Long-term sustainable data platform(s), enabling measurements, responses, in real time

Leading partner	IULM University
Type	Deliverable
Dissemination level	Confidential
Deliverable	D5.1
Due date	31 May 2019
Version	1.0

Project	Healthy corridor as drivers of social housing neighbourhoods for the co-creation of social, environmental and marketable NBS
Acronym	URBiNAT - Urban inclusive and innovative nature
	This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 776783

List of Authors and Reviewers

Authors	
Guido Ferilli	IULM
Orfeo Morello	IULM
Reviewers	
Beatriz Caitana	CES
Gonçalo Canto Moniz	CES
Tom Mackenzie	ITEMS
Marcel Cardinali	OWL
Alexandre Ornelas	UC
Cristina Barros	UC

Document history

Date	**Version**	**Author**	**Summary of change**
7 June 2019	1.0	IULM	Version submitted

Table of Contents

Introduction

The purpose of this document is to define needed steps to complete internal processes and achieve its goals according to URBiNAT long-term sustainable data platform, which as a living document will be updated during project development when the need arises in order to improve the internal processes.

This deliverable 5.1 is the first deliverable of Work package 5 on Observatory for Urban Inclusive and Innovative Nature. As such the deliverable supposes the existing information and tools for developing a sustainable data platform being used. Furthermore, it describes a practical methods and approaches to apply in the process of the platform. It also provides a detailed description of how to employ the existing tools that's available from open-source software to help construct a long-term sustainable data platform

1. Data Management Plan

1.1. General definition

URBiNAT will collect/generate data in order to achieve its objectives, namely:

- To promote social cohesion through the activation of Living Labs and an inclusive community of practice.
- To achieve new models of urban development addressing NBS for inclusive urban regeneration through an innovative public space: Healthy Corridors.
- To create value around urban development, of inclusive and innovative nature, through widespread knowledge sharing and replicability, as well as the monitoring, dissemination and market replication of knowledge produced and demonstrated.

1.2. Objectives

The objective of WP5 is to implement an Observatory for Urban Inclusive and Innovative Nature - OURBiNAT. The aim of this observatory is to devise, collect, structure and continuously maintain and upgrade data pertaining to the URBiNAT project, and specifically its use in support of the front-runner and follower cities. The OURBiNAT will ensure efficient and consistent data gathering within the Living Labs of the project on the development of NBS. It will provide content for the improvement of knowledge sharing among partners, and serve to verify the impact of health corridors in citizens' daily lives based on indicators measured within the Living Labs in different climate zones, cultural settings and socio-economic situations. Ultimately, it will serve to ensure the sustainability of URBiNAT's results beyond the project to contribute to the establishment of an EU-wide reference framework for NBS.

In the scope of WP5, the Data Platform in task 5.1 will be implemented to organize data collection to support the implementation of NBS. A specific focus on NBS in terms of well-being, social and economic effects, and local governance will be dedicated through tasks 5.3, 5.4 and 5.5. During task 5.6, action-research, additional quantitative and qualitative data will be collected/generated in order to systematize the project's trajectory, recognition and evidence of the combination of new conceptual approaches such as NBS and participation, NBS and social and solidarity economy.

The goal of the WP5 is to use and process all data formats so as to ensure convergence in the monitoring, evaluation and simulation of local contexts in an interdisciplinary approach. Ad-hoc formats, like .nna, .var, .wtx and .hid will be generated for the specific use of Artificial Neural Network computational tools. In a massive complex data set, the tools of artificial neural networks in particular are able to identify similar patterns (classifications) and hidden connections within the data. As a result, complex relationships between individuals, neighborhood, society, behavior, and environment can be identified and contribute to a significant increase in knowledge about the effects of NBS.

The objectives of the Data Platform, as described in this deliverable, is to organize data collection in support of the implementation of NBS. This process is structured around the following strategies:

- to design standard procedures for the management of data collected or produced by the project and to be adopted by all those involved in the project,
- to develop an easy to use data collection, view and exchange platform that is able to evolve and adapt according to the needs that arise within the project to ensure the sharing of information in a short time within the URBiNAT´s cities and among partners to foster a knowledge based design of the upcoming actions within the living labs in its different stages.
-

Considering that the URBiNAT Data Management Plan - DMP (D1.3) provides general guidelines regarding the processing, use and management of data in the project, the next section 1.3, 1.4 and 1.5, recovers the indications outlined in D1.3.

1.3. Types and formats of generated data

The OURBiNAT is responsible for process\management\following the data generated by the different WPs and to ensure compliance with established parameters.

