Peer-to-Peer Networking and Applications*, 3(1):67–79, 2010.

[67] Yerach Doytsher, Ben Galon, and Yaron Kanza. Querying socio-spatial networks on the world-wide web. In *Proceedings of the 21st International Conference on World Wide Web*, WWW '12 Companion, pages 329–332. Association for Computing Machinery, 2012.

[68] Kostas Drakonakis, Panagiotis Ilia, Sotiris Ioannidis, and Jason Polakis. Please forget where i was last summer: The privacy risks of public location (meta)data, 2019.

[69] Matt Duckham and Lars Kulik. Location privacy and location-aware computing. *Dynamic & mobile GIS: investigating change in space and time*, 3:35–51, 2006.

[70] Frederick Ojiemhende Ehiagwina, Nurudeen Ajibola Iromini, Ikeola Suhurat Olatinwo, Kabirat Raheem, and Khadijat Mustapha. A state-of-the-art survey of peer-to-peer networks: Research directions, applications and challenges. *management*, 14:19–22.

[71] Ahmed Eldawy and Mohamed F Mokbel. Spatialhadoop: A mapreduce framework for spatial data. In *2015 IEEE 31st international conference on Data Engineering*, pages 1352–1363. IEEE, 2015.

[72] Sarah Elwood. Grassroots groups as stakeholders in spatial data infrastructures: challenges and opportunities for local data development and sharing. *International Journal of Geographical Information Science*, 22(1):71–90, 2008.

[73] Nate J Engler, Teresa Scassa, and D R Fraser Taylor. Mapping traditional knowledge: Digital cartography in the canadian north. *Cartogr. Int. J. Geogr. Inf. Geovisualization*, 48(3):189–199, September 2013.

[74] Martin Erwig and Markus Schneider. Developments in spatio-temporal query languages. In *Proceedings. Tenth International Workshop on Database and Expert Systems Applications. DEXA 99*, pages 441–449. IEEE, 1999.

[75] Eu. Directive 2007/2/EC of the european parliament and of the council of 14 march 2007 establishing an infrastructure for spatial information in the european community (INSPIRE). *Official Journal of the European Union*, Oj L(32007l0002), 2007.

[76] European Data Protection Regulation. General data protection regulation (GDPR), 2016.

[77] David S Evans. *Economic Aspects of Bitcoin and Other Decentralized Public-Ledger Currency Platforms*. Social Science Research Network, 2014.

[78] El-Sayed Ewis. Spatial data sharing: From theory to practice. 2007.

[79] El-Sayed Ewis, Arnold Bregt, and Joep Crompvoets. Spatial data sharing: A cross cultural conceptual model. *Research and Theory in Advancing Spatial Data Infrastructure Concepts*, 2006.

[80] Joshua A T Fairfield. Smart contracts, bitcoin bots, and consumer protection. *Washington and Lee University School of Law*, page 17, 2014.

[81] Mahdi Farnaghi and Ali Mansourian. Blockchain, an enabling technology for transparent and accountable decentralized public participatory GIS. *Cities*, 105:102850, 2020.

[82] R A Finkel and J L Bentley. Quad trees a data structure for retrieval on composite keys. *Acta Inform.*, 4(1):1–9, 1974.

[83] First Nations Information Governance Centre. *Ownership, Control, Access and Possession (OCAP): the Path to First Nations Information Governance (Paper).* Online, 2014.

[84] Ian Foster and Carl Kesselman. Globus: a metacomputing infrastructure toolkit. *The International Journal of Supercomputer Applications and High Performance Computing*, 11(2):115–128, 1997.

[85] The Gordon Foundation. Blockchain technology put to new use in freshwater health, 2018.

[86] Jonathan Fox. The uncertain relationship between transparency and accountability. *Development in Practice*, 17(4):663–671, 2007.

[87] Jonathan L French and Matthew P Wand. Generalized additive models for cancer mapping with incomplete covariates. *Biostatistics*, 5(2):177–191, 2004.

[88] Marco Furini and Valentina Tamanini. Location privacy and public metadata in social media platforms: attitudes, behaviors and opinions. *Multimedia Tools and Applications*, 74(21):9795–9825, 2015.

[89] Dov Gabbay, A. Kurucz, F. Wolter, and M. Zakharyaschev. Applied modal logic. In *Many-Dimensional Modal Logics - Theory and Applications*, Studies in logic and the foundations of mathematics, pages 41–109. Elsevier, 2003.

[90] Prasanna Ganesan, Beverly Yang, and Hector Garcia-Molina. One torus to rule them all: multi-dimensional queries in p2p systems. In *Proceedings of the 7th International Workshop on the Web and Databases: colocated with ACM SIGMOD/PODS 2004*, pages 19–24, 2004.

[91] Sören Gebbert and Edzer Pebesma. A temporal gis for field based environmental modeling. *Environmental Modelling & Software*, 53:1–12, 2014.

[92] Yola Georgiadou, Ourania Kounadi, and Rolf A. de By. Digital earth ethics. In Huadong Guo, Michael F. Goodchild, and Alessandro Annoni, editors, *Manual of Digital Earth*, pages 785–810. Springer, 2020.

[93] Mohammad Gharesifard and Uta Wehn. To share or not to share: Drivers and barriers for sharing data via online amateur weather networks. *J. Hydrol. (Amst.)*, 535:181–190, April 2016.

[94] Gabriel Ghinita, Panos Kalnis, Ali Khoshgozaran, Cyrus Shahabi, and Kian-Lee Tan. Private queries in location based services: anonymizers are not necessary. In *Proceedings of the 2008 ACM SIGMOD international conference on Management of data*, Sigmod '08, pages 121–132. Association for Computing Machinery, 2008.

[95] Gabriel Ghinita, Panos Kalnis, and Spiros Skiadopoulos. MobiHide: A mobilea peer-to-peer system for anonymous location-based queries. In Dimitris Papadias, Donghui Zhang, and George Kollios, editors, *Advances in Spatial and Temporal Databases*, volume 4605, pages 221–238. Springer Berlin Heidelberg, 2007.

[96] Gabriel Ghinita, Keliang Zhao, Dimitris Papadias, and Panos Kalnis. A reciprocal framework for spatial k-anonymity. *Information Systems*, 35(3):299–314, 2010.

[97] github.com/alexbol99. flatten-interval-tree, 2022.

