Geographical Information Management and Applications

The current state of open standards for spatial data in Solid environments for secure spatial personal data and use by third parties.

MSc. thesis

Juraj Zvolenský

7300905

j.zvolensky@students.uu.nl

Supervisors:

Drs. Wilko Quak (C.W.Quak@tudelft.nl)

Dr. Linda van den Brink (l.vandenbrink@geonovum.nl)

Responsible Professor:

Prof.dr.ir. P.J.M. van Oosterom (P.J.M.vanOosterom@tudelft.nl)

Deadline: 18/08/2023

Universiteit Utrecht

TU Delft

University of Twente.

Wageningen UR

Preface and acknowledgements

This master's thesis is part of the Geographical Information Management and Applications (GIMA) program. The objective has been to examine the current state of open standards for spatial data in the Solid environment.

During my studies I have had the chance to explore new concepts and possibilities in the world of Geography. I have completed a smaller project on Solid, and I thought it was very interesting and wanted to explore it more.

I would like to thank my supervisors Linda, Wilko, and Peter for their enthusiasm and guidance throughout this journey. Linda, thank you for teaching me more about standards and your ideas and input during the research. Wilko, thank you for helping with the use case design and providing feedback for the research. Peter, thank you for all your valuable feedback and suggestions. It would not have been possible to complete this project without all your incredible support and supervision. I would also like to say thanks to dr.ir. Lukasz Grus, Han Wammes, and Michiel De Jong for their comments and support throughout the thesis.

Furthermore, thank you to my supportive family, friends, and other GIMA students.

Juraj Zvolenský

Abstract

Data ownership and privacy has become a big talking point in the last decade. The popularity of the online world created several data privacy issues. In the European Union, several legislative changes have been developed, such as GDPR, to improve the privacy and safety of online users. However, this has not completely stopped the apparent abuse of personal data collection and analysis. Solid is a technology developed by Sr. Tim Berners-Lee, the inventor of the World Wide Web, to regain data ownership and privacy and put it back into the users' hands. The heart of Solid is what is called a Pod. A Pod is a file based personal data storage server. Pods also support structured Linked Data resources. This paper examines how spatial data interacts with Pods.

The main research question "*What open standards and protocols can be used to store and interact with spatial data in a Solid environment and a GI system?*" has been developed to research different spatial data standards and ways that Solid can be used to interact with. To help with answering the main question, four sub questions have been developed: *"What technologies enable Solid to create an open and interoperable environment?", "Is it possible to limit the sharing of spatial data to specific parties?", "Is it useful to share this kind of data via a pod?", "What are the challenges with regards to the usability of Solid?"*

To answer these research questions, the project is separated into two parts. First, a literature review has been conducted to understand the concepts behind Solid, its motivation, and the fundamentals of how it works. Several existing Solid-based projects have been examined to understand its possibilities and the variety of use cases it can be used for. The second part involves experimentation with the Pods themselves. Five location points have been formatted according to five open spatial data standards which were chosen based on literature. They are GeoJSON, GML, GPX, KML, and Moving features.

The literature review and discussion provide insight into the current state of Solid and its support for spatial data and its applications. It has been found that while it is not specifically built for spatial data in mind, its open nature is able to accommodate a wide range of data types in file formats. Furthermore, its wide range of access rules mean that users can accurately set up access to certain files or folders within their Pod. Retrieving data from a file located in a Pod is limited to obtaining the whole dataset and performing the queries on the client side, which is not efficient.

Some design principles of Solid remain unanswered and have been recommended for further research. While interoperability is a key principle in Solid environments, they are not a necessity and therefore there are concerns with its interoperability claims. Utilizing Linked Data to its full potential requires extra steps in terms of data preparation. How exactly this affects the usability needs to be researched further. It would be interesting to explore querying of Linked Data compared to file based datasets as it is not currently possible to query data stored in a file. Data in the form of files can only be retrieved as the whole file and then queried by the client. Querying the data before receiving it from a Pod was tested with server-side filtering, with unsuccessful results.

Table of Contents

List of abbreviations

GDPR - General Data Protection Regulation

EU - European Union

GIS - Geographic Information System

ISO - International Organization for Standardization

OGC - Open Geospatial Consortium

W3C - World Wide Web Consortium

RDF – Resource Description Framework

API – Application Programming Interface

URI – Uniform Resource Identifier

HTTP – Hyper Text Transfer Protocol

1.0 Introduction

Solid is a specification which “adds to existing web standards to realize a space where individuals can maintain their autonomy, control their data and privacy, and choose applications and services to fulfil their needs” (Capadisli et al., 2021). Solid is created with the aid of open standards and linked data, which encourages the creation of adaptable and interoperable applications (Sambra et al., 2016). It enables users to fully own and control their data, which is a positive development, especially considering the numerous scandals involving data theft in the past. More than 87 million users' personally identifiable information was engaged in the Facebook and Cambridge Analytica incident, which was made public in 2018 (Isaak & Hanna). Solid offers a decentralized and secure data storage and access platform to stop these kinds of incidents from happening in the future, especially given the growing reliance on the internet and data.

Currently, large scale implementation of Solid is limited. SolidLabs Flanders is researching the possibilities of Flanders becoming the first region in the world to implement it among its citizens (SolidLab, 2022). The SolidLabs research echoes the Solid project’s key aims and motives, which are to give users the true ownership of their data. Other kinds of early implementations of Solid include applications such as Cimba, a microblogging social media application, or Zagel, which is a chatting and messaging application (Mansour et al., 2016). On paper Solid has potential to be a useful platform for citizens. Experimentation and research into possible use cases and the technology behind Solid needs to be explored to understand it more.

The focus of this research is to explore and understand how spatial data can be used in Solid. A sample dataset of five location points has been designed. The points are located around the Utrecht University campus. They have not been chosen based on any criteria as the data is only used as a dummy payload to interact with Solid. The points have been formatted based on five spatial data standards: GML, KML, GeoJSON, XML, and OGC moving features. They are described in more detail in the next chapter, together with a couple of other commonly used spatial data standards. This data has been used to understand if and how spatial data flows through Solid to determine whether it is a feasible platform for future applications.

1.1 Academic relevance

Research on new data provision and management platforms has been growing. It is not a secret that over the last decade, users of various applications have been slowly losing the true ownership over their data. All the current popular platforms owned by companies like Google, Facebook, Apple etc., do not provide enough transparency and control over one’s data. Laoutaris (2018) discusses the concerns of companies exploiting personal data for marketing and advertising purposes and calls for transparency in data handling. The European Union took steps towards more transparency when the General Data Protection Regulation (GDPR) was put in effect in May 2018. It involves all entities which collect, store, and process personal data of EU citizens. It highlights seven key principles: 1. Lawfulness, fairness, and transparency, 2. Purpose limitation, 3. Data minimization, 4. Accuracy, 5. Storage limitation, 6. Integrity and confidentiality, 7. Accountability (GDPR, 2018). Blacklaws (2018) discusses the limitations when it comes to what “personal data” is and concludes that not all data falls into the category. While policies such as GDPR force companies to make the collected data more transparent, it does not completely solve all the issues. The data is still located in closed private environments without the ability to easily transfer them elsewhere, or to manage who exactly has access to it. Therefore, research into platforms such as Solid could move data privacy a step further by letting citizens handle their data.

Solid offers a decentralized approach which is a means of storing and distributing data through multiple locations such as servers, systems etc. The objective is to spread the data around the network and store the data in multiple locations. The benefits include safe data storage with flexible access management, reliability, availability of information, and scalability (Kryukov & Demichev 2018). Instead of hosting and providing data from one point or a datacenter, the data is spread around multiple locations.

1.2 Research objectives

The aim of this research is to contribute to the growing knowledge base of Solid and decentralized data Pods. The focus is to understand how Solid handles spatial data to determine its feasibility for future implementation. The research is composed of two main parts. First, a literature review is conducted to define and understand the current state of Solid and spatial data. Open standards, protocols, and interoperability, which are the key points of Solid are defined. Second, a working test environment is set up to implement a collection of spatial data into a Solid Pod to examine its spatial data support firsthand and perform a set of data handling experiments. There are multiple spatial data formats used to examine whether there are any differences between file formats. This research consists of one main research question and four smaller sub questions presented below.

Main research question:

What open standards and protocols can be used to store and interact with spatial data in a Solid environment and a GI system?

Sub questions:

1. What technologies enable Solid to create an open and interoperable environment?
2. Is it possible to limit the sharing of spatial data to specific parties?
3. Is it useful to share this kind of data via a pod?
4. What are the challenges with regards to the usability of Solid?

The main research question tries to understand what technologies are behind Solid. What kinds of standards is it built with, what standards are supported in terms of data, which protocols are used to communicate in Solid? The main objective is to define the architecture of Solid, to help with assessing its feasibility with future spatial data applications. The supporting sub questions attempt to touch more on the feasibility and usefulness of the Solid environment. The first sub question is closely related to the main research question. It attempts to cover the technological and architectural side of Solid. The second question dives into data access management within Pods to analyze how it works. The third sub question attempts to put Solid in context with other existing web technologies and data storage options. The last sub question is concerned with the general challenges in terms of usability which may have been discovered during the research.

2.0 Literature review

This chapter provides a brief overview of spatial data in Geographic Information Systems (GIS). This brief introduction is followed by reviewing existing literature on Solid, its motivation, and architectural background. Some existing applications of spatial data use are presented. An overview of common spatial data open standards is provided to explore them and their uses in more detail.

2.1 Brief overview of spatial data and GIS

Spatial data is a key component of Geographic information systems (GIS). Unlike classical data, which contains information such as names, numbers etc., spatial data also contains information about objects, such as the shape, size, and location on Earth expressed with geographic coordinate systems (Ordnance Survey, 2022). Spatial data allows features to be represented in space. This expands the potential for research or government operations to work with data which has its place in the world, and can be represented by points, lines, or polygons. Furthermore, spatial data can also contain the dimension of time. Whether we are interested in knowing when a building has been built, or how certain features changed over time (such as forests, cities etc.) the time attribute can capture various stages of development as timestamps. Trajectories are a great way to expand on simple location coordinates by adding additional attributes. This data can either be a Z value for height monitoring, speed, time, and heading. These additional attributes enhance the data and create opportunities for further analysis. For example, trajectory data can be automatically classified to detect which mode of transport has been used (Biljecki et al. 2013) to conduct travel behavior research.

Spatial database systems involve linking attribute data which contains coordinates to a map, and information can be retrieved in the form of a spatial query (Usmani et al., 2020). Relational databases consist of several tables with no hierarchy, which can be linked together by identifying common fields, to process even more complex queries, statistics, and reports of the attribute data (Reddy, 2018).

Apart from relational databases, new technologies are emerging which focus on creating decentralized data storage systems for spatial data. Decentralization is not a new concept in the context of spatial data. Mansberger (2003) discussed data distribution through linked web servers two decades ago. He argued that this method of data distribution has an impact on actuality of the data (how often the datasets are updated). Additionally, single server data distribution also experiences unavailability periods due to maintenance schedules or unexpected issues.