WP2 will use the existing CAD, GIS, statistical databases and collect new data (qualitative and quantitative) through fieldwork. Existing and surveyed data will be used to produce technical reports - the local diagnostics.

Stage 1 - collection of existing data: GIS, CAD, and statistical databases; spatial, social and economic reports/data; previous experience in participatory processes; previous experiences in NBS design and implementation.

Stage 2 - collection of new data: Includes territorial, social and economic data, namely identifying local needs and expectations, behavioural mapping, as well as the survey of indicators on

participatory planning and governance, social justice and social cohesion, public health and well being, economic opportunities and green jobs.

WP3 will gather personal, social and economic data using various tools (including digital) and methods, applying surveys, participatory training workshops, digital communication support actions and relayed in reports, project presentations, and a participatory handbook for NBS. Data will be generated in the Word, PPT, Excel and PDF formats. Furthermore, some data will be packaged in a form that can be read by statistical programs like SPSS or Max Qda for analysis the qualitative data.

WP4 will generate data related to the impact of the NBS in the social (related to the profiles of the communities and the ways in which the communities use the area), economic (related to the activities produced by the citizens and to the economic profile of the citizens) and environmental context (related to the quality of the water, air and soil); and also urban plan data (related to former urban plans, architecture projects, landscape project) and iconography (drawings, photographs, models, videos, posters, publications). The transformations observed in the territory will produce/generate quantitative and qualitative data. While the data will have different formats - word, excel, pdf, png, jpg, dwg, audio and video - mp3 / mp4 / mov - WP5 will structure the data with a view to facilitating compatibility.

WP5 will process all types and formats of data collected/generated by the others WPs. The goal of the WP5 is to use and process all formats that will converge in the monitoring, evaluation and simulation of the local context in an interdisciplinary approach. Ad-hoc formats, like .nna, .var, .wtx and .hid will be generated for the specific use of Artificial Neural Network computational tools [1]. Definition and the application in the URBiNAT?

WP6 will collect contact details \ personal data (e.g. e-mails, phone, social networks, identification data) of partners, participants and any person interested in the project and its progress, in accordance with URBiNAT DMP (deliverable 1.3), compliance with the Ethics Issues (Codes of Ethics and Conduct, and D1.6 Guide/manual on preliminary ethical guidelines, and communication and reporting procedures and GDPR requirements. Respecting the Data Protection Principles, individually and jointly (proportionality, purpose, assessment, accuracy/reliability, time-limited identifiable information, security of storage and access rights D1.6.). Word/Excel files will be produced with this information.

WP7 will generate qualitative analysis in Word files, including audio recordings, transcriptions, and notes from the interviewer. Initially, desktop research will be conducted to analyse the market of the front-runner and follower cities. In this process word files will be used to gather all necessary information (see local diagnostic document). During the market survey, local community members will be invited in an open workshop that will generate ideas for NBS, local supply demands, and existing biotechnological solutions. The findings from these workshops will be documented in Word and audio files. In an online survey, NBS will be rated according to their popularity among citizens. Participation in this activity will be anonymised. The results will be presented in a comparative statistic that lists the winners of the rating. This data will be archived with a soware that has not been determined yet. Key stakeholders (start-ups, citizens, city administration, corporate clients, NBS experts, researchers and technical experts) will be interviewed in a multi-stakeholders analysis. The in-depth one-to-one interviews are used to

[1] An artificial Neural Network (ANN) is a computational mode. Information that flows through the network affects the structure of the ANN because a neural network changes - or learns, in a sense - based on that input and output. ANNs are considered nonlinear statistical data modeling tools where the complex relationships between inputs and outputs are modeled or patterns are found.

calculate the costs, benefits, roles and responsibilities of each NBS project delivery. These interviews will be stored as anonymised transcriptions and audio files. The participants will also have to sign a consent form. In the end of the empirical research, business cases for the most marketable NBS will be scaled and replicated. This information will be handled as open-source within the URBiNAT project and stored in a Word file. Policy recommendations for replication and scaling the best-practice NBS business cases will be shared with all follower and front-runner cities involved in the process in a PDF document.

WP8 will ensure compliance with ethics requirements set out in the project, including the ethical issues for data collection, protection and conservation, and research processes. This includes guidelines and procedures to be adopted throughout the project and for future reference, including those related to the participation of vulnerable groups (specificities) in activities, guidelines for dealing with risks in fieldwork, as well as incidental findings. The procedures and criteria that will be used to identify/recruit research participants. The design of informed consent\assent templates forms and information sheets (in language and terms intelligible to the participants). Also, includes the demonstration of compliance and/or authorization required under national law for collecting and processing personal data. Since social, ethics and legal problems are expected to arise. WP 1 with requirements from WP8, will continuously monitor all ethics aspects of the project and identify issues that may require additional attention and are able to edit and update the ethics data on the URBiNAT platform.

