eMOTIONAL Cities

Mapping the cities through the senses of those who make them

DELIVERABLE 3.5

Data Management Plan II

DECEMBER | 2024

BYTEROAD connecting data

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement nº945307. This document reflects only the author's view and the Commission is not responsible for any use that may be made of the information it contains.

Project Title	eMOTIONAL Cities: mapping the cities through the senses of those who make them
Deliverable	**D3.5** - Data Management Plan II
Work package	**WP3** - Data management and infrastructure
Task	**T3.4** - Deployment and maintenance of the SDI for the different use cases
Number of pages	27
Dissemination level	Confidential, only for members of the consortium (Including the Commission Services)
Leader	Byte Road
Main authors	Joana Simões (BR), jo@byteroad.net **Antonio Cerciello (BR), ant@byteroad.net**
Peer review	IGOT - Instituto de Geografia e Ordenamento do Território da Universidade de Lisboa
Contributors	Paulo Morgado (IGOT), paulo@campus.ul.pt
File name	eMC_2024.12_D3.5_DMPII_RF.pdf
Versions	V1: 30-12-2024 (Peer review) V2: 07-02-2025 (Submitted version)
Revision	07-02-2025

General Disclaimer

This deliverable may be subject to final acceptance by the European Commission. The information and views set out in this document are those of the authors and do not necessarily reflect the official opinion of the European Commission. Neither the Commission nor any person acting on the Commission's behalf may be held responsible for the use which may be made of the information contained therein.

Copyright message

Copyright message ©eMOTIONAL Cities Consortium, 2021-2025. This document contains original unpublished work or work to which the author/s holds all rights except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

Acronym List

ABM	Agent-Based Modeling
AoI	Area of Interest
API	Advanced Programming Interface
CAMB	The Chancellor masters and scholars of the University of Cambridge
CAS	Central Authentication Service
CRS	Coordinate Reference System
CSV	Comma Separated Value
CSW	Catalogue Service for the Web
DMP	Data Management Plan
EC	European Commission
EEG	Electroencephalography
EO	Earth Observation
ERC	European Research Council
FAIR	Findable Accessible Interoperable Reusable
FGDC	Federal Geographic Data Committee
fMRI	Functional Magnetic Resonance Imaging
FOSS	Free Open Source Software
GDPR	General Data Protection Regulation
GeoJSON	Geographical JSON
GIS	Geographic Information System
GPS	Global Positioning System
IETF	Internet Engineering Task Force
I/O	Input/ Output
ISO	International Standards Organization
JSON	JavaScript Object Notation
JWT	JSON Web Token
LDAP	Lightweight Directory Access Protocol
MSU	Michigan State University
NIST	National Institute of Standards and Technology
OGC	Open Geospatial Consortium
OpenAIRE	Open Access Infrastructure for Research in Europe
ORDP	Open Research Data Pilot
Pull Request	PR
REST	Representational State Transfer
SDI	Spatial Data Infrastructure
SOS	Sensor Observation Service
STA	Sensor Things API
STAC	SpatioTemporal Asset Catalog
W3C	World Wide Web Consortium
WCS	Web Coverage Service
WFS	Web Feature Service
WMS	Web Map Service.
WP	Work Package
XML	Extensible Markup Language
XSL	Microsoft Excel file format

Index

LIST OF FIGURES

1. Introduction

Open access to research data plays a critical role in modern science, to ensure that researchers can build their work on top of previous efforts and advance in their research fields, instead of having to spend time and resources re-creating existing experiments. While the digitization of data and the advent of big data have transformed research methodologies, significant obstacles persist in making research data discoverable, or at least easily accessible, as data publishing is not yet a generalized practice, even amongst public institutions. This reality could be the result of different factors, such as a culture of "fear" of becoming obsolete or losing competitive advantage by sharing data assets, or simply a lack of knowledge about common practices and technologies for the storage, management and sharing of research data. Moreover, the lack of standardized data infrastructures and the absence of comprehensive data-sharing practices among institutions contribute to these challenges (Kitichin and Lauriault 2018, Ref. 1). As a result, such data may remain stored in offline locations, such as hard disks, CDs or other ad-hoc storage solutions used by the researchers, which often are not associated with backup policies, and therefore can raise concerns regarding their preservation.

To address these issues, the European Commission (EC) has created the ODRP pilot, which "aims to improve and maximise access to and re-use of research data generated by Horizon 2020 projects (...)" (EUROPEAN COMMISSION Directorate-General for Research & Innovation 2016, Ref. 2), by establishing open access as the default setting for any data generated. From 2017, participation in this pilot became the default setting, although it is possible to opt out at any stage of the project. In this project, we decided not to opt out of this pilot, as we subscribe to this vision of open science based on open data, which fully aligns with our objective of making geospatial information FAIR, by implementing a Spatial Data Infrastructure (SDI). Therefore, we are legally bound to the obligations associated with participation in this pilot:

- Formulating a Data Management Plan (DMP), up to 6 months after the start of the project (D3.1, DMP I), defining briefly which data assets the project will generate/process, how these data will be made accessible and how they will be curated, stored and preserved.
- Providing open access to research data, if possible.
- Producing a final DMP (D3.5, DMP II), before the end of the project (this document), indicating the data produced within the project, their nature and the methodologies used to collect and store them.

It is important to add, that according to the ERC (European Research Council 2017, Ref. 3) "The data management plan will also have to specify if certain datasets remain closed and the reasons for not giving access should be given (for instance, if one of the project objectives would be jeopardised by providing open access to certain data or some of the generated data are considered sensitive, etc.”. In the current project, some datasets will indeed remain closed due to their sensitive nature, and this will be fully justified in this document. It is important to add, that according to the ERC (European Research Council 2017, Ref. 3) “The data management plan will also have to specify if certain datasets remain closed and the reasons for not giving access should be given (for instance, if one of the project objectives would be jeopardised by providing open access to certain data or some of the generated data are considered sensitive, etc.”. In the current project, some datasets will indeed remain closed due to their sensitive nature, and this will be fully justified in this document.

2. Data summary

The eMOTIONAL Cities project aimed to investigate how the urban environment influences the well-being of humans who inhabit it, conducting experiments and collecting data to support the research. Most of these experiments have a precise relationship with the urban spaces where they are performed; for this reason, they require an infrastructure to store, manage, relate and share the geospatial data they originate.

Some experiments, especially those related to the neuroscience domain, involve human volunteers and, for this reason, proper legal and ethical procedures associated with this purpose.

In addition to the experiments and deliverables indicated in the grant agreement, other assets were produced in the scope of this project:

- Software code and documentation.
- Raw experiments data (non-spatial data).
- Articles published in open-access scientific journals.
- Deliverables, as part of the project's Grant Agreement.
- Conference and Workshop abstracts/articles.

In the following sections, we describe these assets in more detail, and the procedures in place to store and publish them, including licensing, when applicable.

Software Code

This project produced software code, either by reusing existing tools or by creating new tools. In addition, it also produced documentation, demos/workshops, and other resources that support software usage. Project partners were encouraged to publish their code in repositories under the project's GitHub organisation (GitHub, 2024, Ref. 4) and to publish software releases on Zenodo (Zenodo, 2024, Ref. 5).

Raw Experiments Data

Although the large majority of the data collected, curated, analysed and produced in the project were made available through the SDI, there are a few datasets collected through the experiments which do not have spatial information associated with them. Even if in digital format, these datasets are not suitable for loading into geospatial software. They were stored on disk online and/or offline and on the project's data lake.

Project Documents

The project deliverables are an important output of the research produced, and they can be used to disseminate the research. Other documents include internal and external presentations (e.g., scientific and project meetings, and conferences). They are available on open access at the project's website (eMOTIONAL Cities, 2024, Ref 6) and on the project's community on Zenodo (eMOTIONAL Cities, 2024, Ref 7).

Articles

This project released publications in Open Access journals from relevant fields (e.g.: Neuroscience, Urban Planning, Geographical Information Systems (GIS)). They are listed on the project's website (eMOTIONAL Cities, 2024, Ref 6).

Consent Forms

Subjects who participated in experiments conducted within the project signed consent forms, in which they were informed about the nature of the experiments and the use that would be made of the data derived from them. Each partner conducting experiments is responsible for using the most appropriate format for the nature of the experiments, following local laws. This documentation is kept in physical form and is stored with all necessary precautions by the partner who conducted the experiment.

Datasets

The *e*MOTIONAL Cities project puts together data from the "traditional" geospatial domain (e.g.: EO images, census, social media streams, climate and environmental mobile and ground-sensors) and from the neuroscience domain (e.g.: fMRI, EEG, Eye tracking, biosensors, Clinical data). In general, neuroscience data is not geotagged as it is analysed based on its temporal dimension, rather than its geospatial dimension. However, geotagging neuroscience datasets was a requirement of *e*MOTIONAL Cities, as we wanted to index them in a common SDI, enabling researchers to leverage the power of geospatial analysis and overlay them with other geospatial datasets. The following sections describe the requirements that drove the design of the SDI, and the summary of the datasets that were published at this point in time.

2.1 Data requirements

At the beginning of the project, we conducted a requirement analysis (T3.1) to inform the design of the SDI. This included a data survey sent to all partner organizations, to capture the characteristics of the datasets they intended to collect and generate. The survey was structured into six sections—Overview, Spatial Attributes, Temporal Attributes, Data Format, Metadata, and Data Delivery—as described in Annex 9.1.

	Partner	Area of Interest	Description	Does the dataset contain PII?	File Size	Operation	Software/ programming language?	Spatially enabled?	Geotag	CRS	Spatial resolution	Temporal resolution
0	FMUL	Lisbon	Eye tracking/ gaze data	No	around 3 MB	W	Proprietary software that comes w/ hardware	No				< 1 ms
1	FMUL	Lisbon	Clinical information	Yes	very small	W	spreadsheet	Yes	Postcode		postcode	
2	DTU	Copenhagen	Road network	No	few hundred MBs	R	python w/\nOverlassAPI	Yes	coordinates	WGS84/ Danish -> EPSG:25832	1 meter/ OSM	
3	IGOT	Lisbon	Twitter data	Yes	multiple files of 10 MB	R	Python	Yes	coordinates	WGS84	coordinates	streaming
4	IGOT	Metropolitan area of Lisbon	Modis/ Sentinel 5 - Air pollution	No	20 - 40 GB each file	R	Google earth engine - Python	Yes	Georeferenced	WGS84	250-2000 m	Daily

Figure 1 - Screenshot of a sample of the survey results.

Because partners held varying levels of familiarity with GIS and the survey's terminology, we decided to run web-conference interviews. During these sessions, we introduced some GIS concepts and filled in the survey together for at least one dataset. This approach aimed at speeding up the completion of the survey, while ensuring accuracy.

From that initial effort, we identified 46 dataset types, including eye-tracking data, U.S. Census data, social media streams and Earth Observation (EO) images. The complete results were documented in Annex 9.2. At the time (month 3 of the project), CAMB expressed the intention to contribute the highest number of datasets, followed by NGR and MSU. As anticipated, data came from several European cities and one U.S. city, with London having the highest number of datasets.

