cities2030

D6.1 S2CP design and reference architecture

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101000640

Deliverable D6.1

Prepared by P20 | Edited by P19, P20, P21, P35, P37 | Checked and reviewed by ExeComm | Approved by P20

Rev 1.4 - September 2022

Document information

Key information	**Data**
Project reference number	101000640
Project acronym	'Cities2030'
Project title	'Co-creating resilient and sustainable food systems towards FOOD2030'
Project start date	October 1st, 2020
Duration	48 months
Project Coordinator	Mr Nicola CAMATTI
Project website	cities2030.eu
Work package (WP)	WP6
WP leader	UPM (P20)
WP co-leader	WIT (21)
Contributors and authors	SINNO (P19): Francisco Royo, Daniel Molina UPM (P20): Borja Bordel, Ramón Alcarria, Jesús Sánchez López, Tomás Robles, Miguel Manso, Diego Martín, Ionut Cira WIT (P21): Micheal Crotty ITC (P30): Tomaz Zadravec, Tomaz Bokan, Daniel Copot, Sasa Straus Uni.Lu (P35): Aurel Machalek, Marharyta Aleksandrova, Latif Ladid PRIM (P37): Pedro Caridade
Peer reviewers	SINNO (P19; Uni.Lu (P35); ITC (P30); WIT (P21)
Peer approval date / revision	07/09/2022
Document type [1]	R
Document/file name	Cities2030_D6.1_S2CP_Rev1.4
Document title	D6.1 – S2CP design and reference architecture
Deliverable number	D6.1
Abstract	The objective of this document is to present in a consistent way the S2CP design and reference architecture, as well as the set of technological

[1] R: document, report (excluding the periodic and final reports); DEM: Demonstrator, pilot, prototype, plan designs; DEC: websites, patents filing, press & media actions, videos, etc.; OTHER: software, technical diagram, etc.

	requirements extracted from the interaction with the use cases and user requirements of Task 3.7.
Project delivery date	September 30th, 2022
PM approval date/version	29/09/2022 | Version 1.4
Submission date	September 29th, 2022
For public dissemination Yes/NO	YES

Disclaimer

Copyright notice

Acknowledgement

The project 'Co-creating resilient and sustainable food systems towards food2030' (Cities2030) has received funding from the European Union Horizon 2020 programme under grant number 101000640

Citation

Be so kind as to cite this work as:

'Co-creating resilient and sustainable food systems towards food2030', Cities2030 Consortium, Guidelines for a data-driven CRFS management platform – Update Cities2030 consortium under the supervision of the project's coordinator.

Legislation

Legislation H2020 Framework Programme – Regulation (EU) No 1291/2013 of the European Parliament and of the Council of 11 December 2013 establishing Horizon 2020 – The Framework Programme for Research and Innovation (2014-2020) (OJ 347, 20.12.2013, p. 104).

Euratom Research and Training Programme (2014-2018) – Council Regulation (Euratom) No 1314/2013 of 16 December 2013 on the Research and Training Programme of the European Atomic Energy Community (2014-2018) complementing the Horizon 2020 – The Framework Programme for Research and Innovation (OJ L 347, 20.12.2013, p. 948).

H2020 Specific Programme – Council Decision 2013/743/EU of 3 December 2013 establishing the Specific Programme Implementing Horizon 2020 – The Framework Programme for Research and Innovation (2014-2020) (OJ L 347, 20.12.2013, p. 965).

Rules for Participation (RfP) – Regulation (EU) No 1290/2013 of the European Parliament and of the Council of 11 of December 2013 laying down the rules for the participation and dissemination in Horizon 2020 – the Framework Programme for Research and Innovation (2014-2020) (OJ L 347, 20.12.2013, p.81).

Financial Regulation (FR) – Regulation (EC, Euratom) No 966/2012 of the European Parliament and of the Council of 25 October 2012 on the financial rules applicable to the general budget of the European Union (OJ L 298, 26.10.2012, p.1).

Rules of Application (RAP) – Commission Regulation (EC, Euratom) No 1268/2012 of 29 October 2012 on the rules of application of l Regulation (EC, Euratom) No 966/2012 of the European Parliament and of the Council on the financial rules applicable to the general budget of the Union (OJ L 298, 26.10.2012, p.1).

Document history

Version (v)	**Date** [2]	**Comment**	**Author**	**Status** [3]
0.1	1/11/2020	Initial structure	UPM	-
0.4	1/06/2021	Reference architecture landscape first draft	All partners	Drafted
0.6	5/12/2021	Adding Combined development methodology (Section 6)	UPM	Drafted
0.7	07/03/2022	Initial component description (Section 5.2)	All partners	Drafted
1.0	15/05/2022	Adding Sections 2 and 3	UPM	Drafted
1.1	11/07/2022	Deliverable reformat. Modified document structure	UPM	Drafted
1.2	30/07/2022	Adding Sections 7 and 8. Full deliverable revision.	UPM	Drafted
1.3	15/08/2022	Revision of SINNO, Uni.Lu, ITC, WIT	All partners	Drafted
1.4	07/09/2022	Final revision by WP leaders (UPM) before PMO validation	UPM	Completed
1.4	29/09/2022	PMO layout check, validation and submission	P01UNIVE	Completed

[2] As per the project's cloud storage if applicable, or per email date if applicable

[3] Drafted, completed or validated

			+ P02EPC	

List of Figures

List of Tables

Table of Contents

Contents

Glossary and abbreviations

CRFS	City Region Food Systems
AI	Artificial Intelligence
ICT	Information and Communication Technologies
KPI	Key Performance Indicators
UFSE	Urban Food Systems and Ecosystems
WP6	Work Package 6
D6.1	Deliverable 6.1
FAO	Food and Agriculture Organization
IoT	Internet of Things
IPv6	Internet Protocol version 6 (IPv6) RFC 8200. https://www.rfc-editor.org/info/rfc8200
NLP	Natural Language Processing
SSRI	Social Space for Research and Innovation
MAA	Multi-Actor Approach
SOCS	Service-based Open Collaboration Space
BDVA	Big Data Value Association
DSBA	Data Spaces Business Alliance
WiFi	Wireless Fidelity
LoRa	Long Range (Wide Area)
GFDx	Global Fortification Data Exchange
NER	Named Entity Recognition
ML	Machine Learning
CL	Conformal Learning
SFSC	Short Food Supply Chain

BC	Blockchain
IPFS	InterPlanetary File System

1 Introduction

In this first section we analyze the context of Task 6.1 and describe its structure and main lines of action globally. We focus especially on the relationship of document D6.1 with the rest of the work packages and tasks of the Cities2030 project. This document is the main result of the working group in charge of defining the use cases, technological components and development methodologies necessary for the Cities2030 project to achieve the expected impact.

1.1 Short description of the CITIES 2030 project

The main goal of Cities2030 is to create a future proof and effective UFSE via a connected structure centered in the citizen, built on trust, with partners encompassing the entire UFSE. Cities2030 commit to work towards the transformation and restructuring of the way systems produce, transport and supply, recycle and reuse food in the 21st century. Cities2030 vision is to connect short food supply chains, gathering cities and regions, consumers, strategic and complement industry partners, the civil society, promising start-ups and enterprises, innovators and visionary thinkers, leading universities and research across the vast diversity of disciplines addressing UFSE, including food science, social science and big data.

1.2 Short description of the WP6 Package

This work package will gather, design, and develop the main components and technological tools to establish a data-driven CRFS management platform for data collection, analysis and representation in multiple interfaces. An initial requirement acquisition will lead to the proposal of a common technical architecture for Cities2030, for with supporting data set will be incorporated to be considered for data analysis and representation. Particularly, a service-based open collaboration space will be incorporated, to be used by Cities2030 participants to improve their multistakeholder dialogue processes. In this space, blockchain technology will be employed to provide some proof of concepts of token-based monetization processes, and reflect multi-stakeholder interaction in a reliable and transparent way. Documentation and software repositories will be available for policy labs and living labs to develop their own solutions with assistance from WP6.

1.3 Purpose of D6.1 report

The objective of this document is to present in a consistent way the S2CP design and reference architecture, as well as the set of technological requirements extracted from the interaction with the use cases and user requirements of Task 3.7.

Along with the design of the architecture of the S2CP platform, the proposed API interfaces to be developed in the rest of the WP6 tasks will be described.

In order to carry out each and every one of these actions in an organized manner, a specific Gantt chart is proposed for the WP, as well as an appropriate development methodology integrated with the progress foreseen in the rest of the WPs.

1.4 Relation of the D6.1 Report with other WPs and deliverables

This deliverable has a direct relationship with WP3 and in particular with task T3.7. The link between tasks T3.7 "Data-driven CRFS management system ideation" and T6.1 is the establishment of final user requirements and development guidelines that has become technical requirements to be taken into account in the development of the architecture and that has determined list of components to be developed, following a co-creation methodology with WP4 and WP5. Additionally, components and tools to support the dissemination of innovations and good practices collected in task T3.6 "Examination of the innovation landscape and living lab prototype" are considered in this document.

Regarding WP2, there has been continuous communication between WP2 and WP6 teams, in order to promote that the specifications and design of the WP6 components detailed in D6.1 comply with ethical, RRI and gender-based aspects aligned with the European Union idea of an inclusive, innovative and reflective society.

This task also is very related to WP4 and WP5. Development methodology proposed in task T6.1 is closely integrated and related to working methodologies in WP4 and WP5. In that way, this document is co-created together with task leader and work package leaders in tasks T4.3 "Deployment and multiplication" and T5.2 "Facilitators and guidelines for innovation and pilots developments".

Finally, regarding WP7, many tools and components identified in T6.1 and to be developed in WP6 will be employed as dissemination instruments and solutions to communicate the project results to stakeholders and society.

1.5 D6.1 structure

The structure of this document is as follows:

In Section 2, we describe the creation of a specific working group to discuss and carry out the activities foreseen in Task T6.1 and WP6 in general. The methodology used to execute the actions by this working group is also described in this section. As main results, the Gantt chart planned for the work within WP6 is introduced.

Section 3 focuses on capturing and analyzing the requirements of the S2CP platform, both from internal users of the project itself (mainly policy labs and living labs), and from the stakeholders of the Cities2030 project (provided through the task T3.7 and deliverable D3.8).

In Section 4, the most technologically relevant architectures are reviewed, and from which requirements and proposals that are applicable to the Cities2030 ecosystem and the S2CP platform can be extracted. This includes both Internet of Things systems, cloud, interoperability frameworks, platforms from previous European projects and, given the importance that Blockchain-type solutions have in the context of the Cities2030 project, solutions based on this innovative technology are also reviewed.

In light of all the previous discussions, Section 5 proposes a specific architecture for the S2CP platform. In this section, each of the components that are part of said architecture and platform are also succinctly described.

The architecture also includes the proposal of different open APIs, according to the requirements of the Cities2030 project.

In Section 6, the methodology and development designed by the working group is presented to successfully complete the architecture planned in Section 5. The relationship of this methodology with the proposals in other work packages is also analyzed.

Section 7 presents a general summary of the status of the work, and the evolution of the methodology, as of the date of delivery of this deliverable (M24). This section also includes an analysis of the status of the different KPIs that affect WP6.

Finally, Section 8 presents the global conclusions of T6.1 at the date of delivery of this document, as well as a brief forecast of the next steps to be addressed from M24 to M48.

1.6 Roles and responsibilities

Lead partner UPM (P20) coordinates the activities, provide guidance, steer implementation and secure alignment, implement activities to deliver planned outcomes.

The rest of task participants: WIT(P21), Uni.lu (P35), ITC (P30), PRIM (P37) and SINNO(P19) develop the task simultaneously at EU level and beyond and contribute to the focus group.

2 Group and work methodology

WP6 is structured into various tasks, the results of which must converge in the design and implementation of a technological platform for the Cities2030 project, the Single-Click CRFS Platform or S2CP. In order to coordinate the actions of the different tasks, a working group is defined in which all the organizations that participate in WP6 are represented.

The result of these works is a cross-cutting development methodology for all the tasks of the WP, as well as a design for the S2CP platform that adapts to the activities of all the tasks and a set of open APIs that allow the integration of modules external to the project with the Cities2030 ecosystem. The necessary management tools are also developed for the coordination of the different tasks in a fluid and coherent way.

This section presents the work methodology of this group, which should allow the capture of the necessary requirements for the design of the platform both from the Cities2030's labs, as well as from external stakeholders, as well as the implementation of the methodology of development and deployment of the necessary impact monitoring mechanisms to assess the performance of the WP and the way the works enable the expected KPIs to be met.

2.1 Composition and members

In order to ensure that the work developed reflects the requirements, needs and limitations of each of the tasks involved in the WP, all WP members were invited to participate in this group. In this way, the working group was made up of representatives from 6 organizations: WIT(P21), Uni.lu (P35), PRIM (P37), SINNO(P19), ITC (P30) and UPM (P20). UPM, as WP leader, acts as president and coordinator of this group.

In addition, each one of these members mainly represents one of the technical tasks described within the WP, in such a way that we guarantee a balanced representation of all the activities in the group. Each representative member is responsible for ensuring that the activities foreseen in the task are adequately reflected and covered both in the general design of the platform and in the proposed development methodology. In addition, it is their responsibility to ensure that the task's work schedule is properly integrated into the overall Gantt chart provided for the WP. Table 1 shows the assignment of each one of these members to each of the tasks in WP6.

Table 1. Partner responsible for WP alignment per task

Task	6.2	6.3	6.4	6.5
Responsible for WP alignment	WIT	UPM (SINNO, ITC)	Uni.lu	PRIM

2.2 Work methodology

The objective of the working group is to create a set of mechanisms that allow the coordination of the different tasks within the WP, and that enable the participating organizations to implement the necessary developments to put the S2CP platform into operation, in accordance with the requirements expressed by both internal users and users external to the Cities2030 project.

The working group meets every two weeks on a regular basis, to have a constant exchange of knowledge and allow progress and impact to be monitored. WP leader proposes the agenda before each meeting. All materials are available on the Correlate project management platform, including the recording of meetings for those members who are unable to attend for any reason.

To achieve the implementation of the aforementioned monitoring, control, requirements capture and development management mechanisms, the following work methodology was followed:

1. The first phase consisted of capturing requirements from potential users of the S2CP platform, both internal and external to the Cities2030 project. The results of this capture of requirements can be consulted in Section 3 of this document. To identify potential users, we mainly use the results obtained in T3.7, which can be consulted in deliverable D3.8. To these potential users we added the different laboratories that have been established in the context of the Cities2030 project during the first months of operation. It is necessary to know not only the requirements of each individual user, but also of the different laboratories as an autonomous and independent work team, and with specific objectives to meet. This phase of capturing requirements must also be integrated in a coherent manner into the component development methodology, as it is the initial input of information that must allow the technological work to start. Section 6 shows how this integration has been carried out. Although the analysis of document D3.8 can be carried out in a relatively simple way and only requires the participation of organizations internal to the project, the capture of information related to the different laboratories is more complex, since on some occasions they are involved both members of the Cities2030 project, and other interested external agents. Therefore, this phase was organized in two stages:
    a. In a first stage, each of the participants in the working group reviewed the proposals included in D3.8. considering the conclusions of the WP3 focus group that are reflected in said deliverable. Each member of the working group defined a set of requirements and use cases that must be reflected in the platform. As reflected in said document, these conclusions and

recommendations already include the vision of both internal and external end users of the project. From the study of these requirements and use cases, during the architecture design phase, it should be possible to extract a set of technologies, functionalities, services or modules that must be present in the S2CP platform. In Section 3.1, the results of this study can be consulted.

   b. Once the catalog of functionalities for S2CP platform has been prepared by indication and recommendation of the WP3 focus group, a working document is composed to inform about the identified functionalities to policy and living labs of the Cities2030 project. In this way, to capture the requirements derived from Cities2030 project labs, bilateral meetings are organized in a small group format, attended by a representative from WP6 (normally the WP leader) and a representative of each of the participating organizations in each one of the Labs. In these meetings, the conclusions reached from the study of D3.8 are presented, considering also the specific objectives that each of the laboratories have set for the improvement of the CRFS and their needs are recorded in terms of support for the different experiments and design of policies that must be carried out. From each of these meetings, it should be possible to complete a descriptive sheet for each Lab indicating the objectives, variables to be monitored and the functionalities or use cases that they expect to need for the implementation of their experiments and policies. Section 3.2 collects all this information.

2. In the second phase of the work methodology, a detailed study of the state of the art is carried out in relation to reference architectures for service provision platforms, similar to the S2CP platform that is intended to be developed in the context of the Cities2030 project. This is an ambitious work, given the large number of proposals and reference architectures that have been reported in the scientific literature and at a commercial level in the last 10 years. In order to speed up the work, this detailed study, whose results can be consulted in Section 4, was developed in two stages:
   a. In a first stage, 5 teams were created within the working group, to analyze the state of the art of the last 10 years with respect to the reference architectures of a technological area. The 5 teams focused on the Internet of Things, Cloud systems, Interoperability Frameworks, Blockchain Systems and specific architectures previously reported in other European and international projects. After carrying out each study, the different teams develop a report summarizing their findings.
   b. The various reports are presented to the work team in a WP6 regular meeting, at the request of the group that developed the report. In this meeting, impressions and comments for improvement are exchanged. The team in charge of the report improves its proposal according to the comments received, being able after this to integrate its contribution in the deliverable. In Section 4 all the reports made in their final form can be consulted.
3. Considering the requirements extracted from WP3, the needs and use cases detected from the bilateral small group meetings with the different Labs and taking into account the state of the art regarding reference architectures for similar platforms to the S2CP ecosystem, the third phase focuses on designing its own reference architecture for the S2CP platform. This architecture must allow satisfying all the previously detected needs and use cases, in addition to being technologically viable as it conforms to the state of the art. The design is carried out collectively among all the members of the working group, who must guarantee that the proposals meet the requirements and commitments of each of the tasks they represent, in addition to being adjusted to the technological arena, reviewed in the second phase. The architecture should identify the different modules, services

and components that must be developed, the task within which the works must be considered, and an organization responsible for each of the functionalities. Section 5 shows the result obtained after completing this phase.

4. The fourth phase focuses on the design of a development methodology that allows the implementation of the functionalities, modules and services committed to in the reference architecture in a reliable and safe manner in the foreseen times, adjusted to the real needs of Labs and users. It is very important that the development methodology implemented in WP6 be adjusted and coordinated with the work methodologies of the Labs themselves (in order to be able to collect the improvements and new needs that arise as they advance in their experiments and policies). To achieve this coordination, monthly meetings are established between the leaders of the WP4, WP5 and WP6. After a period not exceeding 6 months, the development methodology was drawn up and prepared for discussion within the WP6 working group. Section 6 collects the result of this phase, with a detailed description of the methodology finally implemented for the development of the S2CP platform.
5. The fifth and last phase consists of continuous monitoring of the impact, the technological advances achieved, the modifications that need to be applied and the risks that may appear during development. To carry out this monitoring, a pack of tools were developed, which includes a Gantt chart and an analysis of the commitments acquired in the different tasks and their degree of compliance in the S2CP platform. Section 2.3 presents the main elements included in this toolkit. Also, during this phase, periodic reporting mechanisms are implemented, so that the different organizations report on their contribution to the performance of the KPIs included and proposed in the framework of WP3. This report is periodically updated in the meetings of the working group, adding the partial results reported by each of the organizations to represent final achievements. In Section 7 it can be seen the result of this monitoring process up to the delivery date of D6.1 (M24).

The following figure summarizes the different phases of the work methodology and their relationship with the various sections that can be found in this document.

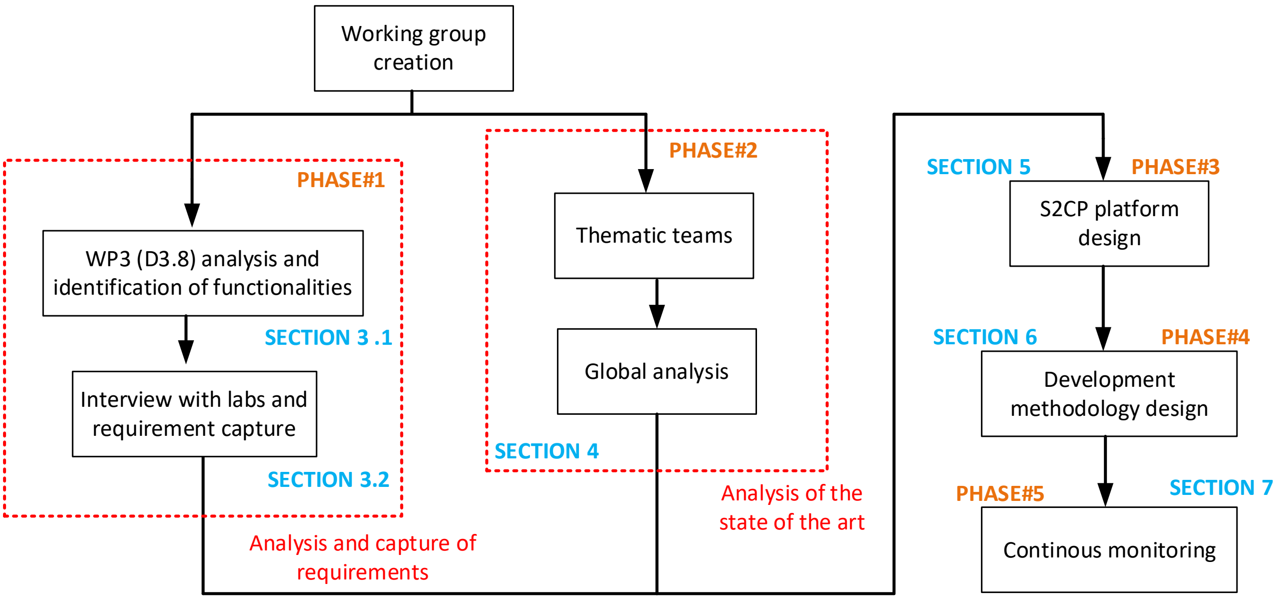


Figure 1. WP6 work methodology diagram

2.3 Internal management tools: Gantt chart

In order to efficiently manage the evolution and impact of the work carried out, WP members have defined a set of monitoring and control tools. These tools are mainly specified in 3 instruments:

- Gantt diagram.
- Double-entry table for monitoring the commitments made, the work carried out and the percentage of compliance with the different obligations.
- Forms for the periodic report of compliance with the indicators defined in WP3 on the S2CP platform.

With regard to the last two tools (double-entry table and form for the periodic report), these are elements with a dynamic evolution, whose final status at the date of delivery of this deliverable is conveniently described in Section 7. Through these tools it has been possible to always identify the state of development of the different components that make up the architecture of the S2CP platform (see Section 5), and the level of success of the different indicators defined in D3.8 (see Section 7 to know the values finally reached in the M24 -date of delivery of this document-).

On the other hand, the Gantt chart is generated with the aim of helping the participants in the work package to achieve the objectives set out in the appropriate times and committed to in the Grant Agreement. This Gantt chart was agreed upon by all the members of the work package, and its structure responds to the development methodology of the S2CP platform (which can be consulted in Section 6).

The following figures represent the WP Gantt chart throughout the duration of the Cities2030 project.

Phase	Task	Relevant events	M8 May-21	M9 Jun-21	M10 Jul-21	M11 Aug-21	M12 Sep-21	M13 Oct-21	M14 Nov-21	M15 Dec-21	M16 Jan-22	M17 Feb-22	M18 Mar-22	M19 Apr-22	M20 May-22	M21 Jun-22	M22 Jul-22	M23 Aug-22	M24 Sep-22
0 - Planning and component definition	0.1 - Initial definition																		
	0.2 - Information																		
	0.3 - Final component																		
1 - First release (M24)	1.1 - Iterative development	R111 - Fist workshop with pilots						R111											
		R112 - Second workshop with pilots									R112								
		R113 - Third workshop with pilots												R113					
	1.2 - Testing period	R121 - Comments collection (end)																R121	
		R122 - D6.1, D6.2A, D6.3A and D6.4A																	R122

Figure 2. Gantt diagram of Phases 0 and 1

Phase	Task	Relevant events	M25 Oct-22	M26 Nov-22	M27 Dec-22	M28 Jan-23	M29 Feb-23	M30 Mar-23	M31 Apr-23	M32 May-23	M33 Jun-23	M34 Jul-23	M35 Aug-23	M36 Sep-23	M37 Oct-23	M38 Nov-23	M39 Dec-23	M40 Jan-24	M41 Feb-24
2 - Final release (M48)	2.1 - Iterative development	R211 - Fist workshop with pilots					R211												
		R212 - D6.5A						R212											
		R213 - Second workshop with pilots									R213								
		R214 - Third workshop with pilots													R214				
		R215 - Fourth workshop with pilots																R215	
	2.2 - Testing period	R221 - Comments collection (middle)																	
		R222 - Comments collection (end)																	
	2.3 - Close actions	R231 - D6.2B, D6.3B, D6.4B and D6.5B																	

Figure 3. Gantt diagram of Phase 2 (tasks 2.1 and 2.2)

Phase	Task	Relevant events	M42 Mar-24	M43 Apr-24	M44 May-24	M45 Jun-24	M46 Jul-24	M47 Aug-24	M48 Sep-24
2 - Final release (M48)	2.1 - Iterative development	R211 - Fist workshop with pilots							
		R212 - D6.5A							
		R213 - Second workshop with pilots							
		R214 - Third workshop with pilots							
		R215 - Fourth workshop with pilots							
	2.2 - Testing period	R221 - Comments collection (middle)		R221					
		R222 - Comments collection (end)				R222			
	2.3 - Close actions	R231 - D6.2B, D6.3B, D6.4B and D6.5B							R231

Figure 4. Gantt diagram of Phase 2 (task 2.3)

3 Requirements, uses cases and labs' definition and roles

This section discusses the requirements that must be considered in the development and design of the S2CP platform. In the first of the subsections, the requirements from document D3.8 are analyzed to extract a set of generic and high-level functionalities that must be present in the platform. The second of the subsections summarizes the results of the interviews between the WP6 members and the participating Labs in WP4 and WP5. The objectives and interests expressed by Labs must be crossed with the functionalities identified from D3.8 to generate the final design of the S2CP platform.

3.1 Requirements and use cases: integrating the WP3 guidelines and KPI framework

From the analysis of the requirements and use cases reported in D3.8, the WP6 working group extracts the needs to be covered by the S2CP platform, making a relationship between these needs and the functional recommendations of D3.8 and including the needs in the following subject areas.

3.1.1 AI tools to automatically identify products and opinions of CRFS issued on social networks or web 2.0

The aim of these instruments is to infer the opinion of consumers and their behavior regarding food and food systems. To achieve this, this component gathers information related to the relevant actors from social network or web pages and performs specific analysis (sentiment analysis, data mining, natural language processing) of the associated information (text and associated images). Generated results can be aggregated and summarized and finally represented in a form of plots and tables for further analysis by a user. This information can later be used to correct action plans and perform accordingly.

The functional recommendations of D3.8 most related to this topic area is:

FR#2: Provision of innovative technologies for the capture of information and creation of repositories (data mining, distributed learning, casual, conformance learning techniques), in the form of components.

Use case 1: What do people think about sustainable food practices? What are their consumption habits?

A sentiment analysis can be provided to understand the opinion of consumers about sustainable food practices and to what extend these practices are adapted in their daily life. To conduct this study, it is possible to use survey data generated by some local authority such as town hall, but also can be combined with related data belonging to the twitter social network. Through this analysis we can conclude whether an analysis of social media is a reliable source of useful information.

3.1.2 AI-specific tools

Tools must allow the use of AI for the analysis of information from the labs, both textual and graphic for the generation of statistics and indicators, but also for the detection of consumption trends.

In this regard, one case to be considered is the application of AI – to be more specific, NLP techniques and sentiment analysis – in order to track how good or bad certain food groups are perceived, such as vegetables, meat or cereals, using information and data gathered from internet websites, such as social network or blogs.

To perform the analysis, first, the reliability of the source of the text to be processed is automatically taken into account. Then, those parts of the text related to food are searched and the groups to which each mentioned food belongs are identified. Finally, sentiment analysis is performed on those parts of the text related to food, and the results obtained from this analysis applied to each one of the appearing groups is saved to see how the perception of the food groups evolves over time.

This field is directly linked to the recommendations of D3.8, specifically with the following functional recommendations:

FR#1: Access and integration of information repositories in multiple formats, such as open data portals, databases, organized thanks to a data governance framework.

FR#2: Provision of innovative technologies for the capture of information and creation of repositories (data mining, distributed learning, casual, conformance learning techniques), in the form of components.

Use case 1: How are the different food groups perceived over time? How can City counsels know which food groups are more urgent to promote or discourage?

This tool would allow municipalities to put greater emphasis on the promotion of healthy food groups such as fruits or vegetables when they are losing popularity, as well as discouraging the consumption of groups considered harmful such as pastries or sweets if they are perceived to be gaining popularity.

Another use case to consider is the application of Artificial Intelligence for the Development of chatbots to assist farmers of urban gardens with answers to their questions, giving advice and recommendations on specific farm problems. In addition to natural language processing or NLP, these intelligent systems produce recommendation based on semantic knowledge of the best type of crops to grow, according not only to the environmental conditions of the city and growing season, but also according to nutritional needs of the population in terms of CRFS.

3.1.3 Tools for CRFS on open data integration and management

Integration and management tools are responsible for data acquisition from different resources, data fusion into Cities0203 data model format and controlling data access permissions.

The **integration and analysis service** creates a central database view to put data from all the sources in data warehouse cloud location. Here the function of the *Open Data storage* is relevant, which is implemented as an open data repository to also meet the requirements of the Open Research data Pilot (ORDP) initiative, in which Cities2030 participates.

The collated data then goes through two modular components to achieve the underlying objective and provide insights, and business intelligence to different actors:

- *Data processing*: Data processing consists of transformation required on the data before it is passed in the pipeline to analytical models. The methodology here consists of approaches such as data clustering, featurization and feature engineering.
- *Data Analysis*: The data analytics pipeline consists of approaches such as federate and ensemble learning to deal with the heterogeneity in data to address different objectives of all the actors involved.

The **data governance service** is responsible for ensuring data security and privacy aspects of the platform. The service included the following features as part of the platform:

- *Multi-party incentivized data sharing* – this component aims towards incentivizing data exchange between different actors to achieve a common objective and showing the positive impact of data sharing in terms of data rich insights generated from the analytical models.

The functional recommendation of D3.8 most related to this topic area are:

FR#1: Access and integration of information repositories in multiple formats, such as open data portals, databases, organized thanks to a data governance framework.

Use case 1: Reducing Food Wastage – Sustainable an Efficient Food Supply Chain

Using an AI enabled small camera that can perform real-time calculations at the edge. An AI model can be trained to determine the volumetric detection of produce in a display bin. Such a system can determine the amount of produce available in real-time in display bins. If a shopper comes and takes the produce, the AI model can automatically update to reflect the fact that there is now less produce in the bin, and if the produce is restocked (or the shopper puts the selected item back), the model automatically gets updated. This kind of system can be used to perform two primary outputs:

1. One is for the store personnel, they can be alerted if there is an outage or depletion of produce in the display bin, and they can re-stock
2. By taking time series data snapshots of the produce over time, the store personnel can improve the supply chain. They can have deep analytical insights around - hours to sell, days to sell, replenishment rates, restocking rates, or purchasing rates. This can be used to tune the supply chain on behalf of the retailer.