[98] github.com/automerge. automerge, 2022.

[99] github.com/CorentinTh. quadtree-js, 2022.

[100] github.com/DocNow. Hydrator [computer software]., 2020.

[101] github.com/opengeospatial. Geoparquet, 2022.

[102] github.com/vasturiano. d3-octree, 2022.

[103] Marcus Goetz and Alexander Zipf. *The Evolution of Geo-Crowdsourcing: Bringing Volunteered Geographic Information to the Third Dimension*, pages 139–159. Springer, Dordrecht, 2013.

[104] Jianya Gong, Lite Shi, Daosheng Du, and Rolf A de By. Technologies and standards on spatial data sharing. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 34(Part XXX), 2004.

[105] Michael F. Goodchild. Citizens as sensors: the world of volunteered geography. *GeoJournal*, 69(4):211–221, 2007.

[106] Michael F Goodchild. The future of digital earth. *Ann. GIS*, 18(2):93–98, June 2012.

[107] Michael F. Goodchild. Big geodata. In *Comprehensive Geographic Information Systems*, pages 19–25. Elsevier, 2018.

[108] Michael F. Goodchild, Pinde Fu, and Paul Rich. Sharing geographic information: An assessment of the geospatial one-stop. *Annals of the Association of American Geographers*, 97(2):250–266, 2007.

[109] Alan Grainger. Citizen observatories and the new earth observation science. *Remote Sensing*, 9(2):153, 2017.

[110] Carlos Granell and Frank O. Ostermann. Beyond data collection: Objectives and methods of research using vgi and geo-social media for disaster management. *Computers, Environment and Urban Systems*, 59:231–243, 2016.

[111] P Groth. Transparency and reliability in the data supply chain. *IEEE Internet Comput.*, 17(2):69–71, March 2013.

[112] The HDF Group. Hierarchical data format (hdf). *https://www.hdfgroup.org/*, 2023.

[113] Ehud Gudes and Nadav Voloch. An information-flow control model for online social networks based on user-attribute credibility and connection-strength factors. In Itai Dinur, Shlomi Dolev, and Sachin Lodha, editors, *Cyber Security Cryptography and Machine Learning*, volume 10879, pages 55–67. Springer International Publishing, 2018.

[114] Joe Neparrja Gumbula. Exploring the gupapuynga legacy: Strategies for developing the galiwin'ku indigenous knowledge centre. *Australian Academic & Research Libraries*, 36(2):23–26, 2005.

[115] Matthew Harren, Joseph M Hellerstein, Ryan Huebsch, Boon Thau Loo, Scott Shenker, and Ion Stoica. Complex queries in DHT-based peer-to-peer networks. In *Peer-to-Peer Systems*, Lecture notes in computer science, pages 242–250. Springer Berlin Heidelberg, Berlin, Heidelberg, 2002.

[116] Yahya Hassanzadeh-Nazarabadi, Sanaz Taheri-Boshrooyeh, and Öznur Özkasap. Dht-based edge and fog computing systems: Infrastructures and applications. In *IEEE INFOCOM 2022-IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, pages 1–6. IEEE, 2022.

[117] Zhenwen He, Chonglong Wu, Gang Liu, Zufang Zheng, and Yiping Tian. Decomposition tree: A spatio-temporal indexing method for movement big data. *Cluster Computing*, 18(4):1481–1492, dec 2015.

[118] Sabine Hennig and Mariana Belgui. User-centric SDI: Addressing Users Requirements in Third- Generation SDI. The Example of Nature-SDIplus. *Geoforum Perspektiv*, page Vol 10 No 20 (2011): Spatially Enabled Society, September 2013.

[119] Frank Hofmann, Simone Wurster, Eyal Ron, and Moritz Bohmecke-Schwafert. The immutability concept of blockchains and benefits of early standardization. In *2017 ITU Kaleidoscope: Challenges for a Data-Driven Society (ITU K)*, pages 1–8. IEEE, 2017.

[120] Majid Hojati, Carson Farmer, Rob Feick, and Colin Robertson. Decentralized geoprivacy: leveraging social trust on the distributed web. *Geogr. Inf. Syst.*, 35(12):2540–2566, December 2021.

[121] Majid Hojati, Rob Feick, Steven Roberts, Carson Farmer, and Colin Robertson. Distributed spatial data sharing: a new model for data ownership and access control. *Journal of Spatial Information Science, under review*, 2022.

[122] Majid Hojati, Colin Robertson, Steven Roberts, and Chiranjib Chaudhuri. GIScience research challenges for realizing discrete global grid systems as a digital earth. *Big earth data*, pages 1–22, January 2022.

[123] Luke Hutton and Tristan Henderson. "i didn't sign up for this!": informed consent in social network research. In *Proceedings of the 9th International AAAI Conference on Web and Social Media (ICWSM)*, pages 178–187. AAAI Press, May 2015. This work was funded by the Engineering and Physical Sciences Research Council (EPSRC) through a Doctoral Training Grant. ; The 9th International AAAI Conference on Web and Social Media, ICWSM-15 ; Conference date: 26-05-2015 Through 29-05-2015.

[124] Apache iceberg. Apache iceberg, 2022.

[125] Andrea Isogai, Daniel D McCarthy, Holly L Gardner, Jim D Karagatzides, Skye Vandenberg, Christine Barbeau, Nadia Charania, Vicky Edwards, Don Cowan, and Leonard J S Tsuji. Examining the potential use of the collaborative-geomatics informatics tool to foster intergenerational transfer of knowledge in a remote first nation community. *Aust. J. Indig. Educ.*, 42(1):44–57, October 2013.

[126] Chris L Jackins and Steven L Tanimoto. Oct-trees and their use in representing three-dimensional objects. *Comput. Graph. Image Process.*, 14(3):249–270, November 1980.

[127] H.V. Jagadish, Beng Chin Ooi, Quang Hieu Vu, Rong Zhang, and Aoying Zhou. Vbi-tree: A peer-to-peer framework for supporting multi-dimensional indexing schemes. In *22nd International Conference on Data Engineering (ICDE'06)*, pages 34–34, 2006.

[128] Priyank Jain, Manasi Gyanchandani, and Nilay Khare. Big data privacy: a technological perspective and review. *Journal of Big Data*, 3(1):25, 2016.