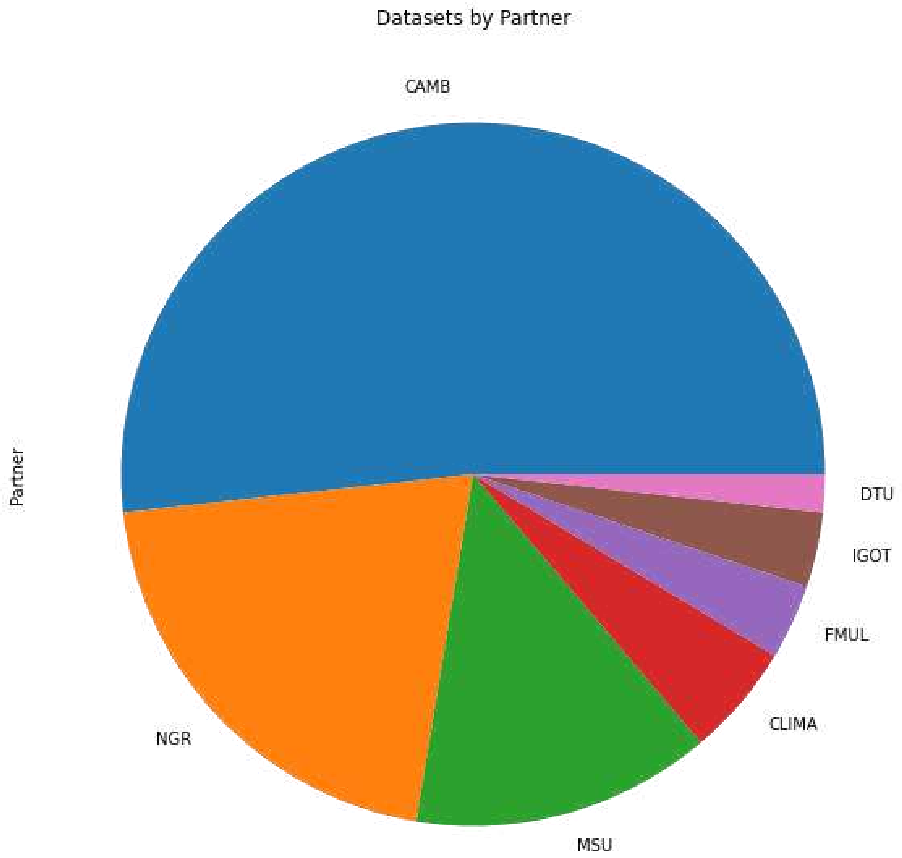


Figure 2 - Proportion of contributed datasets, by partner.

About 13% of the data was expected to contain Personally Identifiable Information, which raised ethical considerations covered in section 6. Approximately 95% of those datasets had geospatial attributes (with the remainder linkable to a geospatial dataset), though some required geocoding. WGS84 was the predominant coordinate reference system. Spatial resolutions ranged from GPS-level detail to broader census tracts, while temporal resolution varied from milliseconds to decades.

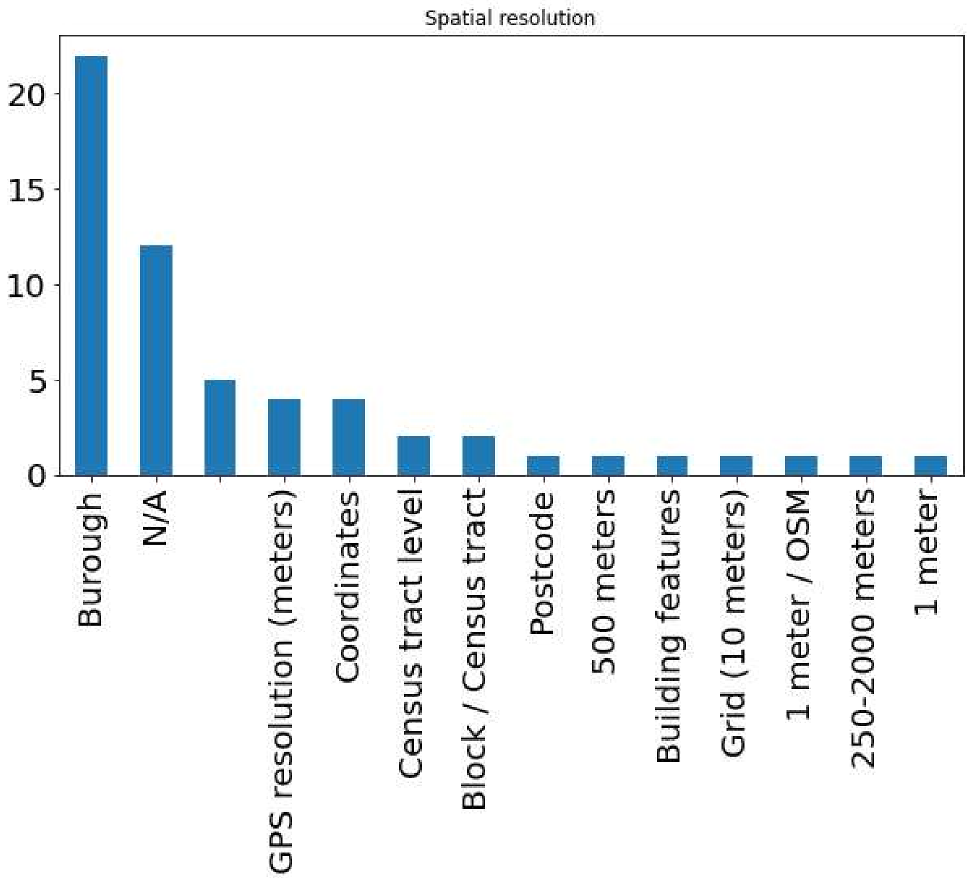


Figure 3 -Spatial resolution/ minimum unit of spatial analysis of the contributed datasets.

Most datasets were under 10 MB, though a few could reach hundreds of gigabytes and potentially grow over time. Data were primarily acquired through direct downloads or GIS software, with many datasets updated during the project's lifetime.

Roughly 72% of the data was in non-proprietary formats—often tabular or standard geospatial formats such as GeoJSON, Shapefile, and GeoTIFF. Metadata formats had not always been specified. The datasets were intended for various uses, including internal project research, public Open Data release, GIS analysis, interactive mapping, and specialized techniques such as linking brain signals to behaviour, sentiment analysis, and spatial clustering.

This initial census allowed us to define on the one hand the functional and infrastructural requirements of the SDI (see SDI Design) and, on the other hand, to appropriately model the data collection, immediately addressing the discussion on the treatment of sensitive data and other ethical concerns.

2.2 SDI Design

The SDI for the eMOTIONAL Cities project has evolved into a robust and flexible system designed to ensure the efficient ingestion, management, and publication of diverse geospatial datasets. The SDI supports both modern and legacy OGC standards, such as OGC API Features, Tiles, Records and Web Map Services (WMS), Web Features Services (WFS), ensuring compatibility with a wide range of tools and user requirements.

Data ingestion begins with a **data lake**, hosted on AWS S3, which acts as a secure staging area for all datasets. Partners (data producers) upload their data in preferred formats (e.g., GeoJSON, CSV, or GeoPackage), accessing the dedicated folders with proper identity through access controls. Once data is in the data lake, a series of automated and manual processes prepare the datasets for publication in the SDI. This task is curated by SDI maintainers (WP3 Team). In this way, the data lake handles raw and intermediate data, while the SDI focuses on public-facing, harmonised and standardised datasets.

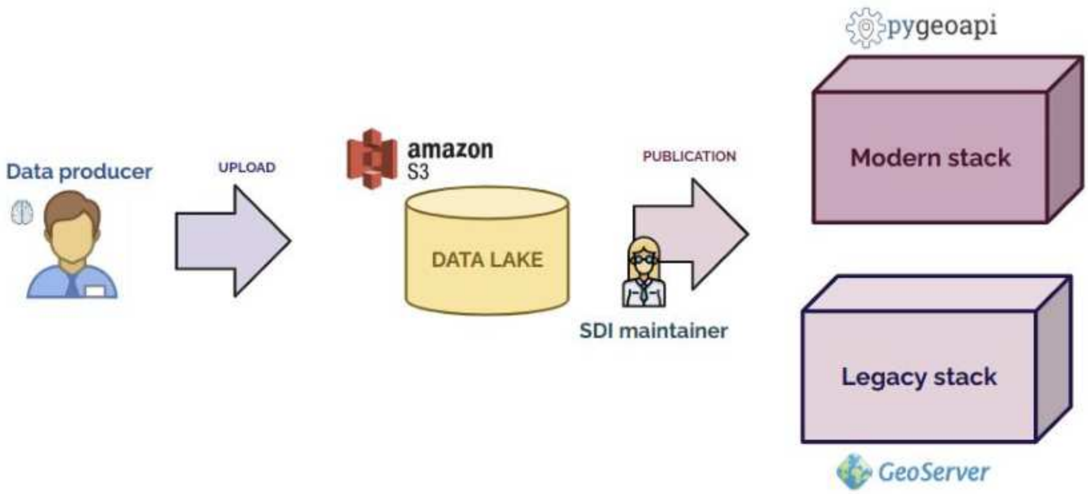


Figure 4 - Data ingestion/publication schema.

The metadata ingestion process runs parallel to dataset uploads, allowing metadata to be submitted independently and linked to its corresponding dataset. This flexibility accommodates diverse workflows and supports project-wide consistency. Partners can provide metadata through different channels, such as user-friendly Google Forms or Sheets for manual entry, and programmatic REST APIs for automated submissions.

At the end of the ingestion pipelines, both datasets and their metadata are seamlessly integrated into the SDI and published using standardized OGC formats, such as OGC API - Features, Tiles, Records, and legacy services like WMS and WFS. These formats ensure the data is interoperable and accessible to a diverse range of users, supporting various use cases from GIS analysis to web-based visualizations; for more details about the Standards used, please refer to section 3.2. Tools like QGIS, Leaflet, and custom web applications can consume the published data, enabling researchers, developers, and decision-makers to easily access and utilize the project's resources for their specific requirements. Among the tools included in the SDI, there is a webGIS, Mapstore (see Figure 5), that enables creating dashboards and even stories about the project's data, without writing one line of code.

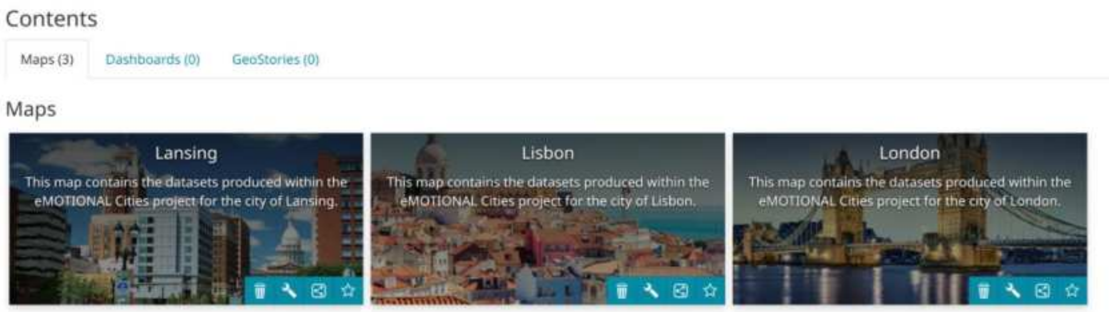


Figure 5 - Screenshot of the MapStore instance, deployed in the SDI.

Furthermore, through links in the metadata records, it is possible to download the published datasets in GeoJSON, GeoPackage and GeoParquet formats, to facilitate offline processing that requires rapid access to the files. Specifically, it should be noted that the adoption of the GeoParquet format has been the subject of a dedicated webinar (Simões Cerciello, 2024, Ref. 8) and an article (doublebyte, 2024, Ref. 9), given its emerging nature among the OGC data formats.