Now such systems working in parallel at different geographical locations in a city (extended to national and international level) can update and exchange parametric information via federated learning, which can further tune and make the supply chain efficient – at regional and global level as part of decision support system provided.

3.1.4 Data Sharing enablers

Participating organizations (data owners) can register their data sources with data sharing enablers and attach metadata describing the data and policies concerning the rules around data discovery and sharing. This is linked to the following D3.8 functional recommendations:

FR#3: Possibility of registration (identification in user accounts). Policies for access to repositories (authorization).

FR#8: Policy-based management system, considering authentication, authorization (security) tools in data access. Modification of authorization policies in accessing or downloading data.

Use Case 1: Commercial Data Search:

A commercial data consumer seeks data for analysis and knowledge insight. They can use data sharing enablers to search the metadata repository to match the search with available data sources. Once discovered, the data consumer can negotiate the terms of access with the target data owners using the registered policies as a basis for the negotiation. The target data owners receive notifications about the proposed negotiation and enter a process of negotiation with the data consumer until both parties are satisfied with the terms or the negotiations reach a stalemate. Where both parties agree on the terms of access and the value exchange a smart contract is automatically generated by the platform. Data access from the data consumer side is governed by the terms of the smart contract and the platform enables the data flow from data source to data sink. All access attempts (successful or unsuccessful) are logged and can be accessed by both parties.

Use Case 2: Regulatory Data Request:

A regulator requires certain data and registers this data request in a data sharing platform. Participating data owners who have registered the data receive notifications of the requests from the platform and align their registered data source with the request. This results in a smart contract being generated which includes the terms of access and any value exchange that is part of the transaction. Data access from the data consumer side is governed by the terms of the smart contract and the platform enables the data flow from data source to data sink. All access attempts (successful or unsuccessful) are logged and can be accessed by both parties.

3.1.5 Communication and collaboration mechanisms with privacy and confidentiality requirements

Communication and collaboration mechanisms are integrated into the so-called Service-based Open Collaboration Space, which supports CRFS Labs' cross-fertilization and communication, findings and results dissemination, and events organization. The purpose of these tools is to facilitate community building among Cities2030 partners, to facilitate focus group creation, around CRFS topics, and facilitate giving and receiving feedback and comments, with credential management and access policies. There is also the possibility of protecting communications, making them anonymous, for a maximum level of privacy.

The functional recommendations of D3.8 most related to this thematic area are:

FR#3: Possibility of registration (identification in user accounts). Policies for access to repositories (authorization).

FR#8: Policy-based management system, considering authentication, authorization (security) tools in data access. Modification of authorization policies in accessing or downloading data.

Use case 1: Open innovation through project communication tools

A lab wants to establish a working group on a topic. Instead of using tools to manage their work with shared documents such as GoogleDocs that limit open innovation and the contributions of external stakeholders, they decide to register their lab in the Service-based Open Collaboration Space, which gives them the possibility of noting events in their calendar, contributing to disseminate the events publicly if they so wish, developing working groups through a forum where they can make contributions and submit them to reviews by other experts and citizens.

3.1.6 Tools for dissemination and classification of results

For a correct dissemination of results, in addition to being able to use the communication and collaboration tools described in the previous subsection, we establish the need for a set of tools that allows us to record innovation activities. These tools generate a catalog of activities and allow their management through moderation interfaces, and their visualization through geospatial elements, filtering, search and consultation.

These activities may be classified through the CRFS pillars of the FAO[4] or MUFPP (44 indicators), to ensure a correct ordering of the information collected.

The functional recommendations of D3.8 most related to this thematic area are:

FR#5: Geospatial data download, visualization and processing services.

FR#6: Generation of cartography adapted to different types of data. Cartography display controls of various types (heat maps, clusters, etc.).

Use case 1: Collection and classification of innovations in the field of organic recycling in Europe.

An innovation center wants to collect all the activities or events carried out in relation to organic recycling in various countries. The center wants to classify actions according to the FAO CRFS indicator framework. To record all this information using the S2CP platform, a form can be used to assist and select the location of the event on a map, and in providing multimedia information about the event. Once this desk research work is done, all the findings will be available to the public when they search for initiatives in the field of food waste.

3.1.7 Tools for managing experiments and innovation

The objectives of these tools are in line with the monitoring of innovation actions, within the framework of an open innovation concept. For this we can use the concept of Social Space for Research and Innovation (SSRI) coming from the Multi-Actor Approach (MAA) promoted in project "EIP-AGRI[5]": Iterative innovation model for the development of research into practical applications and the creation of new ideas thanks to interactions between actors ("cross-fertilisation") and the sharing of knowledge.

The tools for the management of experiments and innovation display a catalog of actions and plans and monitor stakeholders, their representatives and the related connections among them. An important aspect to consider is the monitoring of the impact of the activities, the SSRI maturity levels and progress and well as Context, Policy & Performance indicators catalog and monitoring (KPIs).

The functional recommendations of D3.8 most related to this thematic area are:

FR#3: Possibility of registration (identification in user accounts). Policies for access to repositories (authorization).

FR#4: Modular dashboard (possibly based on widgets) that allows customization and operations by users. Information analysis tools and statistics generation, averages, regressions, etc.

Use case 1: Example-based learning for improving public perception of healthy food

A CRFS Lab wants to determine how the implementation of certain events within a city can improve the citizen's perception of healthy food. To do this, it wants to consult similar innovation actions that have occurred in another city to determine if the impact produced in that city can be applied to its own. To do this,

[4] https://www.fao.org/3/ca6144en/CA6144EN.pdf

[5] https://ec.europa.eu/eip/agriculture/en/about

it accesses the experiment management tool with its credentials and locate cities with demographic and CRFS maturity characteristics like its own one. It accesses the list of actions carried out within the city and identifies the stakeholders who have carried out the actions, their frequency and view the results. The tool allows a CRFS lab to compare the current impact results with those that occurred before the actions were carried out, so that the improvement in the perception of healthy food is clearly identified.

With this information, the person in charge of a Lab can create a similar schedule for his lab more efficiently.

3.1.8 Digital twins of the supply chain for the traceability of food products

S2CP provide a solution for optimization of multi-stakeholder dialogue processes, in which blockchain will be employed to provide some proof of concepts of monetization processes, in a reliable and transparent way. In order to ensure transparency, traceability and trust of the local food production, blockchain technology (BC) presents the natural technology fit in the so-called SFSC (Short Food Supply Chain).

The functional recommendations of D3.8 most related to this thematic area are:

FR#7: Marketplace for data exchange, with blockchain support in the establishment of "contracts"

FR#8: Policy-based management system, considering authentication, authorization (security) tools in data access. Modification of authorization policies in accessing or downloading data.

Use case 1: Digital supply chain for food transparency and authenticity

A nutrition company produces their own food supplements and wants to have a brand image avoiding counterfeits and differentiating itself from other lower quality products. The use of blockchain for food security can allow the validation of the nutritional characteristics of those products, provide proof of origin and traceability of food supplements from the production of their ingredients to the consumer.

3.1.9 Real-time monitoring instruments

Real-time monitoring can receive information and events produced sensors (IoT HW platform), process the information in context-based systems and finally generate information repositories for real-time visualization or monetization. This topic area is linked to the following D3.8 functional recommendations:

FR#4: Modular dashboard (possibly based on widgets) that allows customization and operations by users. Information analysis tools and statistics generation, averages, regressions, etc.

FR#9: Communications broker for the management of information in real time that may come from sensors.

FR#10: Publication of real-time information in some type of control widget.

Use case 1: Realtime garbage collection monitoring

A garbage collection company wants to optimize the actions for the collection of organic waste containers. To do this, it installs a set of sensors that provide information on the level of occupation of the containers over time.

The company can generate notifications from these sensors through the S2CP platform for presentation in a control panel. In turn, another entity interested in these data, such as the person in charge of the Cleaning and Waste Collection Service, can subscribe to this information to, once analyzed, improve decision-making regarding the distribution of containers in the city.

3.1.10 Tools for the sale of traceable digital assets and food products

Data-based enablers must be provided to optimize multi-stakeholder dialogue processes, in which blockchain will be employed to provide some proof of concepts of contract agreements, in a reliable and transparent way. Mechanisms for the acquisition and access to data will be implemented, relying on blockchain technology that offers us evidence of transactions and contracts and proof of ownership.

The functional recommendations of D3.8 most related to this thematic area are:

FR#7: Marketplace for data exchange, with blockchain support in the establishment of "contracts".

FR#8: Policy-based management system, considering authentication, authorization (security) tools in data access. Modification of authorization policies in accessing or downloading data.

Use case 1: Marketplace for enhancing food public procurement

In food procurement processes, there must be an interaction between the companies that offer dining services for public institutions such as schools or hospitals and the centers that receive that service. The particular circumstances of a public procurement contract can be modeled as a blockchain Smart contract, so that contract compliance can be automatically verified. If specific conditions of the supplied products have been agreed, such as a guarantee of origin, or a level of quality in certain products, the blockchain will be able to check how these conditions are met and grant a seal of quality to those products or to the food service.

3.1.11 Mechanisms for customized data visualization

Visualization of data and analytics are of extreme importance in the decision-making process, where multiple layers of information and search and export tools can show stakeholders needs in an easily readable format suitable for analytics and, when possible, project visual information on the map to visually assist the usage of the information provided.

Determining functionalities in the development of these mechanisms include information filtering, timeline analysis and spatial analysis, which allows the user to customize every single aspect required for their search.

From the point of view of the user experience we must consider the following indicators:

- Build an easy-to-use interface.
- Guarantee the usability of the platform on multiple environments (desktop and mobile).
- Guarantee access to every single layer of information.
- Provide useful information and explanations regarding every layer of information provided by all partners.
- Provide the possibility to cross information from different information services.
- Provide the possibility to download information with visual and analytical value.

The functional recommendations of D3.8 most related to this thematic area are:

FR#4: Modular dashboard (possibly based on widgets) that allows customization and operations by users. Information analysis tools and statistics generation, averages, regressions, etc.

FR#5: Geospatial data download, visualization and processing services.

FR#6: Generation of cartography adapted to different types of data. Cartography display controls of various types (heat maps, clusters, etc.).

Use case 1: Geospatial decision support for food distribution activities

An entrepreneur wants to study which is the best place in a city to set up a new greengrocer. To do this, she uses the S2CP platform to study the distribution of fruit and vegetable consumption among the population of that city. It determines the neighborhoods where the consumption of fruits and vegetables is above the average for the city and displays, in turn, on a map of those neighborhoods, the distribution of businesses that currently sell fruit. Through a heat map functionality, she gets to see clearly what the optimal places could be to provide this food distribution service, considering the analysis carried out.

3.2 Living and policy labs' requirements, objectives, roles and definitions

After generating a catalog of functionalities based on the requirements stablished in D3.8, a series of bilateral meetings are held with Labs, with the aim of recognizing the usefulness of these proposals with respect to their needs. The purpose is to identify with a greater level of detail the mechanisms, technologies, algorithms, variables, etc., that must be integrated into the S2CP platform. For this reason, the functionalities of Section 3.1 have been specified and landed on the precise problems of the labs in possible functional components. These **functional components** are listed below with their description and later they are linked to each of the labs according to the needs gathered in the interviews.

Communication space: An integrated digital space including forum and blog functionalities, to communicate different agents in the City Region Food System. Different topics, roles and permissions could be defined, according to the profile, interests and needs of the different agents.

Data governance: This component allows the CRFS Lab to collect specific data for further analyses about the current situation of the ecological agriculture and local commerce (from commercial partners, for example) through a "purpose & consent" model but in an automatic way.

Data prediction: This component analyzes data from social network, especially images, with conformal learning in order to recognize some important elements and calculate the promotion intensity. For example, a Lab can be interested into identifying how often three products are in photos (and if the context if positive, negative or neutral): orange, pears and onions.

Digital twin for supply chain: It allows CRFS labs to have a prototype of a digital twin for a food supply chain, enabling food transparency with specific applications for securely consulting product information and blockchain certification.

Geospatial services: Some information such as the location of supermarkets, urban gardens, social services, other food actors, poverty levels, etc. needs to be displayed into a Geographical Information System. This component allows the CRFS Lab managers to navigate across the territory using a map, evaluating geographical variables they need to know. Even, this information may be later connected with geographical analyses from other previous components.

Good practices dashboard: Using a simple and lightweight web 2.0 system, the different agents in the CRFS could share their actions and good practices regarding the global objectives proposed by the CRFS Lab. This system incorporates a map, where it will be possible to locate all good practices according to their geographical origin and impact. Besides, a global catalogue is displayed. New users can be created to publish good practices from the private sector. For each practice, besides, a short description is also included.

Innovation management tool: This component allows a CRFS lab to improve the internal management of all innovative policies, by monitoring its evolution, success points. This, basically, is an application where new policies are uploaded, so all actors involved in the implementation of that policies can see all the information, share their opinions, problems, etc. and the evolution of the final implementation of this innovative policy can be monitored.

Intelligent tool for agriculture / urban gardening planification: Using this tool, the CRFS lab will be able to plan and schedule the organization of its agriculture / urban gardening policies. Different types of vegetables may be introduced, according to the climate, the nutritional needs of people, the amount of available land, etc. Using public repositories describing important information, this possible component suggests a plan of vegetables to be cultivated to guarantee a balanced diet and good nutrition of population.

Blockchain-based marketplace solutions for supply chain support: Tool that allows guaranteeing compliance with contracts in relation to the quality and price of the food, and provides integration with current business information systems, ensuring transparency of data within the short food supply chain, and allowing citizens to access to additional information about the products.

Open Access Data repository: This component connects different open data sets such as statistical data (national information, FAO, etc.), and other project generated data such as pilot surveys, studies, etc. This open data repository will also meet the requirements of the Open Research data Pilot (ORDP) initiative, in which Cities2030 participates.

S2CP dashboard: From different sources (national institute, European statistics, regional level if possible), all the required generic information (population, incomes, economic activities, etc.) is shown in a customizable dashboard. Collection of food system indicators at country level coming from international databases (FAO, FAOSTAT, World Bank, GFDx). A web interface could allow users to upload their own data to the platform and request the analysis they are more interested into.

Sentiment analysis: This component provides correlation and casual analysis to process data from social networks (detecting user's positive, negative and neutral opinions), to find if there is any correlation or causal relation among the promotion (local commerce, agriculture, etc.) and some independent variables (such as the geographical area, date, users' age).

Web mining component: Using a web-based tool, managers are able to process digital sources and webpages related to the CRFS lab (such as Facebook groups or tourism social networks). The tool creates a report about the most referred words, places, etc.

The following tables summarize the results obtained and the information captured in the aforementioned bilateral meetings.

Table 2. Collected information from 1-Brugge lab

Lab name	1-Brugge
Lab objectives, roles and definitions	
The main objective of Brugge lab is to improve the nutrition of elder people in their city.	
Required technological instruments	
- Communication space - S2CP dashboard - Digital twin	

Table 3. Collected information from 2-Velika Gorica lab

Lab name	2-Velika Gorica
Lab objectives, roles and definitions	
The main objective of this lab is to develop and improve the economic situation in the Velika Gorica region. Besides, they want to improve the nutrition of people in their city.	
Required technological instruments	
- Communication space - S2CP dashboard - Geospatial services	

Table 4. Collected information from 3-Trodos lab

Lab name	3-Trodos
Lab objectives, roles and definitions	
The main objective of Troodos lab is to develop and improve the economic situation in the Troodos region (very mountainous).	
Required technological instruments	
- Sentiment analysis - Intelligent tool for agriculture / urban gardening planification - Geospatial services - Innovation management tool - Good practices dashboard	

- S2CP dashboard

Table 5. Collected information from 4-Vejle lab

Lab name	4-Vejle
Lab objectives, roles and definitions	
The main objective of Vejle lab is to improve the current local CRFS in as many as possible directions, according to the Milan indicators (Milan urban food policy pact monitoring framework). Among all these indicators, those related to local markets and food are the most relevant	
Required technological instruments	
- Communication space - S2CP dashboard - Good practices dashboard	

Table 6. Collected information from 5-Seinäjoki lab

Lab name	5-Seinäjoki
Lab objectives, roles and definitions	
The main objective of this lab is to improve the nutrition of young people in their city.	
Required technological instruments	
- Communication space - S2CP dashboard - Digital twin	

Table 7. Collected information from 6-Bremerhaven lab

Lab name	6-Bremerhaven
Lab objectives, roles and definitions	
The main objectives of Bremerhaven lab are the following: • Support socially disadvantaged neighborhoods and districts • Networking to improve community catering (children (school) and/or hospital, elderly care) • Increasing the supply of nutrients, vitamins, and minerals, for example through the targeted cultivation of plants/herbs in urban gardening • Improving the nutrition education of people	

• Combating the nutritional poverty of children, the elderly and everyone who suffer from nutritional poverty • Expand regional supply chains (promotion local products)
Required technological instruments
- Sentiment analysis - S2CP dashboard - Communication space - Intelligent tool for urban gardening planification - Geospatial services - Data prediction

Table 8. Collected information from 7-Quart de Poblet lab

Lab name	7-Quart de Poblet
Lab objectives, roles and definitions	
The main objective of Quart de Poblet lab is to increase the national and international promotion and relevance of the Valencia's ecological agriculture and local commerce.	
Required technological instruments	
- Innovation management tool - Sentiment analysis - Geospatial services - Open Access Data repository - S2CP dashboard - Digital twin	

Table 9. Collected information from 8-Vidzeme lab

Lab name	8-Vidzeme
Lab objectives, roles and definitions	
The main objective of Vidzeme lab is to promote short food supply chains, in order to support the local social development and the green regulation implementation.	
Required technological instruments	
- Sentiment analysis - Good practices dashboard - Geospatial services	

- S2CP dashboard

Table 10. Collected information from 9-IASI lab

Lab name	9-IASI
Lab objectives, roles and definitions	
The main objective of IASI lab is to promote short food supply chains and the local products.	
Required technological instruments	
- Sentiment analysis - Web mining component - Geospatial services - S2CP dashboard	

Table 11. Collected information from 10-Murska Sobota lab

Lab name	10-Murska Sobota
Lab objectives, roles and definitions	
The main objective of Murska Sobota lab is to achieve economically sustainable, socially responsible and environmentally friendly business for the benefit of all involved stakeholders and wider community.	
Required technological instruments	
- Geospatial services - S2CP dashboard - Marketplace for short food supply chains	

Table 12. Collected information from 11-Vicenza lab

Lab name	11-Vicenza
Lab objectives, roles and definitions	
The main objective of Vicenza lab is "food for all" and improve the nutrition of citizens in the city and its social development through food policies.	
Required technological instruments	
- Innovation management	

- Sentiment analysis
- Data governance
- Geospatial services
- Open Access Data repository
- S2CP dashboard

Table 13. Collected information from 12-Haarlem lab

Lab name	12-Haarlem
Lab objectives, roles and definitions	
The main objective of Haarlem lab is to improve the performance and security of the supply food chain. CRFS in Harlem is very complex, as internal market in Holland is very compact and many national agents participate in all local CRFS. The creation of communication channels among all companies involved in the CRFS is a key.	
Required technological instruments	
- Innovation management tool - Data governance - Communication space - Geospatial services - S2CP dashboard	

As can be seen, all Labs request two different types of tools. On the one hand, generic communication and data representation tools, which are mostly demanded by all laboratories. On the other hand, specific tools, designed and implemented for particular use cases of each of the laboratories. This fact seems to lead us towards a flexible platform design based on components, which allows each of the laboratories to have one where they stay personalized and adjusted to their needs of the S2CP platform.

To finalize this section, a table is presented where the interest aroused by the Labs for the proposed components can be better seen. This figure allows us to establish a system of priorities for the development of the functionalities of the S2CP platform.

Table 14. CRFS Labs interest in S2CP components

Components	CRFS Labs											
	1	2	3	4	5	6	7	8	9	10	11	12
Communication space												
Data governance												

Data prediction						■						
Digital twin	■				■		■					
Geospatial services		■	■			■	■	■	■	■	■	■
Good practices			■	■				■				
Innovation management			■				■				■	■
Gardening planification			■			■						
Marketplace										■		
OA data repository							■				■	
S2CP dashboard	■	■	■	■	■	■	■	■	■	■	■	■
Sentiment analysis			■			■	■	■	■		■	
Web mining									■			

4 Integrating the Reference Architectural landscape

Once the requirements provided by the end users of the platform (Labs) and those detected by WP3 from the comments received from the stakeholders are captured, it is necessary to review in detail the state of the art in technology, to know which are the viable alternatives of implementation that adjust to the requirements previously raised.

To do this, in this section we review all the technologies related to the functions and tools demanded by users and described in Section 3. Specifically, we review the technologies in the following 5 areas, for which digital platforms similar to the S2CP platform are defined.

IoT Reference Architectures (Section 4.1): Some of the existing IoT platforms may address a specific problem or a limited technical environment, offering a point solution addressing only a part of the IoT stacks. On the other hand, some platforms can be very general purpose and integrate the IoT system in a larger (enterprise) system. These models prescribe platforms providing services for the IoT (devices and data) management.

Big Data / Cloud frameworks (Section 4.2): These frameworks address large and complex data sets, where the issues come from having large volumes of data with a lot of variety, that must be collected, updated and processed very fast, while also verifying their quality and trustworthiness. To address all these issues, we need a range of technologies for capturing, managing, processing, analyzing, visualizing and communicating the data. These frameworks are therefore designed in such a way as to manage the big data collection, processing and analysis chain.

Interoperability Platforms (Section 4.3): These platforms combine a number of technologies together, such as IoT, Big Data or Cloud architectures (e.g., FIWARE). They also deal with data exchanges and management, including how to monetize data services. In fact, some platforms (like IDSA[6], see section 4.3.2) focus and put emphasis on information ownership, with the aim of enabling clear and fair exchanges between data sources and consumers.

Blockchain-based platforms (Section 4.4): These platforms provide an alternative to traditional Blockchain infrastructures, which takes into account the requirements of the GDPR for the storage and removal of information. In this section we present the new initiative of NIST DataBlockMatrix.

Project specific data platforms (Section 4.5): Some of these architectures are based on previously studied reference architectures with some modifications to put special emphasis on the business viewpoint of specific verticals. Overall, in every other respect of note, they follow similar principles to the rest of the models in the previous frameworks, with physical devices sitting at the bottom of the architecture, integration and communication on top of them linking to the software and its functional components, and at the top having the final application which implements and satisfies the business goals of the whole system. However, some of the components, interfaces and data models are different from their reference architectures and it is worth highlighting these differences to address large scale pilots.

4.1 IoT Reference Architectures

4.1.1 AIOTI

The Alliance for the Internet of Things Innovation (AIOTI)[7] has specified a high-level Reference Architecture[8] that maps to several other dominant and/or standardized IoT architectural approaches, such as ITU-T[9], Big Data Value Association (BDVA)[10], National Institute of Standards and Technology (NIST)[11], etc.

Considering its functional model, it describes functions and interfaces between functions of the IoT system. The following figure provides a high level AIOTI functional model, referred to as the "AIOTI HLA functional model".

[6] https://internationaldataspaces.org

[7] Alliance for the Internet of Things Innovation (AIOTI): https://aioti.eu/

[8] AIOTI - High Level Architecture (HLA), Release 4.0, AIOTI WG03 – loT Standardisation,. Retrieved from https://aioti.eu/wp-content/uploads/2018/06/AIOTI-HLA-R4.0.7.1-Final.pdf

[9] ITU-T FG-DPM, ITU-T Focus Group on Data Processing and Management to support IoT and Smart Cities & Communities, https://www.itu.int/en/ITUT/focusgroups/dpm/Pages/default.aspx

[10] Big Data Value Association, European Big Data Value Strategic Research and Innovation Agenda, http://bdva.eu/sites/default/files/BDVA_SRIA_v4_Ed1.1.pdf

[11] NIST big data interoperability framework, http://bigdatawg.nist.gov/V1_output_docs.php

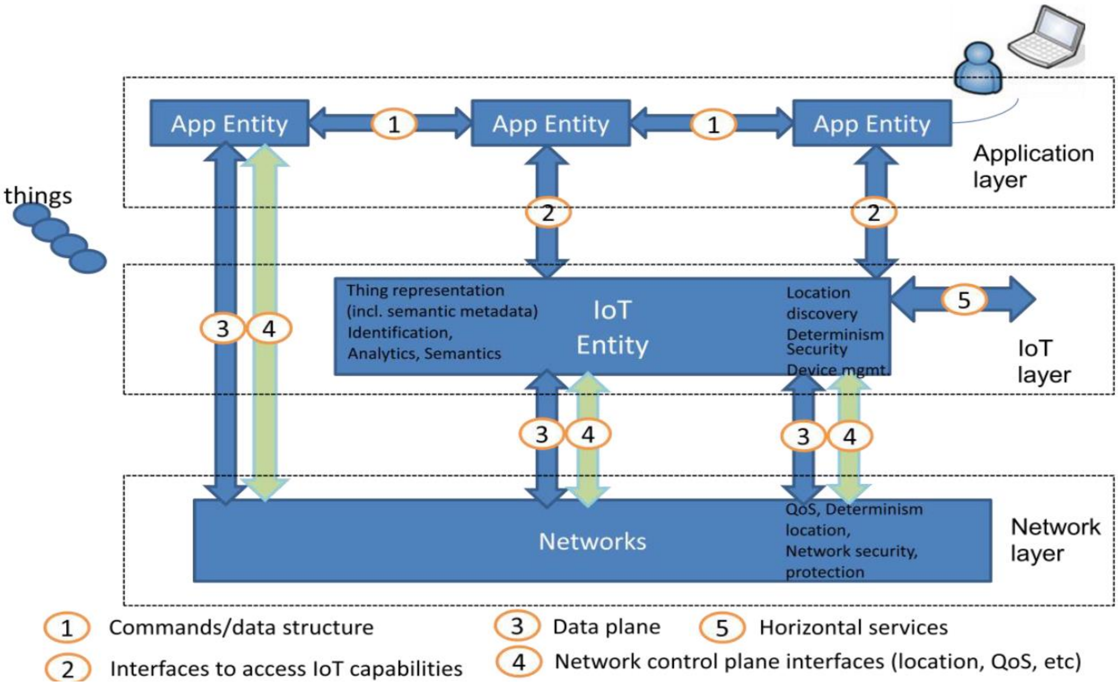


Figure 5. AIOTI HLA functional model (Source AIOTI WG3 IoT High Level Architecture Release 2.1)

Main functions are:

App Entity: is an entity in the application layer that implements IoT application logic. An App Entity can reside in devices, gateways or servers.

IoT Entity: is an entity in the IoT layer that exposes IoT functions to App Entities via the interface 2 or to other IoT entities via interface 5. Typical examples of IoT functions include: data storage, data sharing, subscription and notification, firmware upgrade of a device, access right management, location, analytics, semantic discovery etc.

Networks: may be realized via different network technologies (PAN, LAN, WAN, etc.) and consist of different interconnected administrative network domains.

HLA Security and Management considerations

Security and Management are fully recognized as important features in the AIOTI HLA, considered as intrinsic to interface specifications. **All interfaces must support authentication, authorization, and encryption at hop-by-hop level. End-to-end application-level security could also be achieved** via securing interface 1.

Considering *data management*, a data-centric approach is defined, considering a lifecycle with seven main phases: 1) Obtain/collect, 2) Create/derive, 3) Use, 4) Store, 5) Share/disclose, 6) Archive and 7) Destroy/Delete.

Finally, considering *compliance management*, these security and management domains combined would need be addressed and (re)considered from a compliance point of view, including without limitation accountability, safety, security, data minimization and data retention obligations, security breach notification and disclosure obligations, (personal) data protection compliance, official mandatory policies compliance, etc.

4.1.2 IoT-A

The IoT-A Project[12] addressed the Internet-of-Things Architecture and proposed the creation of an architectural reference model together with the definition of an initial set of key building blocks.

The IoT-A project sets out as one of its missions to provide:

> *"an architectural reference model for the interoperability of Internet-of-Things systems, outlining principles and guidelines for the technical design of its protocols, interfaces, and algorithms"*

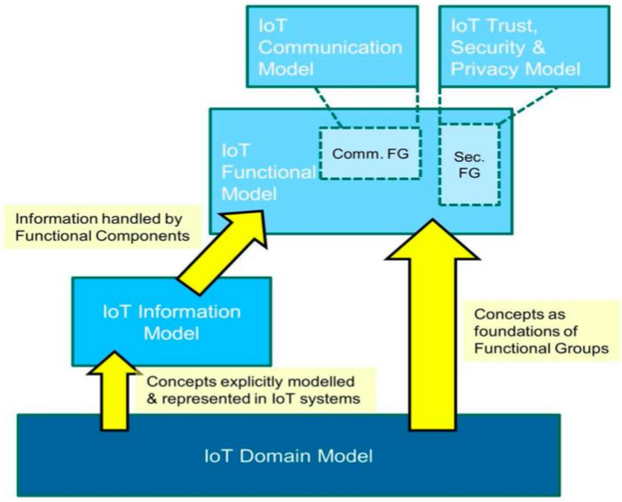


Figure 6. IoT-A Domain Model (Source: IoT-A D1.5)

Here we describe the primary elements of that reference architecture. The reference architecture model is fully described in project IoT-A deliverable D1.5[13].

4.1.2.1 IoT-A Reference Model

The IoT Reference Model establishes a common grounding and language for IoT architectures and IoT systems, shown in the following figure. The arrows indicate how aspects of one model can be used in another.

The foundational block of the Reference Model is the IoT Domain Model which defines a set of high-level components or actors. These components are common to most IoT use-cases and are depicted in Figure 6.

4.1.2.2 IoT-A Information Model

The IoT Information Model defines the structure (e.g. relations, attributes,

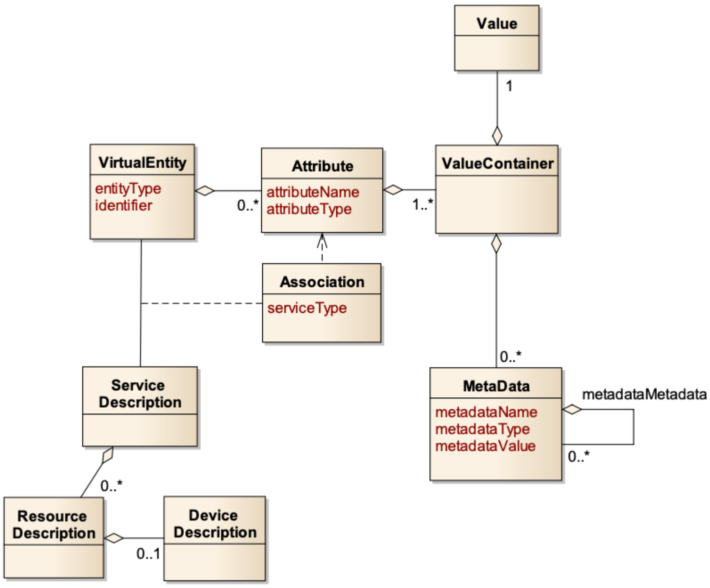


Figure 7. IoT-A Information Model (Source: IoT-A D1.5)

---

[12] Internet of Things Architecture project, https://www.iot-a.eu/

[13] IoT-A D1.5 Final architectural reference model for the IoT v3.0, https://www.researchgate.net/publication/272814818_Internet_of_Things_-_Architecture_IoT-A_Deliverable_D15_-_Final_architectural_reference_model_for_the_IoT_v30

services) of all the information for Virtual Entities on a conceptual level and is shown in Figure 7.