Zichichi et al. (2022) created a decentralized data exchange concept called the Hypercube DHT (Distributed Hash Table). The focus of this system is on efficient querying of decentralized geodata. The system uses keywords which are bound to an identification number. Data can be queried by specifying a set of keywords to look up the data associated with these keywords. Another important feature of this design specification is the fact that it is interoperable with other systems as the Hypercube DHT is completely system independent.

Archiving spatial data, or any digital data in this regard, is also an extensive and difficult task to do and maintain. Decentralization of archives is an approach which provides a couple of key benefits. Locher & Termens (2012) explored the possibilities of alternative preservation methods for spatial data. They discussed several benefits to a decentralized approach, such as data redundancy and variety in software and platform solutions. For data redundancy, it is a safer solution to have data spread out across space in several independent locations. In the case of data loss, data can simply be replaced from another location. The software and platform variety, meaning that data is not stored on identical systems, formats, and management software, provides enough variety in case of technical obsolescence sometime in the future.

2.2 Solid

Solid is a project led by Sir Tim Berners-Lee. The project has been in development since 2016. It aims to create a decentralized method of owning, storing, and exchanging data (Solid, 2022). The idea is based around personal data stores called Pods. Pods are essentially personal storage servers, in which a user can store data. The user is in full control of their Pod in terms of data access and management. By using open and interoperable web standards, Pods can seamlessly exchange information. Solid is not a fully-fledged technology stack, but rather a set of principles which define the integrated use of standards. The Solid Protocol is a document describing the specifications that an application should adhere to, to provide a secure and interoperable exchange medium (Capadisli et al., 2021).

Tim Berners-Lee's vision behind Solid is really to bring balance into the web universe by returning data access control and privacy to the users and owners of data. In his blog post from 2018, when he decided to launch Solid, he talks about how his original vision of the web changed over the years (Berners-Lee, 2018). He openly dislikes the current iteration of the web, which is based on users giving up their personal data to large corporations, which has proven to not be in the best interest of web users. Furthermore, his vision of Solid is to empower through data. Interoperable applications, all being able to talk to each other and streamlining online processes are the innovations which Solid promises to deliver.

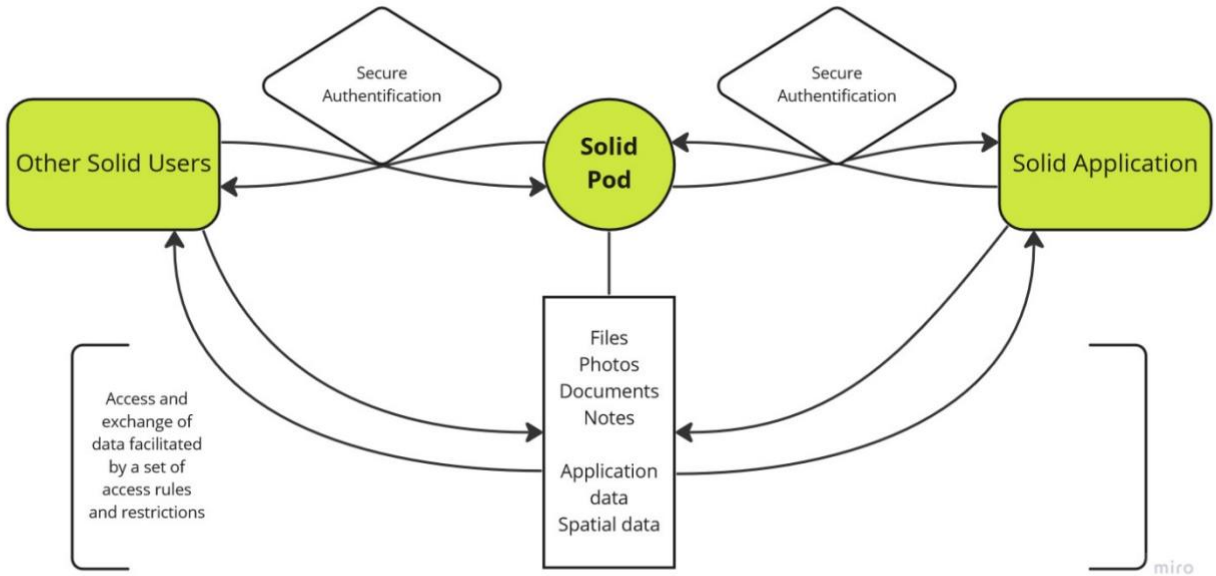


Figure 1. Simple diagram of a Pod and application architecture

Figure 1 shows a simple diagram of the Pod infrastructure. The Pod is a personal server tied to a user which can be used to store a variety of files. The Pod can then be used to connect to a Solid supported application or exchange data with other users. Pods which connect to applications do not actually transfer any data to be stored in the application. The key difference in data interaction here is that the data used in application, or generated while the user is interacting with an application, is stored in the user Pod. The interaction can differ based on the access rules set by the connecting Pods. For example, a person might only want to view certain data from another user. Perhaps they want to add new information to the Pod. The different access rules allow users to change who can view, read, and write data to specific directories in their Pods. Finally, each connection process, either between two Pods or a Pod and an application, is wrapped in a secure authentication method.

Currently, Solid only supports access control on entire files or folders. An interesting use case with regards to structured data is managing access rules on certain parts of a single dataset. This would allow one dataset to be shared with different entities, while only allowing them to see what they have permission for. TrinPod is a company based in the United States, which provides industrial Solid Pods through their platform (TrinPod, 2023). They provide access control at any level, a container, a file, and even individual data points. In a Solid discussion forum, Jeff Zucker, one of the developers of Solid discussed this feature of TrinPod and determined that in their case, each RDF triple is considered a separate resource (Jeff Zucker, 2023). This gives TrinPod the ability to manage access on individual data points.

2.3 Linked Data

Linked Data is a key component of Solid. Linked Data is a method of creating links and relationships between data on the web, which is useful for finding related information to anything a user is looking for (assuming all web data was linked, which is not the case). Linked data follows a few design principles. Resources are described in the form of a Uniform Resource Identifier (URI), and can be queried using either SPARQL (Berners-Lee, 2009), or in the case of working with spatial data GeoSPARQL (Herring & Perry, 2022). A collection of URIs which have defined relationships are specified in an RDF file.

Linked Data is an interesting concept, which in theory provides great opportunities for interoperable environments, applications, and data sharing. The idea of having all the data linked together promised a much simpler navigation around the web, as well as open and interoperable manipulation of the data with systems being able to recognize new concepts which could be resolved by the application (Bizer, 2009). Jain et al. (2010) argue that the current iteration of linked data is shallow and limited in a few ways. First, lack of conceptual description of datasets, which means that datasets cannot be described in relation to a certain topic or a concept. Second, missing links at the schema level, meaning that descriptions of data at the schema level does not support relationship definitions. Third, lack of expressivity, describing the fact that RDF triples are unable to utilize rich expressive features of schemas. Fourth, the SPARQL syntax, which requires precise input details to query a triple can cause issues with querying.

2.4 Research and spatial data applications in Solid

While Solid is still in early stages of application development, several research projects have tried to explore some use cases. Wynckel & Signer (2022) developed a Solid based application monitoring indoor position of users to show how personal location, orientation, and velocity can be stored in a Pod and a simple viewer in which a user can view the tracking data. The end results highlight the advantages of users owning their own storage servers, and the interoperability of the system through open standards allows multiple positioning systems to be simply linked together and exchange data as required.

Figure 2 shows a diagram of how the position is captured and stored in a Pod, as well as how it can be viewed in another application. This concept does not only store the position in simple X and Y coordinates, but it is also able to monitor and save the orientation and velocity. This data is directly uploaded to a Pod in the. ttl format, which is an RDF based representation of data. By using linked data principles, the position data collected can easily be queried and even associated with other data. The process of linking data with other data is very interesting as it can provide new and more in-depth insights. For example, in a corporate office environment, positional data throughout the day could be linked with air quality (humidity, temperature) or light data, to determine whether there is a correlation between these factors.

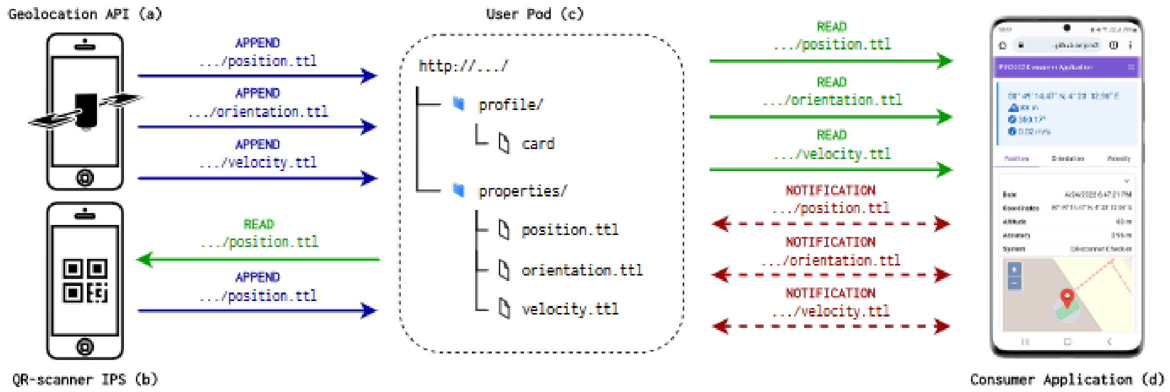


Figure 2. Wynckel & Signer (2022). Position monitoring system connected to a Pod.

Becker et al. (2021) proposed a system for monetizing resources in Solid Pods utilizing the Ethereum blockchain. They have created a demonstration for this concept which involves several steps. First a user uploads a resource to a Pod and defines its cost and time limit for access. Other users can view and sort these resource offers and can purchase them for the set cost for the set time limit of access (a license with expiration). The buyer then sends a payment from their blockchain wallet to the owner of the resource, which is then verified. The buyer now has access to the resource and after their licensing period expires, the access to the resource is removed. *Figure 3* shows the proposed system architecture for this idea.

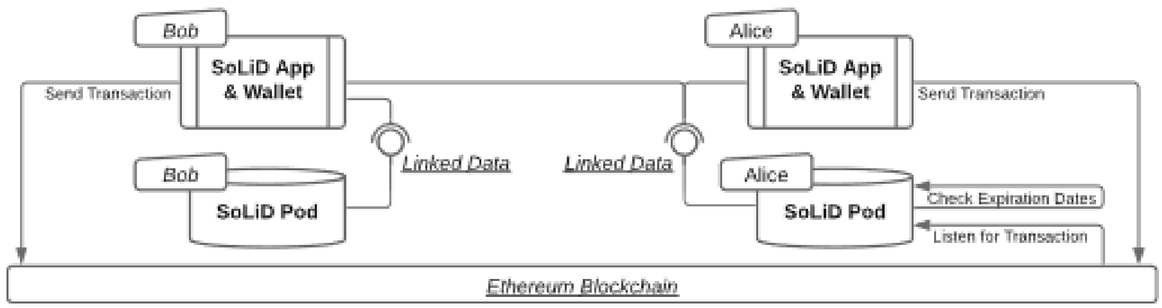


Figure 3. Becker et al. (2021) Design architecture of blockchain monetization

The idea of monetizing resources from Pods is very interesting. It certainly has implications for future development and use of Solid based solutions. Since there is currently no standardized mechanism of monetization, custom solutions would have to be developed to support payment and access. Using the blockchain ensures that the buyer and seller retain a high level of anonymity. There is no personal or banking information shared between the users, and no third-party payment gate system needs to be used. This ensures anonymity when purchasing resources from a digital marketplace which could become a thing in the future.