The main solution implemented in SDI has proven to be adequate for most of the use cases that emerged during the project. However, a use case related to the direct publication of field experiment data, before the downsampling and data cleaning required for publication in SDI, remained relatively unaddressed; in practice, this meant directly publishing the raw data produced by the sensors used in the experiments. To support this need, we added a sensor data server, FROST-Server (Fraunhofer Institut, 2024, Ref. 10), which implements the OGC SensorThingsAPI (STA) Standard (see section 3.2). In this way, it is possible to directly publish data coming from sensors, without the need for any intermediate pipeline.

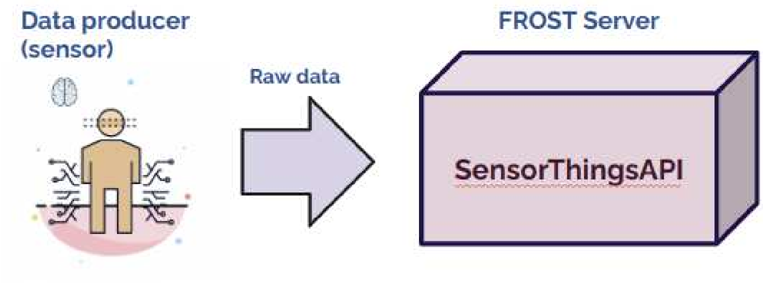


Figure 6 - SensorThingsAPI.

Further details on SDI implementation can be found in the deliverables D3.2 (Simões, Cerciello, 2022, Ref 11) and D3.3 (Simões, Cerciello, 2023, Ref. 12) from eMOTIONAL Cities project.

Such SDI design is well-positioned to support current project requirements and future data-driven research initiatives.

2.3 Data collected

The eMOTIONAL Cities catalogue (eMOTIONAL Cities Consortium, 2024, Ref. 13) contains 114 public metadata records, which describe the data ingested into the SDI to date. Figure 7 describes the visibility status of the datasets: 103 geospatial collections are public (published as OGC API Features, Tiles, WFS and WMS) and 10 geospatial collections are private (published as WFS and WMS, only). One non-geospatial private dataset is stored on a private data lake. The private datasets are safeguarded by security measures (see section 5). It should be noted, that by the time this report was written, some experiments and data collection were still running, to increase sample size, and therefore more data will be ingested in the SDI in the following days.

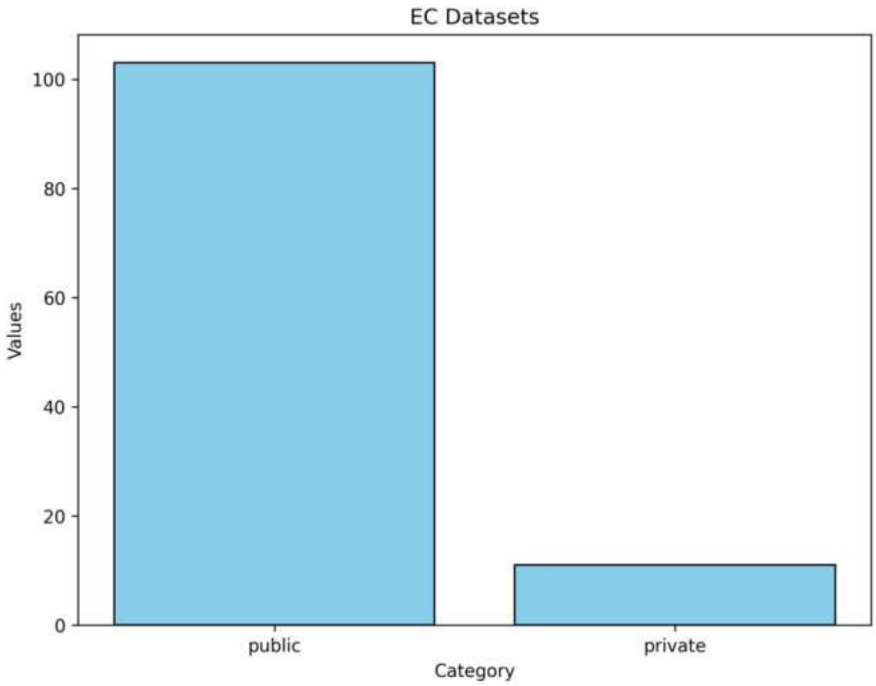


Figure 7 - Number of public/private datasets published on the SDI.

Public datasets were published either using CC BY-NC 4.0 (Creative Commons, 2024, Ref. 14) or Attribution-NonCommercial-NoDerivatives 4.0 International (Creative Commons, 2024, Ref. 15).

The data lake on AWS S3 stores the raw datasets contributed by the partners, as well as the processed datasets that are used by the SDI. The storage size of the S3 bucket is 140.1 GB, while the average object size is 25.2 MB (see Figure 8). Although this average value is not very high, it is still considerably higher than what was expressed during the initial data survey, described on section 2.1.

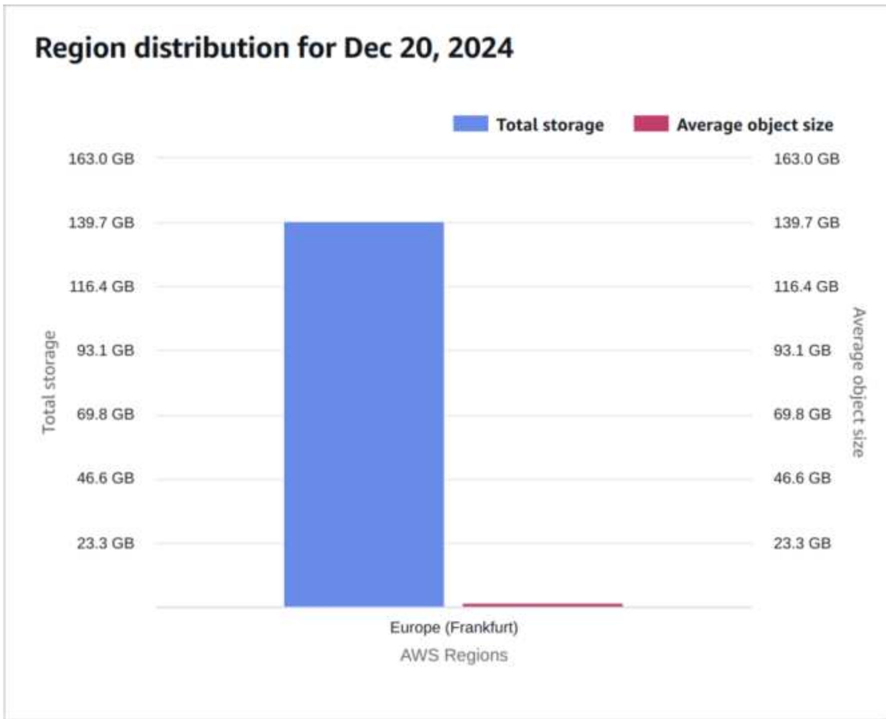


Figure 8 - Size statistics for the S3 Data Lake.

The word cloud (see Figure 9) describes the variety of topics that the datasets comprise, ranging from urban to environment, mental health and pollution. “Urban Health” is the most adopted keyword, which maps well to the project's overarching nature.

Figure 9 - Word cloud of keywords found on metadata records of the eMOTIONAL Cities catalog.

Figure 10 shows the proportion of published datasets per region/case study (e.g.: Lisbon, London, Michigan and Copenhagen). It is worth noting that these values are likely to change, as a data survey conducted in October 2024, flagged that some project partners have the intention to upload more datasets into the SDI before the end of the project (see section 4).

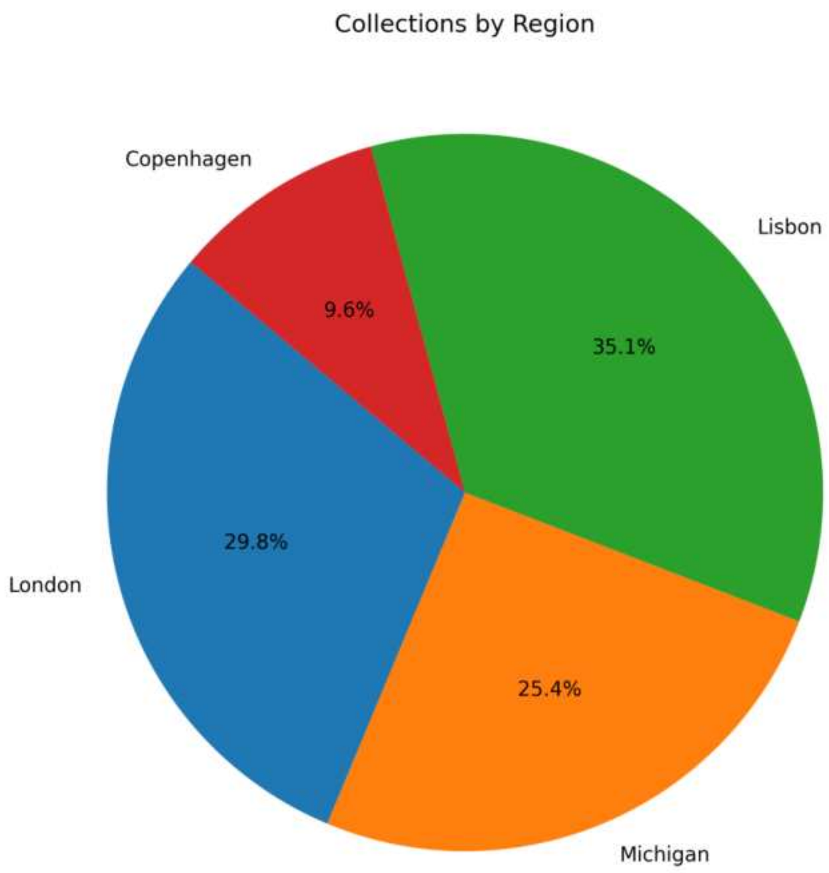


Figure 10 - Published collections, per case study (region).

3. Enabling FAIR data

2.3 FAIR Principles

One of the outstanding challenges of data-intensive science has been to improve knowledge discovery through assisting both humans, and their computational agents, in the discovery of, access to, and integration and analysis of research data (Wilkinson and al. 2016, Ref. 16). In 2014, a workshop named 'Jointly Designing a Data Fairport' took place in Leiden, Netherlands. It brought together a wide group of academic and private stakeholders all of whom had an interest in overcoming data discovery and reuse obstacles. This workshop led to the creation of what became known as the FAIR foundational principles: that data should be Findable, Accessible, Interoperable and Reusable, both for machines and for people.

SDIs play a crucial role in facilitating the implementation of the FAIR principles in the geospatial domain (see section 3.3). In practice, it is not the data itself but the framework within which the data is supplied that will enable data to meet the FAIR data principles (Plunkett 2019, Ref. 17).

2.4 Making data findable, including provisions for metadata

Although often overlooked, metadata is the piece of information which helps users to find data, and once it has been found, it tells them how to interpret and use that data. Publishing metadata facilitates data sharing, and sharing between organizations facilitates cooperation and a coordinated and integrated approach to spatially related policy issues (ESRI 2002, Ref. 18)

During the 80s and 90s, the growing appreciation of the importance of geospatial metadata has led to the development of several initiatives, which established guidelines for the collection of metadata. The International Standards Organization (ISO) undertook the task of harmonizing these standards, resulting in the release of ISO 19115 "Geographic Information – Metadata" in 2003, followed by a revision in 2013. This standard defines how to describe geographical information and associated services, including contents, spatial-temporal purchases, data quality, access and rights to use. Although ISO 19115 is by far the most widespread standard for geospatial metadata, other standards also gain relevance in some parts of the world. This is the case of FGDC in the United States and INSPIRE in Europe.