[129] Marijn Janssen, Yannis Charalabidis, and Anneke Zuiderwijk. Benefits, adoption barriers and myths of open data and open government. *Information Systems Management*, 29(4):258–268, 2012.

[130] Neomi Jayaratne. Enhancing harvester safety and traditional food access through participatory mapping with the ka'a'gee tu first nation of kakisa,

northwest territories. Master's thesis, Wilfrid Laurier University, 1 2021. https://scholars.wlu.ca/etd/2345.

[131] M. Jemel and A. Serhrouchni. Decentralized access control mechanism with temporal dimension based on blockchain. In *2017 IEEE 14th International Conference on e-Business Engineering (ICEBE)*, pages 177–182, 2017.

[132] Henric Johnson, Niklas Lavesson, Haifeng Zhao, and Shyhtsun Felix Wu. On the concept of trust in online social networks. In Luca Salgarelli, Giuseppe Bianchi, and Nicola Blefari-Melazzi, editors, *Trustworthy Internet*, pages 143–157. Springer Milan, 2011.

[133] Elizabeth Judge and Teresa Scassa. Intellectual property and the licensing of canadian government geospatial data: An examination of geoconnections' recommendations for best practices and template licences. *Canadian Geographer / Le Géographe canadien*, 54, 2010.

[134] V Kantere, S Skiadopoulos, and T Sellis. Storing and indexing spatial data in P2P systems. *IEEE Trans. Knowl. Data Eng.*, 21(2):287–300, February 2009.

[135] Verena Kantere and Timos Sellis. Handling spatial data in distributed environments. In *Proceedings of the 15th Annual ACM International Symposium on Advances in Geographic Information Systems*, Gis '07. Association for Computing Machinery, 2007.

[136] Verena Kantere, Spiros Skiadopoulos, and Timos Sellis. Storing and indexing spatial data in p2p systems. *IEEE Transactions on Knowledge and Data Engineering*, 21(2):287–300, 2009.

[137] Kawaljeet Kaur Kapoor, Kuttimani Tamilmani, Nripendra P. Rana, Pushp Patil, Yogesh K. Dwivedi, and Sridhar Nerur. Advances in social media research: Past, present and future. *Information Systems Frontiers*, 20(3):531–558, 2018.

[138] Gordon Keating, Paul Rich, and Marc Witkowski. Challenges for enterprise GIS. *URISA Journal*, 2020.

[139] Patrick Gage Kelley, Michael Benisch, Lorrie Faith Cranor, and Norman Sadeh. When are users comfortable sharing locations with advertisers? In *Proceedings of the 2011 annual conference on Human factors in computing systems - CHI '11*, page 2449. ACM Press, 2011.

[140] Anne-Marie Kermarrec and François Taïani. Want to scale in centralized systems? think p2p. *Journal of Internet Services and Applications*, 6(1):16, 2015.

[141] Carsten Keßler and Grant McKenzie. A geoprivacy manifesto. *Transactions in GIS*, 22(1):3–19, 2018.

[142] E. Mousavi Khaneghah, S. L. Mirtaheri, M. Sharifi, and B. Minaei Bidgoli. Modeling and analysis of access transparency and scalability in p2p distributed systems. *International Journal of Communication Systems*, 27(10):2190–2214, 2014.

[143] H. Kido, Y. Yanagisawa, and T. Satoh. An anonymous communication technique using dummies for location-based services. In *ICPS '05. Proceedings. International Conference on Pervasive Services, 2005.*, pages 88–97, 2005.

[144] Tschangho John Kim. Metadata for geo-spatial data sharing: A comparative analysis. *Ann. Reg. Sci.*, 33(2):171–181, May 1999.

[145] Rob Kitchin and Tracey Lauriault. Small data, data infrastructures and big data, 2014.

[146] Renate Klar and Dirk Lanzerath. The ethics of covid-19 tracking apps – challenges and voluntariness. *Research Ethics*, 16(3-4):1–9, 2020.

[147] Martin Kleppmann and Alastair R Beresford. A Conflict-Free replicated JSON datatype. *IEEE Trans. Parallel Distrib. Syst.*, 28(10):2733–2746, October 2017.

[148] Kaitlin Kok. Monitoring environmental change using a participatory modified photovoice approach with indigenous knowledge holders in kakisa, northwest territories. Master's thesis, Wilfrid Laurier University, 1 2020. https://scholars.wlu.ca/etd/2233.

[149] Alexander Kotsev, Marco Minghini, Robert Tomas, Vlado Cetl, and Michael Lutz. From spatial data infrastructures to data spaces–a technological perspective on the evolution of european SDIs. *ISPRS International Journal of Geo-Information*, 9(3):176, 2020.

[150] Ourania Kounadi and Michael Leitner. Why does geoprivacy matter? the scientific publication of confidential data presented on maps:. *Journal of Empirical Research on Human Research Ethics*, 2014.

[151] Ourania Kounadi and Bernd Resch. A geoprivacy by design guideline for research campaigns that use participatory sensing data. *Journal of Empirical Research on Human Research Ethics*, 13(3):203–222, 2018.

[152] Tahu Kukutai and John Taylor. *Indigenous Data Sovereignty: Toward an agenda*. ANU Press, 2016.

[153] Rabindra Lamsal. Coronavirus (covid-19) geo-tagged tweets dataset, 2020.

[154] Marek Laskowski. A blockchain-enabled participatory decision support framework. In Dongwon Lee, Yu-Ru Lin, Nathaniel Osgood, and Robert Thomson, editors, *Social, Cultural, and Behavioral Modeling*, Lecture Notes in Computer Science, pages 329–334. Springer International Publishing, 2017.

[155] Jae Yong Lee and Mei-Po Kwan. Visualisation Of Socio-spatial Isolation Based On Human Activity Patterns And Social Networks In Space-time: Human Activity Patterns And Social Networks In Space-time. *Tijdschrift voor economische en sociale geografie*, 102(4):468–485, 2011.

[156] Michael Leitner and Andrew Curtis. Cartographic guidelines for geographically masking the locations of confidential point data. *Cartographic Perspectives*, (49):22–39, 2004.

[157] Silvie Levy, Ehud Gudes, and Nurit Gal-Oz. Sharing-habits based privacy control in social networks. In Silvio Ranise and Vipin Swarup, editors, *Data and Applications Security and Privacy XXX*, volume 9766, pages 217–232. Springer International Publishing, 2016.