The Dutch Kadaster (Land Registry) is actively researching new emerging technologies through their various research teams at Kadaster Labs (Kadaster Labs, 2022). Solid Quest is a project developed by Kadaster to research the possibilities of implementing Solid as personal data vaults for citizens (van Andel et al., 2022). One example they have built is a demonstration of a purchase agreement between two users. The process is quite simple. First the seller fills out the agreement with all their details and sends it to the buyer. The buyer fills out all the details. Once the filled-out information is checked and ready to go, both parties sign and each user receives a copy in their personal Pod. Several non-Solid initiatives are in development. For example, GAIA-X, a project focused on developing the next generation of data infrastructure in Europe, with the aim of creating an environment where data and services can be used and shared in a trusted and secure manner (Braud et al., 2021). The vision of the initiative is very similar to what Solid is trying to accomplish. Digital sovereignty, or control of own digital data is a major concept and discussion point within the Gaia-X framework (Autolitano & Pawlowska, 2021).

iShare is a Dutch project which focuses on the challenges of the Dutch logistics sector by creating a framework allowing companies to control their data. The framework attempts to improve the data exchange processes within the logistics sector by creating a scheme with a set of agreements to create a uniform and controlled way of sharing data within the sector (iShare, 2019).

An application developed by Jesús-Azabal et al. (2021) successfully integrated Solid Pods into smartphones. The idea is that smartphones are used as a host for the Pod, so that user information is always available on the phone, with applications requesting access to it. This solution gives the user a precise idea of which application tracks what information and can change access rules for the different applications. This concept could serve as a great platform for future large-scale research involving spatial data. For example, research requiring a lot of participants, such as monitoring movement of people in cities, could become much more streamlined. The data collected, such as location, trajectories, floors travelled etc. can be analyzed to provide insights on public space with a theoretically larger sample of test subjects.

2.5 Responsible use of spatial data

Spatial data has proved to be a powerful tool in all kinds of industries. The ability to track datapoints across space created new possibilities for research. However, as the saying goes “with great power comes great responsibility”, the use of spatial data needs to ensure that collection of personal spatial information is not used in a manner which can compromise the identities of users. More and more data are collected and used in all kinds of industries. The W3C and OGC are actively promoting the ethics for responsible use, as well as legal frameworks which encompass the use of data (Abhayaratna et al., 2021). The previous examples of data mishandling certainly show the need for such advocacy. The recent COVID-19 pandemic also showed that responsible use of data is questionable at best, such as in the case of very little transparency of spread monitoring in China (Ienca & Vayena, 2020). The topic of responsible use is not limited to any industry or research.

Anytime data collection involves individuals such as participants in a study, or users of certain platforms, proper data management is key. Data protection laws exist in many parts of the world; however, they are not always present, or they are not extensive enough. For this reason, data use ethics is (or at least should be) required to prevent any misconduct. The development of these common principles has been proposed in the research world (Sandbrook et al., 2021), where even if a research area is not protected by law, a set of agreed principles would be followed.

2.6 Spatial data standards

Spatial data, databases, and GI systems would not be interoperable or even properly usable, if not for open data standards. “An open GIS system allows for the sharing of geographic data, integration among different GIS technologies, and integration with other non-GIS applications” (ESRI, 2003). To ensure interoperability between different systems, there is a need for those systems to effectively speak the same “language”. Standards are developed for the purpose of maintaining data in formats which can be implemented in various systems, therefore, allowing systems to exchange information.

There are different organizations that deal with creating standards. The International Organization for Standardization (ISO) is an independent entity which develops standards for all kinds of technologies and manufacturing (ISO, 2022). ISO is also involved in creating standards for GIS applications such as metadata (ISO 19115-1:2014). A more GIS focused organization for creating open standards is the Open Geospatial Consortium (OGC). This consortium is responsible for creating geospatial information and services FAIR (Findable, Accessible, Interoperable, Reusable) (OGC, 2022).

2.7 Open spatial data standards

This section provides an overview of the standards which were considered for experimentation in the Solid environment. There are a lot of standards available, so a selection of the key standards has been chosen based on literature (Bermudez, 2017; Kralidis, 2008). The final list of standards used in the research is available in the methodology section. There are several criteria based on which the selection has been made. How widely a standard is use is the first one. Since we are discussing new approaches to data management, it needs to be able to support the popular standards which are commonly used in spatial applications. Second, the use and compatibility with web applications. Not all spatial data standards are commonly used on the web, so a selection of web used standards is important. Third, the standards need to fit the kind of data that is being tested. In this case, location data is being used, so standards which are commonly used and provide support for these points were selected. Each standard involves the use of spatial data; however, they focus on different aspects and technologies for working with this data.

HTTP(S)

The Hyper Text Transfer Protocol, or HTTP, is a lightweight protocol first published in 1996. Its purpose is to be a generic protocol for communication between users, other internet protocols, and allows for information exchange on a request-response basis (Berners-Lee et al. 1996). The communication between a client and a server is established by a client sending a request to a server, with a header containing further information for the server. Once the server verifies the request it responds with a status code and the requested content (Yannakopoulos, 2003). Some common response status codes include 200 for success, 307 and 308 for temporary and permanent redirects, 400 for a bad request to the server, 404 not found, and many others. These codes allow the users and developers to see whether a request has been successful, and if not, it suggests a point at which it failed.

Secure Hyper Text Transfer Protocol, or HTTPS, is a version of HTTP which utilizes Secure Sockets Layer (SSL) for encrypting and authenticating communications. The added security layer does not impact the performance of web communication in any meaningful way (Goldberg at al. 1998), while also being more secure than HTTP (Amann et al. 2017). The web communication of Solid, so information exchange between Pods or Pods-web application also utilizes HTTP(S) as a standard protocol for requesting and accessing online resources.

JavaScript Object Notation (JSON)

JSON is a lightweight data interchange format, which is popular due to its easy-to-read syntax and is also easy to read and parse for machines (json.org, n.d). As the name suggests, its origins can be traced to the JavaScript programming language. The two primary data structures are arrays and objects, which makes the JSON format language-independent and usable in almost all current programming languages (Zunke & D’Souza, 2014).

JSON-LD is JSON based linked data format which is supported by Solid. It basically achieves what RDF and linked data do, however, it does it in a similar fashion to the JSON syntax which is a more widely known way of representing data. Since it is so closely related to JSON, existing JSON parsers and libraries are largely supported (Sporny et al. 2020).

GeoSPARQL

Developed by the OGC, GeoSPARQL is an implementation standard for querying geospatial data on the Semantic web (Herring & Perry, 2022). This standard has been developed to support representing and querying spatial data. It includes a vocabulary for the Resource Description Framework (RDF) and defines how spatial data should be described on the Semantic web. As the name suggests, it is an extension of SPARQL, a W3C query language used in RDF. RDF is a standardized language used for creating linked data models of data, where data resources are defined using URIs (Uniform Resource Identifiers) which are referenced using common vocabularies (Pan, 2009). The documentation from OGC includes information about the structure of RDF, queries, vocabularies, and examples of how to apply this standard in practice (GeoSPARQL Specification, 2022).

GeoSPARQL makes integrating geographical features into RDF much easier as it provides additional classes and methods of defining different features with different attributes. The new distinctions between features can for example be used to find entities within a certain distance of other entities or another specified location (Battle & Kolas, 2011). It also supports a wide range of spatial data types such as points, polygons, arcs, curves, and multicurve. Topological relationships such as overlaps between features are also supported and can be defined. The applications of GeoSPARQL can therefore include a large variety of thematic areas, where querying geospatial datasets is a requirement.

OGC Moving Features

This standard developed by the OGC is responsible for addressing the issues regarding moving objects e.g., location data from mobile phones. The goal of this standard is to provide a standard encoding method for trajectories and moving features. The standard has been submitted to the OGC by Asahara et al. (2015), to create a standard which can handle the integration of location data, trajectories, encoding, and an API to handle large amounts of real time data from moving objects. The root element, mf:MovingFeatures, consists of several elements such as mf:sTBoundedBy, which defines the spatio-temporal bound of trajectories. The mf:Foliation element contains moving trajectories data. Atrributes can be defined with mf:AttrDef, where a user can define custom properties and parameters of an attribute.

It supports three methods of encoding, CSV, XML, and JSON, and binary encoding for simple points. Graser et al. (2020) discuss the differences in encoding. They note that while it is the same standard, the methods do vary in terms of complexity of implementation. Furthermore, the features and attributes that can be encoded with each method differ a bit as well. The more obvious binary encoding only supports basic points. The differences between XML and JSON are a bit more complex. In XML, segments contain a start and end time, and the attributes have static properties. JSON includes points with timestamps and attributes can have independent timestamps as well.

Ryoo et al. (2016) experimented with indoor moving objects. Each indoor object is assigned properties such as an identifier, classification, starting time etc. Combined with another OGC standard Indoor GML, this combination allows for creating a dataset of points and trajectories, in this case, capturing the movement of people in a building.

The GPS Exchange Format (GPX)

Released in 2002, GPX is an XML file format which supports the exchange of GPS data such as waypoints, routes, and tracks between applications and web services (GPX, 2004). It is a commonly used format for working with data gathered from mobile GPS devices. As GPS devices became cheaper and more accessible in the early 2000s, GPX was adopted rather quickly and is used by well-known organizations such as the UK Ordnance Survey or Open Street Map (Haklay & Weber, 2008). The GPX format stores GPS data as points, which can be opened in many GIS applications. It is also a very flexible format, which can be easily converted to other formats with a simple data converter tool available online (Foster, 2022). Furthermore, the main selling points of GPX include the ease of data exchange between applications, simple conversions to other formats, and the XML based architecture allows for easy integration in modern applications. Its relative simplicity and ease of integration make GPX a great way of working with GPS data.

Ferreira & Silva (2013) use the GPX format to integrate GPS data from a walk into Google Earth to visualize the tracks for researching body states using biometric sensors based on the location of participants. Coopmans & Chen (2008) developed a low-cost data logger with Co2 and NH3 sensors and attached it to a moving vehicle to gather air quality readings from a journey. This project also utilized GPX which offers interoperability with a wide range of sensors and using a standard format it is easy to integrate data from a whole fleet of vehicles into a database.

GeoJSON

GeoJSON is a format based on the JavaScript Object Notation (JSON) developed for encoding geometry types such as point, line string, polygon, multipoint, multi-line string, and multipolygon (Butler et al., 2016). It includes a description of geographic features, properties, and spatial extents. The geometry is formatted in the World Geodetic System 1984 (WGS84) and uses decimal degrees as units. GeoJSON is commonly used for creating interactive web maps. For example, Horbiński & Lorek (2020) converted shapefile data into a GeoJSON to create a web map. The advantage of this method is that the location and coordinates of features we want to present on map are encoded and easily usable in different applications without the need for reformatting. However, they do note some disadvantages to the GeoJSON approach, such as its lack of topological information and file size of a GeoJSON which could cause applications with lots of data to load to not perform up to the required specification.