1.4. Reusability of existing data

In the scope of WP1, information on the project, contact details and partner information will be shared with relevant internal and external interlocutors, through internal and external communication tools, such as Basecamp, the project's website and newsletters.

Concerning WP2, the analysis of existing data will be conducted in the first stage of the local diagnostic's data collection, which is based mainly on the reuse of existing data collected at regional and local level, including statistical data, assessment of existing reports and specific studies accredited by the local authorities. The data collected in the local diagnostic in its second stage, namely at the local level, may determine reuse of additional data not included in stage 1, as well as existing info/data stored at NGOs or private archives. The data collected in this WP may be either open to the public or protected, according to the origin and subject.

WP3 features two stages of data reuse. The first stage aims to build on existing knowledge and data collected during the initial desk research and will assist in understanding what was and/or what is being done in the cities from a participatory and digital communication perspective. These due diligence activities will provide a bedrock of existing knowledge that may confirm, complement or contradict the validity and relevance of initial conclusions on participatory practices or digital platforms. The second stage aims to provide context. Data collected from the initial desk research will assist in the formulating of survey questions to pose to target groups in order to provide contextual clarity, trends and other interesting insights stemming from the survey results.

WP4 will reuse data already produced by the NBS partners, in the context of other projects. Existing and applicable best practices will serve as points of reference for the NBS and HC development.

The data collected in different phases of the project's implementation will be used, in the scope of WP5, for comparing the implementation of the activities/NBS. Most of the data will be secondary, namely in the first stage of the task 2.1 (Local diagnostics). The second stage will add newly developed data.

During the development of WP6, the URBiNAT contact list will be further structured and enlarged, partly by integrating updated contact lists of project partners. Additional contacts may be entered from based on the implementation of new social networks, referring to the website and individuals of the partnership.

The activities developed under WP7 will be based on new data, not reuse data since they will be based on exploratory research.

1.5. The origin and sources of the data

Data originated in all the project's Work Packages derive from many, varied sources.

A formal permission will be asked to the institutions that provide non-public data to the project, namely municipalities and private/public archives.

For WP1, data originates from reports, project findings and results, research production, contact details of partners and database of external contacts for communication and dissemination.

In WP2, data originates mainly from the municipalities and local partners. It will also be collected from the National and European Observatories, as well as from online geographic databases (Google Earth, Bing Maps, Apple maps).

In WP3, data will also be collected mainly from the municipalities and local partners. Additional sources include data from: desk research, surveys and one-on-one dialogue with citizens, activity-related meetings, workshops and other participatory experiences in cities covering non-NBS solutions to gain good practice insight from these experiences. It will also use relevant results from WPs 2, 4 and 7.

In WP4, the data is collected in two stages. The first stage uses the data collected in the Local Diagnostic (WP2) from the municipalities GIS platform, surveys, reports, as well as from the statistical institutions. The second stage will collect quantitative and qualitative data from the workshop activities (handout, video recordings, questionnaires) promoted by WP3 to co-design the NBS solutions and the HC urban plan.

WP5 development is divided in two stages. In the first stage the data are originated by the task 2.1 of the project. For this data, primary and secondary, there will be analysis and evaluations from the tasks of the WP. The first stage will use the data already existing from each city. In the second stage, that occurs during the development of the project, the data will be collected from the cities and the partner involved in various aspects of the urban environment and life, and compared with the existing information.

In WP6, the data originates from: information provided by the partners regarding the development of their activities, social media actions and number of visits to the URBiNAT website.

In WP7, the data originates from one-on-one in-depth interviews.

1.6. Size of data

The data originated in URBiNAT derives from many sources and will be provided in various formats. This includes standard resolution for Word documents, Excel spreadsheets and digital documents for communication and dissemination, and also high-resolution images, videos, economic data, social data, and other data, to support the WPs' activities. In this sense, we estimated it from the fact that similar experiments or similar Gis data collection platforms, the dimension of URBiNAT data shall not exceed 1.5 terabyte and the dimension that has been calculated and let available and protected can be 2 terabyte.