[158] Mingke Li and Emmanuel Stefanakis. Geospatial operations of discrete global grid systems—a comparison with traditional GIS. *J. geovisualization spat. anal.*, 4(2), December 2020.

[159] SHL Liang et al. A new peer-to-peer-based interoperable spatial sensor web architecture. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 3(4):1, 2008.

[160] Xueping Liang, Sachin Shetty, Juan Zhao, Daniel Bowden, Danyi Li, and Jihong Liu. Towards decentralized accountability and self-sovereignty in healthcare systems. In Sihan Qing, Chris Mitchell, Liqun Chen, and Dongmei Liu, editors, *Information and Communications Security*, Lecture Notes in Computer Science, pages 387–398. Springer International Publishing, 2018.

[161] Bin Liu, Mads Schaarup Andersen, Florian Schaub, Hazim Almuhimedi, Shikun (Aerin) Zhang, Norman Sadeh, Yuvraj Agarwal, and Alessandro Acquisti.

Follow my recommendations: A personalized privacy assistant for mobile app permissions. In *Follow My Recommendations: A Personalized Privacy Assistant for Mobile App Permissions*, pages 27–41, 2016.

[162] Bin Liu, Wang-Chien Lee, and Dik Lun Lee. Supporting complex multi-dimensional queries in P2P systems. In *25th IEEE International Conference on Distributed Computing Systems (ICDCS'05)*. IEEE, 2005.

[163] S. Liu, Q. Qu, L. Chen, and L. M. Ni. Smc: A practical schema for privacy-preserved data sharing over distributed data streams. *IEEE Transactions on Big Data*, 1(2):68–81, 2015.

[164] Roger A. Longhorn. *1 "The Impact of Data Access Policies on Regional Spatial Data Infrastructure"*.

[165] Hickling Arthurs Low and Hatfield Consultants. Environmental scan on user needs assessments for the arctic spatial data infrastructure, 2019.

[166] Ali Mahdavi-Amiri, Troy Alderson, and Faramarz Samavati. A survey of digital earth. *Computers & Graphics*, 53:95–117, Dec 2015.

[167] Ali Mahdavi-Amiri, Erika Harrison, and Faramarz Samavati. Hexagonal connectivity maps for digital earth. *International Journal of Digital Earth*, 8(9):750–769, 2015.

[168] Miguel D Mahecha, Fabian Gans, Gunnar Brandt, Rune Christiansen, Sarah E Cornell, Normann Fomferra, Guido Kraemer, Jonas Peters, Paul Bodesheim, Gustau Camps-Valls, et al. Earth system data cubes unravel global multivariate dynamics. *Earth System Dynamics*, 11(1):201–234, 2020.

[169] Ahmed R Mahmood, Sri Punni, and Walid G Aref. Spatio-temporal access methods: a survey (2010 - 2017). *Geoinformatica*, 23(1):1–36, January 2019.

[170] Matthew Maloley, Cameron Wilson, and Simon Riopel. Enabling access to arctic location based information. In *Arctic Data Committee Meeting*. Arctic Data Committee Meeting, Arctic Data Committee Meeting, 7 2017.

[171] M. Marsal-Llacuna and M. Oliver-Riera. The standards revolution: Who will first put this new kid on the blockchain? In *2017 ITU Kaleidoscope: Challenges for a Data-Driven Society (ITU K)*, pages 1–7, 2017.

[172] Maria-Lluisa Marsal-Llacuna and Mark Evan Segal. The intelligenter method (i) for making "smarter" city projects and plans. *Cities*, 55:127–138, 2016.

[173] Kirsten E. M. Martin and Helen F. Nissenbaum. What is it about location? *SSRN Electronic Journal*, 2019.

[174] Andrii Martyn et al. The concept of land plot as a combination of smart contracts: A vision for creating blockchain cadastre. *Balt. Surv*, 8:68–73, 2018.

[175] Matthew S. Mayernik. Open data: Accountability and transparency - matthew s mayernik, 2017. *Big Data & Society*, 2017.

[176] Petar Maymounkov and David Mazières. Kademlia: A peer-to-peer information system based on the XOR metric. In Peter Druschel, Frans Kaashoek, and Antony Rowstron, editors, *Peer-to-Peer Systems*, Lecture Notes in Computer Science, pages 53–65. Springer, 2002.

[177] Michael K McCALL and Peter A Minang. Assessing participatory GIS for community-based natural resource management: claiming community forests in cameroon. *Geogr. J.*, 171(4):340–356, December 2005.

[178] Muniba Memon, Syed Shahbaz Hussain, Umair Ahmed Bajwa, and Asad Ikhlas. Blockchain beyond bitcoin: Blockchain technology challenges and real-world applications. In *2018 International Conference on Computing, Electronics Communications Engineering (iCCECE)*, pages 29–34, 2018.

[179] William Michener, Dave Vieglais, Todd Vision, John Kunze, Patricia Cruse, and Greg Janée. Dataone: Data observation network for earth-preserving data and enabling innovation in the biological and environmental sciences. *D-Lib Magazine*, 17:3, 2011.

[180] Gaurav Misra, Jose M. Such, and Hamed Balogun. IMPROVE - identifying minimal PROfile VEctors for similarity based access control. In *2016 IEEE Trustcom/BigDataSE/ISPA*, pages 868–875. IEEE, 2016.

[181] Hossein Mohammadi, Abbas Rajabifard, and Ian P Williamson. Enabling spatial data sharing through multi-source spatial data integration. 2009.

[182] Mohamed F. Mokbel, Thanaa M. Ghanem, and Walid G. Aref. Spatio-temporal access methods. *IEEE Data Eng. Bull.*, 26(2):40–49, 2003.

[183] Anirban Mondal, Yi Lifu, and Masaru Kitsuregawa. P2PR-tree: An r-tree-based spatial index for peer-to-peer environments. In *Current Trends in Database Technology - EDBT 2004 Workshops*, Lecture notes in computer science, pages 516–525. Springer Berlin Heidelberg, Berlin, Heidelberg, 2004.

[184] Francisco Javier Moreno and Santiago Hernández. An algorithm for identifying the best current friend in a social network: Un algoritmo para determinar el mejor amigo actual en una red social. *Ingeniería e Investigación*, 35(2):80–88, 2015.