Chet et al. (2017) developed a web based platform with the aim of analyzing the potential energy efficiency savings of retrofitting existing buildings in a city with better lightbulbs, upgrading HVACs etc. The platform contained a simplified three-dimensional city model utilizing CityGML, GeoJSON and file geodatabases. GeoJSON has been used for the building data, defining building names, properties, and other characteristics.

Keyhole Markup Language (KML)

Originally developed for use with Google Earth, KML became an approved OGC standard in 2008. It is an XML-based programming language for visualizations and annotation of maps and images. The OGC implementation specification also mentions other functions and uses of KML such as specifying icons and labels to identify locations, definition of styles and appearance of KML features, organization of features into hierarchies, writing descriptions of features and many more (OGC KML specification, 2015). An advantage of KML is the flexibility it gives to users in terms of thematic mapping or visualizing statistical data on a map in a browser or an application such as Google Earth (Sandvik, 2008).

KML is an important tool for visualizing data on virtual globes such as Google Earth. Ballagh et al. (2011) discuss the importance of developers and researchers having a way of presenting their work to a much larger audience. The KML formatted data makes science and complex data visualizations very accessible for everyday users to browse through, look at different data layers or time series. Foerster et al. (2009) showcased a method of integrating KML through an OGC web service standard for development of mass market applications for risk management purposes. It involves developing KML files which citizens can import into their application of choice (such as Google Earth) and visualize the latest updates and analysis of risk.

Geography Markup Language (GML)

GML is an OGC standard for encoding geographical features. It is XML based schema, which ensures that this standard provides a framework for describing geographical features which support linked geographic applications, support the exchange of this information, and in general, make it easier for interoperable data and applications to be deployed (Portele, 2007). Similarly to KML, GML can be used to create web-based mapping solutions, for example through Google Maps API (Chow, 2008).

GML is a very flexible format, in terms of its use cases. Alesheikh et al. (2005) showcased an application based on GML, which is used for encoding data from air quality sensors. Their arguments for using GML are that it makes the application "interoperable, platform independent, and vendor neutral". Zhang et al. (2003) proposed an idea for creating GML based databases. The benefits of this approach are interoperability, real time data exchange on the web, scalability of the databases, and reduced data conversion processes.

Features and Geometries JSON (JSON-FG)

Currently in development, this OGC standard is set to build upon the GeoJSON standard. The intended result is to extend the functionality of GeoJSON to support additional concepts for use within spatial data operations and OGC API standards (Portele et al., 2021). In an engineering report, the OGC further positions the standard having a GeoJSON structure with the new additional information being represented on the top of the GeoJSON objects. Furthermore, the report discusses the addition of features such as identifying the feature type, encoding the temporal extent and spatial geometry, and encoding reference systems (Portele, 2021).

2.5 Location tracking

Today, we have the luxury of tracking and finding almost anything on a digital map in a matter of seconds. Whether it is trying to find the way home, looking up the location of a monument, or tracking our fitness progress through different applications, navigating the world has never been easier. The global positioning system (GPS) plays a major role in these developments. Developed in the 60s and 70s by the United States Navy and Air Force programs, this space-based system allows devices to connect directly to a constellation of satellites orbiting Earth to determine precise location (McNeff, 2002). The network consists of 24 satellites which are spaced out so that at least 5 are always in view from any point on the planet (Bajaj et al., 2002).

Since the availability of GPS tracking, countless research fields have utilized this technology. For example, it used to track domestic and wildlife animals (Ramesh et al., 2021), tracking people who suffer dementia (McShane & Skelt, 2009), or a more recent example of developing applications to track citizens during a pandemic to determine the possibility of them having contact with an infected person (Stanley & Granick, 2020). A recent study concerned with bike theft in Amsterdam explored the possibilities of tracking the location of stolen bikes to figure out where they move and study the patterns (Venverloo et al. 2023). The study involved distributing 100 locked bikes with GPS trackers throughout Amsterdam. Of course, with the increase in development of real time tracking applications for citizens, privacy and ethical issues arise which need to be addressed. Michael et al. (2006) discusses the ethical concerns regarding location tracking. They discuss the issues of possible "controlled supervision", accountability and responsibility for the accuracy and errors of the data and highlight the possibility of changing long term effects on health, culture, society, and politics. There are certain methods of anonymizing data, such as inserting fake data, removing identifiable information, or restricting queries which would yield such a small number of results that individual users could be identifiable (Manoharan, 2009).

3.0 Methodology

To assess the usability of Solid in the real world, a simple web application has been developed. The aim of the application was to test how spatial data encoded according to five OGC standards which were selected based on the literature review: GML, KML, GeoJSON, XML, and OGC moving features. The data consist of 5 location points across the Utrecht University campus. The points themselves do not represent any features, however, they do follow a path around the campus which could be thought of as location tracks for e.g., public transport stops. For this study, the points have been manually created without performing any actual location tracking. The sample data and all of the code used within this thesis can be found on GitHub: https://github.com/jzvolensky/solid-pods.

3.1 Use case: Towards interoperable public transport

Currently in many European countries, public transport is not provided by a single company or state organization. This means that a user is required to purchase tickets from multiple sources, often having to download separate applications or use separate websites. Adapting public transport to utilize the Solid ecosystem solves this issue by providing a single Solid-based application which can host multiple providers. Of course, similar reseller websites currently exist such as Trainline, which allows users to buy tickets for trains in multiple countries. However, this still means trusting a third-party with your private information such as personal data, banking information, location etc. Using Pods can solve these privacy concerns. Furthermore, with a single application in your phone you can access public transport anywhere in Europe. This is possible due to the interoperable and open standard nature of Solid. Each European nation can participate and join this open ecosystem.

The aim of this research paper is not to develop a full scale demo application for public transport. Instead, the aim is to look at how spatial data can be used and transferred in and out of a Solid Pod. Before full scale applications can be built, the feasibility of this endeavor needs to be assessed. On paper, it does sound like a solid solution for unifying public transport in a user-safe manner. Some benefits include:

- Integrating public transport of participating countries and providing a single application for all kinds of public transportation anywhere in the EU.
- Providing an easy method of obtaining travel tickets in foreign countries.
- The ability to plan international journeys and directly purchase the necessary tickets.
- The ability to store and view your past journeys on a web map.
- Sharing routes with other Pod users.

3.2 Data

The data consists of five points which have been selected around the Utrecht University campus. The points roughly follow the main road and tram track which go through the campus to simulate a possible route a student might take. These five points, in the form of an X and Y coordinate have been encoded according to the following standards: GeoJSON, GML, GPX, KML, and Moving features. These standards were chosen based on the criteria presented in the literature review. Multiple formats were chosen for a couple of reasons. First, it allows us to test whether Solid has some limitations in terms of supported data formats. Second, it allows us to experiment with different access rules on each of the files to see how it affects the accessibility in a Pod. Third, to see and compare whether there are any unexpected differences in the usability of the data itself once it is in the Pod. The unformatted location data can be seen below:

Name	Coordinates
Point 1	5.163533, 52.084060
Point 2	5.165105, 52.084096
Point 3	5.169003, 52.084723
Point 4	5.173001, 52.084723
Point 5	5.176373, 52.084742

3.3 Spatial data in a Pod

As mentioned before, the focus of this research is to examine how spatial data can interact with Solid Pods. To accomplish this task, a simple application has been developed. The basis of the application is the Getting Started tutorial developed by Inrupt (Inrupt Getting Started, 2023). The guide includes most of the necessary features required for this experimentation, such as logging into a Pod, getting the URL of the Pod, and writing text into a logged in Pod. The final look of the application can be seen below:

Step 1: Logging in
For ease of testing only Pods created and hosted on the Solid Community server are supported. The user is redirected to the Solid Community login page. After successfully logging in the user is returned to the Getting Started application.
Step 2: Retrieving the Pod URL
The second step retrieves the Pod URL from the logged in WebID of a user. The WebID is a front page of a Pod. It is a public link like a link to a Twitter profile, or similar. For example, the WebID of the Pod used in this testing is https://jzvolensky.solidcommunity.net/profile/card#me. From the WebID the Pod URL is retrieved in the form of https://jzvolensky.solidcommunity.net. For simplicity the URL where the files are stored in the next step is hardcoded to be /Spatial%20data%20testing/LocationFiles. This makes the process much faster and for the purpose of this research it does not matter where the files are stored.

Step 3: Send files to the Pod
The first two steps were largely made "by the book" following the tutorial with slight modifications to the URLs and Pod providers. The original code example uploads text by having a user fill in a text box. This is not the ideal way to upload spatial data. Appendix A2 shows a red box where the file uploading logic has been added to the original guide. The data upload is performed by the owner of the Pod for guaranteed upload access.

When the user clicks on the Choose file button, a pop up menu will let them browse their local machine and select a file. Once they select a file and press the Upload file button, an HTTP PUT request is sent to the hardcoded URL of the Pod. If the request has been successful, the user sees a small confirmation "File {File name} uploaded successfully" where the File name is derived from the selected file. Appendix B1 shows the location data stored in the Pod.

3.4 Accessing spatial data in a Pod

Uploading data into a Pod is only half of the equation. The data inside a Pod must also be usable in applications. The usability of data in a Pod was tested with the following experiments to test out the accessibility of data in a Pod as well as testing the file sharing rules of Solid. The file sharing capability, or rather the control over access rules is a key principle of Solid. To test access to data under different rules, some of the uploaded files have been given different permission levels than others. After that, requests to access the data have been sent using curl to monitor the different responses.

A second method of testing data access is through a simple python script. The goal of this script is to load the point data directly from a Pod and plot them on a map. The script uses the Geopandas library for data handling and the Folium library to generate a simple web map from the points. The code and the map can be seen below:

```python
data = gpd.read_file('https://jzvolensky.solidcommunity.net/public/LocationFiles/GeoJSON_locations.geojson')
print(data.describe())

m = folium.Map(location=[data.geometry.y.iloc[0], data.geometry.x.iloc[0]], zoom_start=10)
for idx, row in data.iterrows():
    folium.Marker([row['geometry'].y, row['geometry'].x], popup=row['name']).add_to(m)

m.save('map.html')
```

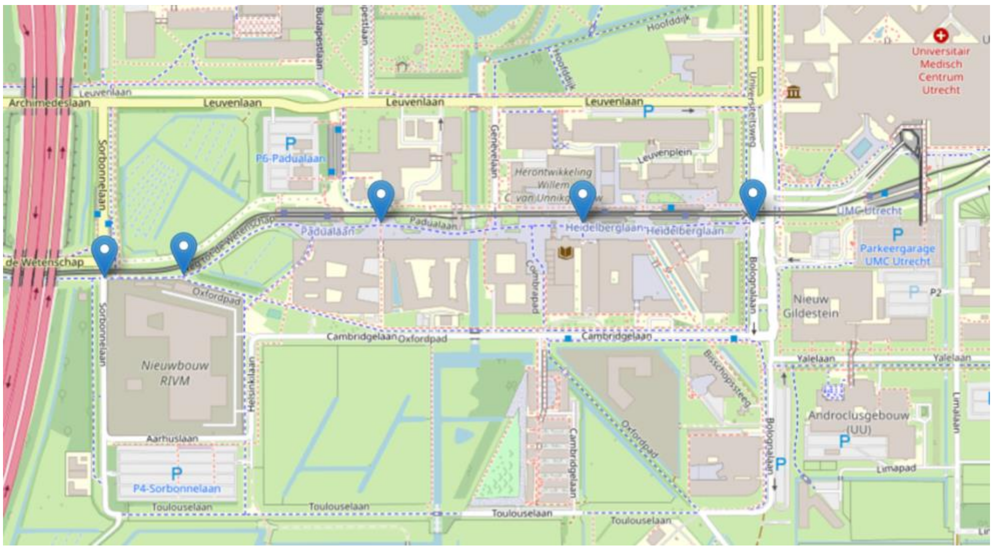


Figure 4. Accessing the dataset directly using Python and plotting it with Folium.