4.1.2.3 IoT-A Functional Model

The architect introduces the FGs (Functionality Groups) as part of the functional model. Each of these FGs is further decomposed into several Functional Components (FC). These are all presented in the functional view diagram (see Figure 8). This figure depicts the Functional Components (FCs) for each Functionality Group (FG) of the IoT-A ARM, which is called the IoT Functional View.

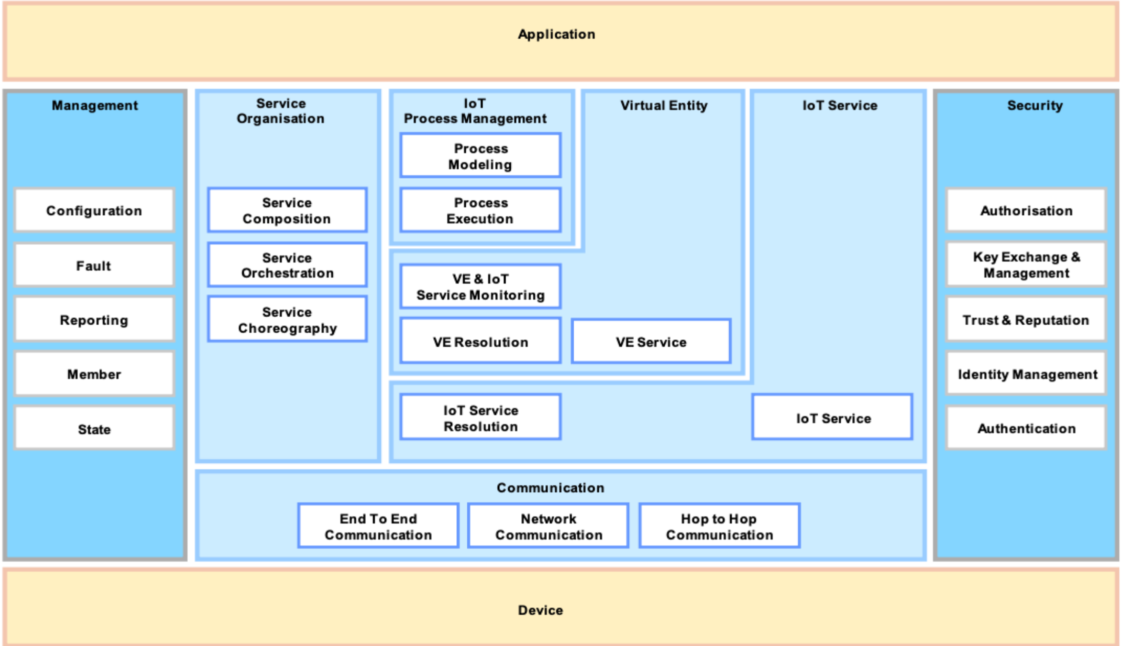


Figure 8. Functional View of the IoT-A architecture

4.1.2.4 IoT-A Communication Model

The IoT Communication Model aims at defining the main communication paradigms for connecting elements, as defined in the IoT Domain Model. The architecture document provides a reference set of communication rules to build interoperable stacks, together with insights about the main interactions among the elements of the IoT-A Domain Model.

4.1.2.5 IoT-A Trust, Security and Privacy Model

Trust is enforced in all layers and elements of the system. IoT-A ARM established concepts such as trust-model domains, trust-evaluation mechanisms, behavioral policies and trust anchors, Federation of trust or M2M support.

Security is considered at 2 different levels:

- Communication, which provides an abstract approach that constrained devices are specifically targeted. These devices are the most challenging elements of an IoT based system. Security for them can be assigned to the Corresponding Gateway, which must implement a constrained device security feature.

- Application. IoT-ARM model covers the basic steps in order to make the analysis that will communicate with the identification of the design policies and actions for mitigating, such as description of elements for protection, categorize risk sources, proposal of design choices and risk assessment.

Privacy issues are enforced through a Functional Component (FC) called Identity Management. This element in collaboration with the Authentication and Authorization components facilitates security policies in the different resources and endpoints of the system.

4.1.3 IPv6 Forum

The number of Internet Connected devices will cross the incredible total of 50 billion by 2025. The connectivity fabric of IP is used to enable more and more efficient context exchange with a broader range of devices and things. Thus, results the Internet of Things.

Projected to increase device counts by orders of magnitude over the next few decades, IoT's impact cannot be overstated. Already enabling a rich set of new capabilities in Smart Cities, Smart Grid, Smart Buildings, and Smart Manufacturing, IoT stands to transform virtually every part of modern life that automation or visibility may improve.

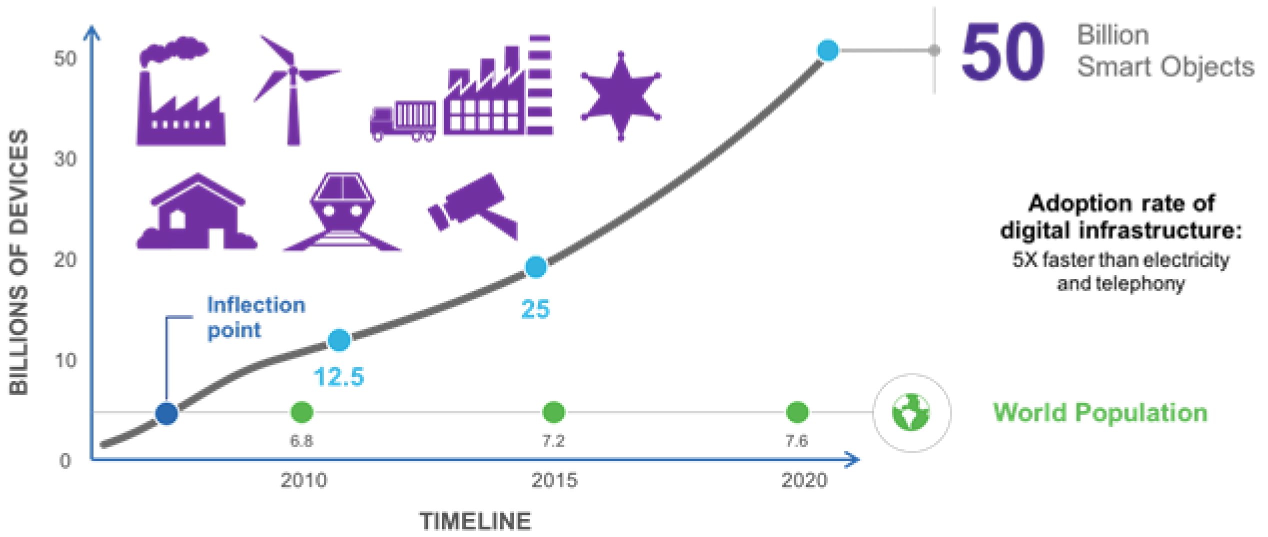


Figure 9. IoT growth (Source Cisco)

No matter the precise forecast, the sheer tsunami of devices coming online in the next months, years, and decades ensures that the future is not exclusively, or even significantly, wired.

Wireless with its adaptability and ease will inevitably dominate the IoT landscape. Exactly which wireless technology or technologies will be used remains relatively unclear, as many new technologies are still emerging, while others are still early in the standards process.

IPv6 can enable and sustain the growth rate of the IoT. It offers a future proof solution. More and more organizations have decided to either transition to IPv6 or to develop new standards only based on IPv6. This is specifically the case for IoT related standards. 3GPP secretary and CTO of ETSI Adrian Scrase has already announced back in April 2019 the move from E.164 for Machine Type communication to IPv6 addressing for larger scale deployment of IoT.

IPv6 does not only enable the scalability required by the IoT Food Supply chain, but also provides enhancement from IPv4 in the field of mobility support, stateless address auto-configuration, support of constraint devices and security to mention only a few of them.

4.2 Big Data / Cloud frameworks

4.2.1 European Cloud Platform, the road to data sovereignty using GAIA-X and IDSA

Digitalization is chewing the world with strong economic and social impacts. Recently, the management of the COVID 19 crisis enlightened the power of digital tools and their impact on stakes such as privacy, surveillance, transparency and censorship. How nations deal with massive digitalization and master the technologies, applications, deployed and used on their soils by their companies and citizens is vividly raised by the US ban on Huawei enforced by the "clean network" strategy. This ban could have set the ground for a technologic war between *"2 blocks: the digital democracies and the techno-authoritarian regimes"*[14][15]

In Europe, Th. Breton the European Commissioner for Internal Market and Services declared that the war for industrial data begins now and Europe will be the main battlefield. Anyhow, the future platforms for harnessing data as close as possible to their place of production, notably at the edge, remain to be built. They will be far different from the ones GAFAM[16] and BATX[17] have developed for private data so far taking into account European regulatory rules[18][19].

The EU launched GAIA-X[20] whose origin stems from the German Federal Government to create the next generation of data infrastructure for Europe, its companies and its citizens. This infrastructure needs to meet the highest standards in terms of digital sovereignty and aims to foster innovation. The targeted infrastructure is regarded as the cradle of an ecosystem, where data and services can be made available, collated and shared in a trusted environment. The goal was to establish a more robust framework in 2020 and to launch the very first use cases by 2021[21].

The International Data Spaces Association (IDSA) had earlier in 2019 defined a reference architecture and a global standard for creating and operating virtual data spaces. The IDS Architecture is based on commonly

---

[14] A. Capri, "Techno-nationalism and Diplomacy: The US-China Race to Reshape Alliances," institutions and Standards, Oct. 2020.

[15] U.S. government, The clean network, https://2017-2021.state.gov/the-cleannetwork/index.html

[16] Google (Alphabet), Apple, Facebook (Meta), Amazon and Microsoft

[17] Baidu, Alibaba, Tencent, and Xiaomi

[18] European Commission, Shaping Europe's digital future: Eurobarometer survey shows support for sustainability and data sharing March 2020.

[19] T. Breton, Future stratégie européenne sur l'intelligence européenne présentée le 19 février dans un livre blanc, Feb. 2020.

[20] https://gaia-x.eu/

[21] GAIA-X, Driver of digital innovation in Europe featuring the next generation of data infrastructure, Federal Ministry for Economic Affairs and Energy (BMWi), May 2020

recognized data governance models facilitating secure exchange and easy linkage of data within business ecosystems[22]. This architecture and components answer very well the requirements of GAIA-X.

Why

Exchanging data, even within a single company is often not straightforward since data is scattered and "siloed" per business process. Exchanging data between partners is even more delicate. The main purpose of a data space is to allow the sovereign and secure exchange of data within a "trusted ecosystem" involving multiple players.

A key capability is to ensure data sovereignty, i.e., providing data owners with full control over their data and their digital identities. This requires the definition of data usage constraints: defining who is allowed to do what in which context with the data shared by the data owner[23].

From a business perspective, a data space aims to allow the building of data-driven ecosystems in which independent partners (from different sizes, ecosystems and financial power) have trust in how their data are handled while allowing the innovative data services to be constructed in a cooperative manner, thereby breaking the information silos.

How

Europe's plan for digital sovereignty uses 2 main axes. The first one is cloud sovereignty, in order to have cloud services that comply with European regulation. The solution for this sovereign cloud infrastructure hinges on the federation of European cloud services along the GAIA-X association. The second one is data sovereignty with the goal of being able to safely share data among participants in a consortium, the foundation of this will be the IDSA's reference architecture model.

IDSA's Reference Architecture (RA) provides an abstract "business view" description of the roles a participant can play in the Data Spaces. "Core Participant" roles are roles assumed by organizations that own, provide and/or consume or use data in the data space. These roles include Data Owner, Data Provider, Data Consumer and Data Application Provider.

They also issue certificates to core software components (e.g., Connectors) that are to be deployed in the data space. To securely exchange and share data in an IDSA-compliant data space, any participant deploys a technical component called the Industrial Data Space Connector. The requirements to be met by a Connector for cross-company exchange of industrial manufacturing data are specified in [24].

GAIA-X focuses on providing a trusted infrastructure to allow secure and sovereign data exchanges by certifying its nodes and actors and by relying on verifiable claims done by nodes. A high-level overview of the GAIA-X architecture[25] with its major elements and functions is provided in the following figure.

---

[22] IDSA, International Data Spaces Association, https://internationaldataspaces.org/

[23] GAIA-X, Policy Rules and Architecture of Standards, Federal Ministry for Economic Affairs and Energy (BMWi), May 2020

[24] DIN SPEC 27070 reference architecture of a security gateway for the exchange of industry data and services; IDSA, Feb. 2020.

[25] GAIA-X, technical architecture, Federal Ministry for Economic Affairs and Energy (BMWi), Release, June 2020.

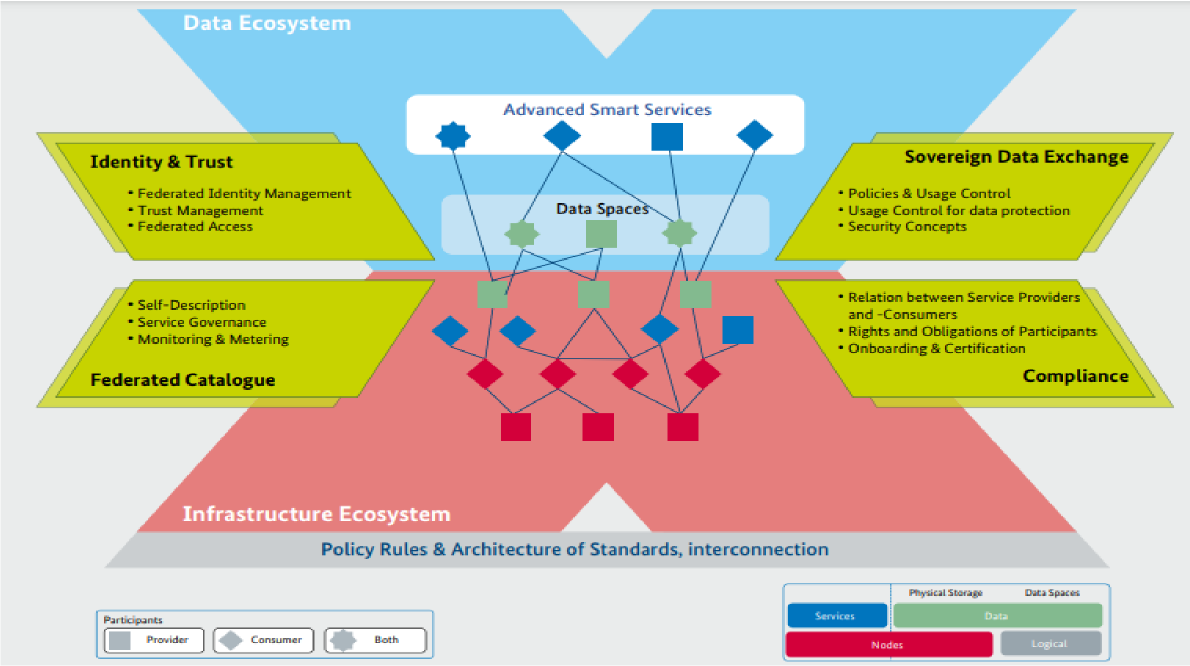


Figure 10. High-level view of the GAIA-X architecture

IDSA defines a federated technical architecture that aims to guarantee data security and protection for all involved participants. It establishes mutual trust among them and ensures data sovereignty for all data providers. Therefore, the data space concepts and components proposed by IDSA are used to support the federated and interoperable infrastructure that the European project GAIA-X aims to address.

The European Digital Single Market intends to ease the transfer and exploitation of data within the Union. In many aspects, the COVID19 pandemic reveals how dependent Europe is on a globalized economy and that a digital infrastructure is central for the continuity of Europeans' social, economic and healthcare activities. Simultaneously, the European Union has acknowledged that data are not simple merchandise and that personal data belong to their data subjects who have fundamental rights to control if and how their data can be used. On the business side, 5G Verticals like Energy, Smart Cities, as well as the Industry 4.0, should lever on a diverse ecosystem of actors by relying on a digital space of trust where data can be collected and processed in full confidence and transparency.

The European Union data strategy affirmed in 2020 has two main streams: data and cloud. The data regulation is set in the Data Governance Act and Data Act. GAIA-X is at the heart of the coordination: to provide use cases and technical architectures for European common dataspace for the data stream; to supply federation and interoperability for the European alliance for industrial data, Edge and Cloud for the cloud stream. Whether its GAIA-X or IDSA both approaches rely on a federation approach as such they are heavily based on current standards that are still being chosen as we write these words. For IEEE members interested this would be a fantastic opportunity to disseminate standards developed within the organization.

4.2.2 NIST Big Data Reference Architecture

The NIST Big Data reference architecture (NBDRA) is defined in the NIST Big Data Interoperability Framework[26] and is intended to enable system engineers, data scientists, software developers, data architects, and senior decision makers to develop solutions to issues that require diverse approaches due to convergence of Big Data characteristics within an interoperable Big Data ecosystem. It provides a framework to support a variety of business environments, including tightly integrated enterprise systems and loosely coupled vertical industries, by enhancing understanding of how Big Data complements and differs from existing analytics, business intelligence, databases, and systems.

The following figure shows the NBDRA and whose description follows:

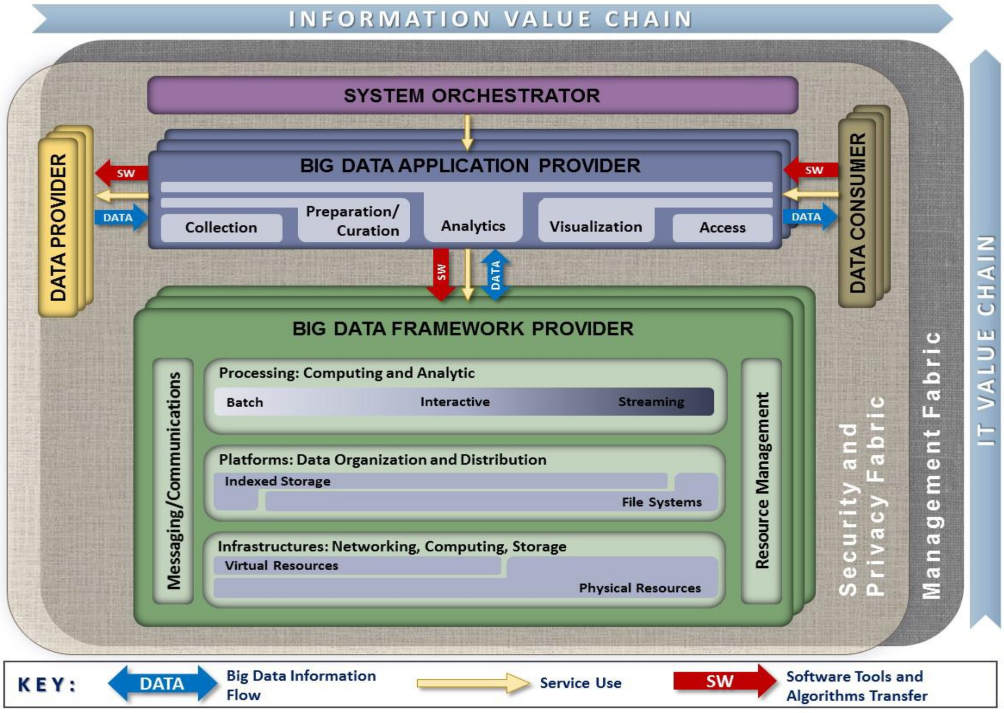


Figure 11. NIST Big Data reference architecture

The NBDRA is organized around five major roles and multiple sub-roles aligned along two axes representing the two Big Data value chains: Information Value (horizontal axis) and Information Technology (IT; vertical axis). Along the Information Value axis, the value is created by data collection, integration, analysis, and applying the results following the value chain. Along the IT axis, the value is created by providing networking, infrastructure, platforms, application tools, and other IT services for hosting of and operating the Big Data in support of required data applications. At the intersection of both axes is the Big Data Application Provider role, indicating that data analytics and its implementation provide the value to Big Data stakeholders in both value chains. The term provider as part of the Big Data Application Provider and Big Data Framework Provider

[26] NIST Big Data Interoperability Framework, Volume 6, Reference Architecture, Version 3: https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1500-6r2.pdf

is there to indicate that those roles provide or implement specific activities and functions within the system. It does not designate a service model or business entity.

The five main NBDRA roles represent different technical roles that exist in every Big Data system (System Orchestrator, Data Provider, Big Data Application Provider, Big Data Framework Provider, and Data Consumer). Also, two other roles are specified, which are Managements and Security and Privacy, providing services to the five main roles.

4.2.3 BDVA Reference architecture

The BDVA Reference Architecture is a reference framework made by the European BDVA (Big Data Value Association)[27] that describes logical components of a generic big data system.

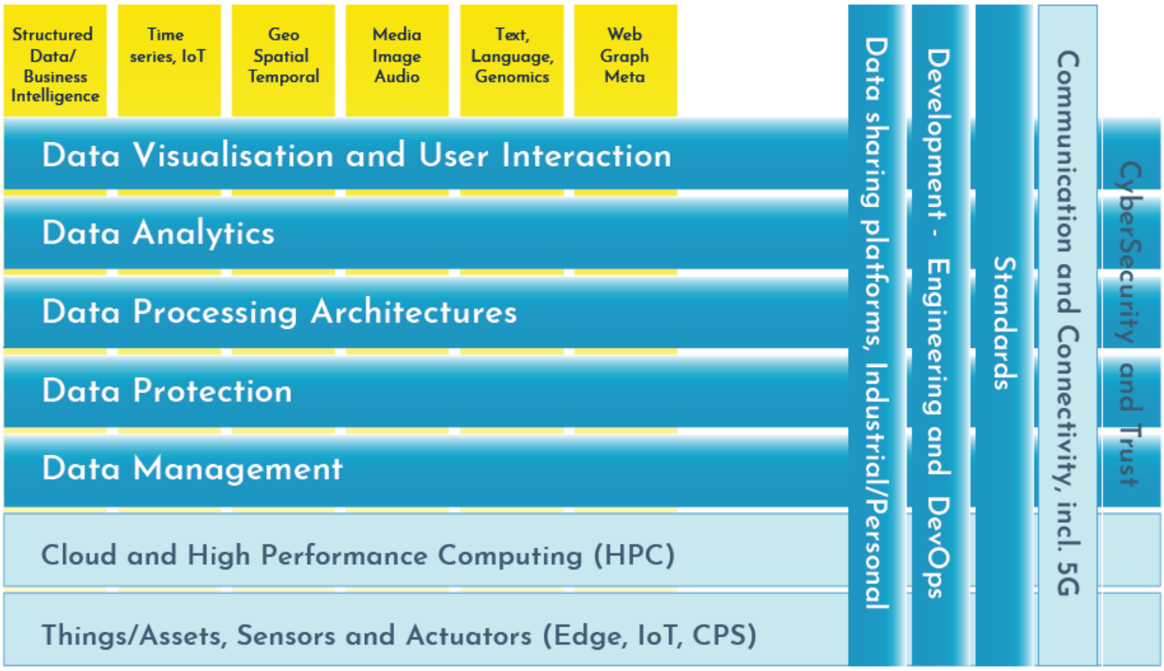


Figure 12. Big Data Value Reference Model[28]

BDVA has proposed their initiative regarding a European Data-Driven Artificial Intelligence and their vision regarding AI and Big Data and how it can drive the European technology and economy[29].

To realize this vision, it will be necessary to address several challenges:

a) Data-driven AI-based solutions for the industry will require new business models.

---

[27] Big Data Value Association webpage: https://www.bdva.eu/

[28] European Big Data Value Strategic research and Innovation agenda http://bdva.eu/sites/default/files/BDVA_SRIA_v4_Ed1.1.pdf

[29] Data-driven Artificial Intelligence For European Economic Competitiveness and Societal Progress. BDVA Position Statement. (2018, November). Retrieved from http://www.bdva.eu/sites/default/files/AI-Position-Statement-BDVA-Final-12112018.pdf

b) Trust in AI and its results must be established; for example, one should be able to explain how AI applications came to a specific result ("Explainable AI"), which would foster responsible technological development (e.g., avoid bias) and enhance transparency in how and why an AI takes a decision.

c) It is necessary to develop an AI and Big Data ecosystem, by developing data for open AI platforms and overcoming the lack of data interoperability.

d) Fuse and develop several technologies, as a successful industrial AI relies on the combination of a wide range of technologies, such as advanced data analytics, distributed AI, and hardware optimized for AI.

Recently **Gaia-X European Association for Data and Cloud AISBL**[30], the **Big Data Value Association (BDVA), FIWARE Foundation**, and the **International Data Spaces Association (IDSA)**[31] joined forces to drive the adoption of data spaces across Europe and beyond. The new collaboration, called **Data Spaces Business Alliance (DSBA)**, is the first initiative of its kind, bringing together the necessary industry players to realize a data-driven future in which organizations and individuals can unlock the full value of their data.

The Alliance is working together in three major areas:

1) Technology and architecture: The Alliance define a common reference model, based on existing architectures and models, leveraging each other's efforts on infrastructure and implementations. It drives interoperability by harmonizing technology components and other elements.

2) Support: The Alliance assists the existing organizations and data spaces by pooling their tools, resources and expertise in a focused way. The support includes handbooks, roadmaps, individual evolution plans, access to Digital Innovation Hubs (DIHs), acceleration programs, etc. and includes a go-to-market toolkit to make data spaces happen.

3) Identification and characterization: The Alliance establish a 'Data Spaces Radar' to actively scout potential data spaces. It provides a unique overview on data spaces evolution on a global level, thanks to an enormous network and unprecedented expertise. The Alliance also promotes frontrunners and identifies best practices used by them, for the benefit of all.

Conclusions

As conclusions of this review, we consider that a platform that carries Big Data components must pay attention above all to the inputs to the outputs (data provider and consumer) and correctly define the interfaces with the labs (WP4 and WP5) as providers and consumers. of information. However, the **importance of the system operators and technicians**, who must be present for the correct functioning of the platform, should not be overlooked. For now, this role is covered by the technological partners belonging to WP6, but within the framework of the **software development methodology** that we will define in Section 6, we will see the need to **incorporate a 'Lean Startup' mechanism** by which the validation of the components generates learning in the Labs that allows them **to acquire enough experience at the end of the project to be operators of their own data management infrastructure**.

[30] https://www.data-infrastructure.eu/GAIAX/Navigation/EN/Home/home.html

[31] https://internationaldataspaces.org/

4.3 Interoperability frameworks

4.3.1 FIWARE

FIWARE[32] is defined as a curated framework of Open Source Platform components to accelerate the development of Smart Solutions. It aims to promote the creation of standards necessary to develop Smart applications in different domains: Smart Cities, Smart Ports, Smart Logistics, Smart Factories, among others. Any Smart application is characterized by collecting relevant information for the application from different sources about what is happening at a given moment. This is known as "background information." Current and historical context information is processed, visualized and analyzed on a large scale. In this way, the expected intelligent behavior occurs.

Context data providers and consumers can interact with the Context Broker, which is a central component of FIWARE architecture, see the following Figure. The Context Broker aims to enable the system to make updates and access to the current state of context.

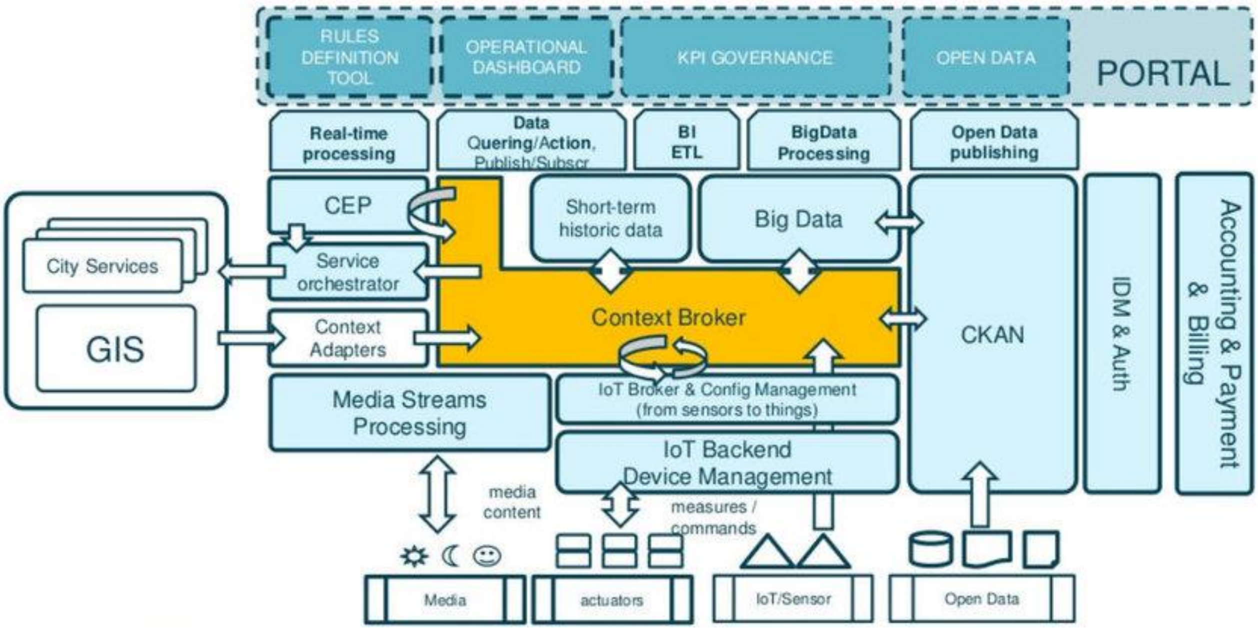


Figure 13. FIWARE Reference Architecture

Building around the FIWARE Context Broker, a rich suite of complementary FIWARE components is available, dealing with:

- **Interfacing with the Internet of Things (IoT), Robots and 3rd Party systems**, for capturing updates on context information and translating required actuations.

[32] www.fiware.org

- **Context Data/API management, publication, and monetization**, bringing support to usage control and the opportunity to publish and monetize part of managed context data.
- **Processing, analysis, and visualization of context information** implementing the expected smart behavior of applications and/or assisting end users in making smart decisions.

Considering this last point of processing, analysis and visualization of context information, it is the most aligned with data management and visualization tasks (T6.2 and T6.5) in Cities2030. In FIWARE, different processing engines, such as Flink, Hadoop and Spark, are used in order to process historical data, so as to extract valuable insights or derive smart actions. Artificial Intelligence or Complex Event Processing functions can be used above the integrated processing engines. Wirecloud web mashup framework is used for Operating dashboards. Extended CKAN portal can offer to third-parties part of the current and historic context data and other Geospatial Visualizations.