It is important to state that these metadata standards are mostly based on the XML format. CSW is the OGC standard which deals with exposing a catalogue of geospatial records in XML over the Internet, and it is still widely used in many SDIs all over the world. However, from 2019 a new standard is being developed - OGC API records (OGC 2021, Ref. 19) - which follows a more modern web approach. Unlike CSW, this standard does not establish mandatory encodings and encourages the use of the linked data format.

The STAC specification is another effort to provide a common language to describe a range of geospatial information, also called 'spatiotemporal asset' (STAC 2021, Ref. 20). There are current efforts to establish a mapping between the record atomic unit of OCG API records and the STAC specification. The possibility of extending the definition of the record with an ISO 19115 profile was also discussed during the OGC API Virtual Code Sprint, in July 2021 (OGC 2021, Ref. 21).

In this project, we took part in these discussions and explored these newer approaches for the encoding and publishing of geospatial metadata.

2.5 Making data accessible and interoperable

The notion of SDIs emerged more than 20 years ago, and has been constantly evolving, in response to both technological and organisational developments. According to the INSPIRE initiative, one of the greatest efforts to establish an infrastructure of this sort in the public sector of Europe, an SDI can be defined as "a framework of policies, institutional arrangements, technologies, data, and people that enables the sharing and effective usage of geographic information by standardizing formats and protocols for access and interoperability" (Tonchovska, Stanley, and De Martino 2021, Ref. 22).

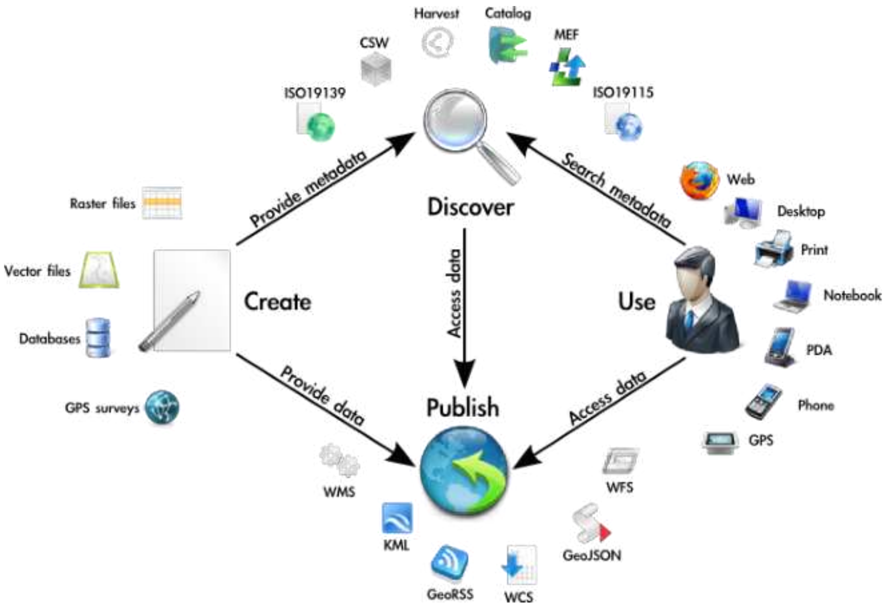


Figure 11 - Different components of an SDI (Benthall and Evans 2021, Ref. 23).

In recent times, we have seen the rise of new architectural approaches, which can be characterized by their data-centrism; in practice, it means that data is an asset and all technologies and approaches should ensure that the access to it is as efficient and easy as possible.

APIs provide an opportunity for developers to easily create new products, by hiding the complexity of upstream infrastructures and offering a set of well-defined and documented methods. Traditionally, SDIs adopt first generation service interfaces from OGC (e.g.: WMS, WFS, WCS and SOS), which are well known and supported by client applications. However, modern web-based APIs go one step further as they (i) provide a simple approach to data processing and management functionalities, (ii) possibly offer different encodings of the payload, (iii) can easily be integrated into different tools, and (iv) can facilitate the discovery of data through mainstream search engines such as Google and Bing (Kotsev et al. 2020, Ref. 24). These APIs often follow a RESTful architecture, which simplify its usage, while minimizing the bandwidth usage. Finally, the OpenAPI specification (OpenAPI Initiative 2011, Ref. 25) allows to document APIs in a vendor independent, portable and open manner, which provides an interactive testing client within the API documentation.

OGC has embraced this new approach in its new family of standards called OGC APIs (OGC 2020, Ref. 25). When this project started, two OGC APIs were published: 'OGC API - Features' (OGC 2021, Ref. 27) and the SensorThings API (OGC 2021, Ref. 28), which provide standardised APIs for accessing to spatial and observation data.

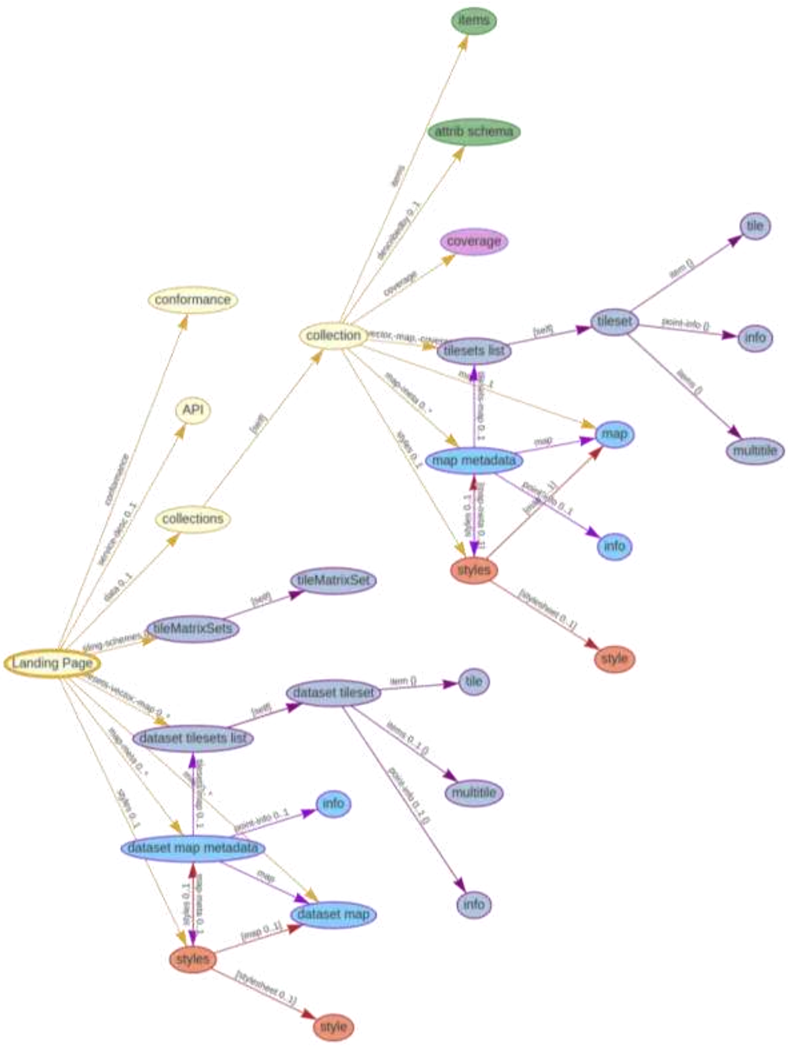


Figure 12 - Unlike other OGC standards, the OGC API family is made of building blocks. The relation between the different blocks is represented on this diagram (Maso 2021, Ref. 29).

During the lifetime of the project, several other OGC APIs Standards were published that allow access, management and querying of different types of geospatial data and metadata: OGC API - Tiles, OGC API - processes, OGC API - Maps, OGC API - EDR, OGC API - Moving Features, OGC API - Common and OGC API – Records (see Figure 13). This shows the rapid pace of evolution of these technologies, which is followed by software implementations and market adoption.

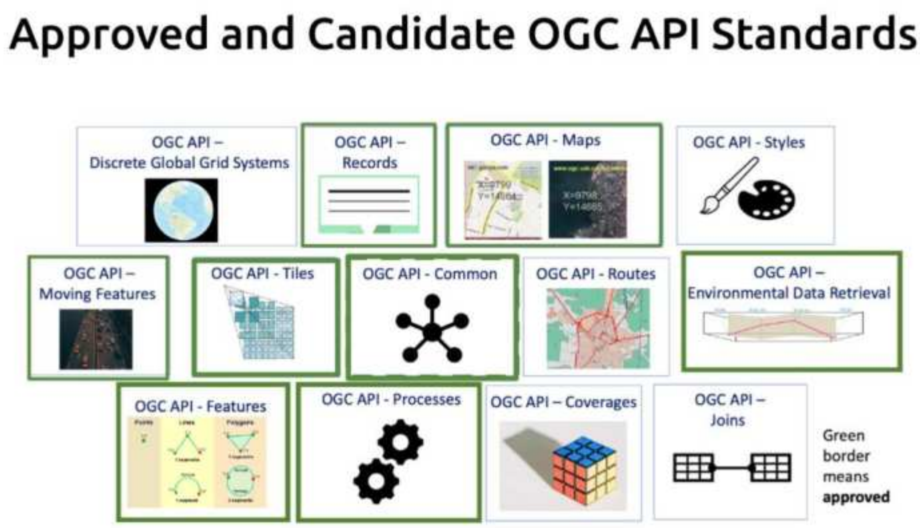


Figure 13 - Approved and Candidate OGC API Standards, as of December 2024.

OGC publishes a Standards Timeline Roadmap (OGC December, 2024, Ref. 30) that shows the current status and predicted development of the different OGC APIs (See Figure 14).

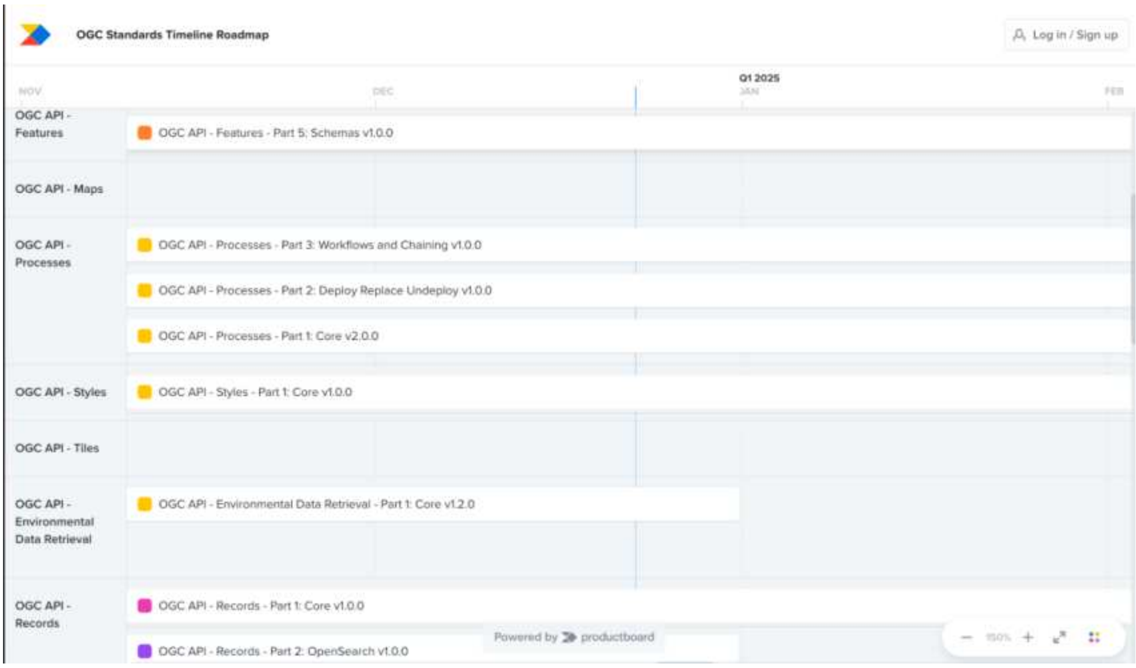


Figure 14 - Screenshot of the "OGC Standards Timeline Roadmap"; updated version at (OGC, December, 2024, Ref. 30).