A complete list of the main experiments performed in Solid can be seen below:

Experiment	Description
Uploading data to a Pod	The first step is to prepare the data and send it to a Pod for further testing. This is done as the owner of the Pod.
Accessing the data	Testing basic access to the data in a publicly available folder without access restrictions using curl. Access is tested both as an owner with access, as well as a third party to examine the restrictions. Data access is also tested with an attempt to query some the data from the Pod as the owner of the Pod.
Changing the location of the data	Moving the data to a non-public folder in a Pod without restricting the access rules.
Changing the access rules	Changing the access rules to certain files to authenticated users only and testing access.

4.0 results

This section presents the outcomes of our experiments performed on Solid Pods in relation to the research questions and sub-questions posed in Chapter 1. The experiments are categorized into three phases, representing the data's journey in Solid: Ingestion, interaction with the Pod, and export/interaction with other applications.

4.1 Data in a Pod

Accessing the data

To examine the accessibility of the data in a Pod, we utilized the curl command. Curl allows us to send a request to the URL where the data is located and print the response into the terminal. Screenshots provided in the text demonstrate this process. For instance, Appendix B2 and 3 show the results when requesting data from the Pod, such as GML_locations and GeoJSON_locations. From these examples, it's evident that the data is accessible. Notably, this data doesn't have any access rules assigned to it.

Until now only the whole files have been requested. What if we wanted to perform a spatial query on the data before retrieving some results? A sample python script has been developed to experiment with querying the points based on the distance between them in the GeoJSON file located in the Pod. There are two parameters which are used as the input. One is the reference point, or the point from which we want to search, and the maximum distance in meters from that point. The script is available in the project’s GitHub repository as ‘dist-extract.py’ (https://github.com/jzvolensky/solid-pods/blob/main/src/dist-extract.py). It is important to note that this example works by fetching the whole dataset and calculating the distances and filtering on the client side. For example, if the reference point is set to Point 3, and the maximum distance is set to 300 meters, the resulting output shows:

Points within 300 meters from Point 3:
Name: Point 2, Coordinates: [5.165105, 52.084096], Distance: 275.32 meters
Name: Point 4, Coordinates: [5.173001, 52.084723], Distance: 273.18 meters

When the distance is set to 600 meters, all the points are available:

Points within 600 meters from Point 3:
Name: Point 1, Coordinates: [5.163533, 52.08406], Distance: 380.96 meters
Name: Point 2, Coordinates: [5.165105, 52.084096], Distance: 275.32 meters
Name: Point 4, Coordinates: [5.173001, 52.084723], Distance: 273.18 meters
Name: Point 5, Coordinates: [5.176373, 52.084742], Distance: 503.59 meters

Server-side filtering was attempted by specifying the query parameters in the request URL. This is not supported by Solid at this time for file based data. Another issue is that unless the distance calculations were implemented and performed in the Solid Server, it is not possible to specify points within 100 meters as the server has no idea what that means. This is not ideal for large datasets. However, with a web application as a middleman, the application could be used to execute the script, query the data from a Pod, and return this data to the user. This is something that should be tested in the future as it is crucial for large datasets to be queried directly without fetching the whole dataset. Another option is to utilize Linked Data and perform SPARQL/ GeoSPARQL queries. This has not been tested, but it would be interesting to dive into in the future.

Changing the location of the data

When the data is placed in a non-public folder, such as our spatial data folder, a Solid user would not be able to access this folder unless a link is shared with them. If the user is provided with a link and the resources are not using restrictive access control, the data is available.

Changing the access rules

Even when the data is in the public folder but is locked behind an access rule, it remains inaccessible without proper authentication. The file itself is visible, but not the contents. In such cases, users are prompted to log in to view the data. We even created a second Pod for testing purposes, and Appendix B5 shows that after logging in, data can be downloaded from the Pod containing the data. To experiment with access rules, we used the KML version of the points and set the access rule to "Authenticated Agent" to view the data. Appendix B4 illustrates that without proper authentication, instead of the data, we receive placeholder HTML code. Attempting to access the data through the Solid web GUI also leads to an error message, as seen in the accompanying Appendix B4.

In summary, Solid Pods can serve as versatile API endpoints for a variety of data types if it is used in file based mode. While not explicitly marketed as a GIS solution, it is possible to upload and access spatial data. Performance of the queries is out of scope for this thesis, and it is certainly a topic which will need to be addressed in the future. For large amount of data, spatial indexing might be necessary to make the queries more efficient.

While the results primarily focused on data interaction and accessibility, the experiments conducted indirectly pointed to usability challenges related to authentication, access rules, and error messages. Further exploration is required to comprehensively address usability challenges.

The outcomes of our experiments align with the research questions and sub-questions, shedding light on the technologies enabling an open and interoperable environment in Solid, the potential for transforming services to use Pods, the feasibility of restricting data sharing, and the utility of Solid Pods for data sharing. The experiments help with answering the usability questions and concerns of Solid.

5.0 Discussion

This section presents comprehensive analysis and interpretation of the experimentation results. The aim of this analysis and interpretation is to extract insights and link them with the research questions. Addressing the implications of using Solid and understanding the broader context of the Solid environment. Moreover, this discussion serves as a platform to reflect on the alignment between our research objectives and the tangible outcomes, while also acknowledging the limitations and offering avenues for future exploration. The structure of this chapter follows the structure of the research questions and objectives.

5.1 Exploring open standards and protocols for spatial data in Solid

The main research question "What open standards and protocols can be used to store and interact with spatial data in a Solid environment and a Geographic Information (GI) system?" is the core question which is concerned with possibilities of Solid within a GIS environment, and its usability with spatial data. There are multiple viewpoints to this question which are all related: technology, data management, and interoperability.

Sub question 1: Technologies Enabling an Open and Interoperable Environment

There are several ways in which Solid promotes interoperability by using open standards. First, the Pod itself. A Pod is a universal storage medium which can accept any kind of data format when it comes to files, documents, images, and other types of media. This applies to spatial data too, so whether a user wants to put e.g., shapefiles into a Pod it is entirely possible. Once the data is in the Pod, it can be made available to other applications. Since the data can be in a wide variety of formats, it means that a user should be able to integrate their data with any application.

Linked Data is another essential Solid feature. By design, Pods completely support structured data. Although Linked Data isn't explored in detail in this study, its potential should be recognized. There are a few Linked Data-specific criteria, but the procedure for uploading structured data is the same as for files. For the information to be represented in the so-called triples, the data must be appropriately formatted in accordance with RDF standards. Although Linked Data has the potential to be transformative, there are obstacles to its practical application, most notably the need for organized data. These complications may require a progressive approach to taking advantage of its benefits.

Sub question 2: Limiting Data Sharing with Specific Parties

Data access and sharing rules is another feature of Solid. The access rules have been thoroughly tested and have been proven to work as expected. Each folder that a user creates, or a file that is uploaded to a Pod, is automatically eligible for access management.

Access control list, or ACL, is the formal title for this. It is a list of rights associated with a specific resource. Each user-created folder and each file added to a pod work flawlessly with access management mechanisms. Each resource has a unique ACL, and any changes made to the access rules or the addition of certain users who are allowed to edit the resource cause the ACL to be adjusted accordingly. In Solid, it is currently only possible to limit the ACL on a file and folder level. Limiting access to datapoints within a file is not officially supported. Some implementations, such as the TrinPod server, go around this by treating the RDF triples as individual resources.

Sub question 3: Usefulness of Sharing Spatial Data via a Pod

The previous sub questions answered a lot of the technical unknowns of Solid. Understanding the technology is the first step in determination of its actual usability. Solid has shown that it can function as a full-fledged ecosystem.

There are several main advantages of Solid, which are also backed by the experimentation. The first and major point in favor of Solid is without the doubt, interoperability. The potential benefits of properly utilizing this feature are immense. Out of the box support for spatial data standards and commonly used web protocols allow Solid to be a strong candidate for projects involving spatial data and people. The support for authentication, identity management, ACL, but also other neat features such as messaging, and discussions create a well-rounded base platform for web applications.

Even by just following the basic introductory tutorial and adapting it slight for files is enough to put together a basic application which can be used to upload and download spatial data from a Pod. In the future Pods are certainly a good consideration when developing applications. Of course, as Solid is meant to be used primarily as a “personal” data store, these applications should be focused on working with data and spatial data on a citizen level.

The usefulness of sharing spatial data through Pods is dependent on the use case. In general, projects which involve handling data of individual users seem to be tailored to the ecosystem well. It is simple to obtain a Pod. However, it is a big question whether Pods can be used in a large commercial enterprise as a database and data access medium. Projects such as TrinPod, do suggest that there is space for this kind of Pod usage. The TrinPod ecosystem and products revolve around collaboration, management of data, and project management.

Sub question 4: Usability Challenges of Solid

The usability of Solid comes down to the implementation. Authentication and user management is straightforward, at least in small testing applications. There does not seem to be any reason why this would change with scaling. The main usability challenge is the use of Linked Data. While Linked Data has many benefits in terms of interoperability and discoverability of data, it is also an extra step of data preparation before it is ready to be uploaded and used. It is also not in the best interest of commercial entities to hop onto Linked Data when they can store and use their data in obscure formats and charge fees for their platform usage.

However, when we put Linked Data on the sideline, Solid is just another data hosting and sharing platform. It does not provide any spectacular features you would not find in comparable commercial alternatives. So, the main advantage (at least in theory) of Solid is also one of, if not its biggest shortcoming.

5.2 Putting the puzzle together

The previous section discussed each sub question in more detail. This section is focused on putting all the pieces together to answer the main research question. The sub questions looked in more detail at the different parts and technologies, which enable Solid to be a personal data store. The focus of this study was to examine how these technologies might enable the use of spatial data specifically. The sub questions address this from several points of view, highlighting the sharing access, data management, and interoperability.