Regarding security considerations, the API/Data access control functions enable access to the context data to parties that own certain privileges. The API management and business support layer can offer auditing of the system and monetize data access.

4.3.2 International Data Spaces (IDS)

The International Data Spaces Association (IDSA)[33] is the evolution of IDS (Industrial Data Space) which itself was an initiative led by Fraunhofer ISST, in cooperation with ATOS, T-Systems, and the idea is promoted by the German Federal Ministry of Education and Research. IDSA is characterized by the focus on information ownership, with the aim of enabling clear and fair exchanges between data providers and consumers.

Data as strategic resource is an enabler for smart services, smart products and our desired lifestyle of the future, which stands for transparency and simplification of everyday tasks and processes amongst others. Data will become an economic asset, and to some extend are already. Therefore, the key focus for a data-driven economy and new evolving business models is enhancing data sovereignty. Data sovereignty is about ownership, security, and value of data and on the other hand about interoperability, data exchange, "sharing economy" and data centric services. IDS brings improvements with regards to these obstacles, e. g. data security is ensured, sovereignty is improved, and processes and cost structures are optimized. Furthermore, IDS can enable access to yet untouched data treasures in companies, while staying in control over its flow and usage, through a set of identified stages:

1. Make data available. This can be done dynamically or on demand. To do so, data have to be described before they can be exposed by ontologies.
2. Link with partners of the ecosystem (see following Figure). First you build a connection, then match data demand and data offer and afterwards start interpreting the data.
3. Control the access to your data by making use of the usage control.
4. Start to create value. This can be done by implementing apps, granting remote software execution or aggregation of information.

[33] https://www.internationaldataspaces.org/

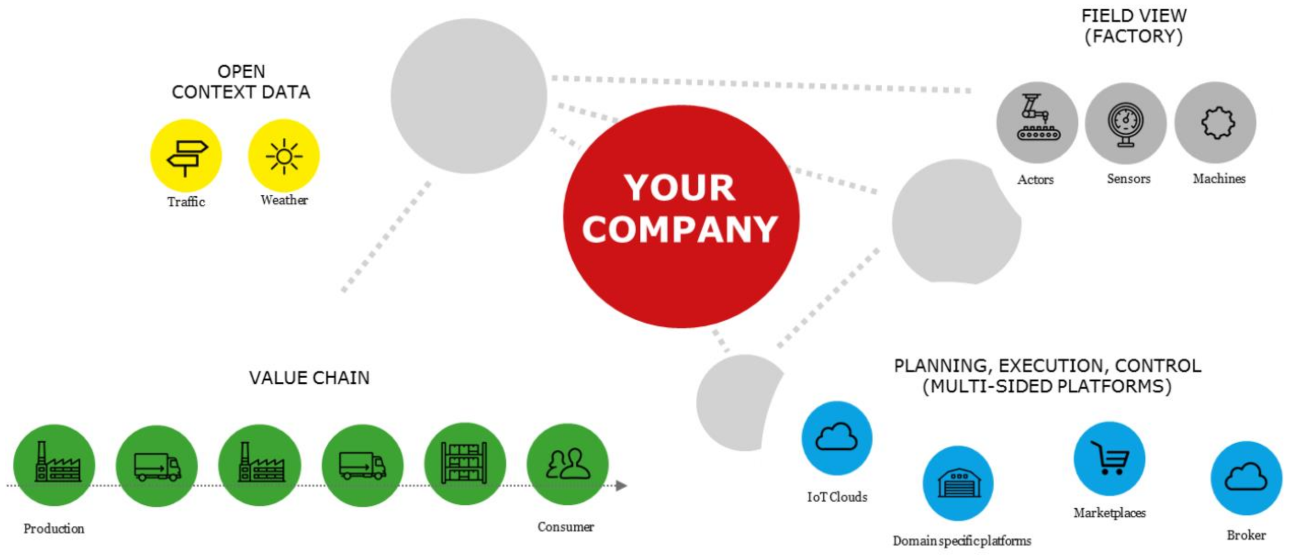


Figure 14. IDS Ecosystem architecture

To conclude this study, the IDS emphasizes **data management and data ownership**, elements that we will take into account in the **design of the Data Integration and Management component and in the Open Access Data repository**.

4.4 Blockchain-based architectures

4.4.1 NIST DataBlockMatrix

Blockchain technology makes it possible to guarantee the reliability of on-chain information, since due to its technological design, the stored information is impossible to alter. This allows many applications in the field of CRFS, from product traceability, guarantee of origin, certifications, to keeping a history of transactions to achieve a digital twin of a value chain.

However, a very important problem that is preventing the official use of blockchain as a repository of official or personal information is the fact that it does not directly comply with the GDPR, by not allowing the deletion of information that may belong to a citizen or legal entity. The difficulty in eliminating the information lies in the technical conception of this infrastructure, where the blocks that make up the blockchain are connected by means of hashes, as shown in the following figure:

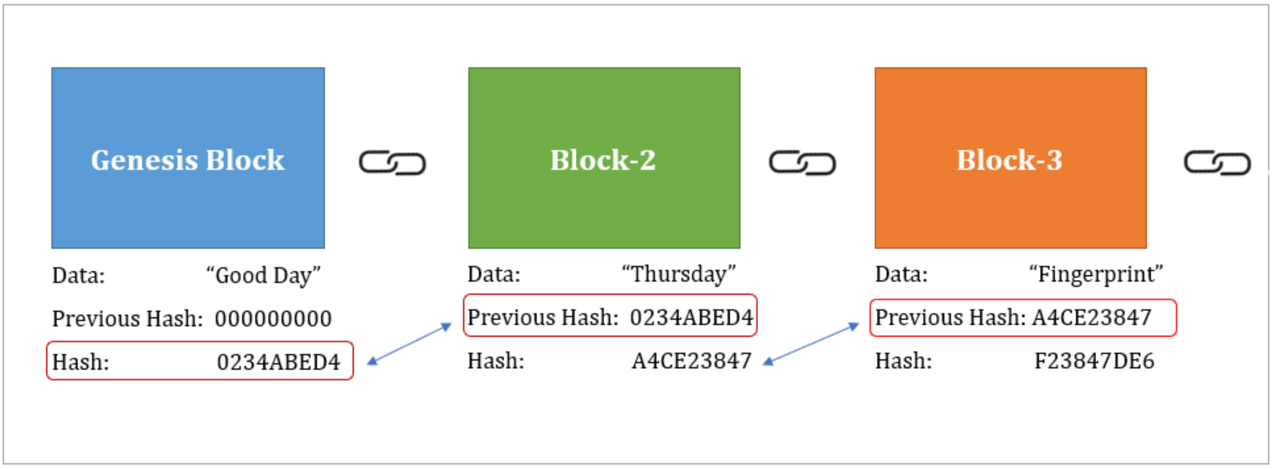


Figure 15. Blockchain hashing example

A substitution of a block to delete or modify the information that is included, generates a change in its hash, and would cause an inconsistency with the information that the other blocks have about it. This would produce the need to change the entire chain, making all the information stored later unusable.

To mitigate these problems, solutions are appearing that allow the deletion of records, such as the one offered by NIST in White paper of May 2022 NIST CSWP 25[34]: DataBlockMatrix.

The datablock matrix data structure may have utility for incorporation into applications requiring integrity protection that currently use permissioned blockchains. This capability could for example be useful in meeting privacy requirements such as the European Union General Data Protection Regulation (GDPR), which requires that organizations make it possible to delete all information related to a particular individual, at that person's request. The data structure provides hash-based integrity protection for a distributed ledger, while also allowing controlled deletion of records. It could be used in a private blockchain, where the goal is to provide integrity assurance that can be verified by all parties. One benefit could be auditability, making it faster and easier to verify transactions and audit the distributed ledger.

In the following table we incorporate all these differences in relation to other distributed management applications:

Cryptocurrency property	Requirement in CRFS
Partial anonymity	ID required for contracts or government regulation
Public access / transparency	Controlled access
Small transaction size	Range of message sizes up to large documents, images

[34] White Paper NIST CSWP 25: A Data Structure for Integrity Protection with Erasure Capability: https://csrc.nist.gov/publications/detail/white-paper/2022/05/20/data-structure-for-integrity-protection-with-erasure-capability/final

Immutable records	Changes and deletions, often required by law
Proof of work	Flexible consensus models
Block ordering guarantees	Timestamps often required
Decentralization	Same in many applications
Replication	Same in many applications
Data integrity guarantee	Same in many applications

We incorporated the NIST DatablockMatrix in the study of deliverable D6.1 because we have identified the following data requirements that can be solved with this technology:

- Privacy: Regulations such as the California Consumers Privacy Act (CCPA) and the General Data Protection Regulation (GDPR) for Europe require much stronger security for personal data, and that system owners must delete all personal data based on a user's request or completion of a transaction.
- Access control complexity: Modern access control often uses rules that depend on data from sources outside the organization, requiring high performance networks with data integrity guarantees.
- "Internet of Things" and ubiquitous sensor nodes: Datasources can include building sensors, smart watches, medical sensors, and many other sensor types.

As a summary, we present the advantages of using DataBlockMatrix over blockchain:

Blockchain – provides integrity, immutability

- No erasure possible, by design
- Double-spend problem solved by distributed timestamp / sequencing guarantees
- Sequencing guarantees require proof of work algorithms
- Proof of work extremely slow, by design

Data block matrix – provides integrity, erasure

- Integrity protection guarantees for all blocks not erased
- Verified timestamps instead of sequencing guarantee
- Greater range of consensus algorithms available, suitable for permissioned distributed ledger
- Very fast consensus algorithms can be used

4.5 Project specific data platforms

4.5.1 DataBio

DataBio Overview

DataBio (Data-driven Bioeconomy)[35] is a H2020 lighthouse project focusing on utilizing Big Data to contribute to the production of the best possible raw materials from agriculture, forestry, and fishery/aquaculture for the bioeconomy industry to produce food, energy and biomaterials, also taking into account responsibility and sustainability issues.

DataBio has deployed state-of-the-art Big Data technologies taking advantage of existing partners' infrastructure and solutions. These solutions aggregate Big Data from the three identified sectors (agriculture, forestry, and fishery) and intelligently process, analyze and visualize them. The DataBio software environment allows the three sectors to selectively utilize numerous software components, pipelines and datasets.

DataBio has been driven by the development, use and evaluation of 27 pilots using market-ready or near-market-ready ICT, Big Data and Earth Observation methods, technologies, tools, datasets and services.

DataBio Platform

DataBio uses the innovative ICTs and information flows centered mostly around the use of proximal and remote sensors to provide a streamlined Big Data Infrastructure for data discovery, retrieval, processing and visualizing, in support of decisions in bioeconomy business operations. The platform supports the forming of reusable and deployable pipelines of interoperable components thus extending the impact of DataBio to new bioeconomy projects as well as to other business areas.

The DataBio platform consists of a development environment, software components used and developed by DataBio partners and pipelines connecting the components to services. DataBio platform provides:

- **Platform**: A big data platform provided as a toolset and framework, which offers functionalities primarily for the domains of agriculture, forestry, and fishery.
- **Toolset**: A big data toolset which offers functionalities as reusable assets primarily for services in the domains of agriculture, forestry and fishery. The functionalities enable new software components to be easily and effectively combined with open source, standards-based big data, and proprietary components and infrastructures.
- **Components and datasets**: Platform components and datasets with links to documented interfaces enabling them to be combined into data pipelines. There are 91 components used in the pilots.
- **Pipelines**: Supports the forming of reusable and deployable pipelines of interoperable datasets and components thus extending the impact of DataBio to new bioeconomy projects as well as to other business areas.
- **Discovery**: Helps to discover new platform assets i.e. pipelines, individual components, datasets, and tools provided by DataBIo platform that you could utilize in your bioeconomy projects. Browse the background documentation provided in DataBio deliverables.
- **Containers**: Operating system level virtualization, known as containerization, is an operating system feature in which the kernel allows the existence of multiple isolated user space instances. DataBio platform assets link to container catalogs enabling federated use of multiple deployment platforms.

DBV Reference model

[35] DataBio webpage: https://www.databio.eu/en/

DataBio has developed a common reference model for Big Data application in Big Data Value Association TF6 SG6 Standardisation group[36]. The BDVA reference model, that in a sense can be considered as a light architecture, serves as a reference for implementing DataBio pipelines, services and pilots. DataBio uses this reference model as its reference architecture, enabling pilots to find right platform components in their implementations.

The BDV Reference Model serves as common reference framework to locate Big Data technologies in the overall IT stack. It addresses the main concerns and aspects to be considered for Big Data Value systems. It is structured into horizontal and vertical concerns:

- Horizontal concerns cover specific aspects along the data processing chain, starting with data collection and ingestion, reaching up to data visualization. It should be noted that the horizontal concerns do not imply a layered architecture. As an example, data visualization may be applied directly to collected data (data management aspect) without the need for data processing and analytics.
- Vertical concerns address cross-cutting issues, which may affect all the horizontal concerns. In addition, verticals may also involve non-technical aspects (e.g., standardization as technical concerns, but also non-technical ones).

Components

DataBio partners provide 91 software components that cover the full range of architectural components in the BDV Reference Model. All the platform components have been categorized and configured according to the BDV Reference Model.

[36] BDVA Task Force 6: Technical: https://www.bdva.eu/task-force-6

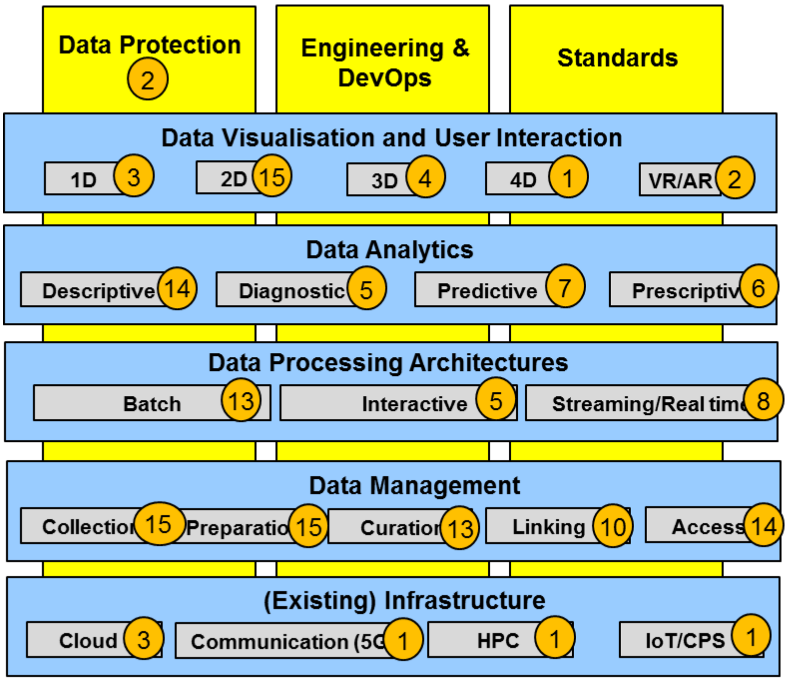


Figure 16. BDV Reference Architecture: number of DataBio components in each class

Most components are hosted by their providers with APIs available. E.g., component A retrieves satellite images, component B stores/fetches the images into/from a database, component C analyses them, and Component D visualizes the analysis results.

A central role of the DataBio Development Platform is to define interfaces of these APIs and to collect the API calls into collections. Such API management is one of the cornerstones for development of Big Data-based applications, where knowledge is obtained from a flood of data.

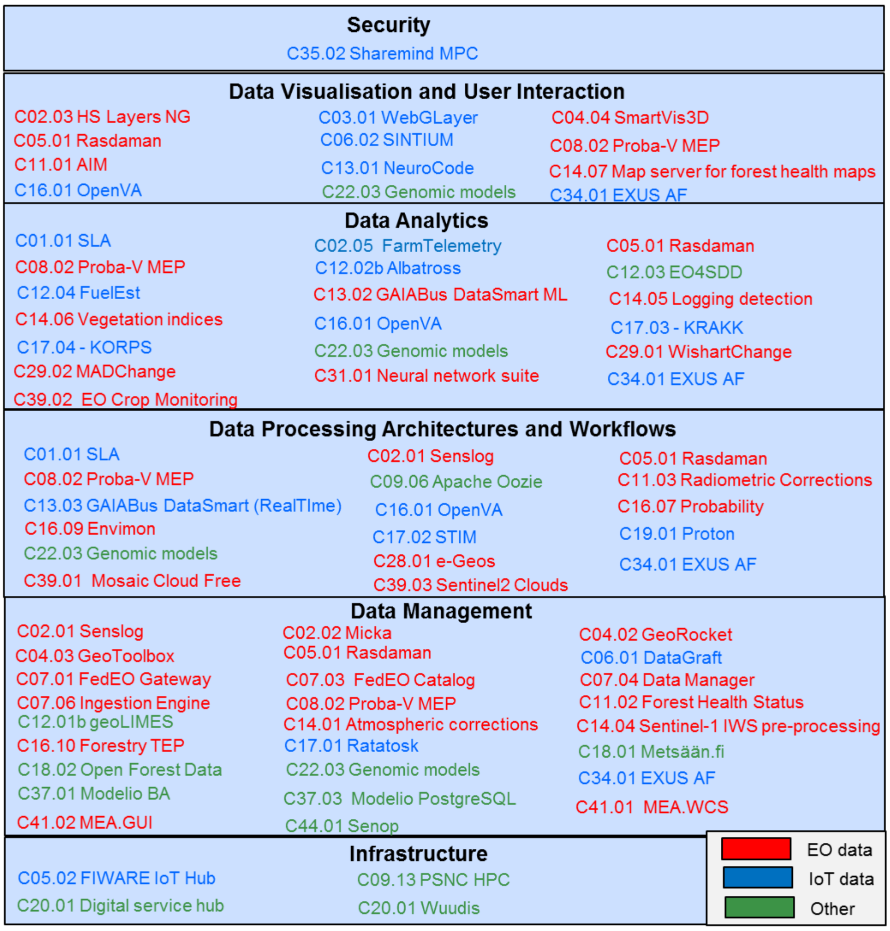


Figure 17. Classification of DataBio components according to the BDVA Reference Model

DataBio is pursuing the use of standards to ensure that its multitude of components can be easily combined and interoperated.

Pipelines

DataBio provides services to users that typically need to know the usability of the service, but do not need to understand the inner wiring, the inner components, nor where the service is deployed. These services are typically accessed via standardized interfaces like APIs, e.g. web services, library interfaces, interactive user interfaces, standard data transfer, or remote call protocols. The services are implemented by pipelines.

Generalized/abstracted pipelines are building blocks used by several pilots. The pipelines are especially applicable in real-time data streams flowing through the sequential stages from source to display of insights.

The pipelines are developed and deployed on a number of infrastructures: bare hardware; virtual machines, containers and other virtualized systems; on private infrastructure or data centres, on public cloud or via cloud services.

Datasets

The datasets were identified based on the needs and the requirements of the platform and of the 27 pilots in the domains of agriculture, forestry and fishery. The datatypes of these datasets are quite diverse; they vary from structured data, semi-structured data, or unstructured data to new generation Big Data. The latter include sensor data (IoT data, Drone data, data from hand-held or from mounted optical sensors), machine-generated data (produced by ships, boats and machinery used in agriculture and in forestry), geospatial data (Earth Observation data from various sources and geospatial data from EU, national, local, private and open repositories), and genomics-relevant data categorized as genomic, transcriptomic, phenomic, metabolomic, farm data, in-situ IoT sensors and other environmental datasets, genomic predictions and selection data from plant breeding efforts and biochemical data collected from tomato fruits and sorghum grains. Historical data have also been used in the pilots.

A significant part of the project pilots uses EO (Earth Observation) data, often coming from the Copernicus Sentinel constellation, as input for their specific purposes, in the context of efficient resource use and increasing productivity in agriculture, forestry and fishery, as well as other types of data.

The datasets have been classified as follows:

- Existing Datasets utilized in DataBio pilots (14 datasets);
- Existing Datasets that have been improved in the DataBio project in terms of easier or better findability, accessibility, interoperability or reusability (6 datasets);
- New Datasets created by the DataBio project by collecting new data or by combining or processing existing data sources (43 datasets).

DataBio Development Infrastructure

DataBio Development Infrastructure offers Hadoop technology that enables collection, organization and analysis of Big Data. The cost-effective storage and scalable processing allow systems to adapt easily-to-fit user requirements, processing scenarios and any data volumes.

The most typical large data processing and analysis scenario involves gathering and storing data for further processing. Therefore, the DataBio Development Infrastructure provides Apache Hive, which implements the well-known and widely used concept of a data warehouse. In addition, the typical warehouse technologies are in the infrastructure complemented by the tools like Apache Storm and Apache Kafka, allowing to process continuous data streams, and the next generation framework Apache Spark that implements a highly efficient data processing model.

Data analysis can be done in the DataBio Development Infrastructure in various ways, taking advantage of the wide range of tools. Apache Pig offers a simple and high-level script language for analyzing large data sets. The efficiency of the script execution is achieved by massive parallelization. Simple statistical analysis, grouping and summaries can be generated using Apache Hive. More sophisticated analysis involving classification, clustering, dimensionality reduction, collaborative filtering and topic models can be done using Apache Mahout. Users can also provide their own software, implementing specific business logics using tools, API's and frameworks (like Spark) provided by the infrastructure.

DataBio Development Infrastructure manages Big Data by providing non-relational, distributed and scalable databases like HBase, Accumulo or Cassandra. Each of them implements a variety of data models and offers a range of functionalities allowing the platform users to choose a solution that is best for their application. The platform also allows relational databases to be used.

DataBio Hub

The components, pipelines and pilots are described online in the DataBio Hub[37] that makes it easy for the developer to navigate among the DataBio entities. DataBioHub supports collaboration between DataBio partners helping to integrate their components as part of DataBio platform and publish reusable pipelines utilizing them as part of DataBio platform for larger bioeconomy community.

The DataBioHub:

- contains descriptions of DataBio components, pipelines and pilots and their mutual relations;
- makes the DataBio data findable as it supports their discovery by i) publishing the metadata according to best practices and standards (geospatial and others), and by ii) applying search keywords (=tags) to the digital objects;
- the actual data are accessible through the DataBioHub by consulting the dataset owner;
- contains information about the APIs, the data model and formats as well as about the access methods;
- promotes interoperability as the metadata (and many times the data) conform to established standards, e.g. in the Earth Observation (EO) field;
- enables reusability, provided that licensing schemes are in place.

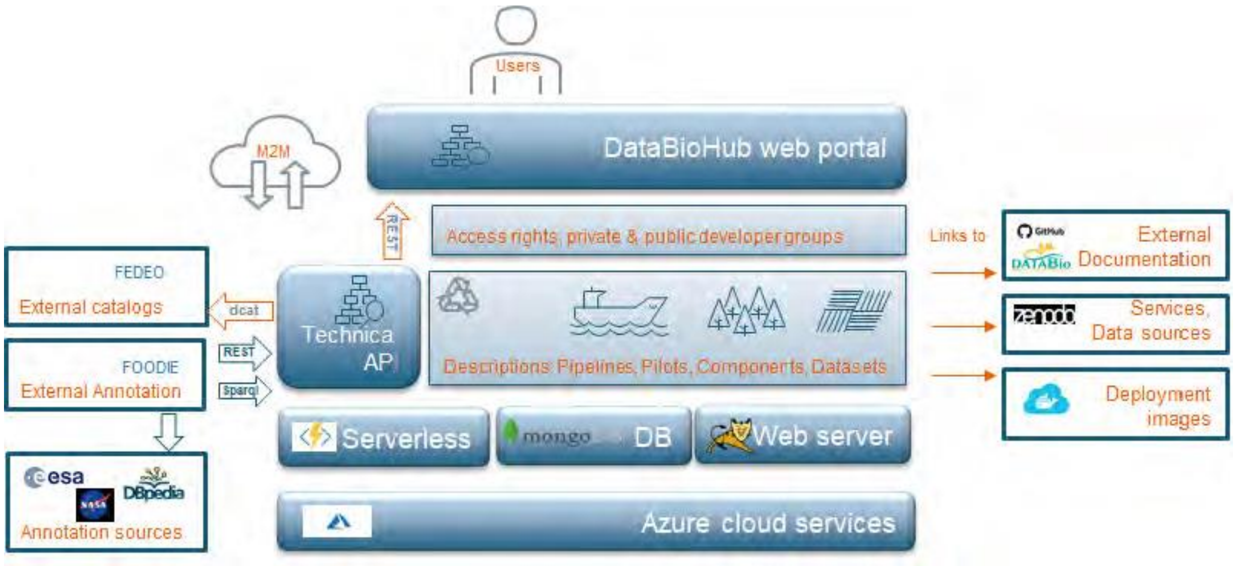


Figure 18. Architecture of DataBioHub

4.5.2 IoF2020 (Internet of Food & Farm 2020)

The IoT has a revolutionary potential. A smart web of sensors, actuators, cameras, robots, drones and other connected devices allows for an unprecedented level of control and automated decision-making. The project

[37] DataBio Hub webpage: https://www.databiohub.eu/

Internet of Food & Farm 2020[38] explored the potential of IoT-technologies for the European food and farming industry.

IoF2020 supported a rich set of use cases across the dairy, meat, arable, vegetables and fruit sectors and provided an interoperability reference architecture to support these specific use case sub-projects[39]. The interoperability approach is shown in the following figure.

The architecture defines a set of interoperability points to enable the use cases to interoperate.

[38] IoF2020 webpage: https://www.iof2020.eu

[39] IoF2020, D3.3 Opportunities and barriers in the present regulatory situation for system development, https://www.iof2020.eu/deliverables/d3.3-opportunities-and-barriers-in-the-present-regulatory-situation-for-system-development-v1.2.pdf

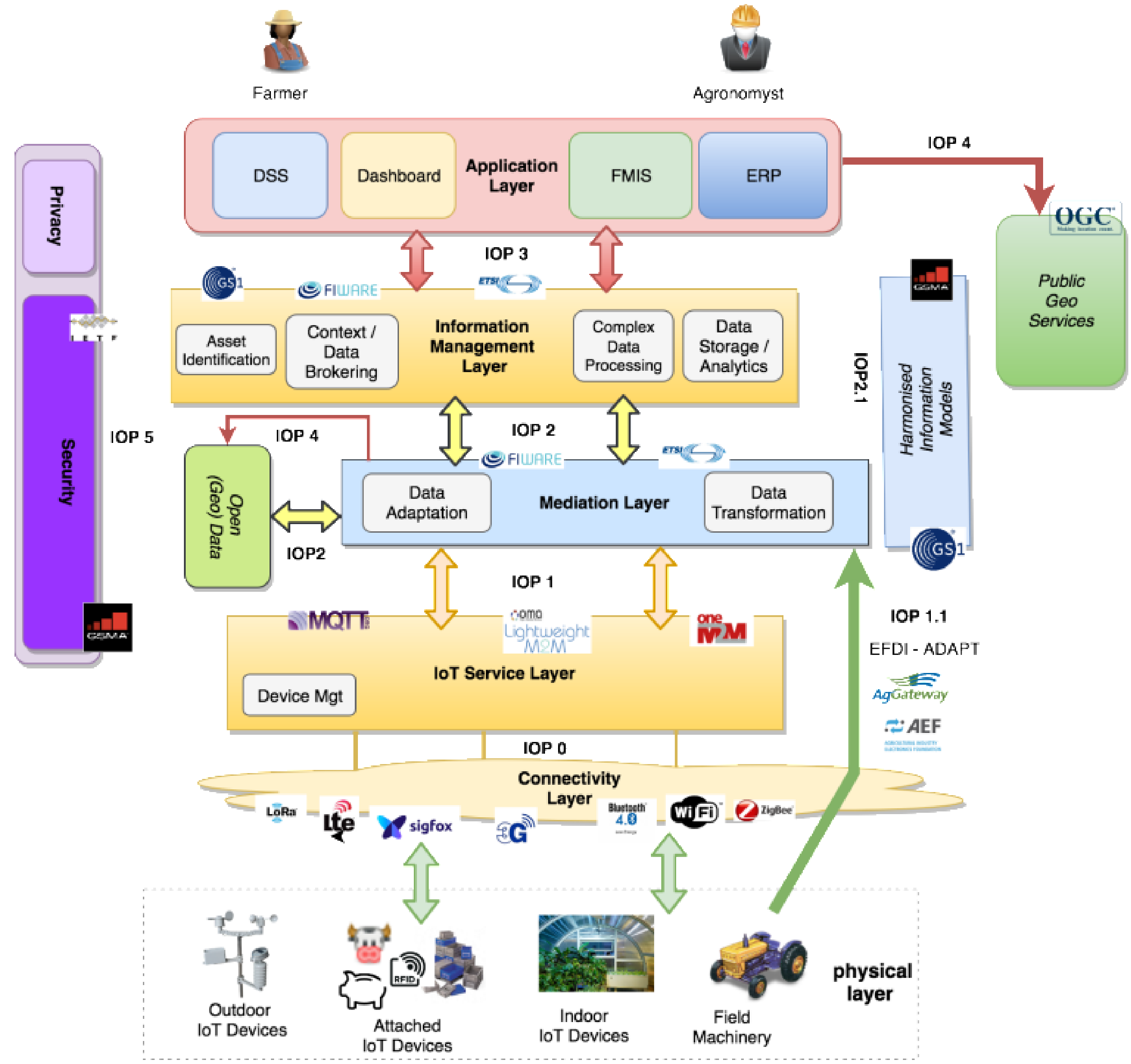


Figure 19. IoF2020 Interoperability reference architecture

IoF2020 IoT Service interoperability Layer

The IoT service layer interoperability is depicted as IOP 1 in the above diagram. The IoT Service Layer exposes the raw data generated from IoT Devices through different application-level transport protocols based on different paradigms (publish / subscribe, request / response, etc.). It exposes different interfaces based on the use-case requirement such as stream-processing or sense-actuate. The service interoperability layer sits between the physical devices and mediation layer. The standard technologies that enable the IOP 1 include MQTT, OMA LWM2M and oneM2M, as these are the most relevant ones for the farming and food domain.

IoF2020 Data Interoperability Layers

Data interoperability points are seen in the above diagram as IOPs 2, 2.1 and 3. The goal being to set standards for data exchange in the Agrifood industry between devices, machinery and cloud services and also between cloud services and external APIs. IoF2020 used the following set of standards to achieve this:

- **FIWARE NGSI** and its evolution to an ETSI standard, **NGSI-LD**, provide a solution to information management and data exchange, enabling the transformation, aggregation, harmonization and publication, as context information, of harmonized data coming from IoT Devices, agricultural machinery or other sources of information.
- The work developed by the **GSMA IoT Big Data Project** on data harmonization (data models) plays a relevant role to achieve data portability, together with **GS-1** standards.