1.7. Ethical issues

Data privacy has been defined as responsibly collecting, using and storing data about people, in line with the expectations of those people, customers, regulations and laws. Data ethics is doing the right thing with data, considering the human impact from all sides, and making decisions based on your brand values. For a balanced 21st century approach, regulations, individual rights, common sense and data ethics needs to be taken seriously (Lawler, 2019).

Local partners shall follow the instructions in the link provided below and make sure local requirements are met in the process.

ethical issues in the participatory processes - link

2. Methodology

This deliverable is the result of the implementation of the platform that will organize data collection to support the NBS implementation. The platform will be constantly updated with information from the Living Labs, to ensure the sharing of information within the URBiNAT's cities.

The primary research method for this study and development was reviewing other similar platforms and technologies that have been used by other institutions and research centers for the publication of data and related metadata, namely Oppla. On the other hand, since the data that is collected by the contributors can be sensitive data that cannot be exported to a cloud environment as a matter of privacy regulations, it is better to use a private server offered by the open source software communities. Therefore URBiNAT observatory platform architecture constitutes two components; DSpace and GeoNode. DSpace is suitable for dealing with terabytes of data and has been used by other similar projects. And GeoNode can offer the possibility to view these layers inside the browser and to be able to edit them if necessary by partners or a group of people.

At the software level, integration is possible thanks to the use of the OAI-PMH protocol. DSpace is able to import data via this protocol. The main idea is to give the possibility to upload geospatial data to GeoNode which integrates an OAI-PMH server to expose the data to our DSpace instance.

2.1. DSpace

DSpace is a platform that allows you to capture items in any format – in text, video, audio, and data. It distributes it over the web. It indexes your work, so users can search and retrieve your items. It preserves your digital work over the long term.

DSpace is suitable for dealing with terabytes of data and has already been used by many research institutes. DSpace provides infrastructure for the management and indexing of any type of document and the publication of metadata.

DSpace has four main characteristics:

1. It is published as a website to give access to all participants who will upload data.
2. Modern RESTful API
3. Built-in search engine
4. OpenURL Support

DSpace component of the URBiNAT project supports FAIR data in trustworthy repositories:

- ❏ **Findable** – Easy to find by both humans and computer systems and based on mandatory description of the metadata that allow the discovery of interesting datasets;
- ❏ **Accessible** – Stored for a long term such that they can be easily accessed and/or downloaded with well-defined licence and access conditions (Open Access when possible), whether at the level of metadata, or at the level of the actual data content;
- ❏ **Interoperable** – Ready to be combined with other datasets by humans as well as computer systems;
- ❏ **Reusable** – Ready to be used for future research and to be processed further using computational methods.

2.2. GeoNode

GeoNode is a web-based application and platform for developing geospatial information systems (GIS) and for deploying spatial data infrastructure (SDI).

GeoNode is built upon a platform of proven open source components including Django, GeoServer, pycsw, OpenLayers and GeoExt. GeoNode implements many Open Geospatial Consortium (OGC) standards, including Web Map Service (WMS), Web Feature Service (WFS), Web Coverage Service (WCS), KML, and Catalogue Service for Web (CSW).

GeoNode development also contributes to the underlying open source projects and software libraries.

Features include:

- **PostGIS Spatial** Databases. PostGIS is an open source software program that adds support for geographic objects to the PostgreSQL object-relational database.
- **GeoServer OGC**(Open Geospatial Consortium) Services. GeoServer is an open-source server that allows users to share, process and edit geospatial data
- **pycsw CSW** Metadata Catalogue. pycsw is an OGC CSW server implementation written in Python. pycsw fully implements the OpenGIS Catalogue Service Implementation Specification [Catalogue Service for the Web].
- **Geospatial Python** libraries
- **OpenLayers** and GeoExt Web Mapping Libraries

GeoNode supports multiple points for application integration.

Features include:

- OGC-compliant web services
- GeoServer REST API
- GeoNode search and REST APIs

3. FAIR Data

3.1. Making data findable, including provisions for metadata

3.1.1. Data acquisition

In the scope of data acquisition, mostly data is obtained by WPs and URBiNAT partners. The other significant source of data is citizens contribution. To simplify the process a website is available to ease people collaboration.

3.1.2. Data transformation

In order to perform defined standard on gathered data, there is a stage that the platform does minor changes in the correct conditions for operation and use.

In order to publish metadata, it is essential to have OAI-PMH (The Open Archives Initiative Protocol for Metadata Harvesting) de facto standard in context of repositories. The platform harvests the documents and data using OAI-PMH.

The observatory and in particular the DSpace component will use the OAI protocol to communicate with the GeoNode geospatial data management module. DSpace will import the metadata about geospatial data layers submitted by partners in GeoNode.