Most of the geospatial data collected/generated in this project was published in the SDI. Although not published in the SDI, some non-geospatial datasets were ingested in the data lake. We did our best effort to create metadata records for every dataset created in the project, regardless of their privacy level or whether they were geotagged.

The SDI enables researchers to:

- Access their own data, transformed to meet some requirements (e.g.: format, CRS).
- Share their data with other researchers from the project.
- Share their data with a wider audience.

Interoperability is at the core of the SDI design (addressed in Task T3.2 of the project). In order to ensure that researchers (and the wider public) can access these data using their desktop GIS, web application or other tools, we adopted Standards from the relevant communities: OGC, Internet Engineering Task Force (IETF) and the World Wide Web Consortium (W3C).

Although the initial data survey (see section 2.1) indicated that we would have both raster and vector datasets, only vector datasets were contributed to the SDI. With those requirements in mind, our initial iteration published data using OGC API - Features and OGC API - (vector) Tiles. Although from the very beginning of the project we embraced the modern OGC API Standards, we decided to also support the first generation of OGC Standards (e.g.: WMS, WFS), as many GIS experts are used to working with them and they are widely supported in GIS tools (Simoes and Cerciello, 2022, Ref. 31). As the project evolved, we also realised that we had another stakeholder that was not completely served by the selected Standards: the biosignals scientist. They collected large amounts of data which, unlike the other geospatial datasets, were dense in the time component, but sparse in the geospatial component; they also had a requirement of storing metadata about the sensors and measurements associated with the data streams. For those stakeholders, we deployed an implementation of STA, as explained in section 2.2. The OGC STA Standard provides an open and unified framework to interconnect IoT sensing devices, data, and applications over the Web (OGC 2021, Ref. 28). It offers both a data model to store the observations and an API to query them.

Finally, towards the end of the project, an emerging Standard candidate started gaining a lot of traction within the Data Science community: GeoParquet (OGC, 2024, Ref 32). It is part of a family of cloud-native geospatial, that leverages cloud technologies to store and stream datasets in a more efficient way. In 2024, GeoParquet was added to the SDI, as an alternative encoding for vector data (doublebyte, 2024, Ref 10).

Figure 15 summarizes the different stakeholders we identified in the SDI, mapped to different Standards. Our goal was to make the SDI more user-friendly by providing each dataset in multiple formats, each format targeting a particular stakeholder.

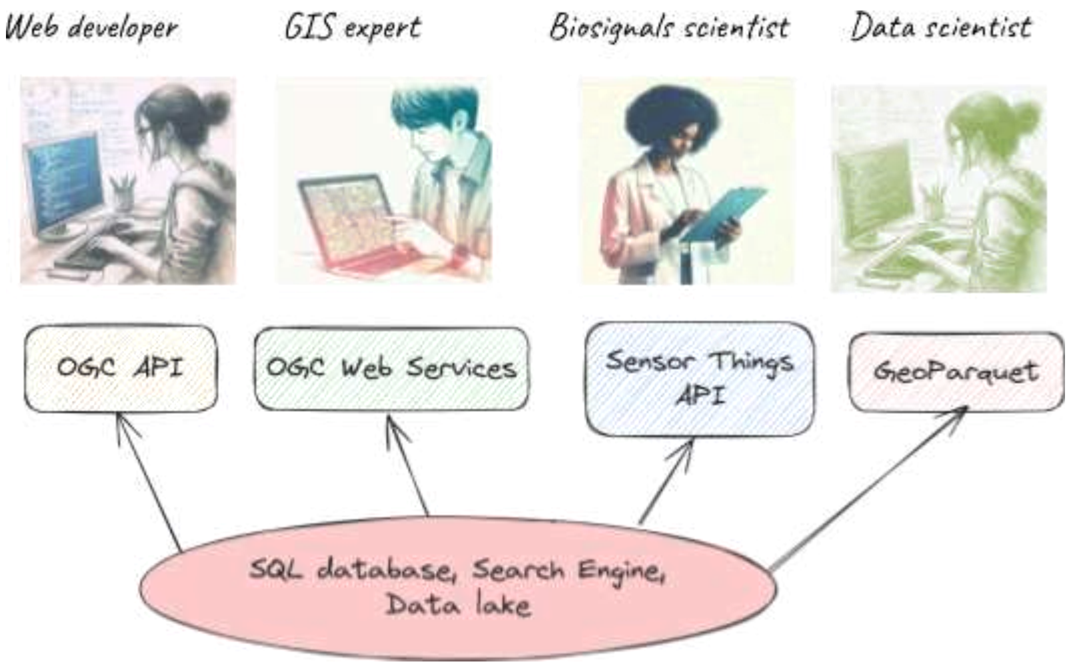
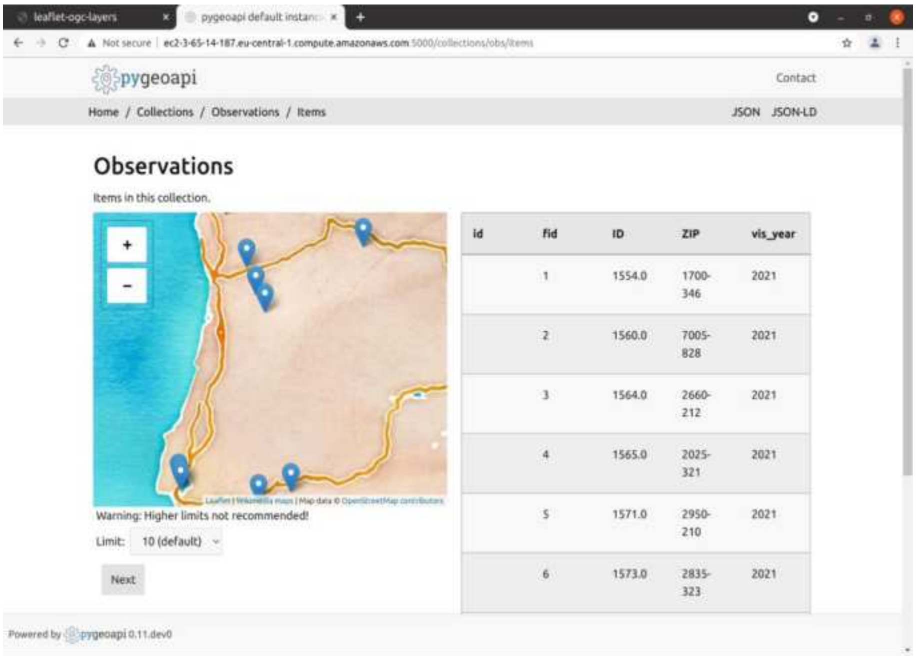


Figure 15 - SDI stakeholders, mapped to Standards.

The code of the SDI is publicly available under an Open Source license, MIT (accomplished in Milestone MS7). The different iterations have resulted in three releases on Zenodo: in May 2022, September 2023 and September 2024 (Cerciello and Simões, 2024, Ref. 33).

The *e*MOTIONAL Cities project has positioned itself as an early adopter and contributor of the new breadth of OGC APIs, by participating in several OGC Code Sprints (OGC, 2024, Ref. 34). A Code Sprint is a collaborative and inclusive event driven by innovative and rapid programming with minimal process and organization constraints to support the development of new applications and candidate standards.



id	fid	ID	ZIP	vis_year
	1	1554.0	1700-346	2021
	2	1560.0	7005-828	2021
	3	1564.0	2660-212	2021
	4	1565.0	2025-321	2021
	5	1571.0	2950-210	2021
	6	1573.0	2835-323	2021

Figure 16 - Pygeoapi deployed on AWS serving a neuroscience dataset, as an outcome of the code sprint (OGC 2021, Ref. 35).

In these code sprints we took a leading role by giving several workshops that showcased the results of the project to other participants (Cerciello, 2021, Ref. 36; Simões, 2022, Ref. 37; Simões, 2022, Ref. 38; Simões, 2022, Ref. 39; Simões, 2022, Ref. 40; Simões, 2023, Ref. 41; Simões, Cerciello and Lopes, 2023, Ref. 42; Simões and Cerciello, 2024, Ref. 43). More information about the outcomes of the code sprints can be found on the engineering reports (for instance, EarthPulse 2021, Ref. 44).

The project also contributed to the implementation of the Standards, by updating and improving pygeoapi, a central component of the SDI Stack (see section 2.2). The eMOTIONAL Cities project has originated 63 Pull Requests (PRs) that were merged into the upstream repository of pygeoapi. Figure 17 shows the pace of contribution on GitHub.

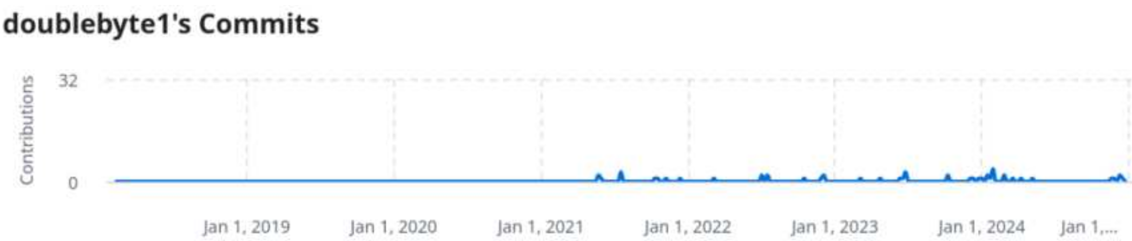


Figure 17 - Commits frequency to the pygeoapi project. This graph shows that the contributions were consistent during the lifetime of the project.

Finally, even if these technologies are sound and ready, they are not relevant until the market adopts them. We gave an important contribution towards that adoption through outreach activities. We disseminated the results of this project in FOSS4G 2022, FOSS4G EU 2024 and FOSS4G 2024 (Kralidis, Bartoli, Cerciello and Simões, 2022 Ref.

45; Cerciello and Simões 2022, Ref. 46; Cerciello and Simões, 2024, Ref. 47; Simões and Sanz, 2024, Ref. 48; Cerciello and Simões, 2024, Ref. 49; Cerciello and Simões, 2024, Ref. 50). FOSS4G is the largest conference of the Geospatial Free and Open Source Community, taking place each year in different parts of the world. In November 2023 we shared best practices learned in this project within the Urban Health Cluster (Cerciello, Simões and Martinez, 2023, Ref. 51) and in October 2024 we gave a webinar about GeoParquet, open to the general public (Simões and Cerciello, 2024, Ref. 9).