The web technologies used in the development of Solid provide a flexible base framework for interoperable applications. It is built entirely on top of existing well know and open standards. This ensures that Pods are extendable in terms of what data we can store in them. It has been shown that it is no more difficult to work with spatial data in a Pod than a classic text file. It is also extendable with Linked Data. Instead of using files it is also possible to use a linked data stack to store the data and use a sparql endpoint for querying and add a triple based access control mechanism. However, this is not an easy task and the amount of work to start using Linked-Data is bigger than the work to start using Solid.

With the previous thoughts in mind, is it worth developing a Solid based application to share spatial data? It seems like it largely depends on the purpose. If the plan is to use it as a file based storage, the only deciding factor is the cost of running a Solid server compared to storing data on similar cloud solutions. These providers also have authentication, user management, and even access management. The interoperability standpoint of Solid is its main selling point, utilizing open standards, especially for spatial data. However, this functionality is not at the level yet where it makes it a clear winner.

Conclusion

In conclusion, the results presented show some promise with Solid Pods. The adaptability to spatial data is a great feature. It is very simple to start integrating spatial data in a Pod. The Access Control Lists (ACL) are almost a great feature. Solid can be used as a file-based pod with ACL but to fully use the potential of Solid a full Linked data stack with a sparql endpoint and an access control mechanism based on triples is needed.

The usefulness of Pods is not very clear and really seems to depend on the needs of individual use cases, rather than being a solution to some problem. Without Linked Data, Pods act as another storage medium with support for open standards, which makes it difficult to justify its use.

It has also been shown that while accessing file based data is a seamless experience, the querying of data remains a big shortcoming. For large datasets, it is simply unacceptable for the client to load the whole dataset and then perform queries. This can be perhaps solved by using a middleman web application to handle complex queries and data transfer requests.

Solid, while being one of the first interoperable web ecosystems, is not the only one. There are initiatives such as Gaia-X in the European Union. While Gaia-X has a different goal in mind (federated and trustworthy data infrastructure) they are built on the same principles of interoperability, data sharing, and cooperation.

Limitations

This research paper tried to shed some light into how spatial data can interact with Solid Pods. The experiments showed some potential for future use. However, this paper was not able to examine every detail with regards to spatial data in a Pod.

First, the sample points used as data, were developed, and uploaded to Solid as individual files. For this reason, Linked Data capability and interactions have not been tested. This is due to the added complexity and time constraints of implementing Linked Data.

Second, the study showed that under certain conditions, which are defined by the owner of the data, it is possible to access it. However, what was not demonstrated is the functionality of appending existing data (to see what happens with and without permission).

Future Research

Solid is an interesting concept with a lot of good ideas put together. Solid has shown that it can support a large variety of spatial data without any issues or constraints (as a file storage). Linked Data is a complex idea which takes time to adopt and implement. From the technological point of view, it would be interesting to do a comparison of running a large scale Solid server and compare it to other existing methods of data storage in terms of performance, efficiency, and computing requirements.

Assessing the feasibility of transforming existing applications and public services to utilize Pods is another point which needs to be addressed in the future. In theory it can make public services more efficient by proving Pods to citizens and users. However, it is unknown how difficult this will be. The Belgian Flanders project will be a great starting point a benchmark in determination of the feasibility of Pods. With the right application, Pods could in theory be also used in the company space with spatial data.

Throughout the thesis, several ideas have been mentioned for future research. For example, experimenting with Linked Data in Solid. This has potentially a big impact in how data can be handled and accessed. It would also be interesting to see how SPARQL and GeoSPARQL queries can be performed on structured data in a Pod. It would also be interesting to see if precise querying of file based datasets can be possible (either on its own or through another application). Server-side filtering and spatial queries are also something that needs to be researched for large spatial datasets. Investigating how TrinPod manages RDF triple level data access seems like an interesting point to start looking into more precise access control.

References

Abhayaratna, J., Daemen E., Janowicz, K., Parsons, E., Smith, R., Verschoor, F. (2021). The responsible use of spatial data. W3C Interest group note. Online. Available at: https://www.w3.org/TR/responsible-use-spatial/ [Accessed 15/11/2022].

About Comunica. (2023). Online. Available at: https://comunica.dev/about/ [Accessed 18/02/2022].

Alesheikh, A. A., Oskouei, A. K., Atabi, F., & Helali, H. (2005). Providing interoperability for air quality in-situ sensors observations using GML technology. *International Journal of Environmental Science & Technology*, *2*(2), 133-140.

Asahara, A., Hayashi, H., Ishimaru, N., Shibasaki, R., & Kanasugi, H. (2015, October). International Standard "OGC Moving Features" to address "4Vs" on locational BigData. In *2015 IEEE International Conference on Big Data (Big Data)* (pp. 1958-1966). IEEE.

Autolitano, S., & Pawlowska, A. (2021). Europe's quest for digital sovereignty: GAIA-X as a case study. *IAI Papers*, *21*(14), 1-22.

Bajaj, R., Ranaweera, S. L., & Agrawal, D. P. (2002). GPS: location-tracking technology. *Computer*, *35*(4), 92-94.

Ballagh, L. M., Raup, B. H., Duerr, R. E., Khalsa, S. J. S., Helm, C., Fowler, D., & Gupte, A. (2011). Representing scientific data sets in KML: Methods and challenges. *Computers & Geosciences*, *37*(1), 57-64.

Battle, R., & Kolas, D. (2011). Geosparql: enabling a geospatial semantic web. *Semantic Web Journal*, *3*(4), 355-370.

Becker, H., Vu, H., Katzenbach, A., Braun, C. H. J., & Käfer, T. (2021). Monetising resources on a solid pod using blockchain transactions. In *The Semantic Web: ESWC 2021 Satellite Events: Virtual Event, June 6–10, 2021, Revised Selected Papers* (pp. 49-53). Cham: Springer International Publishing.

Bermudez, L. (2017). New frontiers on open standards for geo-spatial science. *Geo-spatial information science*, *20*(2), 126-133.

Berners-Lee, T. (2009). *Linked Data*. Online. Available at: https://www.w3.org/DesignIssues/LinkedData.html [Accessed 06/11/2022].

Berners-Lee, T. (2018). One Small Step for the Web. Online. Available at: https://medium.com/@timberners_lee/one-small-step-for-the-web-87f92217d085 [Accessed: 22/11/2022].

Bierman, G., Abadi, M., & Torgersen, M. (2014, July). Understanding typescript. In *European Conference on Object-Oriented Programming* (pp. 257-281). Springer, Berlin, Heidelberg.

Biljecki, F., Ledoux, H., & Van Oosterom, P. (2013). Transportation mode-based segmentation and classification of movement trajectories. *International Journal of Geographical Information Science*, *27*(2), 385-407.

Bizer, C. (2009). The emerging web of linked data. *IEEE intelligent systems*, *24*(5), 87-92.

Blacklaws, C. (2018). *Algorithms: transparency and accountability. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 376(2128), 20170351*. doi:10.1098/rsta.2017.0351

Braud, A., Fromentoux, G., Radier, B., & Le Grand, O. (2021). The road to European digital sovereignty with Gaia-X and IDSA. *IEEE Network*, *35*(2), 4-5.

Butler, H., Daly, M., Doyle, A., Gillies, S., Hagen, S., & Schaub, T. (2016). *The geojson format* (No. rfc7946).

Capadisli, S., Berners-Lee, T., Verborgh, R., Kjernsmo, K., Bingham, J., & Zagidulin, D. (2021). The Solid
Protocol. Solid Project. https://solidproject.org/TR/protocol [Accessed 10/10/2022].

Cassidy, M. J., Jang, K., & Daganzo, C. F. (2010). The smoothing effect of carpool lanes on freeway bottlenecks. *Transportation research part A: policy and practice*, *44*(2), 65-75.

Chen, Y., Hong, T., & Piette, M. A. (2017). City-scale building retrofit analysis: a case study using CityBES. *Build. Simul*.

Chow, T. E. (2008). The potential of maps APIs for internet GIS applications. *Transactions in GIS*, *12*(2), 179-191.

Coopmans, C., & Chen, Y. (2008, September). A general-purpose low-cost compact spatial-temporal data logger and its applications. In *2008 IEEE AUTOTESTCON* (pp. 64-68). IEEE.

ESRI. (2003). *Spatial Data Standards and GIS Interoperability.* An ESRI White Paper.

Ferreira, E., & Silva, M. J. (2013). Using sensors and the geospatial web to map bio-data. *Revista Sensos-e, 0, online journal. Disponível em http://sensos-eese. ipp. pt*.

Foerster, T., Schaeffer, B., Brauner, J., & Jirka, S. (2009, February). Integrating ogc web processing services into geospatial mass-market applications. In *2009 International Conference on Advanced Geographic Information Systems & Web Services* (pp. 98-103). IEEE.

Foster, D. (2022). GPX Explained. Online. Available at: https://www.topografix.com/gpx_for_users.asp [Accessed: 05/11/2022].

GDPR. (2018). *What is GDPR, the EU's new data protection law?* Online. Available at https://gdpr.eu/what-is-gdpr/ [Accessed 25/10/2022].

GeoSPARQL Specification. (2022). Online. Available at: https://opengeospatial.github.io/ogc-geosparql/ [Accessed 31/10/2022].

GPX. (2004). Online. Available at: https://www.topografix.com/gpx.asp [Accessed 15/11/2022].

Graser, A., Zimányi, E., & Bommakanti, K. C. (2020). From Simple Features to Moving Features and Beyond?. *arXiv preprint arXiv:2006.16900*.

Haklay, M., & Weber, P. (2008). Openstreetmap: User-generated street maps. *IEEE Pervasive computing*, *7*(4), 12-18.

Herring, J. Perry, M. (2022). GeoSPARQL - A Geographic Query Language for RDF Data. Online. Available at https://www.ogc.org/standards/geosparql [Accessed 31/10/2022].

Horbiński, T., & Lorek, D. (2022). The use of Leaflet and GeoJSON files for creating the interactive web map of the preindustrial state of the natural environment. *Journal of Spatial Science*, *67*(1), 61-77.

Ienca, M., & Vayena, E. (2020). On the responsible use of digital data to tackle the COVID-19 pandemic. *Nature medicine*, *26*(4), 463-464.

Inrupt documentation. (2023). Online. Available at: https://docs.inrupt.com/ [Accessed 2/1/2023].

Isaak, J., & Hanna, M. J. (2018). User data privacy: Facebook, Cambridge Analytica, and privacy protection. *Computer*, *51*(8), 56-59.

iShare. (2019). Goals and scope of the iShare Scheme. Online. Available at: https://ishareworks.atlassian.net/wiki/spaces/IS/pages/70222185/Goals+and+scope+of+the+iSHARE+Scheme [Accessed: 1/12/2022].

ISO. (2022). *About us.* Online. Available at: https://www.iso.org/about-us.html [Accessed 31/10/2022].

Jain, P., Hitzler, P., Yeh, P. Z., Verma, K., & Sheth, A. P. (2010, March). Linked data is merely more data. In *2010 AAAI Spring Symposium Series*.

Jesús-Azabal, M., Moguel, E., Laso, S., Murillo, J. M., Galán-Jiménez, J., & García-Alonso, J. (2021). Pushed SOLID: deploying SOLID in smartphones. *Mobile Information Systems*, *2021*, 1-13.