The partners can upload and make content available via standard OGC protocols such as Web Map Service (WMS) and Web Feature Service (WFS).

The data is available for browsing, searching, styling, and processing to generate maps which can be shared publicly or restricted to specific users only.

Supported upload formats include shapefile, GeoTIFF, KML and CSV. In addition, it is possible to connect to existing external spatial databases and services.

For name conventions on how to publish data in the platform, we use Linked Data. Linked data is a method for publishing structured data using vocabularies. The structured data can be indexed more easily and makes data findable and optimized for re-use.

Linked data for the repository can be seen as much wider supported OAI-PMH interface with better integration of foreign data and concepts (links, words and ontologies)

The Linked Data principles

- ❏ Use URIs(Uniform Resource Identifier) as names for things
- ❏ Use HTTP URIs so that people can look up those things
- ❏ When someone looks up a URI, provide useful information, using the standards
- ❏ Include links to other URIs, so the can discover more things

It means that linked data makes extensive use of URIs. And URI is unique. URIs are used as names and identifiers non as locators only. If a URI is requested, it delivers data(use content negotiation). The Linked Data uses RDF(Resource Description Framework) as data model and one of it's representations (RDF, XML, …). Link is the glue of the web, create links with our data

3.1.3. Ontology

Ontology is a model of the common relations and objects that are generally applicable across a wide range of domain ontologies. Employing ontologies in the URBiNAT platform could pave the way for making data interpretable for machine and much simpler to find.

RDF(Resource Description Framework) is a data model which makes extensive use of URI and links. Vocabularies and ontologies express semantic of terms and relations between terms used in the Semantic Web. Ontologies and vocabularies are published as Linked Data and following the linked Data principles as RDF. RDF-Schema(RDFS) and Web Ontology Language are used to describe vocabularies and ontologies.

3.1.4. What does the repository store?

The repository ingest workflow involves a multi-stage process that starts with receipt of heterogeneous data collections from a variety of data. The repository will store and publish digital objects and their descriptive metadata. Metadata information describes the digital content in the repository to ensure its long-term accessibility.

- ❏ Provides information about quality aspects or issues of the created object along with its access privileges/rights
- ❏ Ensures smooth data management.

Most of the content will be related to geographic data or data that has a georeferencing. Digital geospatial data consists of digital information that identifies the geographic location and characteristics of natural or constructed features and boundaries on the earth. Such data resources include geographic information systems (GIS) data sets, digitized maps, remote sensing data resources such as digital aerial photography, and tabular data that are tied to specific locations.

Digital objects are comprised of several files such as documents(PDF, text,...), Tables(CSV, Excell,...), Image, Video and GIS data, etc. And descriptive metadata include structured format as key-value for instance the data.title, data.contributor, data.author.

3.2. Making data openly accessible

Data is accessible by the website regarding to the user's level of access. A registered user has no permissions on the system until a DSpace administrator assigns permissions. DSpace create automatic groups; Administrator and Anonymous. Administrators can run and curation activities. Administrators can create groups and collections to manage user working on the system (organization/ role) for instance Archivist, Researcher, WPs, etc. Then when assigning rights to a collection or item, we assign the group rather than individuals. It is the best practice to manage restrictions and access levels.

A short manual guideline will be included for partners to ease the way of accessing data.

In order to make possible the inclusion of relevant software, we tried to use open source software and tools. As a result, any output of the platform source code will itself be open source .

In addition to the users and groups permission discussed above, we can add a new policy to restrict an item or collection for an individual user or a group.

Since we use Semantic Web features to build the platform, the data is machine-readable. The Semantic Web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation.

According to each specific storage use, data and associated metadata, documentation and code will be deposited on the Basecamp platform, URBiNAT's website, the Observatory's server, and in a certified repository for open access to research data. Appropriate arrangements are under

exploration with Oppla. Oppla is a virtual hub where the latest thinking on nature-based solutions is brought together from across Europe.

All content on the Oppla website will be openly accessible. The API will be protected by a system of API keys (although this may only apply to editing), but keys will be easy to obtain, and available to all legitimate consumers of the content. Access to datasets themselves will be controlled by the dataset providers, but they will be encouraged to make the data as open as possible, and to specify any restrictions in the metadata.

In other words, the datasets will be stored online on a semantic web. All the data will be findable both on the web and via API. All contents will be openly accessible. Only widely recognised open formats will be implemented to maximize interoperability. This will allow others to use from remote data and metadata. The use of widely used standards will be encouraged and controlled by the dataset providers.