2.6 Dissemination Policy and Intellectual Property Management

The general principles for handling knowledge and intellectual property rights (IPR) within eMOTIONAL Cities are the ones described in the signed Grant Agreement. These principles are in line with the H2020 IPR recommendations (H2020 rules for participation (Art. 41, 49)) and the model grant agreement (Art. 23-31).

Aligned with the European policy on open data (EU, Ref. 52) and FAIR Principles, the project pushed for the “openness” of information, with public project results that can be reused by other stakeholders. In the cases where some restrictions were put in place (for instance, for security reasons), these were explicitly stated.

4. Allocation of resources

According to the NIST, cloud computing is a “model for enabling convenient, on-demand network access to a shared pool of configurable computing resources that can be rapidly provisioned and released” (Mell and Grance 2021, Ref. 53). Although cloud computing in some form was available before 2000, it has been increasingly adopted from 2010, providing a cheaper and more convenient alternative to using in-house, physical servers.

The SDI is hosted on Amazon Web Services (AWS), a cloud platform whose costs were accounted for in the budget of the project proposal. Task T3.4 covered the publishing and maintenance of the SDI in this platform. Moreover, we leveraged the capabilities of the cloud provider for implementing tasks such as security or backups.

At the core of the SDI is a cloud-based data lake hosted on an AWS S3 bucket (Amazon Web Services, 2024, Ref. 54). This data lake acts as a central repository for all project datasets, which are stored in a variety of formats such as GeoJSON, GeoPackage, GeoParquet, and others. Each project participant has a dedicated folder, with access managed using AWS Identity and Access Management (IAM) to ensure data security and control (Amazon Web Services, 2024, Ref. 55).

AWS Identity and Access Management (IAM) works with S3 buckets by allowing fine-grained control over who can access the bucket and its contents. IAM enabled the creation of users, groups, and roles, each with specific permissions defined in policies. These policies determine the actions that entities can perform on the bucket, such as reading, writing, or deleting objects. IAM policies can be applied at the bucket level or to individual objects, ensuring that only authorized users or systems can access the data.

The use of S3 has offered a scalable solution that accommodated the evolving data storage requirements of the project while supporting diverse data formats.

The entire SDI infrastructure (except the S3 data lake storage) is containerized using Docker, allowing seamless deployment and scalability across different environments. Docker Compose is used to orchestrate the integration of various components, ensuring consistent performance and ease of management.

We have allocated enough resources to cover the hosting of the platform and associated functionality during the lifetime of the project.

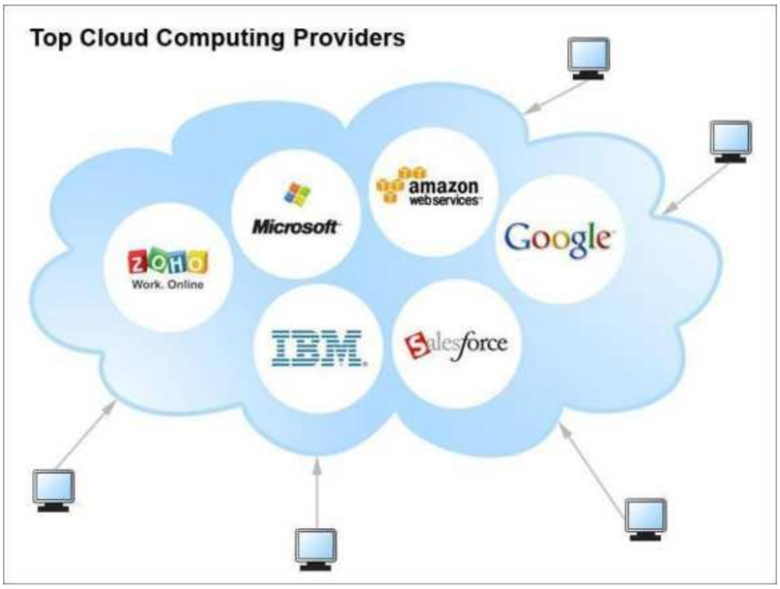


Figure 18 - Example of Cloud Computing Providers (Lengerson 2014, Ref. 56).

Despite this centralised approach, the FMUL partner opted to store specific raw datasets generated from their experiments on a Network-Attached Storage (NAS) system under their management. This decision was driven by the sensitive nature of the data and the need for the usage of neuroscience proprietary formats. FMUL has committed to processing these datasets to ensure the anonymisation of sensitive information before sharing them via the SDI. This ensures that only de-identified data are accessible through the centralised infrastructure, maintaining compliance with data privacy regulations while enabling collaborative research.

To ensure the reliability and performance of the SDI, it has been monitored continuously using GeoHealthCheck (GeoHealthCheck, 2024, Ref. 57), a software solution specifically designed to assess the health of geospatial web services. GeoHealthCheck tracks the availability, response times, and compliance of SDI components with relevant standards. This monitoring enhances the system's robustness by identifying and addressing issues proactively, ensuring consistent performance and user satisfaction.

Git is a distributed version control system, which is generalised as a best practice for collaborative code development. GitHub is a Git platform, which is commonly used to host open source software development projects.

At the beginning of the project, we established a GitHub organisation (*e*MOTIONAL Cities 2021, Ref. 4) to host the code assets of the project, in particular those which are published as open-source software, as it increases the discoverability of the assets. Although we invite all partners to publish their code in these repositories, some are working on existing projects which already belong to other organisations. For logistic reasons, those repositories were kept there.

Aligned with our effort to use and produce Open Source Software, we recommended the use of compatible open-source licences. Our recommendation was to use the MIT licence (OpenSource Initiative, Ref. 58), a permissive software licence, which allows for commercial use. However, in some cases, this licence was not adopted, either because the software is derived from work in a restrictive licence (e.g. GPL), or because it is related to hardware which requires a different licence (e.g.: Apache).

For convenience reasons, we have created a Google Drive folder, shared among the partners of the project, to organise and store working documents. These include: research documents organised by WP, deliverables and milestones, documents related to the Urban Health Cluster, and presentations in internal and external meetings. When completed, some of these documents were published in the “Resources” section of the project’s website, so that they can be properly disseminated or, as explained in the next paragraph, in the Zenodo repository.

Zenodo, developed under the European OpenAIRE program, is a powerful open repository designed to support the principles of Open Science. It provides a versatile platform for hosting a wide range of assets and ensures their longevity and traceability through the provision of persistent Digital Object Identifiers (DOIs) (Zenodo, 2022, Ref. 5).

We have established a community for the eMOTIONAL Cities project, under the EU Open Research Repository on Zenodo (Zenodo eMOTIONAL Cities, 2024, Ref. 7); the goal of this community is to archive and disseminate project-related outputs. This repository is strongly recommended for hosting software, research papers, geospatial datasets, and other outcomes generated by the project. By utilizing this repository, the project ensures that its publications and data are easily discoverable, reusable, and aligned with the FAIR principles (Findable, Accessible, Interoperable, and Reusable).

Moreover, geospatial datasets included in the repository will be accompanied by standards-based metadata, as described in Section 3.2, to enhance discoverability and streamline access through relevant tools. The eMOTIONAL Cities Zenodo repository thus provides a centralized, reliable, and FAIR-compliant platform for the long-term preservation and dissemination of project deliverables, and other scientific outcomes, e.g. posters, presentations, technical reports of specific methodological protocols, etc.

4.1 Life After the Project: Sustaining Tools and Resources

During the project, we selected tools widely recognized for facilitating open data and software sharing at no cost. This includes GitHub to host developed software, workshop materials, and SDI documentation, and the recommendation to upload project deliverables to Zenodo. These platforms ensure accessibility and long-term discoverability in line with open science principles.

The SDI, given its specific technical requirements, required a more detailed discussion. Currently hosted on AWS, the SDI serves as a cloud-based solution integral to the project's data management. Its architecture is built around an S3 bucket for scalable, secure data storage, with containerized services orchestrated through Docker Compose. As mentioned in Section 4, AWS hosting and maintenance costs during the project were fully covered within the budget.

Data or documents not managed through these platforms, such as FMUL's sensitive data stored on its NAS system, remain under the partner's management. FMUL will take measures to ensure the availability of this data for as long as possible, prioritizing privacy and compliance.

Post-project, sustaining the SDI will require a practical and cost-effective hosting solution. A potential plan involves transitioning the SDI to a local server managed by one of the partners, DTU. This shift would reduce dependency on commercial cloud providers and may lower long-term operational costs. However, the transition would require addressing challenges such as scalability, performance, and maintenance previously handled by AWS. Hosting on DTU servers would also provide enhanced control over infrastructure, aligning with the project's commitment to secure and open data sharing.

The fact that the SDI is fully virtualized, based on docker images and orchestrated via Docker Compose, abstracts it from the actual infrastructure; this should minimize the effort to migrate it to a new infrastructure.

5. Data security

Each researcher in the eMOTIONAL Cities project is responsible for the proper handling and maintenance of the data they collect. This includes adopting appropriate tools for tasks involving data management, processing, and the transmission of sensitive information. Sensitive data must remain encrypted and/or access-controlled throughout its entire lifecycle.

Data security has been a fundamental consideration in the project, striking a balance between the openness of publicly accessible data and the protection of sensitive datasets. Most datasets within the SDI are open and publicly available, requiring only basic security measures to ensure integrity and availability. Any dataset not meeting these characteristics is either kept private or stored securely in the data lake, with strict measures to prevent unauthorized access.

Some datasets stored in the SDI, while not containing personal information, are deemed sensitive due to the nature of the topics they address. These datasets are protected with controlled access mechanisms using GeoServer's advanced security features, such as role-based access control. This ensures that restricted datasets are accessible only through GeoServer's services and to authorized individuals.

The finalized SDI design actively facilitates best practices in data security. It safeguards sensitive data against unauthorized access, corruption, or theft across its entire lifecycle.

As mentioned in section 4, certain project partners, such as FMUL, independently manage highly sensitive datasets. These are securely stored on a locally managed NAS system, providing an additional layer of security. FMUL ensures that sensitive data is fully anonymized before sharing it via the SDI, thereby minimizing risks and adhering to privacy regulations.

This comprehensive approach ensures the secure management of data while supporting the project's goals of accessibility and compliance with ethical standards.

6. Sensitive data and Ethics

The eMOTIONAL Cities project is committed to adhering to the highest ethical standards and ensuring compliance with relevant national, European Union, and international legislation. This includes the Charter of Fundamental Rights of the European Union and the European Convention on Human Rights and its Supplementary Protocols. Ethical considerations, particularly regarding data protection and privacy, are addressed in detail in Work Package 9, specifically in Deliverable D9.14 (Ethics Requirements).

The project follows the "data minimization principle," ensuring that all data processed is relevant and strictly limited to the research objectives. This approach is particularly significant for protecting personal data, upholding the principle of proportionality, and safeguarding the right to privacy.

Sensitive data in the eMOTIONAL Cities project is handled with the utmost care to ensure privacy and compliance with ethical standards. Such data is either anonymized before being shared through the SDI or securely stored by partners on dedicated systems with strict access controls.

The transfer of data between EU and non-EU partners is strictly regulated. Data transfer agreements ensure compliance with GDPR and include clear obligations for both transferors and recipients, following the Standard Clauses for data transfers to third countries outlined in COMMISSION IMPLEMENTING DECISION (EU) 2021/914. Similarly, data transfers from non-EU jurisdictions to the EU require proper authorization and documentation to align with applicable regulations.