IoF2020 Geo-Services Interoperability Layer

The geo-services interoperability is depicted in above diagram as IOP 4. It utilizes two interfaces - WFS (Web Feature Service) and WMS (Web Map Service) to achieve geographical data interoperability in the software solution. A brief description of WFS and WMS is as below:

- WFS: It is an interface specified by the Open GIS Consortium (OGC) that allows for the exchange of geographic data across the Web. It supports the vector data model and uses Hyper Text Transmission Protocol (HTTP) along with Extensible Markup Language (XML) transformation for information exchange and operations on geographic features.
- WMS: It is a specification that outlines communication mechanisms allowing disjoint software products to request and provide preassembled map imagery ("compiled" map images, which may contain both vector and raster data) to a requesting client.

In both the interfaces the connectivity between components and other modular layers is enabled via API (application program interface) exposition.

IoF2020 Security Interoperability Layers

The security interoperability points are depicted as IOP 5 in the above diagram. The security interoperability consists of three elements in the IoF2020 solution:

- Data Access (Authorization and Authentication): This ensures that only eligible users have access to the system and there are role-based views for each of the users in the system.
- Infrastructure Security: The infrastructure security is enabled through physical devices by enabling machine level physical protocols combined with application-level protocols with correct user-access.
- Platform Security: Different service developed might need different levels of security measure. This point ensures different security requirements based on their scope, including confidentiality, integrity, authentication, authorization, immutability, trust, and non-repudiation.

4.5.3 DEMETER

DEMETER Overview

The DEMETER Project is a large-scale deployment of farmer centric interoperable smart farming-IoT based platforms delivered through a series of 20 pilots across 18 countries (15 States in the EU). DEMETER aims to lead the Digital Transformation of the European Agri-food sector based on the rapid adoption of advanced technologies, such as Internet of Things, Artificial Intelligence, Big Data, Decision Support, Benchmarking, Earth Observation, etc., to increase performance in multiple aspects of farming operations, as well as to assure the viability and sustainability of the sector in the long term. DEMETER has defined a Reference Architecture[40] that will be used to guide the development of the technologies in all work packages that will be used for the first round of the DEMETER pilots. A complete analysis and description of the reference architecture is provided in the project deliverable D3.1[13].

DEMETER Concepts

DEMETER is built around the following main concepts:

- DEMETER Stakeholders Open Collaboration Space (SOCS)
- DEMETER Agricultural Interoperability Space (AIS),
- DEMETER Enabler HUB (DEH
- DEMETER Dashboard

These are shown together with a description of the concepts in the following diagram.

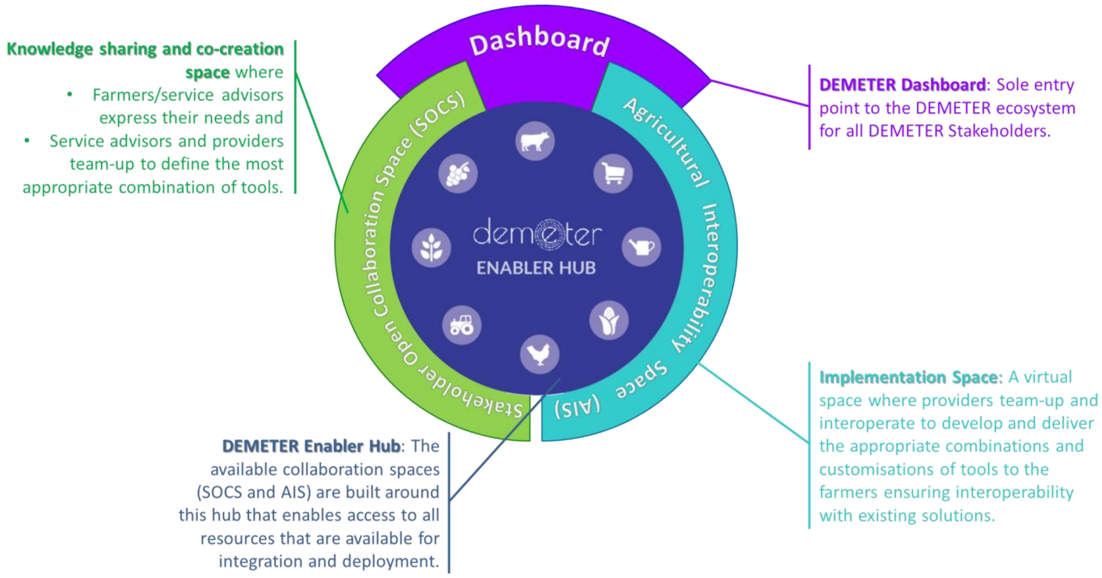


Figure 20. DEMETER Concepts

DEMETER Reference Architecture

DEMETER proposes an overarching approach that integrates heterogeneous technologies, platforms, and systems, while supporting fluid data exchange across the entire agri-food chain, addressing scalability and governance of ownership. The proposed architecture consists of a DEMETER Provider and a DEMETER

[40] D3.1 DEMETER Reference Architecture (Release 1), https://h2020-demeter.eu/wp-content/uploads/2020/10/D3.1-DEMETER-reference-architecture_v1.0.pdf

Consumer service based on the architecture model introduced by the Industrial Data Space (IDS)[41], then further specified by the International Data Space Association (IDSA)[42], which is the continuation of IDS.

As data interoperability is of such high importance, the proposed solution provides the necessary data translation mechanisms combining the use of a semantic data model called Agriculture Information Model (AIM), developed by DEMETER, along with the respective data translation, management and inference mechanisms adopting widespread standardized solutions (e.g., NGSI-LD[43], Saref4Agri[44], ADAPT[45]).

4.5.3.1 DEMETER Functional View

To implement the high-level view of the architecture and to implement the objectives and vision of it, DEMETER provides several facilities & modules that interact with each other, with the necessary stakeholders as well as with existing devices, platforms, systems, and data sources. In the following figure, the main functional blocks of the DEMETER Reference Architecture are shown, along with the external entities involved and their respective interactions; these constitute the functional view of the architecture.

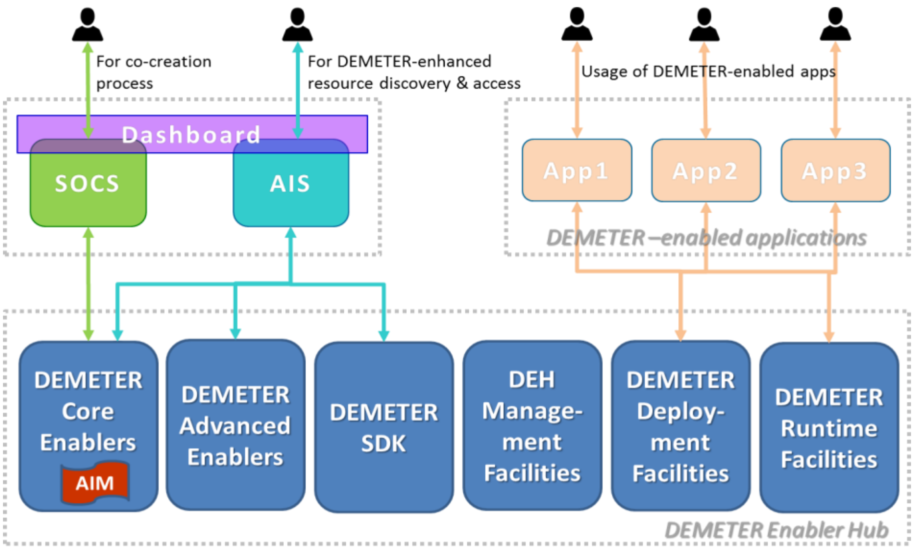


Figure 21. Functional Blocks of DEMETER Reference Architecture

4.5.3.2 DEMETER Architecture Instantiations

DEMETER provides in its reference architecture a set of 19 architecture instantiations, one for each of its first set of pilots ranging across a multitude of scenarios including irrigation, dairy, pest control data brokerage

---

41 http://www.industrialdataspace.org/
42 https://www.internationaldataspaces.org/
43 ETSI. Context Information Management (CIM); NGSI-LD API. Available online: https://www.etsi.org/deliver/etsi_gs/CIM/001_099/009/01.01.01_60/gs_CIM009v010101p.pdf
44 ETSI. SmartM2M; Extension to SAREF; Part 6: Smart Agriculture and Food Chain Domain. Available online: https://www.etsi.org/deliver/etsi_ts/103400_103499/10341006/01.01.01_60/ts_10341006v010101p.pdf
45 ADAPT - Agricultural Data Application Programming Toolkit, Homepage URL: https://adaptframework.org/

among many others.

4.5.3.3 DEMETER Data View

The DEMETER Data View primarily consists of two components:

- **DEMETER Data & Knowledge Repository:** This component targets the problem of the physical representation of the semantic model. It creates a mapping between different communication, networking, and digital resources which generates the information flow between the business processes.
- **DEMETER User & Resource Registry:** This component represents a relational database management system that structures data in tables with certain properties. The database will contain information relating to users, credentials, access controls and views etc.

The diagram below brings together the above concepts, such as entities, information and their interactions within business processes, in order to structure the data flows between them:

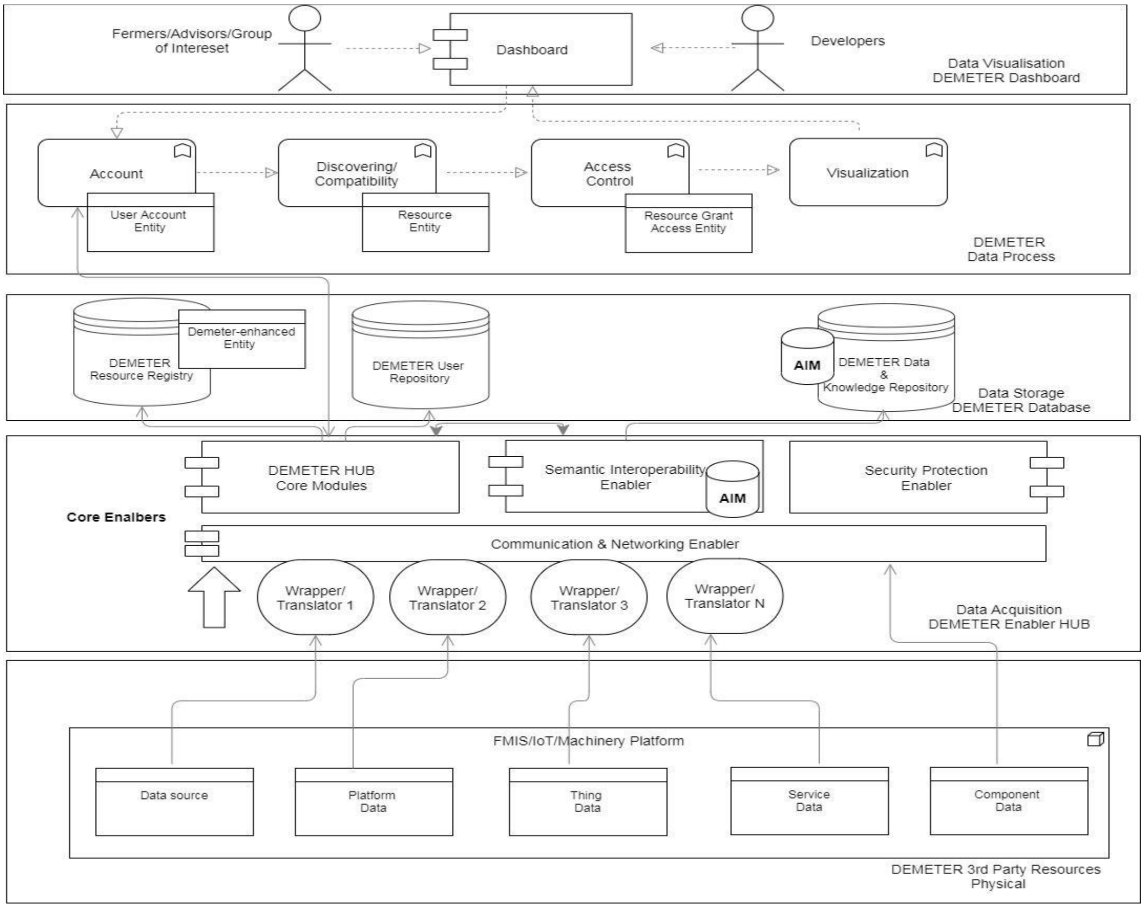


Figure 22. DEMETER Main Data Flows

4.5.4 PoliRural

PoliRural provides a set of knowledge resources including an inclusive learning environment where rural populations, researchers and policymakers come together to address common problems. The project uses text mining to assess the perceived effectiveness of past or planned policy interventions and studies the development trajectory of agriculture and its allied sectors until 2040 using several scenarios in which the evolution of rural populations occupies a central place.

As a result, PoliRural leaves decision makers at different levels of government better equipped to tackle existing and emerging rural challenges, rural populations more empowered and rural areas more resilient.

The following sections detail the different components of PoliRural architecture.

PoliRural Text Mining

This component produces a semantic tree which can be explored interactively on the PoliRural platform (Digital Innovation Hub). The text mining can help to identify significant parameters of change, for example related to population dynamics, value creation, employment, cost of living, quality of life etc. The system dynamic modelling can help people understand the ways in which they interact and evolve over time, and how they can be manipulated by policy intervention and community action.

A generalized pipeline of Text Mining for Foresight Process is given in the figure below.

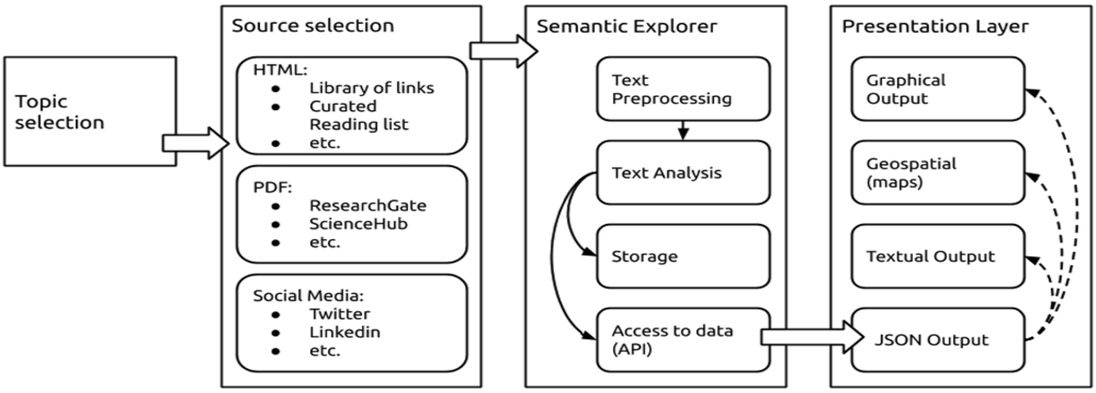


Figure 23.PoliRural text mining pipeline

Topic Selection: produces a collection of topics that define the successive stages of the Text Mining Pipeline. For example, a user wants to examine the dynamics of a set of selected performance indicators in the area of Landscape Development is "Landscape Development" (a topic).

Source Selection: Depending on what the end-user wants to achieve, it selects hypertext or PDF files, uploaded documents or Social Media posts.

Semantic Explorer: provides end-users with textual and quantified information, automatically extracted from the big corpus of texts.

Presentation Layer: The information obtained from Semantic Explorer is presented as responses from a set of API endpoints in JSON format and its derivatives (for example, Geospatial information is presented as GeoJSON). Those documents can therefore be interpreted later as markup textual output (HTML, MD, etc.), graphics (charts, diagrams, etc.) or maps (heatmap, multi-polygonal areas, etc.).

Streaming from Social Media

Streaming from Social Media is a process of registering data from platforms such as Twitter and LinkedIn in the Local Database to be available on the Text Mining Pipeline. A general Source Selection of the pipeline of the Streaming from Social Media is represented in the figure below.

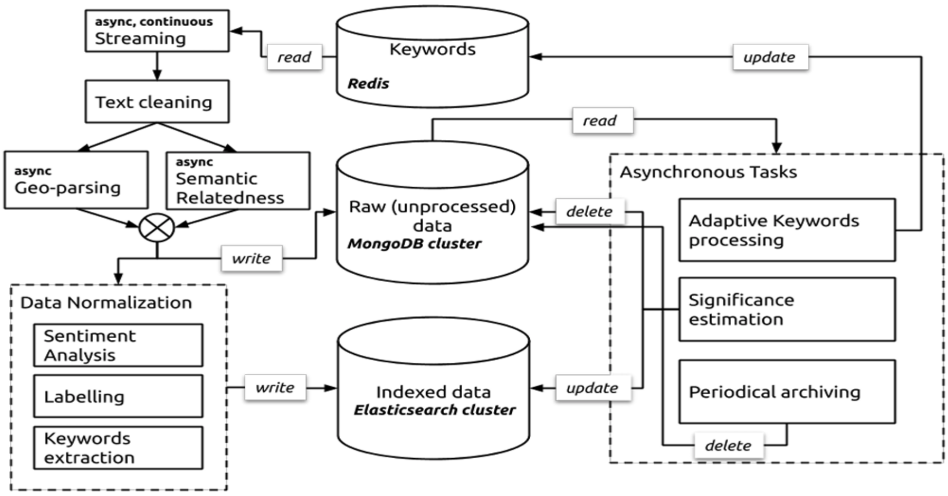


Figure 24. Polirural streaming from social media pipeline

Text cleaning: 1) Removal from the original text of all symbols, stop processing words, etc. that are not relevant to text processing and 2) Removal of personal sensitive information (names, phone, NER numbers, emails, etc.). Techniques: NLP, lemmatization, etc.

Geoparsing: This stage is required to filter posts that do not contain information about geolocation(s). Such posts cannot be used for the analysis of Pilot areas and therefore should be discarded.

Semantic Relatedness Estimation: This stage is necessary for filtering out useless tweets or posts from SM platforms. The number of raw posts streaming from Social Media platforms are forecasted to come in enormous quantities, while only a fraction of them can serve as a valuable input for Semantic Explorer.

Data Normalization: is a process of converting an original post from SM (tweet, post from LinkedIn, etc.) to the unified form for feeding to clients of API or/and live feed (end-users). Normalization includes Sentiment Analysis, Labelling and Keywords Extraction.

Asynchronous Tasks: The main purpose of asynchronous tasks is the selection of the number of SM posts already saved in the Local Database for additional processing, and (if necessary) deletion or update, based on some common information.

Semantic Explorer

Semantic Explorer consists of a set of API endpoints. The outputs of one service can be an input for another, which suggests a concept of a pipeline. For example, the result of NER is used by the Geo-parser, which returns coordinates and geoshapes by geonames. Likewise, a Profiler uses results of both NER and Topic Extraction to define a context, in which it should look for meaningful characteristics of the subject (see Figure below).

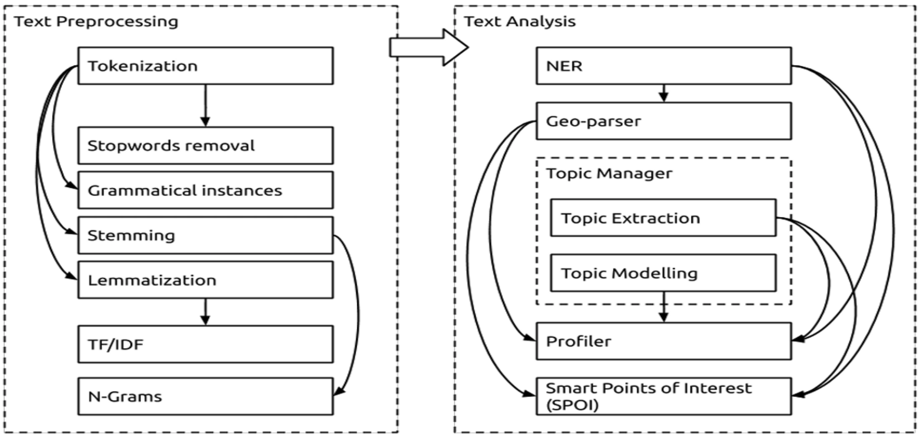


Figure 25. Polirural Semantic Explorer Pipeline

The Deliverable PoliRural D2.1 Text Mining Technical Specifications[46] contains more details about this component.

4.5.5 PoliRural Digital Innovation Hub

The aim of the PoliRural Digital Innovation Hub (DiH) is to offer a public user interface and introduction to the innovations of the project. To this end, the DiH entry point is built on a content management system. Furthermore, the value of the DiH will be to provide four distinctive sections, or spaces, that cater for both internal and external users:

1. An interaction space with forums, dialogue and Wiki capabilities to support stakeholder interaction.
2. A learning space for Massive Open Online Courses to facilitate dissemination and uptake of knowledge and methodology developed through the project.
3. An experimentation space for testing analytics and visualization including text mining and system dynamics based on real data.
4. A development and hosting space for creating virtual instances of the shared reference to be used by each pilot when developing their applications.

The technical platform of the innovation hub is described as a reference architecture that can be realized in three different ways.

1. As a central service provider, shared data and components node.
2. As a centralized pilot node hosted in the PoliRural private cloud.
3. As a decentralized pilot node hosted in a legacy environment.

This approach permits the innovation hub architecture to be applicable to both new and legacy IT environments. We believe this approach is superior to a make-or-break monolithic system that must be identical in every way, shape and form.

[46] Polirural deliverable D2.1: https://polirural.eu/wp-content/uploads/2021/11/D2.1..pdf

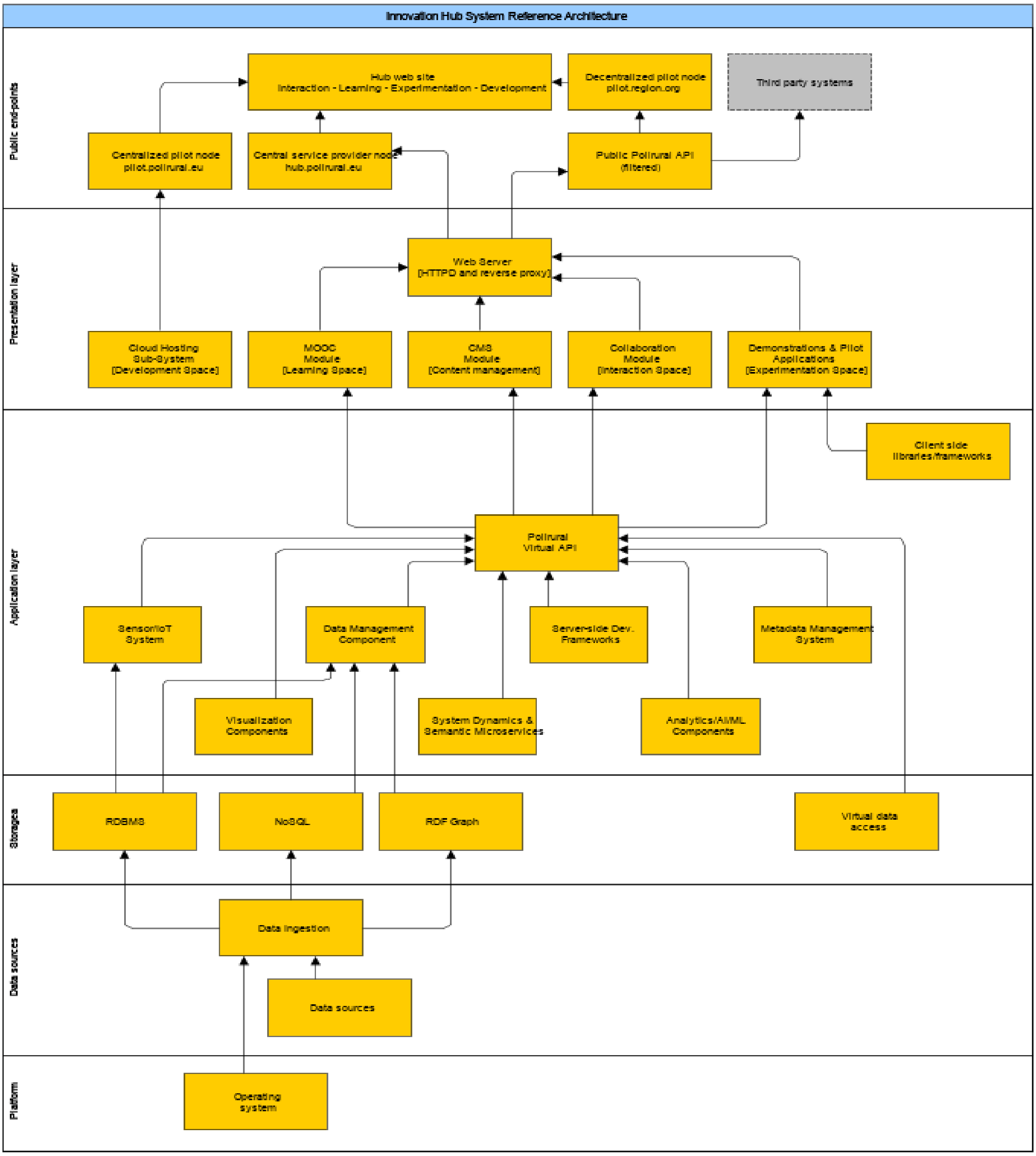


Figure 26.Polirural Innovation Hub System Reference Architecture

The PoliRural Deliverable D3.1 Innovation Hub Technical Specification[47] describes in detail this architecture and deployment options.

4.5.6 REDIRNET

Introduction

The call for greater cooperation and interoperability between different emergency services at scenes of major crisis events has been present for some time. With the increasing level of complexity of events and the increasingly developed specialisms of experts this need for interoperability will continue to grow and require more effective communication and information exchange, facilitated and supported by a range of sophisticated modern technologies and IT systems.

The Socio-professional networking component, shown the following figure, provides a decentralized and self-building opportunity for an interoperability network. It also allows the operation of the interoperability network without major operational costs, since the collaboration rules are set by the agencies themselves according to basic rules established by the REDIRNET project in the light of the interoperability issues that emerge in the first phase of the project. Collaboration rules are set according to mutual agreements between the agencies involved and should cover issues such as the mutual visibility and controllability of data fields, data streams and switches. The concept also covers end-to-end encryption of the data between the agencies, with the agencies themselves choosing encryption algorithms and keys. Thus, interoperability is protected, even if the infrastructure is operated by a third-party commercial entity.

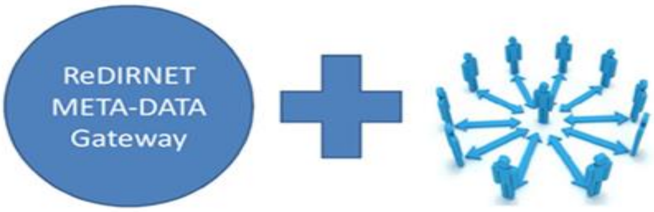


Figure 27. REDIRNET socio-professional networking

The REDIRNET project focused on inter-agency first responder data communication and data sharing. The basic idea was to allow the agency representatives to set up access permissions for their resources and to share their sensor data with partner agencies. The concept allows a partner agency to access the shared sensor data in the same manner as the data owner. Depending on the sensor capabilities, there are two types of data connection between the agencies: a request-response connection is used for batched data, while a stream connection is established for a continuous data flow. It is assumed that the partner agency consuming the data is capable of processing the incoming data.

In the example shown in the following figure, an agency B is offering data stream from a Sensor 2 and a Camera 1 to an agency A. Although it is not shown in the diagram, it has previously been agreed between the two agencies that the Sensor 2 and the Camera 1 of agency B can be made visible to agency A using the REDIRNET socio-professional network, which provides seamless interoperability, allowing (following

[47] Polirural D3.1: https://polirural.eu/wp-content/uploads/2021/11/D3.1..pdf

appropriate configuration) Sensor 2's data and Camera 1's image from agency B to be displayed in agency A's control center as Sensor 4 and Camera 3. The same information can also be displayed on Agency A's field officers' handheld devices.

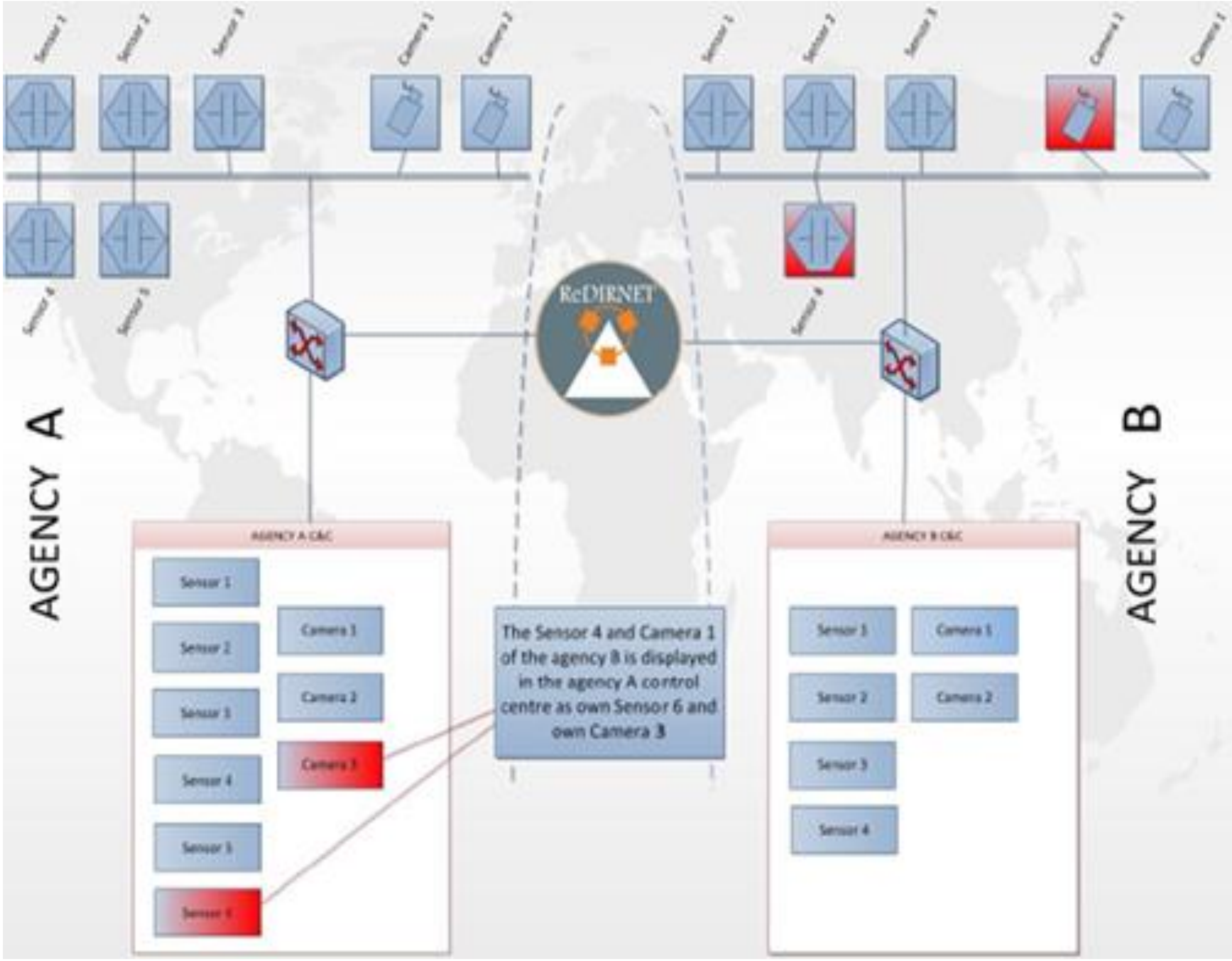


Figure 28. REDIRNET system concept

The second component of the REDIRNET concept was the Gateway, which acts as a middle layer between the third-party plug-ins and REDIRNET Core. Each sensor is required to implement the REDIRNET Gateway plug-in interface to be able to provide and/or consume local or remote resources. The REDIRNET Gateway is supported by REDIRNET Core services. Specifically, the services allow REDIRNET Gateways to be registered and authenticated, the publication of local resources, searching for remote resources and requesting access to remote resources. The REDIRNET Core includes a User, Logs and Permission database component used as a persistence layer for agencies' assets and inter-agency partnership management. This component is connected to an Ontology service, which provides a semantic description of sensors and search capabilities. Both components (collectively named Core Data Storage) serve as the data storage backend for the Main Switch component, which encompasses all the functionality of the REDIRNET Core. The following figure shows the relationships between the components.