3.3. Making data interpretable

URBiNAT Data are geodata, scientific observations and measurements that, once analyzed and interpreted, can be developed into evidence to address a question. To succeed in the task assigned to the project it is required that a vast amount of data is collected in one form or another.

The data analysis will help the user who will use the system to determine the best course of action, and make their decision on the interpretation of the information. Because URBiNAT partners build on their own work and the work of others, it is important that they are systematic and consistent in their data collection methods and make detailed records so that others can see and use the data they collect.

To get help in achieving the goal of standardization, interpretation and data organization, the system will make use of DSpace-CRIS. DSpace-CRIS is an “extended” version of DSpace with a powerful and flexible data model to describe not just publications, but all the entities that populate the research environment and their meaningful links. Creation of CRIS objects as part of the submission process allows users to create new objects (persons, projects, organizations, events, etc.) on demand as needed during the submission of a DSpace item. It can even be used to provide a workflow around the collection of data.

DSpace-CRIS is free, open source, compliant with open source standards and provides a sustainable and effective tool to manage research information such as researcher’s profile, research outputs, metrics, reports and statistics.

3.3.1. DSpace-CKAN (The Comprehensive Knowledge Archive Network)

CKAN, http://ckan.org/, a fully-featured, mature, open-source data management solution. CKAN provides a streamlined way to make data discoverable and presentable.

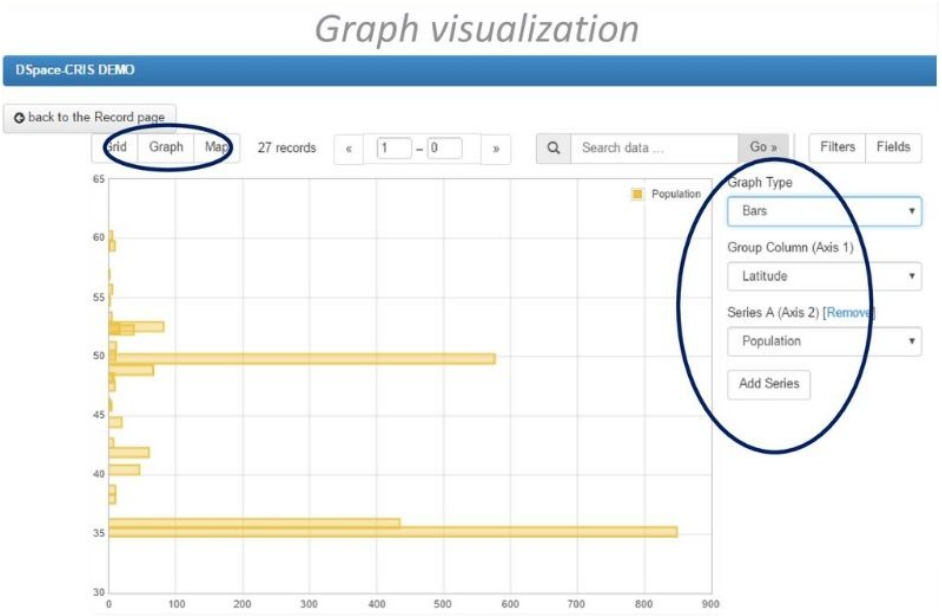
The DSpace-CKAN module running inside DSpace-CRIS allows to connect the publications with the datasets and better describe the context of realization of the dataset by creating detailed records on the equipment used, the related services, the projects concerned, etc.

In the URBiNAT project, the DSpace-CKAN module focuses on issues that should be addressed when presenting numerical data for different audiences and highlights ways that will maximise the impact of such data and ensure that they are easy to read and interpret. It ensures that the audience understands the argument and is not left struggling to interpret data that are poorly presented (raw format) or in an inappropriate format.