The eMOTIONAL Cities project ensures that data ethics remain a foundational element of its research activities, respecting the privacy and rights of all participants. For more comprehensive details, please refer to Work Package 9, where these aspects are thoroughly covered.

7. Final Remarks: Open Data, Open Standards and Open Software

One of the main goals of the data management procedures established in the eMOTIONAL Cities project was to ensure that data that is produced in the context of the project and that is not subject to commercial exploitation or access restrictions can be made available as open data following the findable, accessible, interoperable, and reusable principle. In particular, Open data that can be used by third parties, possibly in different contexts, to generate new beneficial results, including new open data. Accordingly, such open data is one of the sustainable results of the project and was made accessible in the most efficient ways, by adopting open industry Standards.

Earlier on this journey, we have chosen to adopt emergent Standards (OGC API), which evolved during the lifetime of the project to a degree of maturity that makes them completely feasible for a modern SDI (Simoes and Cerciello, 2024, Ref. 59). We are glad that the work conducted in WP3 contributed to this evolution, improving interoperability not only in the scope of the eMOTIONAL Cities project, but in the overall wider geospatial community.

Fundamental to the success of this project was the use of FOSS solutions. This meant that we were able to build software and tools based on a solid foundation and that the work carried out will persist in time, providing a starting point for further future projects.

References

1. **Kitchin, R. and Tracey Lauriault.** 2018. "Digital data and data infrastructures". *Ash, J., Kitchin, R. and Leszczynski, A. (eds) Digital Geographies*. Sage, London. pp. 83-94.
2. **European Commission Directorate-General for Research & Innovation.** 2016. *H2020 Programme Guidelines on FAIR Data Management in Horizon 2020.* https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf.
3. **European Research Council.** 2017. *Guidelines on Implementation of Open Access to Scientific Publications and Research Data.* https://ec.europa.eu/research/participants/data/ref/h2020/other/hi/oa-pilot/h2020-hi-erc-oa-guide_en.pdf.

4. **GitHub.** 2024. "eMOTIONAL Cities Project Organisation." https://github.com/emotional-cities.
5. **Zenodo.** 2024. "Zenodo – Research. Shared." https://zenodo.org/.
6. **eMOTIONAL Cities.** 2024.-a. "eMOTIONAL Cities – Mapping the Cities through the Senses of Those Who Make Them." https://emotionalcities-h2020.eu/.
7. **eMOTIONAL Cities.** n.d.-b. "eMOTIONAL Cities Zenodo OpenAIRE Community." https://zenodo.org/communities/emotional-cities/.
8. **Simões, Joana and Antonio Cerciello.** *Navigating GeoParquet*. Google Slides. Accessed 2024. https://docs.google.com/presentation/d/1Bbv0s72fnGUavAJijX8nPs6irrLbkRSxrUMOk7Av0jk/edit.
9. **doublebyte.** 2024. "Navigating GeoParquet: Lessons Learned from the eMOTIONAL Cities Project." *Doublebyteblog (blog),* August 2, 2024. https://doublebyteblog.wordpress.com/2024/08/02/navigating-geoparquet-lessons-learned-from-the-emotional-cities-project/.
10. **Fraunhofer Institut.** 2024. "FROST-Server" *GitHub.* https://github.com/FraunhoferIOSB/FROST-Server
11. **Simões, Joana, and Antonio Cerciello.** 2022. "D 3.2 – Architecture Definition and Code for the Generic SDI." *Zenodo,* August 31, 2022. https://doi.org/10.5281/zenodo.13365357.
12. **Simões, Joana, and Antonio Cerciello.** 2023. "D 3.3 – Description of the SDIs I." *Zenodo,* April 14, 2023. https://doi.org/10.5281/zenodo.13365412.
13. **eMOTIONAL Cities Consortium.** 2024. "Emotional Cities Metadata Catalogue." *Zenodo,* September 6, 2024. https://doi.org/10.5281/zenodo.13709777.
14. **Creative Commons.** n.d.-a. "CC BY 4.0 – Legal Code." https://creativecommons.org/licenses/by/4.0/legalcode.
15. **Creative Commons.** n.d.-b. "Attribution-NonCommercial-NoDerivatives 4.0 International." https://creativecommons.org/licenses/by-nc-nd/4.0/.
16. **Wilkinson, Mark D., et al.** 2016. "The FAIR Guiding Principles for Scientific Data Management and Stewardship." *Scientific Data* 3, no. 160018 (March): n/a. https://www.nature.com/articles/sdata201618.
17. **Plunkett, Gordon.** 2019. "What Is the Role of Open Data in SDIs?" *Esri Canada.* https://resources.esri.ca/news-and-updates/what-is-the-role-of-open-data-in-sdis.
18. **ESRI.** 2002. *Metadata and GIS.* An ESRI® White Paper. Redlands, CA: ESRI. http://downloads.esri.com/support/whitepapers/ao_/metadata-and-gis.pdf.
19. **OGC.** 2021. "OGC API – Records." *GitHub.* https://github.com/opengeospatial/ogcapi-records.

20. **STAC.** 2021. "Enabling Online Search and Discovery of Geospatial Assets." *SpatioTemporal Asset Catalogs.* https://stacspec.org/.
21. **OGC.** 2021. "July 2021 OGC API Virtual Code Sprint." *GitHub.* https://github.com/opengeospatial/ogcapi-code-sprint-2021-07.
22. **Tonchovska, Rumyana, Victoria Stanley, and Samantha De Martino.** 2021. "Spatial Data Infrastructure and INSPIRE." *Open Knowledge Repository Europe and Central Asia Knowledge Brief* 55 (September): n/a. http://hdl.handle.net/10986/17082.
23. **Benthall, Sebastien, and Galen Evans.** 2021. "Spatial Data Infrastructure Best Practices with GeoNode." *SlideShare.* https://www.slideshare.net/SebastianBenthall/spatial-data-infrastructure-best-practices-with-geonode.
24. **Kotsev, Alexander, Marco Minghini, Robert Tomas, Vlado Cetl, and Michael Lutz.** 2020. "From Spatial Data Infrastructures to Data Spaces—A Technological Perspective on the Evolution of European SDIs." *ISPRS International Journal of Geo-Information* 9, no. 3 (March): 176. https://doi.org/10.3390/ijgi9030176.
25. **OpenAPI Initiative.** 2011. "OpenAPI Specification." *Swagger.* https://swagger.io/specification/.
26. **OGC.** 2020. "OGC APIs – Building Blocks for Location." https://ogcapi.ogc.org/.
27. **OGC.** 2021. "OGC API – Features." *GitHub.* https://github.com/opengeospatial/ogcapi-features.
28. **OGC.** 2021. "OGC SensorThings API." *GitHub.* https://github.com/opengeospatial/sensorthings.
29. **Maso, Joan.** 2021. "OGC API Tiles." http://joanma.uab.cat/temp/graphs/OGCAPItiles.htm.
30. **OGC.** December, 2024. "OGC Standards Timeline Roadmap." https://roadmap.productboard.com/1b15493e-0271-45e6-bbf9-4d7fcb5c22a5.
31. **Simões, J. and Antonio Cerciello**, "Serving Geospatial Data Using Modern and Legacy Standards: a Case Study from the Urban Health Domain", ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. 48W1, pp. 419–425, 2022. doi:10.5194/isprs-archives-XLVIII-4-W1-2022-419-2022.
32. **OGC.** 2021. "GeoParquet." *GitHub.* https://github.com/opengeospatial/geoparquet/.
33. **Cerciello, Antonio, and Joana Simões.** 2024. "Openapi-sdi." *Zenodo,* September 13, 2024. https://doi.org/10.5281/zenodo.13758189.
34. **OGC.** 2024. "OGC Code Sprints." https://developer.ogc.org/sprints.
35. **OGC.** 2021. "May 2021 OGC API Virtual Code Sprint." *GitHub.* https://github.com/opengeospatial/ogcapi-code-sprint-2021-05.
36. **Cerciello, Antonio.** 2021. "Load Feature Data into Your Frontend Application." In *November 2021 Geospatial API Code Sprint.* GitHub, November 16, 2021.

https://github.com/opengeospatial/developer-events/wiki/November-2021-Geospatial-API-Code-Sprint#load-feature-data-into-your-frontend-application.

37. **Simões, Joana.** 2022. "Organize Your Data on the Cloud with a Crawlable OGC API – Records Catalogue." In *Metadata Code Sprint.* GitHub, September 14, 2022. https://github.com/opengeospatial/developer-events/wiki/Metadata-Code-Sprint#organize-your-data-on-the-cloud-with-a-crawlable-ogc-api---records-catalogue-.
38. **Simões, Joana.** 2022. "Share Your Data with OGC API – Features." In *Vector Data Code Sprint.* GitHub, July 13, 2022. https://github.com/opengeospatial/developer-events/wiki/Vector-Data-Code-Sprint#share-your-data-with-ogc-api-features.
39. **Simões, Joana.** 2022. "Serve Vector Tiles with OGC API – Tiles." In *Space Partitions Code Sprint.* GitHub, May 10, 2022. https://github.com/opengeospatial/developer-events/wiki/Space-Partitions-Code-Sprint#serve-vector-tiles-with-ogc-api-tiles.
40. **Simões, Joana.** 2022. "How-to Access OGC API Features without Writing One Line of Code." In *2022 Joint OGC–OSGeo–ASF Code Sprint.* GitHub, March 9, 2022. https://github.com/opengeospatial/developer-events/wiki/2022-Joint-OGC-%E2%80%93-OSGeo-%E2%80%93-ASF-Code-Sprint#how-to-access-ogc-api-features-without-writing-one-line-of-code.
41. **Simões, Joana.** 2023. "Serve Vector Tiles with OGC API – Tiles." In *Tiling Interfaces Code Sprint.* GitHub, June 12, 2023. https://github.com/opengeospatial/developer-events/wiki/Tiling-Interfaces-Code-Sprint#serve-vector-tiles-with-ogc-api---tiles.
42. **Cerciello, Antonio, Joana Simões, and Gonçalo Lopes.** 2023. "Publishing Environmental and Biosignals Data Using OGC Standards." In *October 2023 Open Standards Code Sprint.* GitHub, October 30, 2023. https://github.com/opengeospatial/developer-events/wiki/October-2023-Open-Standards-Code-Sprint#publishing-environmental-and-biosignals-data-using-ogc-standards.
43. **Simões, Joana, and Antonio Cerciello.** 2024. "Publish Vector Data with #OGCAPI." In *2024 Joint OGC–OSGeo–ASF Code Sprint.* GitHub, February 20, 2024. https://github.com/opengeospatial/developer-events/wiki/2024-Joint-OGC-%E2%80%93-OSGeo-%E2%80%93-ASF-Code-Sprint#publish-vector-data-with-ogcapi.
44. **EarthPulse.** 2021. "EarthPulse Engineering Report." *GitHub.* https://github.com/doublebyte1/ogcapi-code-sprint-2021-05/blob/main/SummaryEngineeringReport/results/EarthPulse.adoc.
45. **Kralidis, Tom, Francesco Bartoli, Antonio Cerciello, and Joana Simões.** 2022. "Implementing OGC APIs Using Elasticsearch and pygeoapi." In *FOSS4G 2022* [conference presentation].
46. **Cerciello, Antonio, and Joana Simões.** 2022. "Serving Geospatial Data Using Modern and Legacy Standards: A Case Study from the Urban Health Domain." In *FOSS4G 2022* [conference presentation].