[185] Guy M Morton. A computer oriented geodetic data base and a new technique in file sequencing. *International Business Machines*, 1966.

[186] Abdullah Murad, Brian Hilton, Thomas Horan, and John Tangenberg. Protecting patient geo-privacy via a triangular displacement geo-masking method. In *Proceedings of the 1st ACM SIGSPATIAL International Workshop on Privacy in Geographic Information Collection and Analysis - GeoPrivacy '14*, pages 1–9. ACM Press, 2014.

[187] Satoshi Nakamoto. Bitcoin: A peer-to-peer electronic cash system. *Decentralized Business Review*, page 21260, 2008.

[188] Natural Resources Canada. The spatial data infrastructure (sdi) manual for the arctic, 2018.

[189] Muqaddas Naz, Fahad A. Al-zahrani, Rabiya Khalid, Nadeem Javaid, Ali Mustafa Qamar, Muhammad Khalil Afzal, and Muhammad Shafiq. A secure data sharing platform using blockchain and interplanetary file system. *Sustainability*, 11:7054, 2019.

[190] Pascal Neis, Dennis Zielstra, and Alexander Zipf. The street network evolution of crowdsourced maps: Openstreetmap in germany 2007–2011. *Future Internet*, 4:1–21, 2012.

[191] N. Nizamuddin, K. Salah, M. Ajmal Azad, J. Arshad, and M. H. Rehman. Decentralized document version control using ethereum blockchain and IPFS. *Computers & Electrical Engineering*, 76:183–197, 2019.

[192] Michael Nofer, Peter Gomber, Oliver Hinz, and Dirk Schiereck. Blockchain. *Business & Information Systems Engineering*, 59(3):183–187, 2017.

[193] Matthieu Noucher, Françoise Gourmelon, Pierre Gautreau, Jade Georis-Creuseveau, Adeline Maulpoix, Julie Pierson, Nathalie Pinède, Olivier Pissoat, and Mathias Rouan. Spatial data sharing: A pilot study of french sdis. *ISPRS International Journal of Geo-Information*, 6(44):99, 2017.

[194] Timothy Nugent, Fabio Petroni, Benedict Whittam Smith, and Jochen L. Leidner. An on-chain method for automatic entitlement management using blockchain smart

contracts. In Witold Abramowicz and Rafael Corchuelo, editors, *Business Information Systems Workshops*, Lecture Notes in Business Information Processing, pages 255–266. Springer International Publishing, 2019.

[195] Ogc. *OGC Arctic Spatial Data Pilot: Phase 1 Report - Spatial Data Sharing for the Arctic*. Number 1 in 1. Ogc, 2017.

[196] Antoine Olbrechts. Statement by the edpb chair on the processing of personal data in the context of the covid-19 outbreak. text. european data protection board-european data protection board, 2020.

[197] Svein Ølnes, Jolien Ubacht, and Marijn Janssen. Blockchain in government: Benefits and implications of distributed ledger technology for information sharing. *Government Information Quarterly*, 34(3):355–364, 2017.

[198] Md Mehedi Hassan Onik, Chul-Soo Kim, Nam-Yong Lee, and Jinhong Yang. Privacy-aware blockchain for personal data sharing and tracking. *Open Computer Science*, 9:80–91, 2019.

[199] Maya Oppenheim. People are starting to trust 'traditional' media over social media, report finds, 2018.

[200] Jun Pang and Yang Zhang. A new access control scheme for facebook-style social networks. *Computers & Security*, 54:44–59, Oct 2015.

[201] Christopher Parker, Andrew May, and Val Mitchell. The role of vgi and pgi in supporting outdoor activities. *Applied ergonomics*, 44, 2012.

[202] Alireza Partovi, Wei Zheng, Taeho Jung, and Hai Lin. Ensuring privacy in location-based services: A model-based approach. *ArXiv*, 2020.

[203] Vishal Patel. A framework for secure and decentralized sharing of medical imaging data via blockchain consensus. *Health Informatics Journal*, 25(4):1398–1411, 2019.

[204] Nikos Pelekis, Babis Theodoulidis, Ioannis Kopanakis, and Yannis Theodoridis. Literature review of spatio-temporal database models. *The Knowledge Engineering Review*, 19(3):235–274, 2004.

[205] Z.-R. Peng. A proposed framework for feature-level geospatial data sharing: a case study for transportation network data. *International Journal of Geographical Information Science*, 19(4):459–481, 2005.

[206] Donna J. Peuquet and Niu Duan. An event-based spatiotemporal data model (estdm) for temporal analysis of geographical data. *International Journal of Geographical Information Systems*, 9(1):7–24, 1995.

[207] PostGIS. Postgis clustering data, 2022.

[208] Gunnar Prause. Smart contracts for smart supply chains. *IFAC-PapersOnLine*, 52(13):2501–2506, 2019.

[209] Matthew B. J. Purss, Robert Gibb, Faramarz Samavati, Perry Peterson, and Jin Ben. The OGC® discrete global grid system core standard: A framework for rapid geospatial integration. In *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 3610–3613. IEEE, 2016.

[210] Chunyao Qian, Chao Yi, Chengqi Cheng, Guoliang Pu, Xiaofeng Wei, and Huangchuang Zhang. Geosot-based spatiotemporal index of massive trajectory data. *ISPRS International Journal of Geo-Information*, 8(6):284, 2019.

[211] Qiang Qu, Ildar Nurgaliev, Muhammad Muzammal, Christian S Jensen, and Jianping Fan. On spatio-temporal blockchain query processing. *Future Generation Computer Systems*, 98:208–218, 2019.

[212] Sriram Ramabhadran, Sylvia Ratnasamy, and Joseph M. Hellerstein. Prefix hash tree an indexing data structure over distributed hash tables. In *PODC 2004 conference*, 2004.

[213] G. S. Ramachandran, R. Radhakrishnan, and B. Krishnamachari. Towards a decentralized data marketplace for smart cities. In *2018 IEEE International Smart Cities Conference (ISC2)*, pages 1–8, 2018.

[214] Canada Natural Resources. The spatial data infrastructure (SDI) manual for the arctic, 2018.

[215] Colin Robertson, Chiranjib Chaudhuri, Majid Hojati, and Steven A Roberts. An integrated environmental analytics system (IDEAS) based on a DGGS. *ISPRS J. Photogramm. Remote Sens.*, 162:214–228, April 2020.