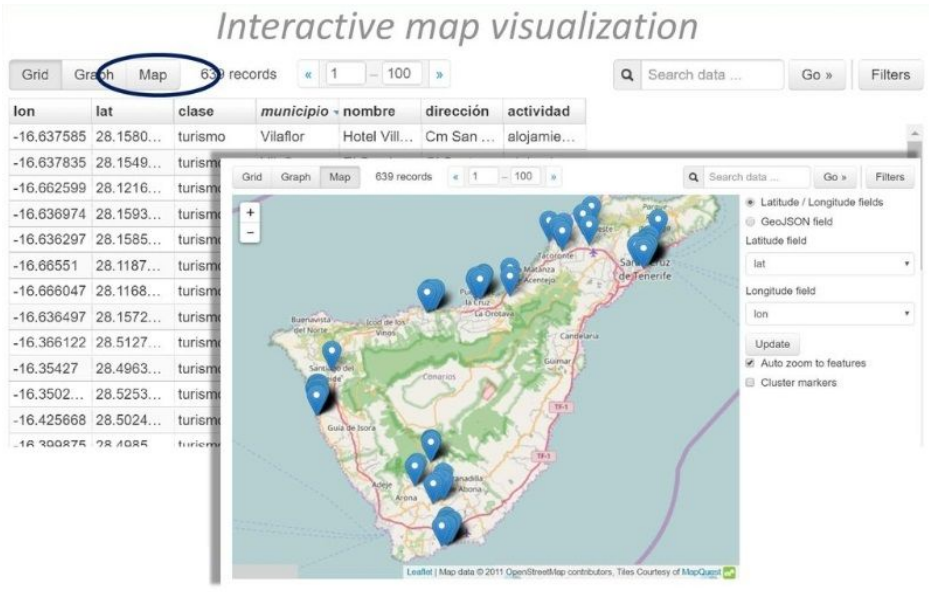
One of the main advantages of the Module is that it allows the management of dataset access via DSpace, without the need of getting into a CKAN instance. The CKAN datastore API is proxied by the DSpace-CRIS, enhancing the access condition sizes set on the original DSpace bitstream (e.g.: the preview function is available under the same conditions of the uploaded bitstream: open access, embargo, etc.).

If structured data is uploaded or linked to CKAN as a .csv or Excel table, the DataStore loads it into a database, allowing CKAN to give a range of ways to view and process the data. Initially the data is displayed as a table.

The user can sort the data on particular columns, filter or facet by values, or hide the columns entirely. The data can also be displayed on a graph, choosing the variables on the axes and comparing a number of variables by graphing them together on the same y-axis.



A map view can be associated with the same data, being geo-referenced



3.4. Increase data re-use

The observatory sends data to Oppla platform and do not allow Oppla to take data from URBiNAT platform. The observatory only functions as a hub, concentrator of the data collected by the URBiNAT partners. The URBiNAT provides a service that sends specified data, DSpace URL and some metadata to the Oppla platform.

The availability of data for re-use depends on the specifics data and methods used. The same applies to data quality assurance processes.

The time during which NBS data will remain re-usable depends on the partners offering these solutions, to be decided when implementing the Healthy Corridors in front-runner cities.

4. Integration

4.1. Architectural Drivers

Table 1 summarizes the identified software that can provide required capabilities of URBiNAT.

Project	**URL**	**Description**
CKAN	http://ckan.org/	Open source project, maintained by Open Knowledge Foundation
GeoNode	http://geonode.org/	platform for the management and publication of geospatial data; mature open source project
DSpace	http://www.dspace.org/	turnkey institutional repository application; mature open source project, maintained under the stewardship of DuraSpace, a not- for-profit organization

Table 1. Open source frameworks reviewed for project use.

A major consideration in the evaluation is the long-term sustainable data platform for URBiNAT. Building on an existing, active and widely used open-source project is considered a major asset to assure long term viability of the application.

1. Other significant factors that were considered of high value in the development framework include: adaptability of the user interface to be compatible with an independently developed user experience concept for URBiNAT users;
2. ease of extensibility, with a plug-in architecture that allows the addition of functionality without having to modify the core codebase;
3. support for geographic data and map-based search and data browsing;
4. support for administrative activities like user management, access control, and activity logging.

4.2. Schematic Representation

Figure 1 depicts how we will realize DSpace-CKAN features and GeoServer. As can be seen GeoServer and CKAN access a single Postgres database, and an URBiNAT extension manages the GeoServer instance according to the data that is uploaded to DSpace-CKAN. Other URBiNAT extensions described above are represented in a single box labeled “Various other URBiNAT

extensions". The full text indexing extension is represented separately because it will use SOLR as an additional external component.