47. **Cerciello, Antonio, and Joana Simões.** 2024. “Adding GeoParquet to a Spatial Data Infrastructure: What, Why and How.” In *FOSS4G 2024* [conference presentation].
48. **Simões, Joana, and Jorge Sanz.** 2024. “Rendering OGC API Compliant Vector Tiles on the Fly with pygeoapi + Elasticsearch.” In *FOSS4G 2024* [conference presentation].
49. **Cerciello, Antonio, and Joana Simões.** 2024. “A Spatial Data Infrastructure Using Modern Standards: Lessons Learned from the eMOTIONAL Cities Project.” In *FOSS4G 2024* [conference presentation].
50. **Cerciello, Antonio, and Joana Simões.** 2024. “An eMOTIONAL SDI—What Makes an SDI User Friendly?” In *FOSS4G EU 2024* [conference presentation].
51. **Cerciello, Antonio, Joana Simões, and Ricard Martinez.** 2023. “WEBINAR: Building a Spatial Data Infrastructure—Lessons Learned from eMOTIONAL Cities and GDPR Strategies from WELLBASED.” Live webinar, November 23, 2023.
52. **EU Commission.** n.d. “European Legislation on Open Data.” https://digital-strategy.ec.europa.eu/en/policies/legislation-open-data.
53. **Mell, Peter, and Tim Grance.** 2021. “The NIST Definition of Cloud Computing.” *Computer Security Resource Center.* https://csrc.nist.gov/publications/detail/sp/800-145/final.
54. **Amazon Web Services**. “Amazon S3.” Accessed December 2024. https://aws.amazon.com/en/s3/.
55. **Amazon Web Services**. “AWS Identity and Access Management (IAM).” Accessed December 2024. https://aws.amazon.com/en/iam/.
56. **Lengerson, Diana**. 2014. “3 Factors to Consider When Choosing a Cloud Provider.”PHXnews. https://www.phxnews.com/www/3-factors-consider-choosing-cloud-provider/
57. **GeoHealthCheck.** 2024. “GeoHealthCheck – OGC Web Services Monitoring.” https://geohealthcheck.org/.
58. **OpenSource Initiative.** 2024. “The MIT License.” https://opensource.org/licenses/MIT.
59. **Simões, Joana, and Antonio Cerciello.** 2024. “A Spatial Data Infrastructure Using Modern Standards: Lessons Learned from the Emotional Cities Project.” In *FOSS4G 2024 – Academic Track.* https://doi.org/10.5281/zenodo.14223773.

ANNEX I. DATA SURVEY FIELDS

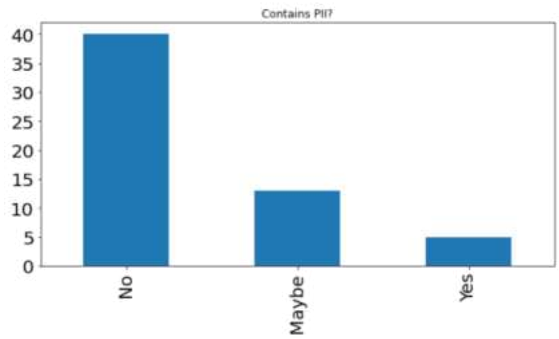
Field	Description
Timestamp	Timestamp of the survey
Email Address	Email address of the contact point
Name	Name of the contact point
Partner	Consortium partner

Area of Interest	Area covered by this dataset
Description	High-level description of the dataset
Does the dataset contain PII?	Contains Personal Identifying Information
File Size	Approximate size of the file
Main operation	Is this data going to be read or written through the project (or both)
Software/ programming language?	Which tool was used to acquire this data
Spatially enabled?	Is there a geospatial attribute
Geotag	Define geospatial attribute
CRS	Coordinate Reference System
Spatial resolution	Minimum spatial unit
Temporal resolution	Minimum temporal unit
Frequency of data	Latency of data collection
File format	Choose from a list
File format (Other)	Format not contained in the list
Is it proprietary?	Is it a proprietary format
Are you going to provide metadata?	Are there plans of providing metadata
Metadata standard	Choose metadata standard from a list
Metadata standard (Other)	Format not contained in the list
User	Identify the user of this dataset
Use Cases	Identify the use cases for this dataset
Specific SDI Requirements - Notes	Identify what is expected from the SDI
Expected availability date	Expected availability for the dataset in the SDI

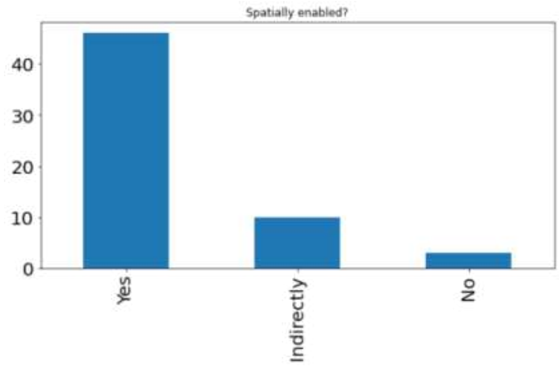
ANNEX II. DATA SURVEY RESULTS

Results available here:
https://github.com/emotional-cities/survey-analysis/blob/master/survey_results.csv

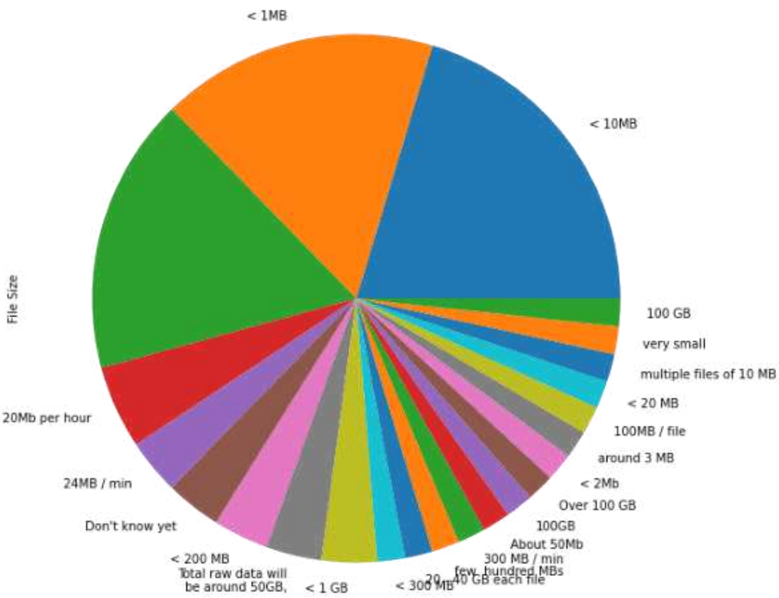
II.1 Personal Identifiable Information



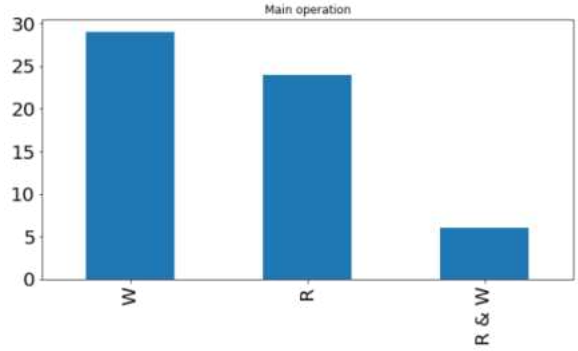
II.2 Spatially Enabled Information



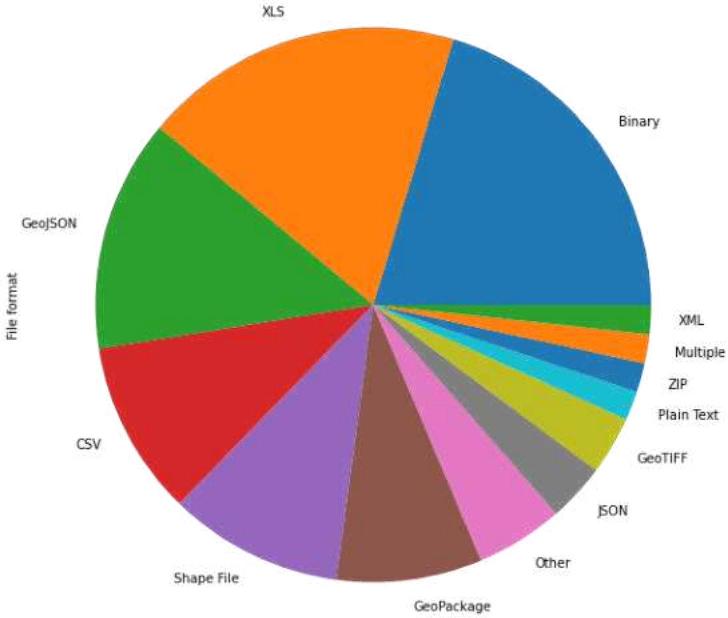
II.3 File Size



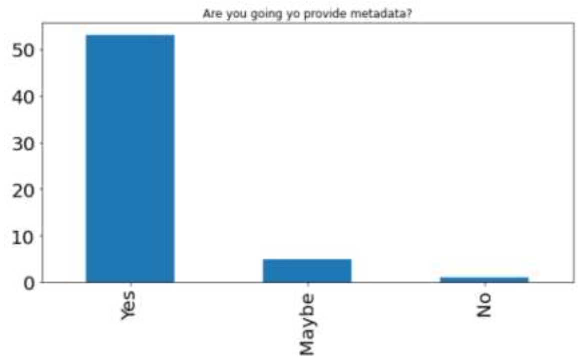
II.4 Main Operation



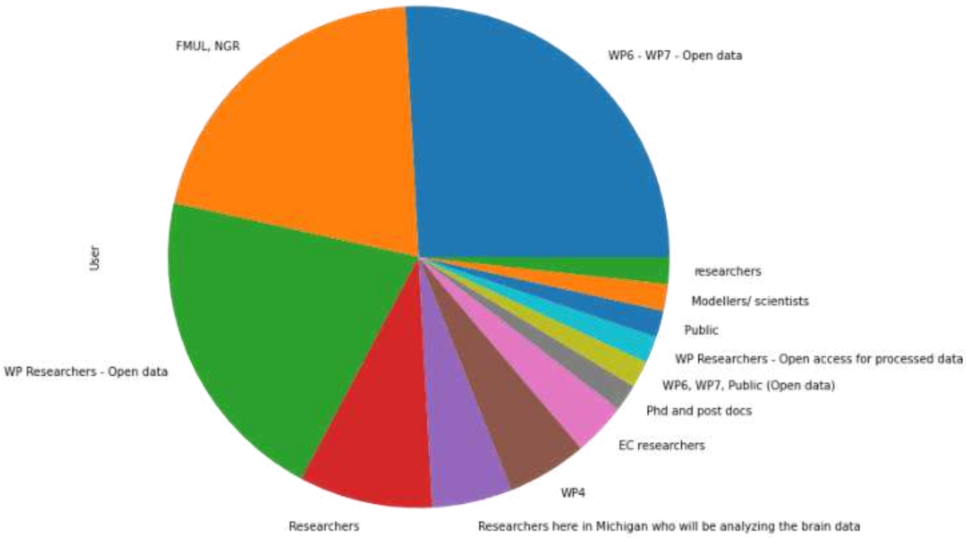
II.5 File Format



II.6 Metadata Provision



II.7 Data Users



eMOTIONAL Cities
Mapping the cities through the senses of those who make them