[216] Carmen Ruiz Vicente, Dario Freni, Claudio Bettini, and Christian S. Jensen. Location-related privacy in geo-social networks. *IEEE Internet Computing*, 15(3):20–27, 2011.

[217] Robert A. Rundstrom. GIS, indigenous peoples, and epistemological diversity. *Cartography and Geographic Information Systems*, 22(1):45–57, 1995.

[218] Ozgur D Sahin, Shyam Antony, Divyakant Agrawal, and A El Abbadi. Probe: Multidimensional range queries in p2p networks. In *International Conference on Web Information Systems Engineering*, pages 332–346. Springer, 2005.

[219] Kevin Sahr. Central place indexing: Hierarchical linear indexing systems for mixed-aperture hexagonal discrete global grid systems. *Cartographica: The International Journal for Geographic Information and Geovisualization*, 54(1):16–29, 2019.

[220] Kevin Sahr, Denis White, and A. Jon Kimerling. Geodesic discrete global grid systems. *Cartography and Geographic Information Science*, 30(2):121–134, 2003.

[221] Pierangela Samarati and Latanya Sweeney. Protecting privacy when disclosing information: k-anonymity and its enforcement through generalization and suppression. 1998.

[222] Dipto Sarkar, Clio Andris, Colin A. Chapman, and Raja Sengupta. Metrics for characterizing network structure and node importance in spatial social networks. *International Journal of Geographical Information Science*, 33(5):1017–1039, 2019.

[223] Dipto Sarkar, Renee Sieber, and Raja Sengupta. GIScience considerations in spatial social networks. In Jennifer A. Miller, David O'Sullivan, and Nancy Wiegand, editors, *Geographic Information Science*, Lecture Notes in Computer Science, pages 85–98. Springer International Publishing, 2016.

[224] Teresa Scassa. Legal issues with volunteered geographic information. *The Canadian Geographer / Le Géographe canadien*, 57(1):1–10, 2013.

[225] Stephan Schlosser, Daniele Toninelli, and Michela Cameletti. Comparing methods to collect and geolocate tweets in great britain. *Journal of Open Innovation: Technology, Market, and Complexity*, 7(1), 2021.

[226] Lorenz Schwittmann, Christopher Boelmann, Matthäus Wander, and Torben Weis. SoNet – privacy and replication in federated online social networks. In *2013 IEEE 33rd International Conference on Distributed Computing Systems Workshops*, pages 51–57, 2013.

[227] Sukhjit Sehra, Jaiteg Singh, and Hardeep Rai. Assessment of openstreetmap data - a review. *International Journal of Computer Applications*, 76, 2013.

[228] Hossein Shafagh, Lukas Burkhalter, Anwar Hithnawi, and Simon Duquennoy. Towards blockchain-based auditable storage and sharing of IoT data. *arXiv:1705.08230 [cs]*, 2017.

[229] Taylor Shelton, Ate Poorthuis, and Matthew Zook. Social media and the city: Rethinking urban socio-spatial inequality using user-generated geographic information. *Landscape and Urban Planning*, 142:198–211, 2015.

[230] Charles Shen and Feniosky Pena-Mora. Blockchain for cities–a systematic literature review. *IEEE Access*, 6:76787–76819, 2018.

[231] Renée E. Sieber. Spatial data access by the grassroots. *Cartography and Geographic Information Science*, 34(1):47–62, 2007.

[232] Djoko Sigit Sayogo, Jing Zhang, Theresa A Pardo, Giri K Tayi, Jana Hrdinova, David F Andersen, and Luis Felipe Luna-Reyes. Going beyond open data: Challenges and motivations for smart disclosure in ethical consumption. *J. theor. appl. electron. commer. res.*, 9(2):3–4, August 2014.

[233] Rachel Charlotte Smith and Ole Sejer Iversen. Participatory design for sustainable social change. *Design Studies*, 59:9–36, 2018. Participatory Design.

[234] Alessandro Soro and Cristian Lai. Range-capable distributed hash tables. In *Proceedings of the 3rd ACM Workshop On Geographic Information Retrieval,*, 2006.

[235] Alessandro Soro and Cristian Lai. Range-capable distributed hash tables. In *Gir*, 2006.

[236] Andrew Spring. Capitals, climate change and food security: Building sustainable food systems in northern canadian indigenous communities. Master's thesis, Wilfrid Laurier University, 1 2018. https://scholars.wlu.ca/etd/2034.

[237] Anthony Stefanidis, Andrew Crooks, and Jacek Radzikowski. Harvesting ambient geospatial information from social media feeds. *GeoJournal*, 78(2):319–338, 2013.

[238] Mathis Steichen, Beltran Fiz, Robert Norvill, Wazen Shbair, and Radu State. Blockchain-based, decentralized access control for IPFS. In *2018 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData)*, pages 1499–1506. IEEE, 2018.

[239] Makere Stewart-Harawira. Challenging knowledge capitalism: Indigenous research in the 21st century. *Social. stud.*, 9(1), June 2013.

[240] Ion Stocia. Chord: A scalable peer-to-peer lookup service for internet applications. In *Proc. of ACM SIGCOMM, 2001*, 2001.

[241] Christoph Sturm, Klaus R Dittrich, and Patrick Ziegler. An access control mechanism for P2P collaborations. In *Proceedings of the 2008 international workshop on Data management in peer-to-peer systems - DaMaP '08*, New York, New York, USA, 2008. ACM Press.

[242] Yuxiang Sun, Tianyi Zhao, Seulgi Yoon, and Yongju Lee. A hybrid approach combining r∗-tree and k-d trees to improve linked open data query performance. *Applied Sciences*, 11(5), 2021.

[243] Ali Sunyaev. *Distributed Ledger Technology*, pages 265–299. Springer International Publishing, 2020.

[244] Melanie Swan. *Blockchain: Blueprint for a New Economy.* "O'Reilly Media, Inc.", 2015.

[245] David Swanlund and Nadine Schuurman. Resisting geosurveillance: A survey of tactics and strategies for spatial privacy. *Progress in Human Geography*, 43(4):596–610, 2019.

[246] Latanya Sweeney. k-anonymity: a model for protecting privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 10(5):557–570, 2002.

[247] Andrew S. Tanenbaum and Maarten van Steen. *Distributed Systems: Principles and Paradigms.* Pearson Prentice Hall, 2 edition edition, 2016.