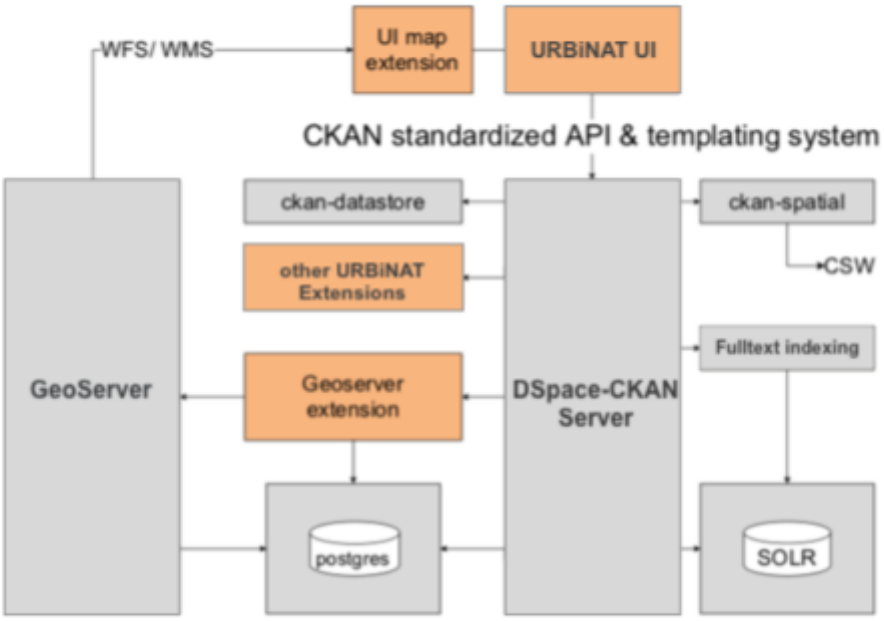


Figure 1: URBiNAT Component Architecture.

5. Data Security

5.1. Provisions for data security

It is important that the repository of choice is trustworthy. Such repositories guarantee to safeguard your data and make sure that the data will remain findable, accessible and reusable for either an indefinite or a specific period. Important criteria for the trustworthiness of repositories are that the data can be found via a persistent identifier such as a DOI (Digital Object Identifier), URN (Uniform Resource Names) or Handle (handle is an abstract reference to a resource, handles contain the information necessary to locate, access, and otherwise make use of the resources).

Having a persistent identifier means that the link to the data is maintained, even when something changes in the location of the data (i.e., if the web-address changes). The trusted repository also guarantees a standard way of describing the data (with metadata and additional documentation), and that there is a license specifying the access conditions for sharing and reuse.

5.2. Storage in certified repositories

From the hardware point of view, the data itself is stored in a Network Attached Storage (NAS) provides an array of storage devices to provide storage and act as a file server. The model used is QNAP TS-432XU-RP with the following characteristics

- ❏ Snapshot Support
- ❏ RAID (Redundant Array of Inexpensive Disks) Support
- ❏ AES 256-bit encryption (Hardware encryption)
- ❏ Two-step verification options
- ❏ IP and protocol access control

The hardware that makes the system work is located in the IULM University server room

5.2.1. Procedures for data conservation (from D1.6).

All collected data, including sensitive information (from interviews, ICT digital platform and observation), will be stored in wiki, cloud and external drivers to be secured by passwords and kept by CES. Data will be stored for a period of 5 years after the end of the project, to safeguard the possibility of external audits and the use of URBiNAT's data in future publications or as a basis for future research. Since collected data will be kept anonymous, the coding key linking participants to pseudonyms will be destroyed 5 years after the project is concluded.

Participants have the right to:

- ❏ access, rectify or correct and complete personal data, including by means of providing a supplementary statement.
- ❏ erasure ("right to be forgotten") from URBiNAT datasets, if they wish. Conditions for erasure of personal data without undue delay are outlined in article 17 of the GDPR.
- ❏ restriction of processing. Data will continue to be stored in URBiNAT's servers, but processing will happen only when: there is subject consent to do so; if it is necessary for acting or defending the project in a legal claim; if the protection of the rights and interests of another human being or legal person is at stake; for demonstrated reasons of unavoidable public interest.
- ❏ data portability, meaning that data subject has the right to receive his or her personal data, any time, in a structured, commonly used and machine-readable format for purposes of data transfer.
- ❏ object, on grounds relating to his or her particular situation, at any time to processing of personal data concerning him or her, including profiling based on those provisions.

In case a data breach is likely to result in a risk for the rights covered by URBiNAT's policy, participants will be notified in a proper timing.

Bibliography

CKAN Integration, available at: https://wiki.duraspace.org/display/DSPACECRIS/CKAN+integration (retrieved: May 29, 2019)

Lawler, B. (2019), “Five Global Trends in Data Ethics and Privacy in 2019”, *Looker,* available at: https://looker.com/blog/big-data-ethics-privacy (retrieved: May 24, 2019).

Oppla welcome pack, available at: https://oppla.eu/sites/default/files/docs/Oppla-WelcomePack_0.pdf (retrieved: May 29, 2019)