Kadaster Labs. (2022). Online. Available at: https://labs.kadaster.nl/ [Accessed 25/11/2022].

Kralidis, A. T. (2008). Geospatial open source and open standards convergences. In *Open source approaches in spatial data handling* (pp. 1-20). Springer, Berlin, Heidelberg.

Kryukov, A. P., & Demichev, A. P. (2018). Decentralized data storages: Technologies of construction. *Programming and Computer Software*, *44*, 303-315.

Laoutaris, N. (2018). Data transparency: Concerns and prospects [point of view]. *Proceedings of the IEEE*, *106*(11), 1867-1871.
Locher, A. E., & Termens, M. (2012). Exploring alternatives for geodata preservation.

Manoharan, S. (2009, April). On GPS tracking of mobile devices. In *2009 Fifth International Conference on Networking and Services* (pp. 415-418). IEEE.

Mansberger, R. (2003). Geoinformation in support of decentralization.

Mansour, E., Sambra, A. V., Hawke, S., Zereba, M., Capadisli, S., Ghanem, A., ... & Berners-Lee, T. (2016,
April). *A demonstration of the solid platform for social web applications*. In *Proceedings of the 25th international conference companion on world wide web* (pp. 223-226).

McNeff, J. G. (2002). The global positioning system. *IEEE Transactions on Microwave theory and techniques*, *50*(3), 645-652.

McShane, R., & Skelt, L. (2009). GPS tracking for people with dementia. *Working With Older People*.

Michael, K., McNamee, A., & Michael, M. G. (2006, June). The emerging ethics of humancentric GPS tracking and monitoring. In *2006 International Conference on Mobile Business* (pp. 34-34). IEEE.

OGC KML Specification. (2015). Online. Available at: http://docs.opengeospatial.org/is/12-007r2/12-007r2.html#1 [Accessed 15/11/2022].

OGC. (2022). *About OGC*. Online. Available at: https://www.ogc.org/about [Accessed 31/10/2022].

Open data Bratislava. (2023). Online. Available at: https://opendata.bratislava.sk/ [Accessed 2/1/2023].

Ordnance Survey. (2022). *What is spatial data?* Online. Available at: https://www.ordnancesurvey.co.uk/business-government/our-partners/what-is-spatial-data [Accessed 31/10/2022].

Pan, J. Z. (2009). Resource description framework. In *Handbook on ontologies* (pp. 71-90). Springer, Berlin, Heidelberg.

Portele, C. (2007). Opengis® geography markup language (GML) encoding standard. Version 3.2. 1.

Portele, C. (2021). OGC Testbed-17: OGC Features and Geometries JSON Engineering Report. Online. Available at http://docs.ogc.org/per/21-017r1.html#toc0 [Accessed 02/11/2022].

Portele, C., Vretanos P. P., Simmons, S. (2021). OGC Features and Geometries JSON Standards Working Group Charter. Online. Available at: https://portal.ogc.org/files/?artifact_id=97517 [Accessed 02/11/2022].

Ramesh, G., Sivaraman, K., Subramani, V., Vignesh, P. Y., & Bhogachari, S. V. V. (2021, January). Farm animal location tracking system using arduino and gps module. In *2021 International Conference on Computer Communication and Informatics (ICCCI)* (pp. 1-4). IEEE.

Reddy, G. P. O. (2018). *Spatial Data Management, Analysis, and Modeling in GIS: Principles and Applications. Geospatial Technologies in Land Resources Mapping, Monitoring and Management, 127–142.* doi:10.1007/978-3-319-78711-4_7

Ryoo, H. G., Kim, S., Kim, T., Seo, D., & Li, K. J. (2016, October). Generation of moving object trajectories in indoor space. In *Proceedings of the Eighth ACM SIGSPATIAL International Workshop on Indoor Spatial Awareness* (pp. 43-46).

Sambra, A. V., Mansour, E., Hawke, S., Zereba, M., Greco, N., Ghanem, A., ... & Berners-Lee, T. (2016).
Solid: a platform for decentralized social applications based on linked data. MIT CSAIL & Qatar Computing Research Institute, Tech. Rep

Sandbrook, C., Clark, D., Toivonen, T., Simlai, T., O'Donnell, S., Cobbe, J., & Adams, W. (2021). Principles for the socially responsible use of conservation monitoring technology and data. *Conservation Science and Practice*, *3*(5), e374.

Sandvik, B. (2008). Using KML for thematic mapping. *Institute of Geography School of GeoSciences. Edinburgh, University of Edinburgh. MSc in Geographical Information Science*, *22*.

Solid. (2022). About Solid. Online. Available at: https://solidproject.org/about [Accessed 15/11/2022].

SolidLab. (2022). *Welcome to SolidLab*. Online. Available at https://solidlab.be/ [Accessed 20/10/2022].

Sopjani, L., Stier, J. J., Hesselgren, M., & Ritzén, S. (2020). Shared mobility services versus private car: Implications of changes in everyday life. *Journal of Cleaner Production*, *259*, 120845.

Sporny, M., Longley, D., Kellogg, G., Lanthaler, M., & Lindström, N. (2020). JSON-LD 1.1. *W3C Recommendation, Jul*.

Stanley, J., & Granick, J. S. (2020). The limits of location tracking in an epidemic. *American Civil Liberties Union*, 1-9.

TrinPod. 2023. TrinPod Server. TrinPod. Online. Available at: https://TrinPod.us/

Usmani, R. S. A., Hashem, I. A. T., Pillai, T. R., Saeed, A., & Abdullahi, A. M. (2020). *Geographic Information System and Big Spatial Data. International Journal of Enterprise Information Systems, 16(4), 101–145.* doi:10.4018/ijeis.2020100106

Van Andel, M., Bussemaker, S., Michielsen, J. (2022). Solid Quest Github Repository. Available at: https://github.com/kadaster-labs/solid-quest [Accessed 1/12/2022].

Van de Wynckel, M., & Signer, B. (2022). A Solid-based Architecture for Decentralised Interoperable Location Data. In *12th International Conference on Indoor Positioning and Indoor Navigation*. CEUR Workshop Proceedings.

Venverloo T, Duarte F, Benson T, Leoni P, Hoogendoorn S, et al. (2023) Tracking stolen bikes in Amsterdam. PLOS ONE 18(2): e0279906. https://doi.org/10.1371/journal.pone.0279906

Welch, T. F., & Widita, A. (2019). Big data in public transportation: a review of sources and methods. Transport reviews, 39(6), 795-818.

Yannakopoulos, J. (2003). Hypertext transfer protocol: A short course. *University of Crete. August*.

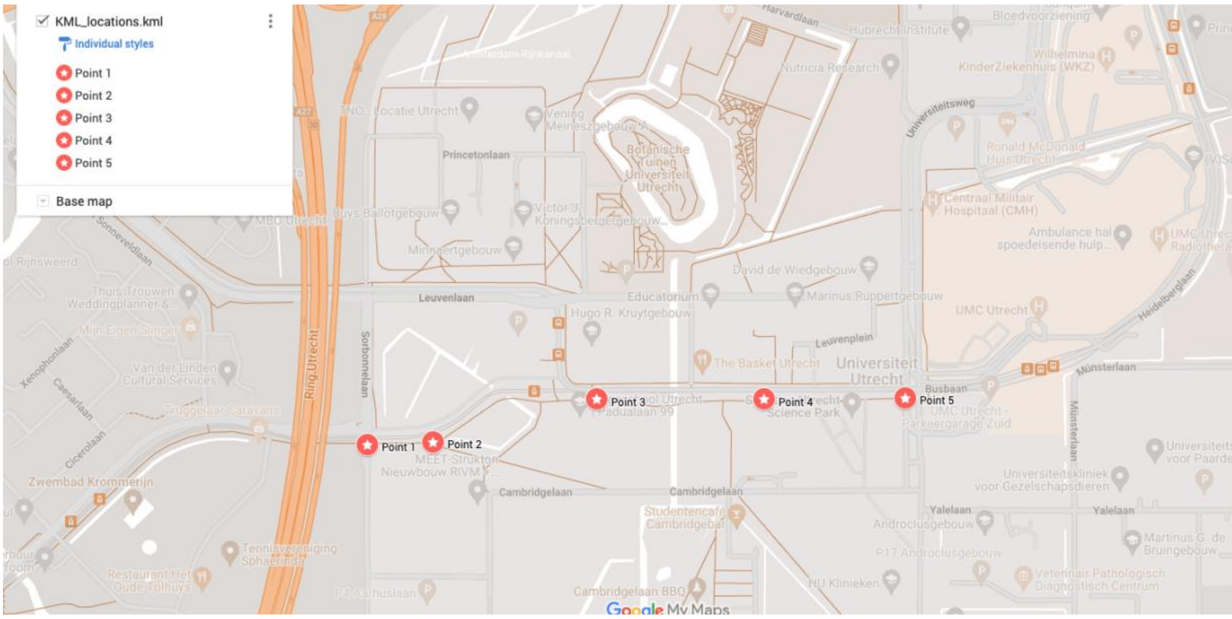
Zhang, C., Li, W., Day, M. J., & Peng, Z. R. (2003). GML-based interoperable geographical databases. *Cartography*, *32*(2), 1-16.

Zichichi, M., Serena, L., Ferretti, S., & D'Angelo, G. (2022). Complex queries over decentralised systems for geodata retrieval. *IET Networks*.

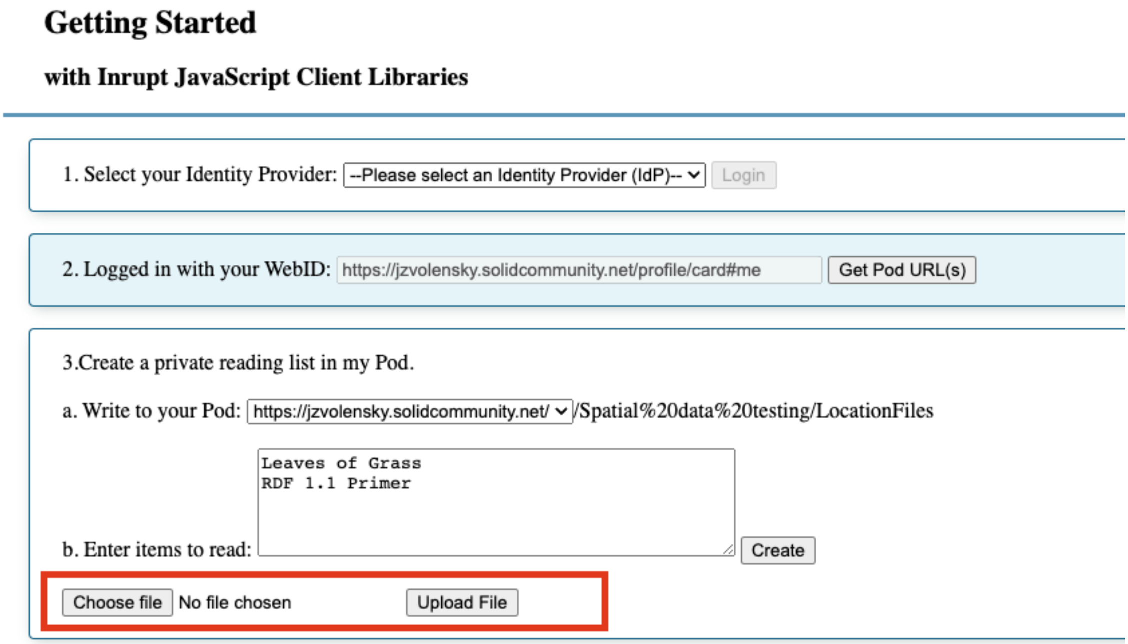
Zucker, J. (2022, September 14). Re: Granting access control to resources? Solid Project Forum. Available at: https://forum.solidproject.org/t/granting-access-control-to-resources/5617/7

Zunke, S., & D'Souza, V. (2014). Json vs xml: A comparative performance analysis of data exchange formats. *IJCSN International Journal of Computer Science and Network*, *3*(4), 257-61.