[248] Chunqiang Tang, Zhichen Xu, and Sandhya Dwarkadas. Peer-to-peer information retrieval using self-organizing semantic overlay networks. In *Proceedings of the 2003 conference on Applications, technologies, architectures, and protocols for computer communications*, pages 175–186, 2003.

[249] Yufei Tao and Dimitris Papadias. Mv3r-tree: A spatio-temporal access method for timestamp and interval queries. In *Proceedings of the 27th International Conference on Very Large Data Bases*, Vldb '01, page 431–440, San Francisco, CA, USA, 2001. Morgan Kaufmann Publishers Inc.

[250] Yannis Theodoridis, Michael Vazirgiannis, and Timos Sellis. Spatio-temporal indexing for large multimedia applications. In *Proceedings of the Third IEEE International Conference on Multimedia Computing and Systems*. IEEE, 1996.

[251] Felicity Thomas. *Handbook of Migration and Health*. Edward Elgar Publishing, 2016.

[252] Davies Tim, Walker B, Stephen, and Rubinstein Mor. *The State of Open Data: Histories and Horizons*. African Minds, 2019.

[253] Peter Triantafillou and Theoni Pitoura. Towards a unifying framework for complex query processing over structured peer-to-peer data networks. In *Databases, Information Systems, and Peer-to-Peer Computing*, Lecture notes in computer science, pages 169–183. Springer Berlin Heidelberg, Berlin, Heidelberg, 2004.

[254] Aggeliki Tsohou and Eleni Kosta. Enabling valid informed consent for location tracking through privacy awareness of users: A process theory. *Computer Law & Security Review*, 33(4):434–457, 2017.

[255] Twitter. Twitter trends FAQs, 2020.

[256] Uber. Uber h3-js, 2020.

[257] Sarah Underwood. Blockchain beyond bitcoin. *Communications of the ACM*, 59(11):15–17, 2016.

[258] Jennifer M. Urban, Chris Jay Hoofnagle, and Su Li. Mobile phones and privacy. *SSRN Electronic Journal*, 2012.

[259] USA.gov. Application of fincens regulations to certain business models involving convertible virtual currencies, 2019.

[260] Marshall Van Alstyne, Erik Brynjolfsson, and Stuart Madnick. Why not one big database? principles for data ownership. *Decision Support Systems*, 15(4):267–284, 1995.

[261] Siri Veland, Amanda Lynch, Zachary Bischoff-Mattson, Lee Joachim, and Noor Johnson. All strings attached: Negotiating relationships of geographic information science. *Geographical Research*, 52(3):296–308, 2014.

[262] Susan Verducci and Andreas Schröer. *Social Trust*, pages 1453–1458. Springer US, 2010.

[263] Jeroen Verplanke, Michael K. McCall, Claudia Uberhuaga, Giacomo Rambaldi, and Muki Haklay. A shared perspective for PGIS and VGI. *The Cartographic Journal*, 53(4):308–317, 2016.

[264] Akrivi Vlachou, Christos Doulkeridis, Kjetil Nørvåg, and Yannis Kotidis. Peer-to-peer query processing over multidimensional data.

[265] Nadav Voloch, Priel Levy, Mor Elmakies, and Ehud Gudes. An access control model for data security in online social networks based on role and user credibility. In Shlomi Dolev, Danny Hendler, Sachin Lodha, and Moti Yung, editors, *Cyber Security Cryptography and Machine Learning*, Lecture Notes in Computer Science, pages 156–168. Springer International Publishing, 2019.

[266] Nadav Voloch, Priel Nissim, Mor Elmakies, and Ehud Gudes. A role and trust access control model for preserving privacy and image anonymization in social networks. In Weizhi Meng, Piotr Cofta, Christian Damsgaard Jensen, and Tyrone Grandison, editors, *Trust Management XIII*, IFIP Advances in Information and Communication Technology, pages 19–27. Springer International Publishing, 2019.

[267] Radoš Šumarada. Legal issues regarding spatial data. *FIG XXII International Congress*, page 10, 2002.

[268] Smita Wadhwa and Pawan Gupta. Distributed locality sensitivity hashing. In *2010 7th IEEE Consumer Communications and Networking Conference*, pages 1–4, 2010.

[269] Donggen Wang, Fei Li, and Yanwei Chai. Activity spaces and sociospatial segregation in beijing. *Urban Geography*, 33(2):256–277, 2012.

[270] Fangju Wang and G. Brent Hall. Fuzzy representation of geographical boundaries in gis. *International Journal of Geographical Information Systems*, 10(5):573–590, 1996.

[271] Hua Wang and Lili Sun. Trust-involved access control in collaborative open social networks. In *2010 Fourth International Conference on Network and System Security*, pages 239–246, 2010.

[272] Shangping Wang, Yinglong Zhang, and Yaling Zhang. A blockchain-based framework for data sharing with fine-grained access control in decentralized storage systems. *IEEE Access*, 6:38437–38450, 2018.

[273] Shaowen Wang, Mary Kathryn Cowles, and Marc P. Armstrong. Grid computing of spatial statistics: using the teragrid for g(d) analysis. *Concurrency and Computation: Practice and Experience*, 20(14):1697–1720, 2008.

[274] Shaowen Wang and Yan Liu. Teragrid giscience gateway: Bridging cyberinfrastructure and giscience. *International Journal of Geographical Information Science*, 23(5):631–656, 2009.

[275] Shuai Wang, Yong Yuan, Xiao Wang, Juanjuan Li, Rui Qin, and Fei-Yue Wang. An overview of smart contract: Architecture, applications, and future trends. In *2018 IEEE Intelligent Vehicles Symposium (IV)*, pages 108–113, 2018.

[276] Uta Wehn. The distribution of spatial data: Data sharing and mediated cooperation. In *Inside the Communication Revolution: Evolving Patterns of Social and Technical Interaction*, pages 186–205. Oxford University Press, 2002.

[277] Uta Wehn. In search of rigorous models for policy-oriented research: A behavioral approach to spatial data sharing. *URISA Journal*, 15:19–28, 2003.

[278] Uta Wehn. *Mapping the Determinants of Spatial Data Sharing*. Routledge, 2017.

[279] Aaron Wright and Primavera De Filippi. Decentralized blockchain technology and the rise of lex cryptographia. *Social Science Research Network*, 1(Id 2580664), 2015.