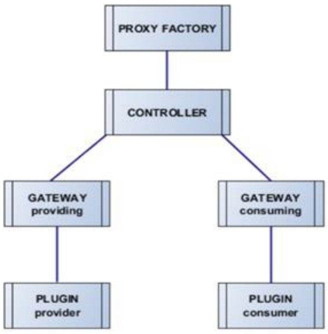


Figure 29. REDIRNET architectural components

The Gateway and Plugin components can act with two roles: providing and consuming. Currently, the Gateway is implemented as a hybrid, i.e. both functionalities in one executable file. Complementary providing and consuming plugins are implemented separately in two independent executables.

How Components Work Together

This section shows how components and commands work together when a resource is transmitted.

The process of resource transmission is split into two separate phases. The first phase (next Figure left) is dedicated to command flow, while the second (next Figure right) handles data flow.

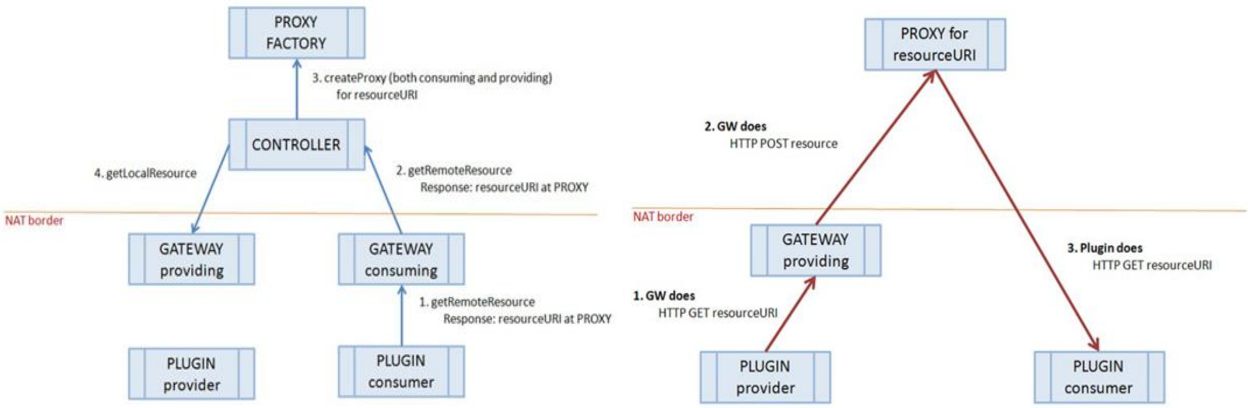


Figure 30. (Left) Command flow. (Right) Data flow

1. Consuming Plugin sends rdn:GetRemoteResource command to Gateway.
2. Consuming Gateway pass rdn:GetRemoteResource command to Controller.
3. Controller evaluates resource request and sends response back to Gateway, which propagates response down to Plugin. If the request is evaluated as valid, the Controller then:

- orders Proxy Factory to create Proxy for specific URI, which accepts data posted from providing Gateway and simultaneously exposed resource for consuming counterpart.
- sends URI of the providing resource in getRemoteResource response back to the consuming Plugin.

- sends URI for posting resource in rdn:GetLocalResource method into providing Gateway.

PHASE 2 - Data flow

1. Providing Gateway retrieves resource from providing Plugin via standard HTTP GET method.
2. Providing Gateway pushes resource via HTTP POST method of Proxy to URI specified in command getLocalResource in step 3 of command flow phase.
3. Consuming Plugin retrieves resource from Proxy via standard HTTP GET method from URI specified in response to getRemoteResource command performed in step 1 of command flow phase.

Conclusion

The research projects outlined above have demonstrated without a shadow of doubt that using IP technology and more specifically IPv6 for critical communications can benefit immensely from the extensibility of IP networks, with new features such as auto-configuration and end to end access to name a few. The IPv6 implementations are relatively new in this sector but the American Public Safety Forum has already mandated IPv6 over 4G for the US safety networks. Europe has a few safety initiatives such as BAPCO and the European Safety Forum but they have not been very vocal in this area.

REDIRNET project finished before new technologies such as Blockchain were developed. The project successfully reached the aim to secure data exchange between various critical infrastructure organizations. Data were not stored in one place; special secure agents were able to find appropriate data in each database. Data interoperability can be added value for Cities2030 project.

4.5.7 Fed4IoT

Fed4IoT[48] project focused on providing scalable and interoperable Smart Cities applications, by introducing novel IoT virtualization technologies. It presented and delivered VirIoT, the Fed4IoT platform that achieved a set of distinctive virtualization goals as listed below:

- Virtual Things: This extends the concepts of virtual machine in the cloud counterpart to IoT. Virtual Things are a smart proxy to IoT sensors or actuators, capable of complex interaction with real things or other producers/consumers of information.
- IoT interoperability and virtualization: This in the platform is achieved through utilizing the concept of broker-based system architecture. It provides access to the data items managed by the Virtual Things through existing, standard IoT broker technologies, among which the developers can freely choose, rather than doing through the overhead of understanding a new IoT layer.
- Virtual Silos/IoT systems-as-a-service: This is the combined isolated environments made of Virtual Things and an IoT broker.

The project was a step forward in the area of IoT virtualization and made novel contributions in the field. The project finished before upcoming technologies such as Blockchain could be experimented in the Fed4IoT

[48] Fed4IoT Deliverable D2.3: https://fed4iot.org/wp-content/uploads/2020/11/D2.3.pdf

framework. It successfully delivered the underlying Fed4IoT platform showcasing the need of IoT virtualization, and its benefits combined with its cloud counterpart.

5 S2CP platform architectural design

Once the requirements that the S2CP platform must satisfy and the technological mechanisms available in the current state of the art to carry out the corresponding implementations are known, in this section we describe the final design proposed for the platform. In Section 5.1 we give a general and high-level description of the proposed design for the platform. Section 5.2 analyzes in detail each of the modules that are part of the general architecture. Finally, Section 5.3 reviews how these components correspond to the work foreseen in each of the WP tasks, giving a prioritized order for the development of the different software modules.

5.1 Proposed architecture: general overview

Given that different Labs have different needs, at the same time that all Labs demand certain functionalities across the board, for the S2CP platform we are going to choose a component-based architecture (similar to those reviewed in Section 4.5: DataBio, DEMETER, PoliRural and REDIRNET).

Each component is defined as an independent software unit, although all components may share common mechanisms that facilitate their use by Labs, such as a registration and login service with the same credentials.

In this way, Labs can obtain a personalized instance of the platform, composed only of those components that are of interest to them and cover the technological needs of their activities.

To avoid duplication, some components may be linked, such as the data analysis components whose results may be displayed on the dashboard together with the general statistical data requested by Labs.

In addition, the platform is linked to external sources, able to obtain data from social networks or web repositories, directly or indirectly feeding (through the data repository component of the S2CP platform itself) the different analysis and visualization components.

On the other hand, the platform is not only customized for each Lab through the selection of differentiated components, but also through the adaptation of the components to the specific situations and particular needs of each of the laboratories that require it. This is the case, for example, of Blockchain-based Marketplaces that can be adapted according to the use cases that arise in each of the laboratories.

The following figure shows the proposed architecture for the S2CP platform, where components belonging to four subsystems are identified (Data integration and analysis, Service ecosystem, Privacy and Security and Adaptive Visualization Dashboard), which correspond to the four development tasks of WP6 (T6.2 to T6.5 respectively). The S2CP platform defines 17 components whose detailed description can be found in Section 5.2.

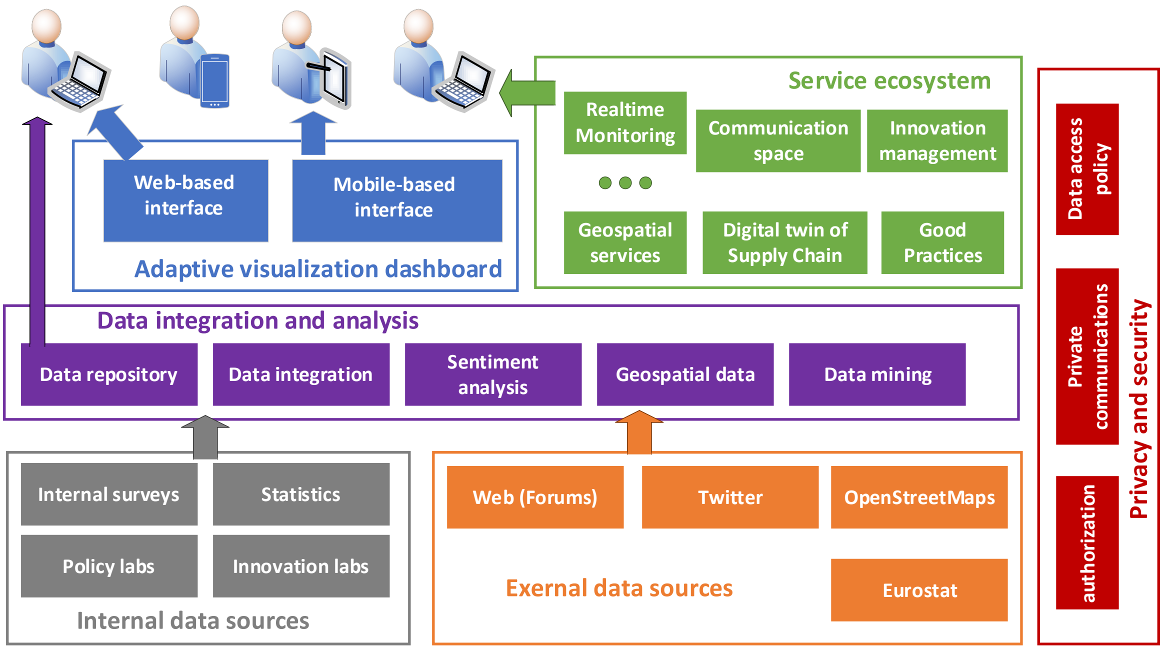


Figure 31. S2CP proposed functional architecture

5.2 S2CP components

In this functional view we detail the main components that are included in the S2CP platform along with the technologies, algorithms, etc. that are supported.

5.2.1 Sentiment analysis: Identification of correlation and causal relationships in data

Analysis of social networks data can be very useful to understand opinions, requirements and intentions of the users and consumers. In this component, we perform sentiment analysis of data collected from Twitter. To achieve reliable results, this component uses techniques from the theory of causality and Causal Machine Learning (causal ML). The high-level architecture of the component is shown in the figure below.

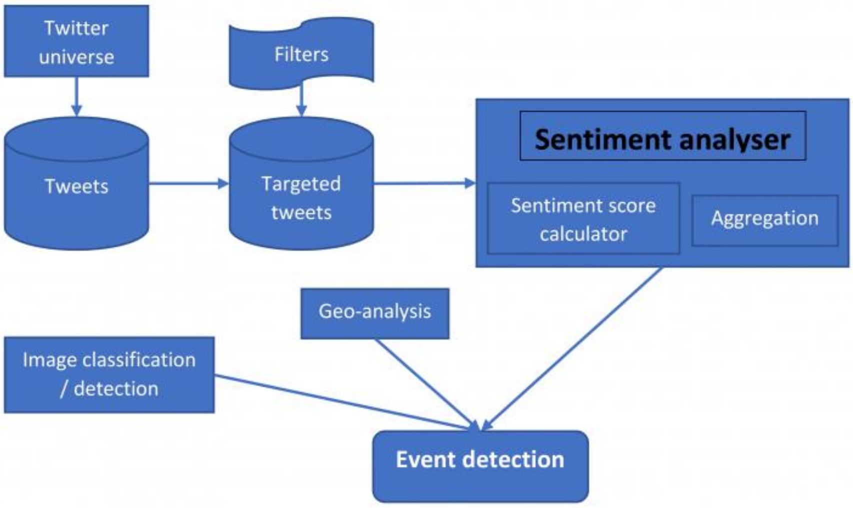


Figure 32.Architecture of the sentiment analysis tool.

First, we collect the data that corresponds to the queries of interest from Twitter. After that, the collected data is filtered and processed with sentiment analysis module. We use several libraries to perform sentiment analysis of data in different languages and aggregate results at the end. As the result, we get a number representing the relevant sentiments. This can be visualized using historical and geographical distribution plots. Additionally, we perform image classification to identify the presence of certain items in the images. The tool can also track the occurrence of keywords of interest in the texts of the tweets. All this information is at the end combined to detect events that can be potentially useful for understanding the interests of the users and to shape the policies of the tool operators.

Traditional Machine Learning (ML) techniques, including those designed for event detection, are based on analysis of correlations. To reduce this problem, we additionally employ the theory of causality at the event detection stage. Among such techniques are the analysis of causal graph with *do*-calculus, PC-algorithm for constructing the skeleton of the causal graph, and Additive Noise-based modes for orientation of certain type of edges in the graph.

We also aim to use image detection techniques to identify face expressions and further infer the relevant sentiments and opinions of the users. The latter task can be accomplished with traditional image classification algorithms based on artificial neural networks and trained on relevant datasets.

The designed functionality can be extended to analyze information from other social networks as well, such as Facebook. To achieve this, one needs to connect to another social network through the relevant Application Programming Interface (API) and transform the obtained data into the format accepted by the component.

5.2.2 Data prediction with conformal learning

Conformal prediction is a framework that produces predictions with accuracy guarantees. For a given value of significance level ϵ, a conformal predictor is guaranteed to make exactly ϵ errors in the long run. That is, the resulting region prediction contains the true value with probability 1-ϵ. Practical value of such a tool can hardly be overestimated, as the generated predictions also provide reliable uncertainty quantification based on solid mathematical theory.

A schematic representation of a Conformal predictive component is presented in Figure of the right. As in the traditional predictive framework, the dataset is split into training and test sets. The latter set is used as in the traditional ML, that is for evaluating the predictor. The training dataset, however, is further split into proper training and calibration sets. The proper training set is used to train a predictive model, and the calibration set is used for its calibration and estimation of uncertainty of the predictor. The trained model and the user-defined nonconformity function are used to calculate calibration scores on the calibration set. These calibration scores are at the end used to generate region predictions that will produce no more than ϵ errors.

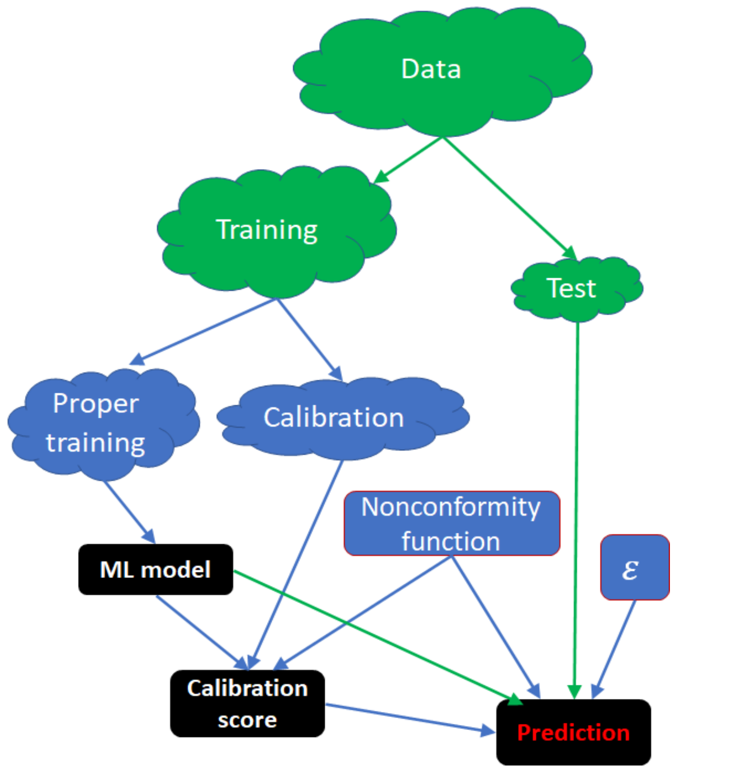


Figure 33. Architecture of a Conformal predictive component.

As we can see, the Conformal Learning framework has only 2 parameters: the nonconformity function and the allowed error rate ϵ. The value of ϵ is usually application specific. However, the choice of nonconformity function has a significant impact on the efficiency and performance of a conformal predictor. Theoretically, any function, even a random function, can be used. But the proper choice of nonconformity function will result in Conformal predictors with the desired characteristics.

Conformal prediction exists both for classification and regression tasks and it can be wrapped around any predictive algorithm. Thereby, the user is required to provide a dataset and the value of ϵ, choose the type of prediction problem and the underlying prediction model (or models), and a nonconformity function. As the result, the component generates region prediction with predefined accuracy guarantees. The documentation of the component provides guidance on the choice of different parameters with respect to the objectives of the user.

5.2.3 Data mining platform

Monitoring trends in the perception of how desirable food groups are allows municipal authorities to know how they should best target their food-related campaigns. To assess such a trend, it is necessary to perform sentiment analysis on food-related information obtained from Internet sites such as social networks, personal blogs or Internet forums. The algorithm that allows extracting these trends for each food group is summarized in the following figure:

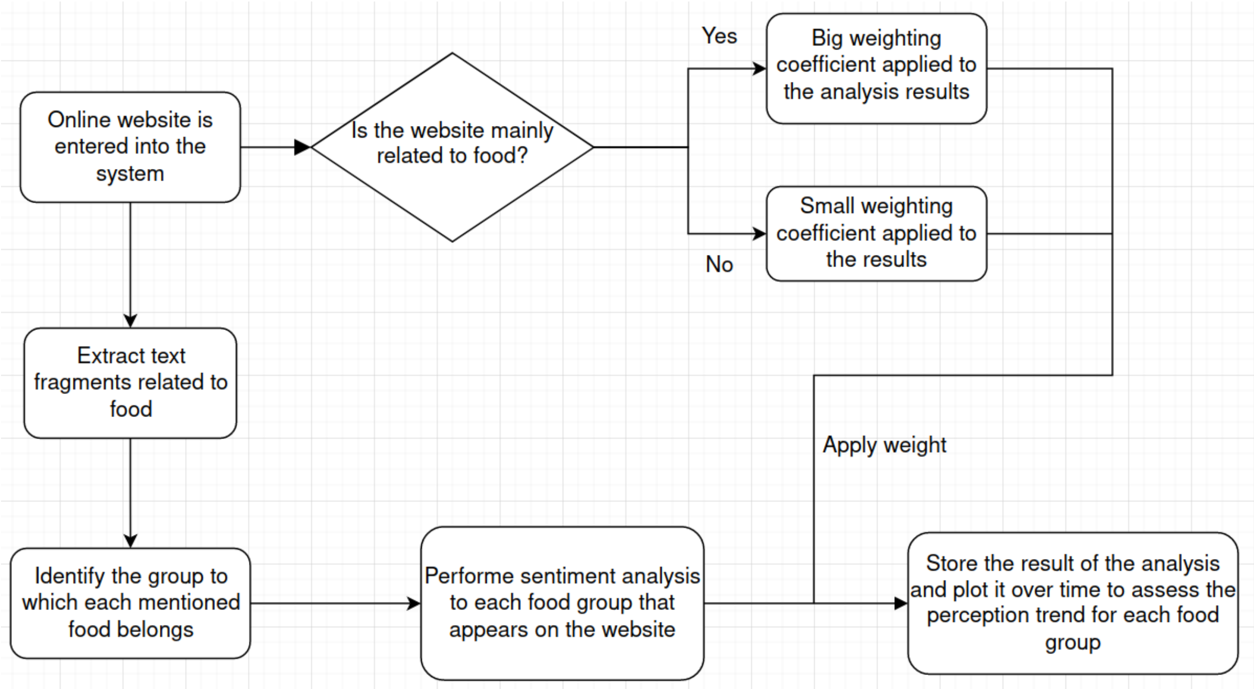


Figure 33: food perception trend follower algorithm flowchart

As we can appreciate in the figure, the online website link from which we are obtained data must be introduced into the tool, and then the analysis is automatically performed.

After the website has been entered, we can see that there is a fork in the figure. On the right part of the fork, an analysis is fulfilled that permits the website to be classified in two different options, according to its relevance for this tool. In this manner, how much food-related content is on the website is identified and based on this parameter, the website is classified as high relevance or low relevance, which implies giving a different weighting coefficient that will give it more or less importance to the result of the sentiment analysis.

So, on the other part of the fork, the sentiment analysis itself is performed. To do this, first those parts of the text in which content related to food appears are extracted. These mentioned foods are then categorized into their corresponding food group, such as meat, fruit, or fish. Consecutively, sentiment analysis is applied to the extracted food-related text fragments. Now, the weighting coefficient obtained from the aforementioned other part of the fork is applied to the result of the sentiment analysis, and finally the results are stored in variables corresponding to the food group to which the food that appears in the text belongs. The results of each new website entered into the tool are always stored and added to the old ones, allowing a trend to be evaluated with each new iteration.

5.2.4 Data Integration and Management

The integration and analysis service is responsible for data acquisition from different resources and putting it in Cities2030 data model format. The data coming for analysis would have different data models, and the integration service will create a central database view to put data from all the sources in data warehouse cloud location.

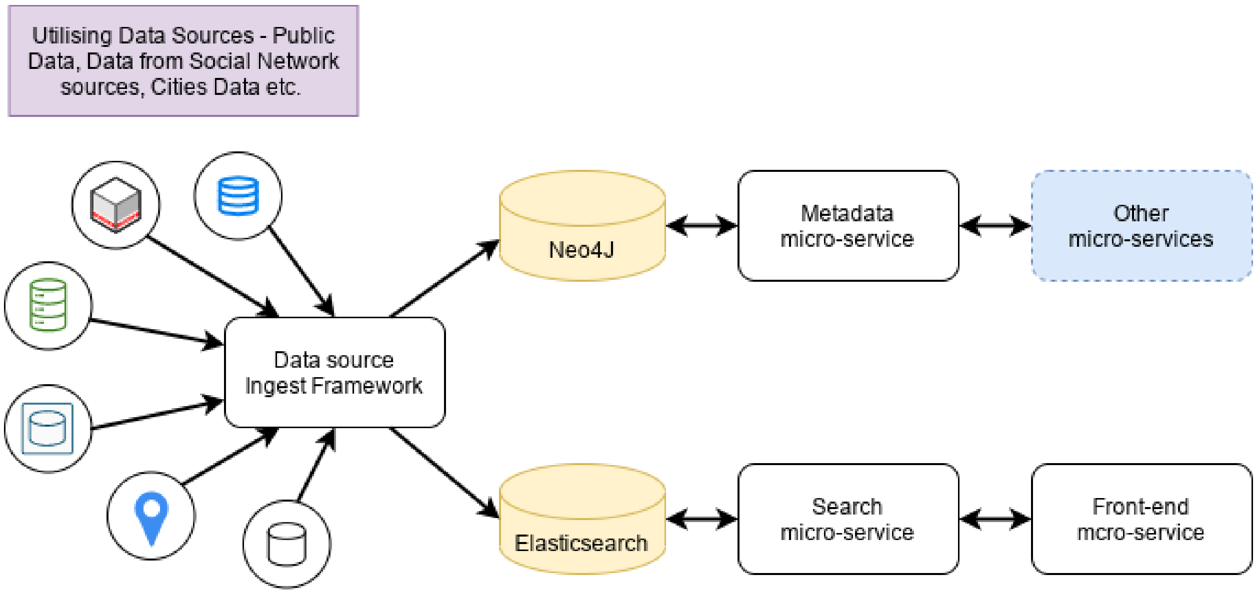


Figure 34: Data integration and Management architecture

The collated data then goes through two modular components to provide insights, and business intelligence to different actors:

- **Data processing:** Data processing consists of transformation required on the data before it is passed in the pipeline to analytical models. The methodology here consists of approaches such as data clustering, featurization and feature engineering.
- **Data Analysis:** The data analytics pipeline consists of approaches such as ensemble learning to deal with the heterogeneity in data to address different objectives of all the actors involved.

As the appropriate data-processing and data analysis algorithms are very domain specific, pre-baked pipelines focused of specific types of questions and handling specific types of data have been chosen from the use-cases. These are specialized and optimized versions of the generic data processing and analysis services, that are pre-baked and ready to be used. These are outlined further in the other sections. The focus of this component is on collating the various data sources, enabling data source cataloguing, providing a Data View to access the data and promoting good data handling by providing supporting functionality for Data Governance component.

5.2.5 Data access repository

Publishing and consuming open data is a cornerstone for the development of applications and the creation of an innovation ecosystem. In this regard, this section explains how Cities2030 data repository has been created, and how users can expose their data publishing it in this S2CP component.

Cities2030 data repository is an open-source solution based on the well-known Open Data Publication network CKAN[49] (Comprehensive Knowledge Archive Network), most widely used by cities, public authorities, and organizations. This repository enables the publication, management, and consumption of open data, usually, but not only, through static datasets (CSV, XLSX, etc.). This allows to catalogue, upload, and manage open datasets and data sources, while supports searching, browsing, visualizing, or accessing open data.

Easy data access

This component incorporates a central keyword search, that can be faceted by tags, location, format, license, publishing organization, etc. The embedded browser allows searching by groups, keywords, and publishers, enabling a standardized interface for viewing datasets and downloading them through links and direct access. Preview and data explorations is possible with some data formats.

Private datasets

This S2CP component allows users to control the visibility of their datasets, enabling the creation of private datasets that only certain users can access. This is core to support the access control and GDPR compliance.

Geospatial information

Data repository has advanced geospatial features, covering data preview, search, and discovery. Where structured data with location information is loaded into the Data Store, it is possible to plot the data into an interactive map. The screenshot shows a map view of a sample dataset, with markers showing individual data points and full details shown for records as they are selected. A user searching for datasets can filter the results by geographical location, specifying a bounding box to limit the interesting areas. Different coordinate geometries and formats are supported. To integrate datasets with other systems, metadata can be coded in INSPIRE standard and major metadata schemas (ISO19139[50] and GEMINI[51] 2.1), including OGC's CSW[52] standard. The architecture is extensible, making it easy to support other standards and distribution services.

Single Sign On

The OAuth2[53] extension allows site visitors to login using an authentication server. In this way, other S2CP component such as Cities2030 Community can be used as the identify provider leveraging the single sign-on approach required for the access control management of CKAN datasets. *This feature is still under development.*

Multilingualism

Thanks to widely stablished CKAN framework, the Cities2030 data repository is translated by into over 10 languages, supporting all international characters and supporting multilingual search, string translations & more for the European Commission Open Data Portal.

[49] https://ckan.org/

[50] https://www.iso.org/standard/32557.html

[51] https://www.agi.org.uk/uk-gemini/

[52] https://www.ogc.org/standards/cat

[53] https://oauth.net/2/

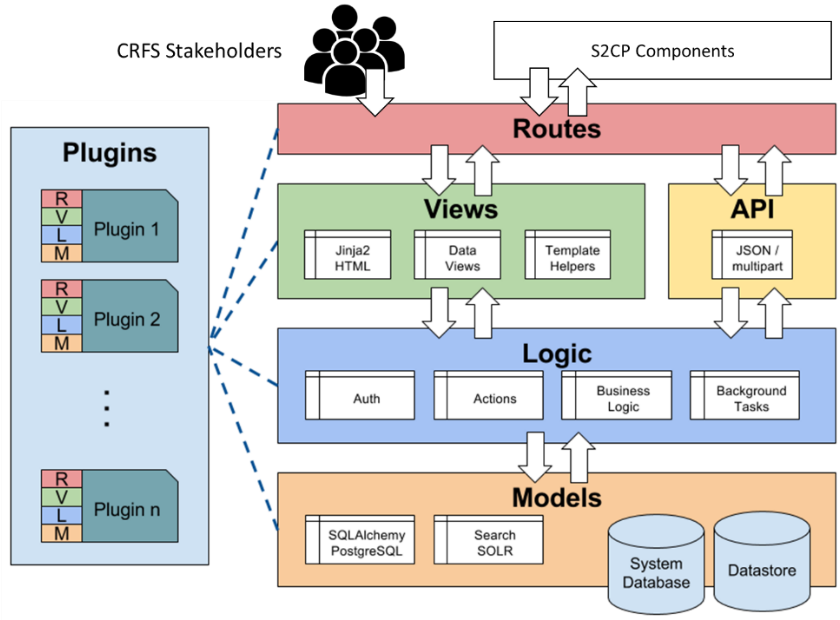


Figure 35.Architecture of the data access repository, adapted from [54].

5.2.6 Geospatial information and services

The Geospatial information and services component is based on state-of-the-art technologies and common and standardized application interfaces in commercial products and in the world of research. This component shows the geographical distribution of the different agents throughout the different CRFS on a map system and can be divided into two entities.

An architectural diagram for this component is shown below:

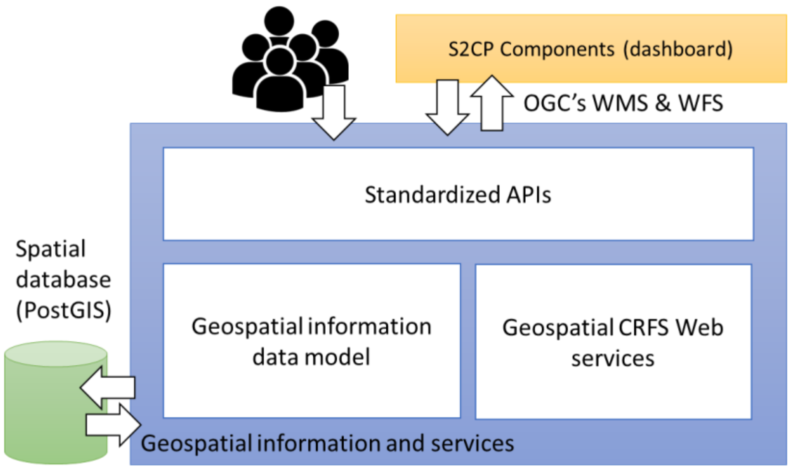


Figure 36. Architecture of Geospatial information and services component

[54] https://docs.ckan.org/en/2.9/contributing/architecture.html

Geospatial Information

This entity collects relevant information from external geospatial sources and stores geographical distribution of various CRFS agents in a relational database, considering many attributes related to CRFS. The following functionalities have been taken into account for the development of this component:

- Consideration of geographical places at city level, currently supporting 10 cities (CRFS labs).
- Definition of workflow for the update of information. The workflow takes into account currently available information from Open Street Maps public database.
- Filtering by amenities and shops, divided into categories, to obtain a spatial vision of the distribution and relevance of locations or points of interest.
- Possibility of personalized visualizations, thanks to the integration into the Geospatial CRFS Web Services of T6.3 and the S2CP dashboard of T6.5.