Appendix A: Methodology details



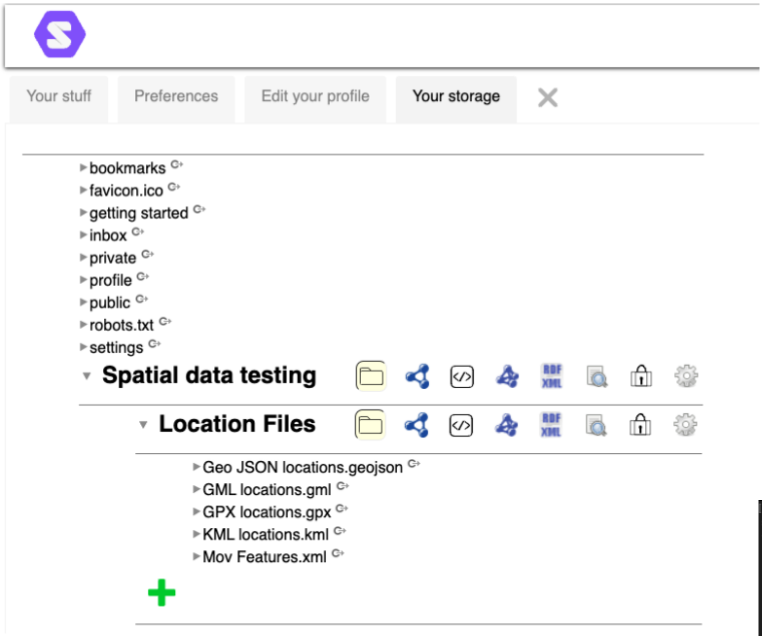
Appendix A1: Five sample points around the Utrecht University campus



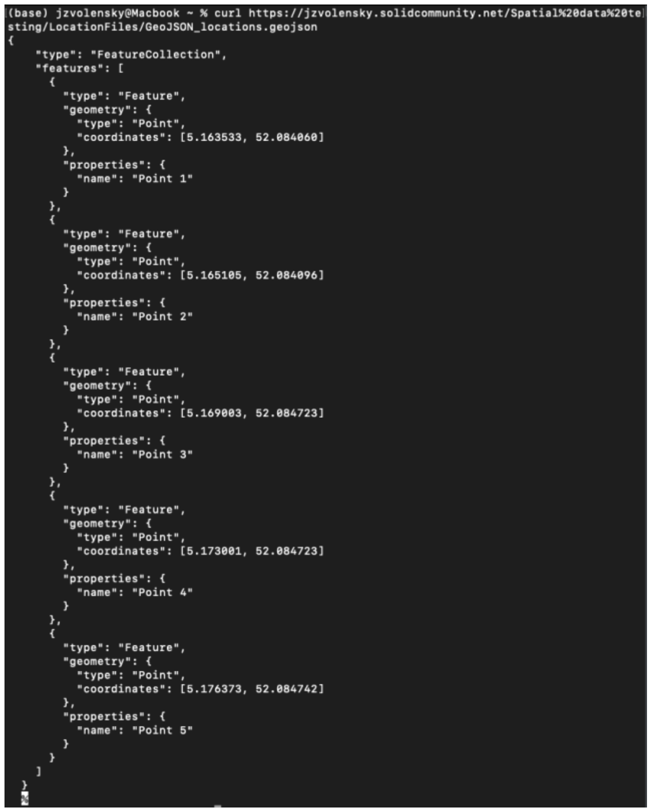
Appendix A2: Added functionality to the Solid tutorial application.

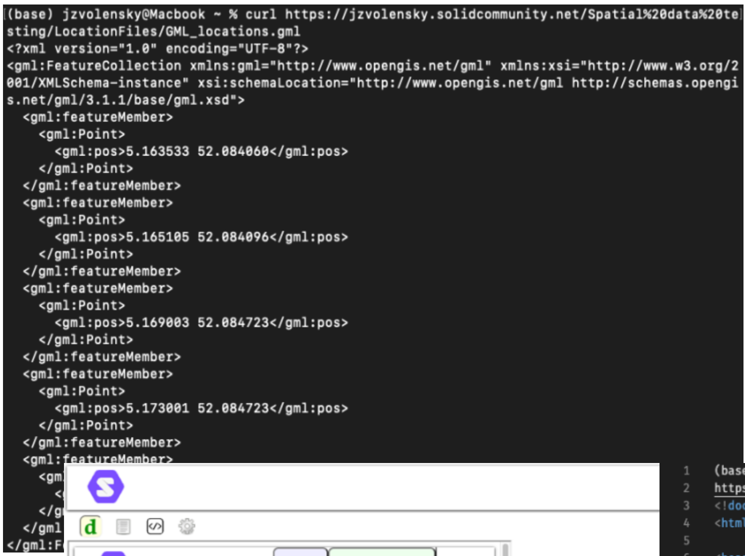
Appendix B: Results and analysis

Appendix B1: Location of the data in a Pod viewed on the web.

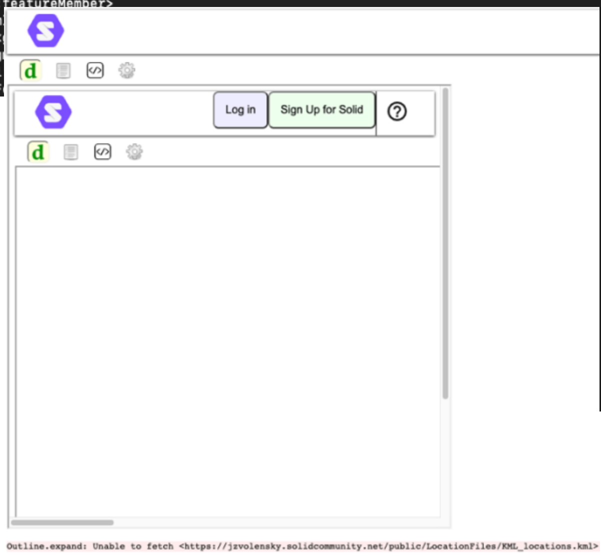


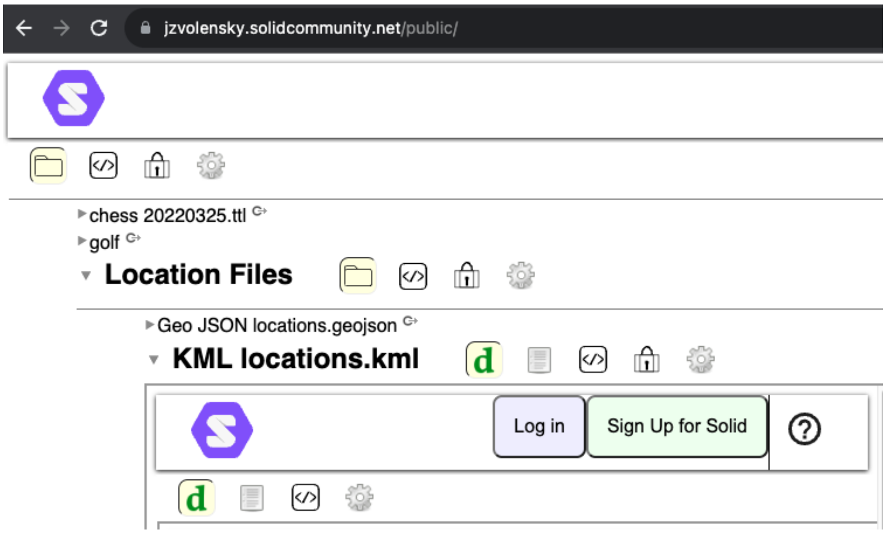
Appendix B2: Successfully requesting GeoJSON data with the command line.



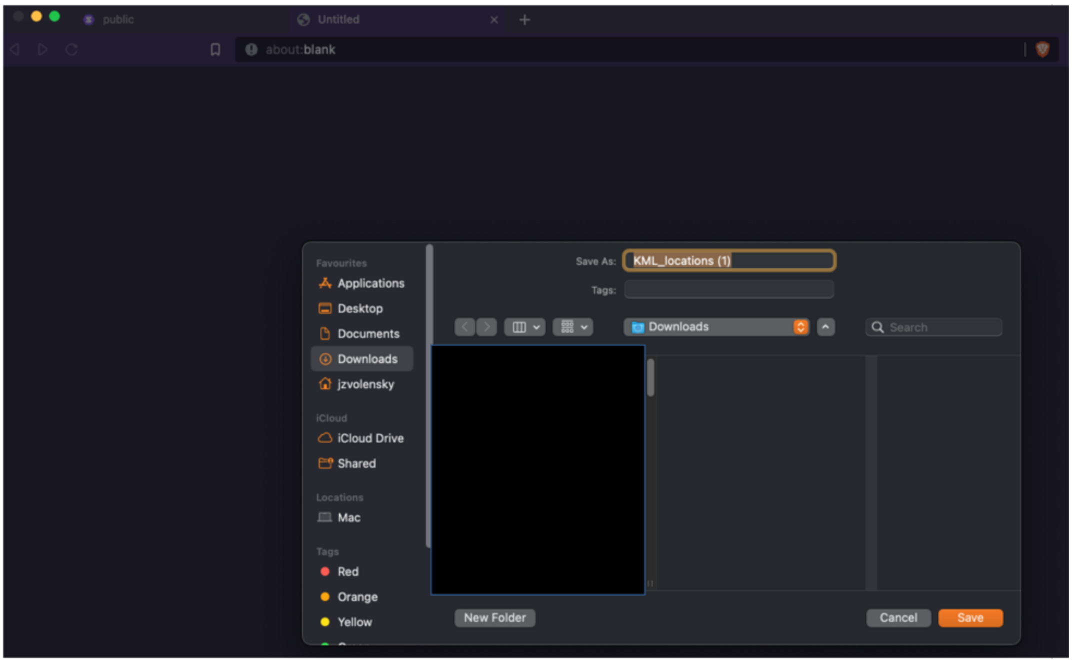


Appendix B3: Successfully requesting GML data with the command line.





Appendix B4: Login requirements and errors when accessing data, the user is not authorized to see. Top left and bottom are using the web GUI, while the top right is a request through the command line.



Appendix B5: Successful authorization allows data download.