[280] Jizhe Xia, Chaowei Yang, and Qingquan Li. Building a spatiotemporal index for earth observation big data. *International journal of applied earth observation and geoinformation*, 73:245–252, 2018.

[281] Junfeng Xie, F. Richard Yu, Tao Huang, Renchao Xie, Jiang Liu, and Yunjie Liu. A survey on the scalability of blockchain systems. *IEEE Network*, 33(5):166–173, 2019.

[282] William E. Yancey, William E. Winkler, and Robert H. Creecy. Disclosure risk assessment in perturbative microdata protection. In Josep Domingo-Ferrer, editor, *Inference Control in Statistical Databases: From Theory to Practice*, Lecture Notes in Computer Science, pages 135–152. Springer, 2002.

[283] C. Yang and R. Raskin. Introduction to distributed geographic information processing research. *International Journal of Geographical Information Science*, 23(5):553–560, 2009.

[284] Chaowei Yang, Wenwen Li, Jibo Xie, and Bin Zhou. Distributed geospatial information processing: sharing distributed geospatial resources to support digital earth. *International Journal of Digital Earth*, 1(3):259–278, 2008.

[285] Chaowei Phil Yang and C. Vincent Tao. *Distributed Geospatial Information Service*, pages 103–120. Springer, 2006.

[286] W Ye, A I Khan, and E A Kendall. Distributed network file storage for a serverless (P2P) network. In *The 11th IEEE International Conference on Networks, 2003. ICON2003*. IEEE, 2004.

[287] David Yermack. Corporate governance and blockchains. *Review of Finance*, 21(1):7–31, 2017.

[288] Man Lung Yiu, Gabriel Ghinita, Christian S. Jensen, and Panos Kalnis. Outsourcing search services on private spatial data. In *2009 IEEE 25th International Conference on Data Engineering*, pages 1140–1143, 2009.

[289] Man Lung Yiu, Christian S. Jensen, Xuegang Huang, and Hua Lu. SpaceTwist: Managing the trade-offs among location privacy, query performance, and query accuracy in mobile services. In *2008 IEEE 24th International Conference on Data Engineering*, pages 366–375, 2008.

[290] Tun-Hao You, Wen-Chih Peng, and Wang-Chien Lee. Protecting moving trajectories with dummies. In *2007 International Conference on Mobile Data Management*, pages 278–282, 2007.

[291] Jia Yu, Jinxuan Wu, and Mohamed Sarwat. Geospark: A cluster computing framework for processing large-scale spatial data. In *Proceedings of the 23rd SIGSPATIAL international conference on advances in geographic information systems*, pages 1–4, 2015.

[292] May Yuan. Temporal gis and spatio-temporal modeling. In *Proceedings of Third International Conference Workshop on Integrating GIS and Environment Modeling, Santa Fe, NM*, volume 33, 1996.

[293] Victor Zakhary, Cetin Sahin, Theodore Georgiou, and Amr El Abbadi. LocBorg: Hiding social media user location while maintaining online persona. In *Proceedings of the 25th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pages 1–4, 2017.

[294] C. Zhang, W. Li, and T. Zhao. Geospatial data sharing based on geospatial semantic web technologies. *Journal of Spatial Science*, 52(2):35–49, 2007.

[295] Chengyuan Zhang, Lei Zhu, Jun Long, Shuangqiao Lin, Zhan Yang, and Wenti Huang. A hybrid index model for efficient spatio-temporal search in HBase. In *Lecture Notes in Computer Science*, Lecture notes in computer science, pages 108–120. Springer International Publishing, Cham, 2018.

[296] Chi Zhang, Arvind Krishnamurthy, and Randolph Y Wang. Skipindex: Towards a scalable peer-to-peer index service for high dimensional data. *Department of Computer Science, Princeton University, New Jersey, USA, Tech. Rep*, pages 703–04, 2004.

[297] Chuanrong Zhang and Weidong Li. The roles of web feature and web map services in real-time geospatial data sharing for time-critical applications. *Cartography and Geographic Information Science*, 32(4):269–283, 2005.

[298] Su Zhang, Scott M. Freundschuh, Kate Lenzer, and Paul A. Zandbergen. The location swapping method for geomasking. *Cartography and Geographic Information Science*, 44(1):22–34, 2017.

[299] Yan Zhang, Bing Li, Ben Liu, Jiaxin Wu, Yazhou Wang, and Xia Yang. An attribute-based collaborative access control scheme using blockchain for IoT devices. *Electronics (Basel)*, 9(2):285, February 2020.

[300] Yun Zhang, Zhi Tang, Jing Huang, Yue Ding, Hao He, Xiaosheng Xia, and Chunhua Li. A decentralized model for spatial data digital rights management. *ISPRS International Journal of Geo-Information*, 9(2):84, 2020.

[301] Bo Zhao and Daniel Z. Sui. True lies in geospatial big data: detecting location spoofing in social media. *Annals of GIS*, 23(1):1–14, 2017.

[302] Kaiqi Zhao, Lisi Chen, and Gao Cong. Topic exploration in Spatio-Temporal document collections. In *Proceedings of the 2016 International Conference on Management of Data*, New York, NY, USA, June 2016. Acm.

[303] Zibin Zheng, Shaoan Xie, Hong-Ning Dai, Xiangping Chen, and Huaimin Wang. Blockchain challenges and opportunities: A survey. *International Journal of Web and Grid Services*, 14(4):352–375, 2018.

[304] Ge Zhong and Urs Hengartner. A distributed k-anonymity protocol for location privacy. In *2009 IEEE International Conference on Pervasive Computing and Communications*, pages 1–10, 2009.

[305] Mohammed Zia. B-drive: A blockchain based distributed iot network for smart urban transportation. *Blockchain: Research and Applications*, 2(4):100033, 2021.

[306] R. Zimmermann, We-Shinn Ku, and Haojun Wang. Spatial data query support in peer-to-peer systems. In *Proceedings of the 28th Annual International Computer Software and Applications Conference, 2004. COMPSAC 2004.*, volume 2, pages 82–85 vol.2, 2004.

[307] Matthew A. Zook and Mark Graham. The creative reconstruction of the internet google and the privatization of cyberspace and DigiPlace. *Geoforum*, 38(6):1322–1343, 2007.