For each city a couple of datasets are supported according to OpenStreetMaps formats[55]:

- Point-type set (OSM nodes), which represents the point-based elements of the establishments according to some categories.
- Linestring type set (OSM Ways), which are presented in open or closed polygons, generally demarcating an area of an establishment, such as a supermarket.

This entity is described in detail in D6.2.

Geospatial CRFS Web Services

This entity allows that information about different projects managed by the cities within their CRFS can be published in OGC standardized services and cartography. It will therefore be possible to generate maps with different types of visualization, from heat maps to evaluate the density of elements in a certain zone, and the area of influence, to the provision of marker controls such as clusters, to improve the spatial visualization of markers that they are highly concentrated. The following functionalities have been taken into account for the development of this component:

- Filtering by categories, to obtain a spatial vision of the distribution and relevance of locations or points of interest.
- Integration of spatial datasets coming from the Geospatial Information entity.
- Provision of information in standardized APIs, using OGC's standards Web Map Service[56] and the Web Feature Service[57].
- Personalized visualizations, thanks to the integration into the S2CP dashboard (T6.5). Possibility of performing spatial operations or complex analysis in the S2CP dashboard.

5.2.7 Communication space: Cities2030 Community

The Cities2030 Community is an open collaboration space used by Cities2030 participants to improve their multi-stakeholder dialogue processes. The main functionalities of this component are:

- Facilitate community building among Cities2030 partners

[55] https://wiki.openstreetmap.org/wiki/OSM_file_formats

[56] OpenGIS Web Map Service (WMS) Implementation Specification: https://www.ogc.org/standards/wms

[57] OGC Web Feature Service (WFS) Implementation Specification: https://www.ogc.org/standards/wfs

- Facilitate focus group creation (around CRFS topics) and cross-fertilization
- Facilitate giving and receiving feedback and comments among partners
- Disseminate and share information on events, activities and results to target groups
- Gather digital files into Labs' calendar of events, workshops, meetings etc.
- Control the amount and quality of information shared externally

An architectural diagram for this component is shown below:

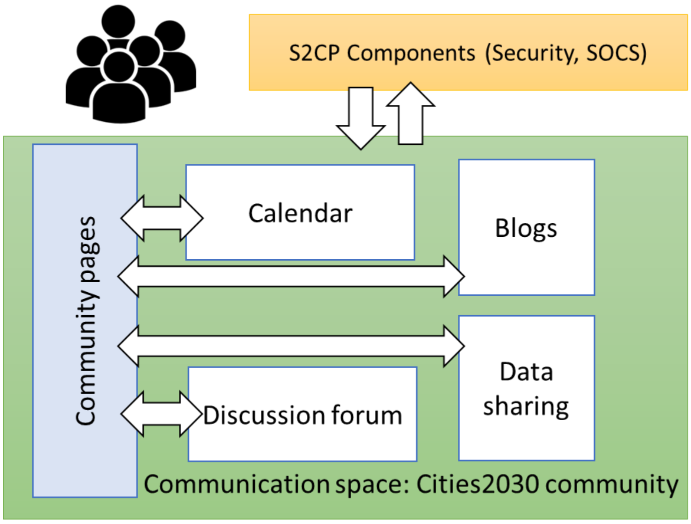


Figure 37. Architecture of communication space

As we can see in the figure, the main functionality, which is transversal to the rest of them is the lab community pages. This allow the CRFS lab to create a social network profile and manage the provision and access to content. This component is accessible directly by CRFS actors, but also have interconnections with other components such as secure communications and other SOCS components.

Connected to this functionality we have the rest of features of the communication space. This connection represents that lab owners and their delegates have administration permissions, and can accept invitations, structure their lab sections, and post calendar events, forum posts, and graphic evidence of innovation activities.

To help in the dissemination tasks and to connect with the efforts that are taking place from WP7, the incorporation of a Blogs tool is proposed, which facilitates the publication of news related to the activities carried out within the framework of the Labs. Finally, it is proposed to incorporate a data sharing functionality by which users can upload their data repositories and share them among all registered users.

Detailed information about this component can be found in D6.3.

5.2.8 Good practices in Food Systems

The good practices component brings a map-based platform to allow the introduction of innovation activities by any Cities2030 partner, so that a catalog of information can be compiled, for filtering, searching and consulting (considering FAO pillars and MUFPP indicators).

This component is planned for collecting information generated in other tasks, such as T3.6, which results in "The 100 Innovation Framework". This framework aims to explore and map the novel trends in several applicable spheres of the food system, identifying emerging digital and technological solutions, start-ups, practices of local communities, findings of international research projects and other initiatives that promote positive transformation in CRFS. More than 140+ innovations and good practices from more than 20 countries were collected and subject to be introduced into the "good practices" component.

Next, we detail the functionalities of this component:

- Authentication in the form of Registration and Login / Log out.
- Different user profiles: administrator / moderator / publisher user / visitor user
- Provision of map for spatial visualization of contributions. The map shows the points with a cluster-based control, so that the markers are aesthetically more uniform in areas of greater accumulation of points.
- Provision of filters by country. Search bar for textual filtering.
- Provision of controls for article organization by date and title.
- Article validation mechanisms, prior to publication: An administrator checks the articles proposed by other users, being able to delete them if they consider that they are not correct or validating them and adding them to the catalog if not.
- Programmed from scratch at the project development time. Free source code and available on Github.
- Deployed in a virtualization server (Cloud system) in facilities of *Universidad Politecnica de Madrid* (P20).

The following figure represents the functional architecture of this solution:

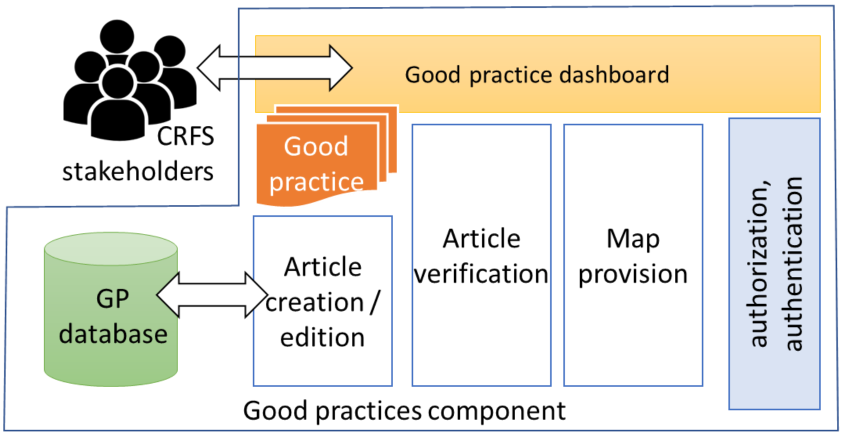


Figure 38. Architecture of good practices component

5.2.9 Innovation management platform (multi-actor approach tool)

The Social Spaces for Research and Innovation – Multi-Actor Approach tool (SSRI-MAA) is a social space monitoring tool. This provides key functionality to manage SSRI and their stakeholders, working groups and actions. It also provides KPIs and data acquisition tools.

SSRI-MAA is built considering a microservices-based architecture, that means it has created independent services for some system modules. The reason for doing this is that sometimes it is better to isolate concrete system components attending to security and scalability purposes. In this section we provide a detailed view of SSRI-MAA Tool architecture, defining each component and giving an overview of the system that lets the tool run.

The implementation of microservices has been done using container technology, owned by Docker Inc. It has been created a container for each microservice and by this way, we control communications and resources consumed by each service. All containers communicate between themselves by Docker private network. Some of them can also send and receive data from the internet (outside Docker network) through open communication ports. For security reasons only required services are connected and accessible from the Internet.

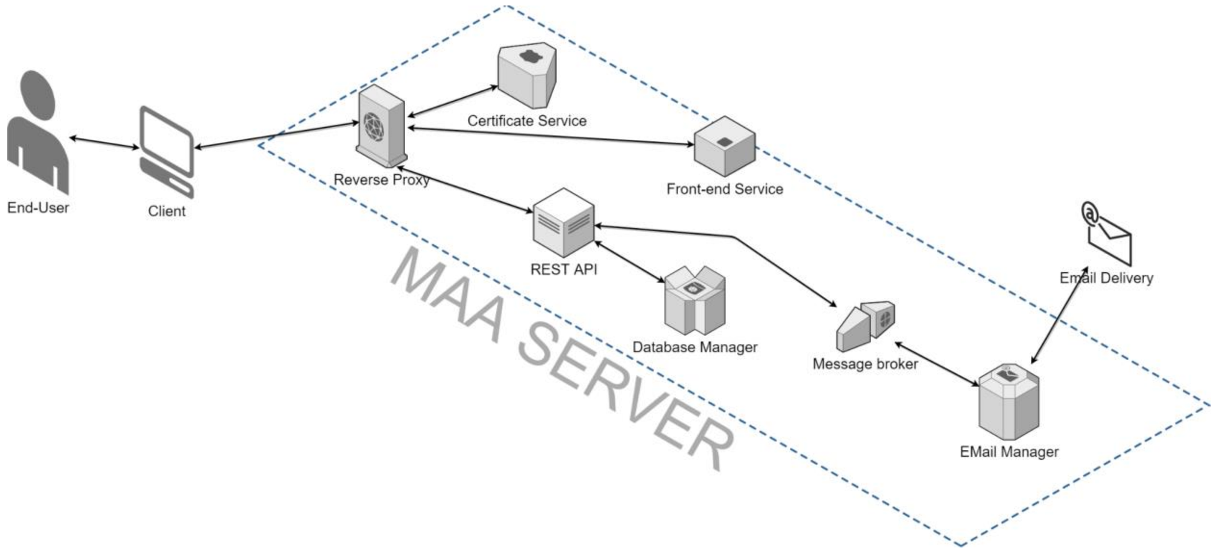


Figure 39. SSRI-MAA Architecture

Architecture components

In this section we describe SSRI-MAA Tool architecture components and their interactions to provide end-user functionality.

- **Reverse proxy component**: Reverse proxy service canalizes the connections to designated services and collaborates with certificate service to enhance the security of communications. When a user enters the URL on its navigator, it connects to the reverse proxy as first step, then the reverse proxy identifies the requested URL and redirects the petition to the required service.
- **Certificate component**: It provides certificates for all domains the tool is using. Once a certificate is generated, this component stores it in a secure directory that is used by the reverse proxy component in order to provide HTTPS secure connections for all domains.
- **Frontend component**: Frontend service provides all client-side implementation and lets end-users execute it in their navigator by accessing the MAA Tool URL. This code includes usability, styles, security and implements the requests to REST API service to exchange user's data with backend.
- **REST API component**: REST API provides a transparent data interface so the frontend service can ask for the backend data to show to the end-users. API REST also connects to the database component to retrieve the business model information.

- **Database manager component**: This component is accessed by REST API in order to get and modify the business model data, that is to say, get and update information about social spaces, stakeholders, actions, users and more. This component is in charge of the persistence of MAA Tool.
- **Message broker component**: There are a set of services that are orchestrated through a message broker as these services need to be scheduled to be processed in a specific date time, like Email System. When a request from the frontend service is calling the REST API, there are some situations where this service could not provide an instant response, for example when a request needs to process a large amount of data. In those situations, we need to add the request to a queue and have a service that is in charge of processing queued requests in order. This method allows REST API to provide an instant response and when the request is finished the system notifies the job is done.
- **Email manager component**: The email manager component is an example of consumer service in this architecture. For example, when a user signs up on our platform, an email must be sent to it to notify that user has been signed up. API REST performs an instant response to frontend request but the task of sending email is queued and the email manager handles the request when it has finished the remaining queued tasks. Email manager uses an external SMTP server to send the emails, and a template system that injects HTML on each email in order to put styles on it.

5.2.10 Blockchain-enabled marketplaces for different purposes

S2CP must provide marketplace solutions, as a way to optimize multi-stakeholder dialogue processes, in which blockchain will be employed to provide some proof of concepts of monetization processes or contract agreements, in a reliable and transparent way.

These types of blockchain developments are very particular to the domain problem they seek to solve and require some difficulty for non-expert users to adopt said technologies. Cities2030 is a unique opportunity to experiment with various types of marketplaces, according to Lab's particular situations. As it can be seen in Table 14, up to the date of writing this deliverable, only the Murska Sobota lab has formally expressed interest in this tool to fulfill its lab goals. For this reason, although we consider other alternatives for other labs (which we describe in more detail in D6.3), we dedicate this section to the solution proposed for this Lab, called Blockchain for Short Food Supply Chain (SFSC).

Blockchain for Short Food Supply Chain (SFSC)

In order to ensure transparency, traceability and trust of the local food production, blockchain technology (BC) presents the natural technology fit in the so-called SFSC.

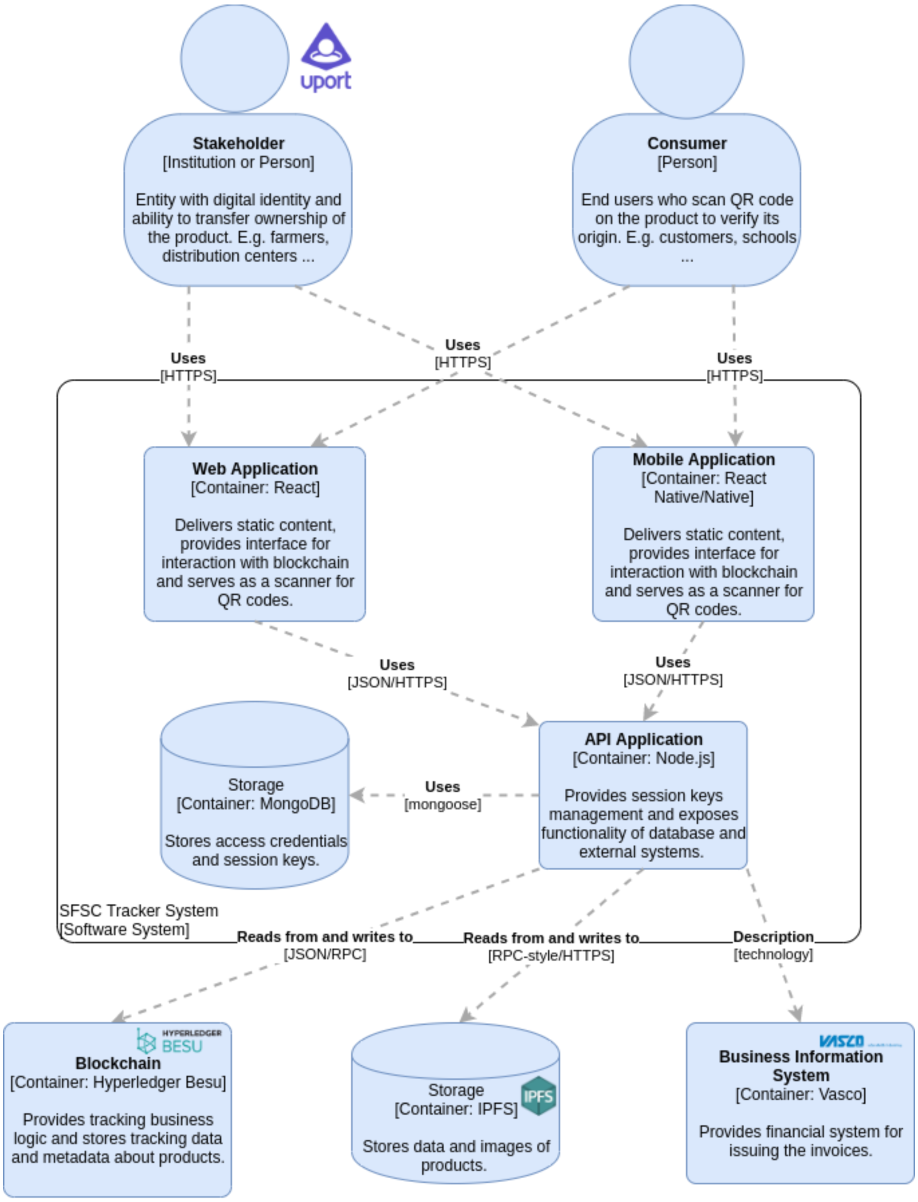


Figure 40. Blockchain-based SFSC architecture

The figure shows the system architecture. The architecture is divided into users at the top, front / backend system with support components in the middle, and external (data) systems at the bottom. Users are split into supply chain stakeholders and end users. The difference is that shareholders have their own digital identity, which in our case is implemented with the uPort library. Both connect to the system through a decentralized application available as a web and mobile application. The decentralized application connects to an API whose role is the connectivity of external systems with the web / mobile application. The supporting component for the API is MongoDB, a database used solely to maintain session keys and record access information. Through the API, the mobile / web application can connect to blockchain, storage (IPFS) and business information system (Vasco). Blockchain with smart contracts is used for the main business logic, since the ERC-721 token is implemented with a smart contract. Hyperledger Besu is used for the blockchain platform. Product data or images are stored in a dedicated storage system, IPFS (InterPlanetary File System), because of the size of the subtasks, since blockchain is not intended to store a large amount of data, but only for transactions. The Vasco Business Information System is used to issue physical invoices to shareholders and is integrated into the automated process of transferring ownership of a digital twin.

Use of this component is mainly linked with activities of the Murska Sobota lab, where they want to provide a blockchain-based marketplace solution to increase local food consumption and ensure that the products are really coming from home environment. The goal is to ensure the transparency of data within the short food supply chain established in the city of Murska Sobota and allow citizens to access to additional information about the products.

5.2.11 Supply chain transparency use case (Digital twin)

The blockchain technology has great potential to be used in food chain cycle. Anyway, there are some challenges. As a new technology is the best way how to prove the technology is to create testing environment with involvement of the practitioners. The consortium of Cities2030 has unique chance to test existing supply chain platform by company UNISOT. This is opening possibility to create transparency use case.

The platform enables to prove that products are sustainable, original, safety and ethically sourced. The architecture of the system enables data capturing and monitoring from all parts of the supply chain; from a farmers smartphone app or extracting data from the production plant to automatic IoT sensors, Computer Vision, ERP plugins, Big Data Analytics and Machine Learning.

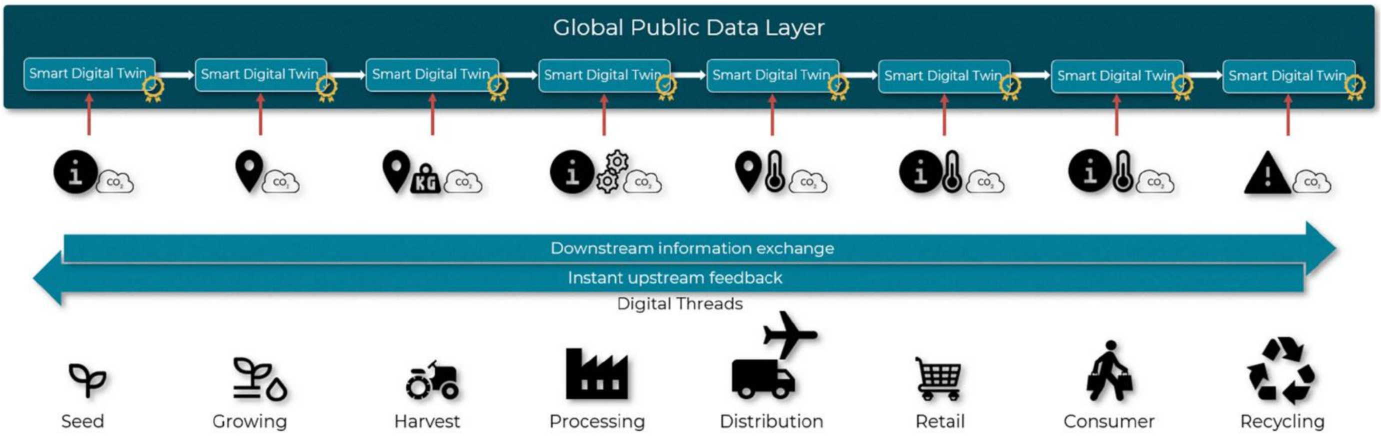


Figure 41.UNISOT platform architecture

Figure 41 shows the platform architecture by UNISOT with the following components:

- Seed: Information regarding seed origin & quality, water consumption, soil moisture, soil temperature, ground areal and fertilization is captured and stored in Smart Digital Twins.
- Growing: All through the growing phase, information about fertilizers, weather conditions and different treatments is captured and added to the Smart Digital Twins.

- Harvest: When the product is harvested, information such as time, locations, batches, weights, methods, returns, waste, etc. can be captured and registered to their corresponding Smart Digital Twins.
- Processing: Important information about the received raw biomaterials is useful for the processor to effectively plan and efficiently handle the production. Production plants consecutively add information about time, locations, processes, resource consumption and waste materials to the respective Smart Digital Twins.
- Distribution: Transportation can be critical for maintaining the quality and value of the products. Information from IoT sensors, Transport Management System and Mobile Apps is securely stored.
- Retail: The retailer now has access to valuable product information such as Sustainability, Provenance, Quality and Brand information. They can add additional information about their handling of the products such as storage temperatures, handling, energy consumption, waste, etc.
- Consumer: By connecting every single product item with its Smart Digital Twin by a unique QR code or RFID chip, consumers can get detailed, verified information about the Composition, Quality, Provenance, Sustainability and Brand for the specific individual item they are holding in their hand.
- Recycling: To close the complete product life cycle, information about recycling, waste management and reuse can be stored in Smart Digital Twins. This enables life cycle-traceability, life cycle-management and sustainable reuse of resources and materials.

How the architecture components are used in concrete demonstrator for Cities2030 and evaluation of the security of such solution is detailed in Deliverable D6.4.

5.2.12 Real-time data monitoring

Monitoring efficiency in food systems is paramount for CRFS actors. The real-time data monitoring infrastructure can receive information and events produced sensors (IoT HW platform), process the information in context-based systems and finally generate information repositories for real-time visualization or monetization.

One use that this component is the monitoring of the environmental conditions of the workers in a food production facility, linked to the Arganda Lab (Spain). The rationale for this use case is that the monitoring of food production processes in terms of energy consumed or sustainability indicators is paramount for food processing companies. A dashboard considering current production processes and their relations to other stakeholders allows to understand inefficiencies for better decision making.

The architecture of the solution can be seen in the following figure:

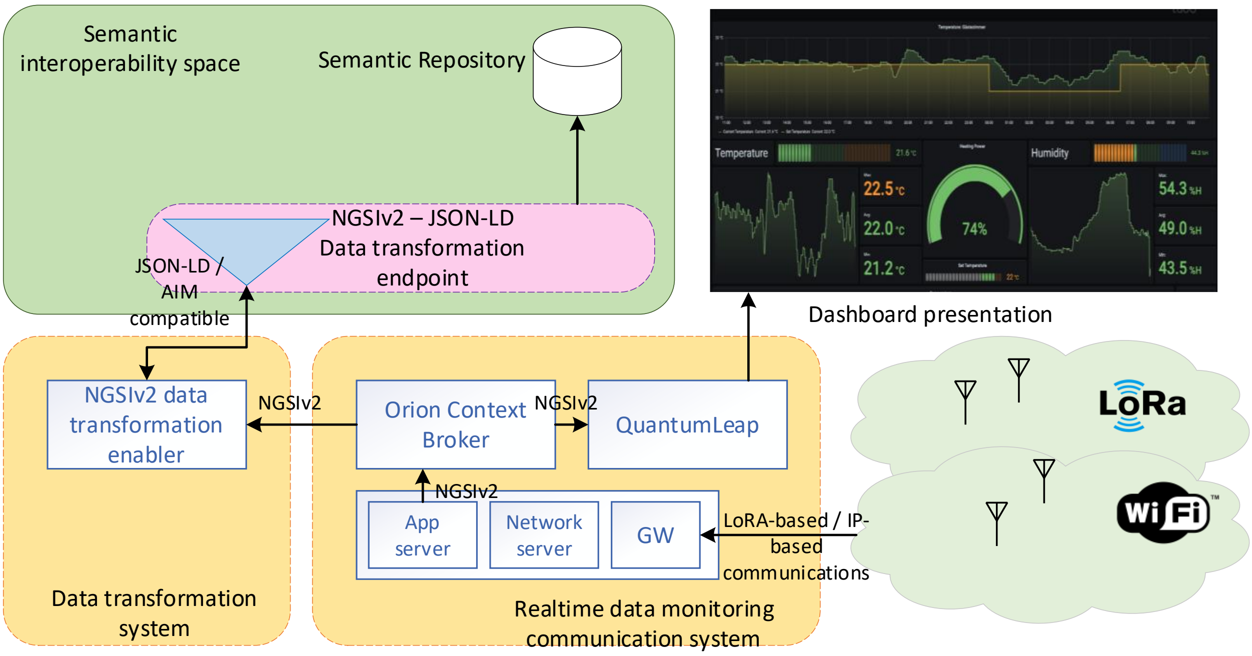


Figure 42. Real-time data monitoring component's architectural diagram

This proposal allows receiving information from a sensor network that is deployed in an intelligent environment. As can be seen, the current sensorization infrastructure is based on the LoRA and WiFi data generation networks. Generated data is registered in an information management infrastructure with FI-WARE (Orion Context Broker and QuantumLeap) components. The Orion Context broker allows receiving Communications from the sensor infrastructure and redirecting the information to other entities, following the publish/subscribe communications paradigm (see IoF2020 IoT Service Interoperability Layer and FIWARE, Section 4.3.1).

The information is sent in the NGSIv2 format, but it is also transformed in the data transformation system to JSON-LD format, which is compatible to existing data models such as IoF2020 and DEMETER's agriculture information model (AIM), as explained in Sections 4.5. 2 and 4.5.3.

Once the information is accessible through a semantic interoperability format, it can be published in various information repositories for easier access and integration in other databases and combined datasets.

5.2.13 Tools for private communications: Privately

Maintaining and developing messaging applications as well as the infrastructure necessary to make them work is costly. These systems must be able to support millions of users simultaneously, and to achieve this more complex and powerful architectures are needed, composed by many servers. In this kind of infrastructures, users do not make payments to maintain it, but are an essential part of the business model. These business models collect information, analyze it, and in some cases sell it. In the end, the user loses full control of their data because they are transferred to other companies in exchange for free services. With the decentralization of information and the use of blockchain we can avoid this problem.

Privately is a decentralized blockchain-based app that allows communication between actors in a private manner, with the maxims of anonymity and integrity.

The main functionalities of this S2CP component are described below:

- Private communications: Users can start a conversation only knowing their public BC addresses.
- Group communications: Users are able to create groups and invite other users. There are read, write, and general administration permissions.
- Secret communications: Allows two users to communicate in such a way that it is as difficult as possible to determine that they are communicating

Finally, a functional architecture of the tools for private communications is described below:

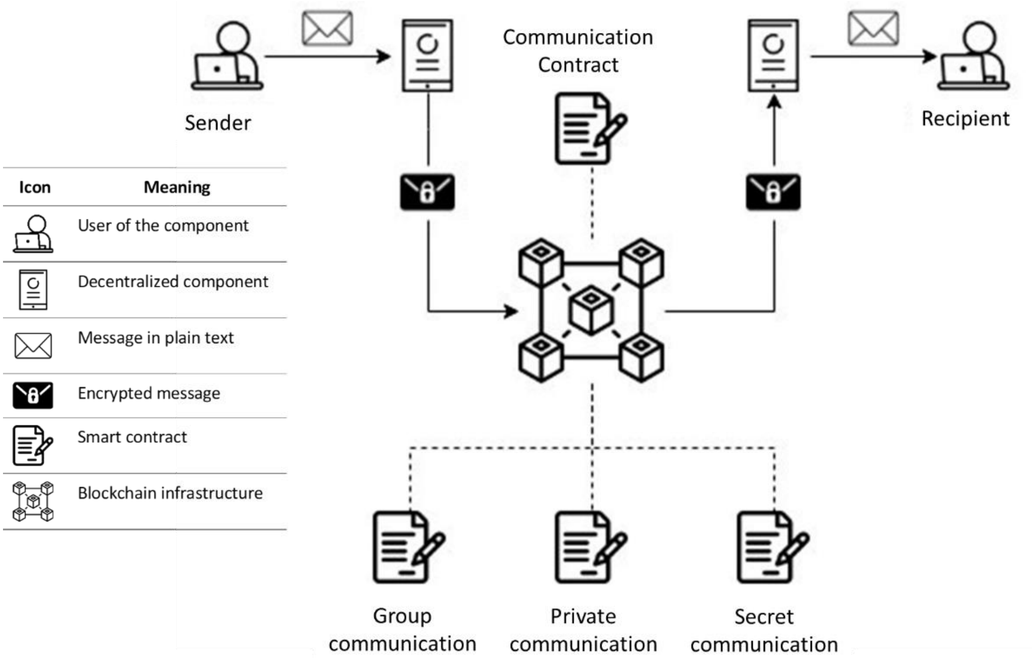


Figure 43. Tools for private communications architectural diagram

5.2.14 Data Governance Framework

The Data Governance component is used to promote data sharing between the users of the S2CP platform. In order to do this, it needs the trust of both the data producers and also consumers of the data. Therefore, it is also responsible for ensuring data security and privacy obligations are enacted when handling the data.

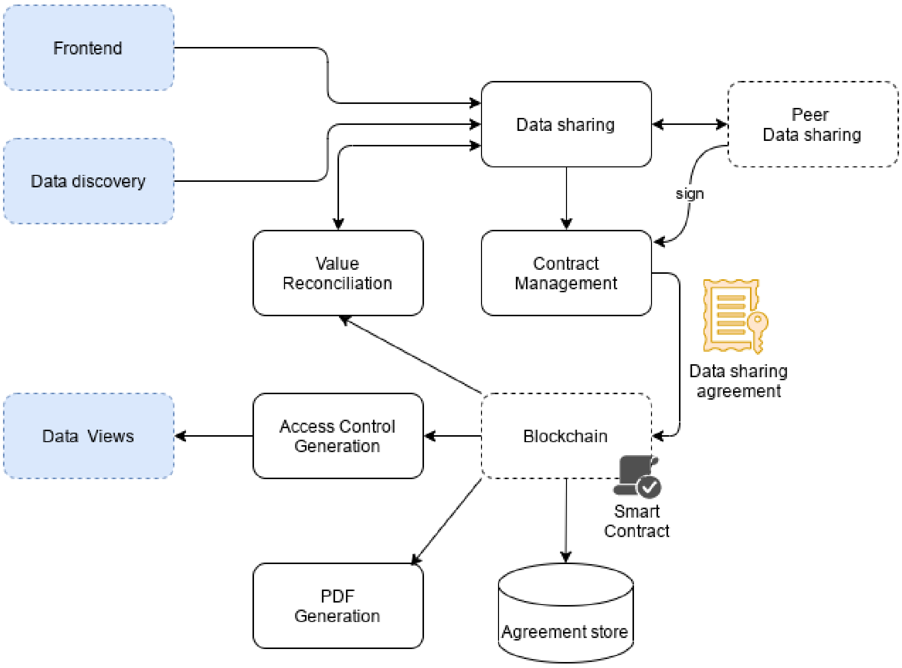


Figure 44: Data governance component architectural diagram

This component provides the following features as part of the platform:

- Data discovery - It facilitates different data formats and model of various data sources to be discoverable by the required components.
- Data value exchange negotiation - This works on the data exchange and data interoperability for the underlying objective and ensuring that the data exchange takes place in the specified format.
- Multi-party incentivized data sharing - This aims to incentivize data exchange between different actors to achieve a common objective and showing the positive impact of data sharing in terms of data rich insights generated from the analytical models.
- Integrated with smart contracts - It provides the facility to connect with the underlying blockchain platform and provide smart contracts to facilitate the data exchange.

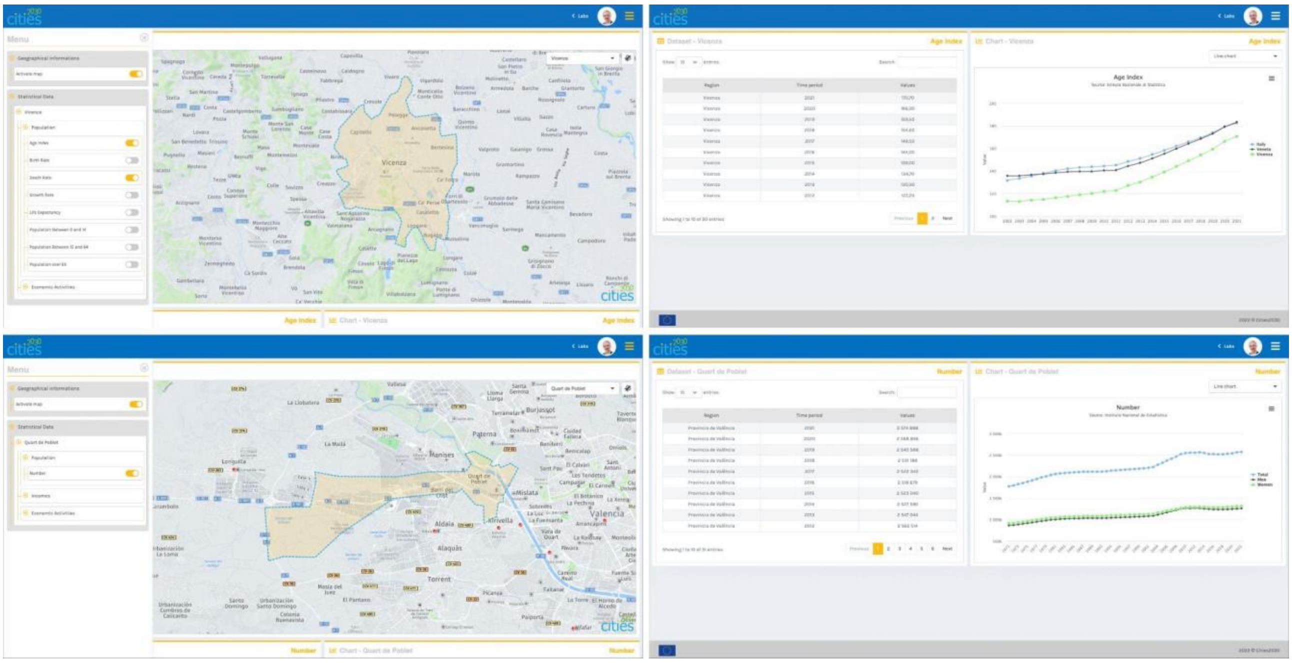
In order to achieve a degree of transparency and fairness:

- Policies are generated using a "purpose & consent" model
- Its operation is underpinned with blockchain enabled smart contracts
- These Smart contracts compliant with COPA-COGECA code of conduct

5.2.15 S2CP Dashboard: data representation and decision making

Geospatial visualization is a well-known tool for decision making activities being defined[58] as "a collection of methods and tools for transforming and combining geographic data and preferences (value judgements) to obtain information for decision making". Visualization is often recommended to explore large, multi-dimensional datasets. The current dashboard will be fed by a multi-property fusion database, receiving the data in API. This a major risk of the traditional approach to decision support: a huge amount of data sits outside a well architected digital record, either not considered or in unstructured fragments like spreadsheets, emails, and documents. Assumptions are not transparent, and decisions often don't have valid supporting data. Many organizations spend time examining the reasons "why" and repeating poor decisions in a perpetual loop.

The dashboard platform has 3 layers of blocks: 1) the core; 2) database; 3) frontend. The core of the platform is an OpenGIS module that converts raw data into geographical-statistical information, aggregating the spatial coordinates and the layers of data. Also at the core is a temporary database that receives information from an API call, digesting further the information and preparing for visualization. The temporary database is a relational database, like PostgreSQL and PostGIS. The visualization module depends on the data that comes from the call and may have different levels of abstraction according to the security level of the data and user. The visualization module is built over html, php and javascript, being interactive according to the user needs. The user (according to the security levels) will have a catalog of information that is available in the database (global and local) and may choose what information he wants to visualize and carry out the statistical analysis. The platform will be available on web and mobile services.



[58] Malczewski, J.; Rinner, C. Multicriteria Decision Analysis in Geographic Information Science; Springer Science + Business Media: Berlin/Heidelberg, Germany, 2015.

Figure 45: S2CP dashboard architectural representation

5.3 Task distribution and time schedule for development

The high number of components identified in the proposed architecture, the complexity of some of them, and the need to time deliveries to end users in order to facilitate training and learning periods, led the WP6 work team to define three differentiated development periods in the project: M1-M18, M18-M36 and M36-M48. Section 6 details in depth the motivation for these different periods and their duration, which were agreed with WP4 and WP5.

On the other hand, the context where the different components identified above must be developed are the tasks that make up WP6. The result of this cross analysis is a distribution of components among the different tasks, and among the three different development periods identified. The following table shows the result of said crossed analysis.

Table 15. Development periods for each S2CP component

	M1-M18	M18-M36	M36-M48
T6.2	Sentiment analysis	Data prediction with conformal learning Open Access Data repository Geospatial information	Data mining tools Data integration and Management
T6.3	Communication space	Blockchain enabled marketplaces Supply chain transparency use case Innovation management tool	Geospatial services Good practices Real-time monitoring
T6.4		Private communications: Privately	Data governance framework
T6.5	[Not started]	S2CP dashboard	S2CP dashboard

6 Combined Development Methodology

Once the technological components that are going to be part of the platform have been identified in the previous section, it is necessary to provide WP6 with a software development methodology that allows the desired modules to be implemented in such a way that they meet the initially detected requirements (see Section 3). In this section we describe the development methodology agreed upon by the WP6 team for this purpose.

The purpose of WP6 is the development of an ecosystem of technological applications and services that allow *Policy Labs* and *Innovation Labs* (mentioned in this deliverable as just 'Labs') to establish themselves and achieve their objectives, following and through the work methodologies defined for labs. These methodologies are the "Co-creation capacity building programme" and the "Extended Innovation Pattern" respectively, and in accordance with the work carried out in WP4 and WP5.

Since labs are going to require technological support from the moment the project begins, it is important to speed up the processes of design, start-up and improvement of technological tools and services as much as possible.

To do this, WP6 defined a novel agile work methodology (which we called "Combined Development Methodology" -CDM-) in which we distinguish 3 different stages: initial design, development and deployment, and improvement and scaling of services and applications. In each of these stages, a specific agile and open sub-methodology is used, which allows at all times to offer the rest of WPs a product and technological solution that is viable and usable by both the policy labs and the living Labs. These methodologies are Design thinking, Lean Startup and Scrum.

On the other hand, the Cities2030 project considers a wide catalog of labs, of different types (policy and innovation labs), with very different starting situations, and with very heterogeneous short-, medium-, and long-term objectives. In some respects, all laboratories requires the same services. For example, both the "Co-creation capacity building programme" and the "Extended Innovation Pattern" include a stage in which laboratories must analyze the current situation of their CRFS, and from there identify the objectives they wish to address, and the indicators of success that they will use to measure their performance. However, in other aspects, each type of laboratory, and even each individual laboratory, requires specific technological tools. This is the case, for example, of the Living Labs, which according to the "Extended Innovation Pattern" must carry out experiments to monitor and validate the good practices they have devised. Each laboratory therefore requires a tool specifically adapted to the experiment that has been carried out and the type of objective and indicator that is being monitored.

Consequently, we apply the Combined Development Methodology through two techniques that intersperses and overlaps depending on the type of application or service that is being developed at any given time: the large group technique, and the small group technique. In the "large group" modality, general workshops are organized, to which representatives of all the laboratories are invited to make a joint presentation about those technological tools that are of general interest and application. In the "small group" modality, bilateral meetings are organized (or possibly with a reduced number of laboratories) in which specific training sessions are held for the type of laboratory that participates in the session, its work methodology, its objectives to reach etc.

Finally, given that both the "Co-creation capacity building programme" and the "Extended Innovation Pattern" are iterative methodologies, the Combined Development Methodology also defines macrocycles in which the sequence of 3 stages is repeated (Design thinking, Lean Startup and Scrum) in order to devise new components and applications that are adapted to the emerging needs of the different laboratories. Roughly, and according to the proposed schedule for the evolution of work in the WP4 and WP5 packages, each macrocycle has an approximate duration of between 12 and 18 months. Below, we describe the duration and purpose of each macrocycle:

- Macrocycle #1: Serves to cover the highest priority needs of the labs. It approximately covers the first 18 months of work.
- Macrocycle #2: Allow the experimentation of the Living Labs and the development of action plans in the policy labs. It covers between months 18 and 36.
- Macrocycle #3: Focused on the public communication of the results and the transfer of results from the project to the agents that must continue with the innovation in each of the CRFS. It covers the last 12 months of the Cities2030 project.

Each macrocycle of the Combined Development Methodology follows the following sequence of phases:

1. The first stage consists of the design of the technological components, applications and services that are provided to the Labs, which play the role of 'clients' of the S2CP solutions. In order to achieve a design that meets the real needs of laboratories, meets their expectations and is focused on their user experience, this first stage of each macrocycle is governed by a design thinking sub-methodology. Task 6.1 is the one that mainly contains the actions derived from the application of this first sub-methodology. This phase of the CDM methodology was carried out coinciding with the development of the "Step 0: Set up the lab" and "#0 CRFS Living Lab for Developments and Innovations" phases of the "Co-creation capacity building programme" and the "Extended Innovation Pattern" respectively. During these phases, the different laboratories only had to organize their efforts to start the work process. Therefore, in general, the laboratories did not need specific technological tools of the Cities2030 project during these phases. We took advantage of this circumstance to use this period of time to identify needs, hypotheses, general designs, etc. At this stage of the CDM we identified the following actions:
    a. **Identification of needs**: Using "small group" technique, the specific and general needs of each of the laboratories is identified. Bilateral meetings are established with each of the laboratories, to which all the participating partners of each one is invited. A form is available in which the following must be identified: laboratory typology, general objective, specific implementation of the work methodology and variables or indicators of interest. The WP6 representatives in each of these meetings in "small group" format filled one of these forms for each laboratory. We also considered as a need to cover the KPIs proposed by the "focus group" in WP3 (Task 3.7).
    b. **Research**: in the context of each of the WP6 development tasks (6.2, 6.3, 6.4 and 6.5) and considering the content of the files developed in the previous action, the state of the art was investigated in search of those tools and technological innovations that have been reported successful in meeting previously identified needs. The results of this research process are collected in each of the deliverables of each task.
    c. **Prioritization**: an analysis of the efforts available within the work package is carried out, and their distribution throughout the entire work period of the Cities2030 project (Gantt diagram). In a plenary meeting with all the partners participating in WP6, and in accordance with the analysis of the efforts previously made and the previously described state of the art, a prioritized list is made of those needs that must be addressed in first place. In this way, and as we detail later, the functionalities and technological applications are addressed in 3 groups:
        i. Macrocycle #1: In this first period, labs are expected to reach the "Step 2 - Pathway Development" and "#1 - Understand CRFS" phases of the "Co-creation capacity building programme" and the "Extended Innovation Pattern" respectively. For this reason, those applications and tools focused on the establishment of communities (including communication mechanisms based on Blockchain) and the analysis of data on the current situation of the CRFS (basically the purpose of the aforementioned phases) have priority.
        ii. Macrocycle #2: It focuses on covering those needs related to phases "#3 - Experimenting" in the "Extended Innovation Pattern" and the "Step 3 Action Plan"

in the "Co-creation capacity building programme". At this point, the technological tools that are developed aree focused on allowing the experimentation of the Living Labs and the development of action plans in the policy labs. The priority components in this second macrocycle are those that make it possible to monitor supply chains, manage innovation processes in action plans, monitor indicators, and in general apply the good practices detected in the previous phases by the different laboratories to their respective CRFS.

iii. Macrocycle #3: During this stage, the Policy Labs reach the "Step 4: Scaling up & Continuity" phase in the "Co-creation capacity building programme"; and the Living Labs begin the development of the "#4 CRFS-LL Dashboard" described in the "Extended Innovation Pattern". These phases in both methodologies are characterized by being a process of closure and summary of the lessons learned, identification of best practices, and preparation of the CRFS and its agents to continue autonomously the innovation process that allows them to achieve their objectives in the future time horizon of 2030. For this reason, the components developed mainly in this last macrocycle are focused on the public communication of the results (such as the innovations or good practices developed); and in the transfer of results from the project to the agents that must continue with the innovation in each of the CRFS.

d. **Ideate**: using the brainstorming technique, the partners participating in each of the WP6 tasks identify the most appropriate technological components to meet the needs of the laboratories, with the best possible user experience and adjusted to the available efforts and to the state of the art. The list of components, both general and specific, must be exhaustive and cover the needs of all laboratories equally (regardless of their type, objectives, etc.). These components must be fully operational within each macrocycle of the CDM methodology, so their development period is between 12 and 18 months. It is important that, although the components focused on community building and analysis of the current situation of CRFS have priority in the first macrocycle, the components developed in this first macrocycle also include functionalities for those laboratories that advance faster in their respective work methodologies. For this reason, there are functionalities, for example, focused on experimentation ("#3 - Experimenting" in the "Extended Innovation Pattern") or the design of action plans ("Step 3 Action Plan" in the "Co-creation capacity-building programme"). The same reasoning applies to the rest of the macrocycles.

e. **Dissemination of conclusions and review** (cyclical process): with the list of components and technological tools devised in the previous action, each of the laboratory records is completed. These files are sent to each of the representatives of the laboratories by email. A period is then opened to collect their comments, which gives rise to an iterative review process until a technical proposal adjusted to the needs of the users and the KPIs proposed by the focus group in WP3 is reached.

2. Once the components, applications and services that must form part of the technological platform have been identified, the initial development and deployment takes place in a second phase. For this, we use the Lean Startup sub-methodology, which allows end users to be put in early contact with the technological solutions under development, guaranteeing a validation adjusted to the

expectations of the laboratories. By exposing the prototypes to clients early, we can review the conclusions drawn from the "Design Thinking" process, if necessary, with minimal cost. This process is followed using the "large group" or "small group" technique depending on the nature of each of the technological components. The process by which a first prototype (mockup, even) is exposed to clients (laboratories), for the first two macrocycles, is carried out mainly in the context of T6.1 (although the rest of the WP tasks are also involved to a lesser extent). While the iterative validation process is carried out by our own specialists in each of the following tasks (6.2, 6.3, 6.4 and 6.5). In order to allow laboratories to reach their intended objectives in each of the phases of their own methodologies, the end of this second Lean Startup phase should not extend beyond the middle of a macrocycle (this second phase in each macrocycle would end at months 9, 27 and 42). In the context of this sub-methodology, the following actions are carried out.

a. **Initial development**: this action includes the technical and programming work aimed at implementing a first minimum viable product capable of broadly satisfying the needs detected in the previous phase, in accordance with the conclusions drawn from the Brainstorming action. To facilitate the management of the development process, each of the technological components or functionalities have a responsible partner.

b. **Deployment and communication**: it is essential that a first minimum viable product is developed for each of the identified components before the laboratories formally begin work on each of the phases of their own methodologies. By that time, an initial functional product should have been available, with which laboratories can work, even if its functionalities are limited. To do this, each partner responsible for the development of each of the components must have a space for its public deployment within the Cities2030 community. Then, using the "large group" (general workshops) or "small group" (bilateral meetings) technique, depending on the component, the functionalities are presented to the Labs synchronously, and they are provided with a user manual.

c. **Validation and learning**: after each communication and deployment action, a period is opened in which the laboratories can send their comments to the partner responsible for each component in order to enrich the future iterative development process, which should lead to the final and complete version of each of the components. Once each responsible partner considers that they have the set of requirements and the necessary information to carry out the complete development process, they can start the next phase of the CDM methodology. If necessary, the Lean Startup phase could be revisited if the product finally developed did not fully meet the expectations and needs of the labs.

3. The Startup Lean sub-methodology allows prototyping the technological components and functionalities in such a way that the essential elements are specified jointly with the end users. However, once the basic principles and functionalities of each component have been established, a process is opened for the evolution, improvement and scaling of each tool, which must allow each of the applications or services to evolve from a basic scheme to a fully functional solution. To carry out this process, in the third stage of the CDM methodology, a Scrum Agile sub-methodology is implemented. In this sub-methodology, new features are implemented through iterative spiral cycles, which can be superimposed to introduce various new features in parallel. These iterative processes must conclude with the end of the macrocycle, in such a way that the Labs can gradually incorporate the new functionalities into their work methodology before reaching and starting a new

phase. All improvement and evolution processes are carried out in the context of the specific tasks (6.2, 6.3, 6.4 and 6.5); and specific descriptions can be found in the corresponding deliverables. Each of these iterative cycles of development includes the following actions:

a. **Planning**: this action includes the theoretical design of each of the new functionalities, which may be proposed independently by one of the partners or arises through brainstorming in a plenary meeting of the WP6 group. Each planning cycle lasts 15 working days.

b. **Implementation**: includes all the technical and programming work necessary to integrate the new functionality into the technological tool and obtain a viable new product.

c. **Product presentation**: depending on the nature of the technological component being developed, the "small group" or "large group" technique is used to present the new functionality to Labs that use each component. It is accompanied by a user manual focused on the development of use cases.

d. **Evaluation**: After each public communication of the new functionalities, a period is opened in which the laboratories can report their comments and possible improvements regarding the new functionality to the partner responsible for the component. During this period, and in parallel, a new cycle can be started with the planning of a new functionality.

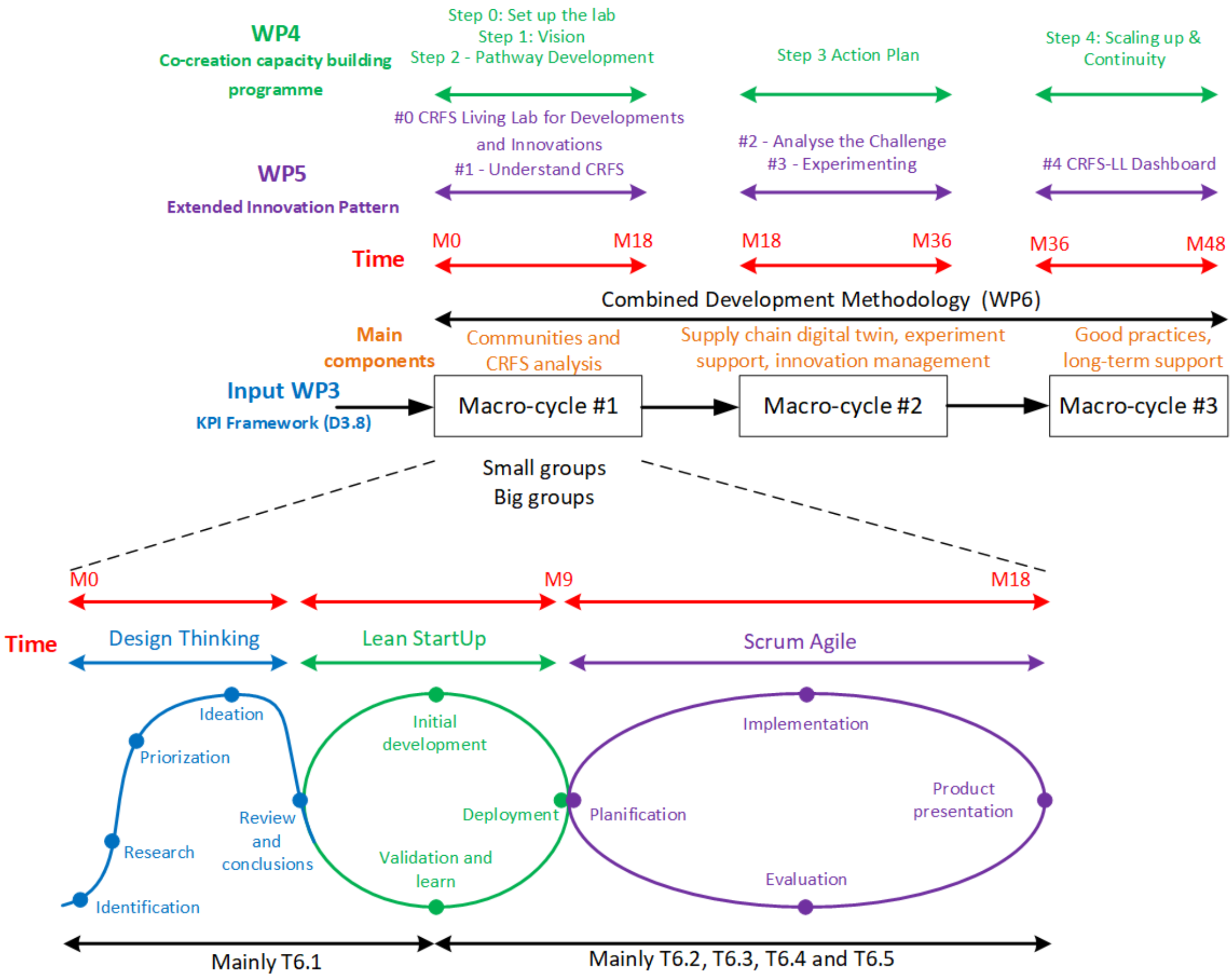


Figure 46. Combined development methodology (CDM)

7 Current development state and impact monitoring

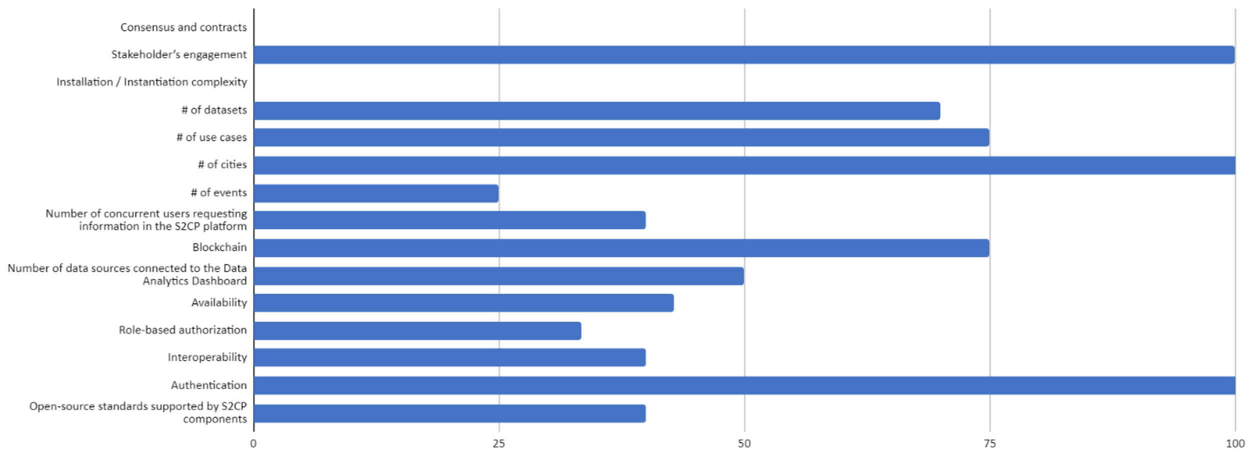
At the time of delivery of this document (M24), *Macro-cycle #1* has been successfully completed, where all the planned components of all the tasks have been released for use by the laboratories. This includes, as stated in Section 5.3, the *communications space* and the *sentiment analysis tool*. In addition, very significant and relevant progress has been made in many of the tools planned for the second macrocycle. In particular, on the date of delivery of this document there are already viable versions released of the following components: *S2CP dashboard*, *Multi Actor Approach tool*, *blockchain for short food supply chain*, and the *tools for private and confidential communications based on blockchain*.

Similarly, other tools such as *Geospatial information and services* have a high level of maturity that we hope will allow their release to users in less than three months from the delivery of this document.

It is important to highlight that the component of *Good practices*, which was initially planned for the last macrocycle of the development methodology, has been advanced and has already been put into operation at the request of the participants in WP3 with the purpose of disseminating the innovations identified as soon as possible.

The works were carried out in the foreseen times according to the Gantt chart. In the period described in this document, the three meetings and workshops were held with the laboratories in large group format, as foreseen in the Gantt chart. In addition, bilateral training meetings were held with the different laboratories, according to the development methodology described in Section 6.

As a result of the continuous monitoring of the indicators foreseen in document D3.8, and whose fulfillment is entrusted to WP6, it has been possible to determine the values defined in the following figure for all the KPIs of the framework created in WP3.



As can be seen in three of the indicators, the expected value has already been reached. These indicators are: (1) Percentage of critical components whose access is protected by credentials (email, password, tokens); (2) Number Cities (Municipalities) incorporating any of the S2CP enablers in pilots and living lab activities within the time period of the Project; and (3) Level of engagement of stakeholders in platform co-creation.

On the other hand, for three other indicators, 75% of the expected value for the total project has already been reached. These indicators are: (1) Number of datasets collected / represented by the platform, corresponding to participating cities; (2) Number of use cases defined as framework for S2CP architectural design; and (3) Number of components using blockchain with benefits and usage in term of revenues, immutability, or consensus.

Furthermore, and although the average value of compliance with the indicators is situated in the M24 at 55%, there are two indicators for which no element of value has yet been provided. These two indicators are: (1) Number of components using blockchain with benefits and usage in term of revenues, immutability, or consensus; and (2) Complexity in terms of IT knowledge required and local adjustments.

The reasons for this low value are the complexity of the components that satisfy these indicators. On the one hand, the blockchain-based components are expected to finish their development in the second macrocycle (which concludes M36). Although for other types of tools it is possible to release a provisional version in advance, the complexity of blockchain-based tools and the need for joint training with end users make it very difficult to carry out intermediate products with this type of component. On the other hand, it is not possible to assess the complexity of the platform until a relevant number of components have been released and used by laboratories. We hope that valuable assets can be counted on for both indicators at the end of the second macrocycle of the development foreseen in the methodology.

The status of the indicators from M24 to M48 will be checked in *Task 6.6 – WP6-specific impact monitoring and assessment implementation*. As the completion date of T6.1 is M24, T6.6 inherits from T6.1 the monitoring and compliance with the proposed methodology, the collection of results, and the analysis of the results from the point of view of the generated impact, through performance indicators.

8 Conclusions and next steps

With the aim of organizing the technological work of WP6 and adjusting it to the needs and requirements of internal and external users of the project, a working group was created for this task whose purpose was the management and monitoring of WP6, and organization of the efforts. In a more technical field, the objective of this task is to provide the general design for the S2CP platform, to which the work developed in the rest of the tasks that make up the WP must be adjusted.

In order to adjust the final design of the platform to the needs and requirements of internal and external users of the project, a detailed review of D3.8 was carried out. Said document was generated in the context of WP3 with the purpose of guiding the technological development and adjusting it to the needs reported by the stakeholders. From the analysis of this document, a set of general technological functionalities are extracted. On the other hand, and using the methodology of the bilateral meeting in a small group, the requirements and needs of the different laboratories were captured. To do this, the functionalities identified in the previous analysis were started, which were presented and discussed with each of the Labs until a list of necessary tools was generated, already with a much more specific technological description.

The captured requirements should be applied equally to all tasks in the WP. Therefore, in the work team, each of the participants had the main role of representing the work carried out in each of the tasks, ensuring that their particularities are taken into account in the decisions made at the WP level and becoming aware of the requirements and characteristics that the activities of the task should satisfy.

Another source of essential information to be able to carry out the design of the architecture of the S2CP platform is the state of the art. Therefore, also within the constituted work team, roles were designated for the study of the five major technological areas in which designs have been reported for digital platforms similar to the S2CP platform that must be developed in this WP.

As a conclusion of these studies, the architecture of the S2CP platform must be based on a component scheme, which act as independent software modules. These components are grouped in a personalized way, to reach instances of the S2CP platform that are specific to each of the laboratories and adapted to their needs. In addition, some components can also be customized, giving rise to different versions that adjust to the different use cases demanded by laboratories. This is more frequent in those components that assist the Labs in the experimentation phases defined by the WP4 and WP5 methodologies.

In this way, the S2CP platform will have at least 17 components, distributed among the four development tasks that make up the WP. Components are classified in various typologies: Data integration and analysis

(Task 6.2), service provision (Task 6.3), privacy and security provision (Task 6.4) and data visualization (Task 6.5).

All these components are distributed in three different development periods, which are adjusted to the periods defined in the methodologies of WP4 and WP5, according to a proprietary development methodology for WP6 called CDM -Combined Development Methodology-.

As of the delivery date of this document (M24), we can conclude that: (1) the *Communication space* and *Sentiment Analysis* components, planned for the first development macro-cycle, have been successfully completed and released for use by the Labs and on schedule, in addition, the component of *Good practices* has been released in advance at the request of the Labs participating in WP3; (2) several components planned for the second development macrocycle (M18-M24) have already been put into operation (although it is a version with reduced functionality) -these components include the *S2CP dashboard*, the tools for innovation management (*Multi-Actor Approach tool*), *blockchain components for private and confidential communications*, and *blockchain systems for short food supply chains*.

Finally, the evolution of the indicators foreseen in D3.8 has been very satisfactory. Average compliance stands at 55%, with three indicators for which 100% of the expected value has already been reached, and another three for which 75% of the success value has been reached.

In the second half (M24-M48) of the project, WP6 plans to complete the two remaining development macrocycles, assisting the laboratories in the experimentation and drawing of conclusions phase. The work will focus on a smaller number of components but with greater complexity, especially in carrying out proofs of concept that require blockchain technology for their operation.