

SEBASTIEN

M14 - Open Data Policy

Milestone Lead	CMCC
Milestone due date	2023/03/31
Status	FINAL
Version	V1.0
Project	SEBASTIEN



DOCUMENT INFORMATION

Title	M14 Open Data Policy
Agreement	INEA/CEF/ICT/A2020/2373580
Action	2020-IT-IA-0234
Creator	Maria Mirto (CMCC)
Milestone Description	Open Data Policy has been formulated
Means of verification	Report shared with the Agency
Contributors	Maria Mirto (CMCC), Marco Milanese (UNITUS)
Requested deadline	M6
Reviewer	Marco Mancini (CMCC), Cinzia Caroli (Cineca)

Index

1. Executive Summary	2
2. What is open data and what is it for	2
2.1 Concept and history of Open Data	3
2.2 Reuse of open data (commercial reuse, accountability, Data Visualization)	4
3. Public data in the PA and Open Data legislation in Italy	5
3.1 Current legislation and full Open Data	7
3.2. The regulatory path in Italy	9
3.3 Open Data Maturity Report	10
4. Open Data: licenses, formats, Raw data and LOD	11
4.1 The licenses	11
4.2 The formats	13
5. SEBASTIEN Datasets	14
5.1 Service 1: Coping with environmental stressors for breeds	16
5.2 Service 2: Intensive farming risk management under climate extremes	17
5.3 Service 3: Extensive farming management and feed availability	18
5.4 Service 4: Livestock farming under risks from combined abiotic and biotic factors	18
6. Conclusions	19
7. References	20

1. Executive Summary

This Milestone (M14) is aimed at examining the data available in the project, analyzing both the potential and the constraints related to data use and re-use, considering in particular:

- Licensing of data (raw, integrated, processed, etc.);
- Type of services that will be developed;
- Technological architecture and related levels of service;
- Agreement on the use and re-use of data and services.

Milestone M3 *“Report on Data Sources”*, that reports the data sources identified in the project, and Deliverable 2.2 *“List of suitable data sources and of newly acquired data”* have also been considered.

The document is organized as follows: Section 2 describes the definition of open data and a brief introduction about its history. Section 3 describes the legislation in Italy, since all the data will be produced by Italian organizations. Section 4 describes the licenses and formats used in the open data (from raw data to linked Open Data – LOD). Section 5 describes the datasets available as input and those produced by the services. Finally, Section 6 draws the conclusions.

2. What is open data and what is it for

Open data is data that can be freely used, re-used and redistributed by anyone - subject only, at most, to the requirement to attribute and share alike [1].

The term open data refers to information held, produced and updated by public administrations made available free of charge for anyone interested in reusing them.

All data held by public bodies not subject to privacy constraints can be displayed in an open format, through the assignment of an open license that allows reuse and the choice of a technological format that is in turn open and standardized.

Access to open government data allows, for example, citizens to know the timetables of public transport and organize their travels better, including information about the budget approved by Municipalities, and the quality of the air we breathe every day. More generally, the availability of open data makes it possible to transform the knowledge held by public administrations into services that can be immediately used by citizens.

The full Open Definition gives precise details as to what this means. To summarize, the most important are:

- **Availability and Access:** the data must be available as a whole and at no more than a reasonable reproduction cost, preferably by downloading over the internet. The data must also be available in a convenient and modifiable form.

- **Re-use and Redistribution:** the data must be provided under terms that permit re-use and redistribution including the intermixing with other datasets.
- **Universal Participation:** everyone must be able to use, re-use and redistribute - there should be no discrimination against fields of endeavor or against persons or groups. For example, 'non-commercial' restrictions that would prevent 'commercial' use, or restrictions of use for certain purposes (e.g., only in education), are not allowed.

Open data means **interoperability**. Interoperability denotes the ability of diverse systems and organizations to work together (inter-operate). In this case, it is the ability to interoperate - or intermix - different datasets.

Interoperability is important because it allows different components to work together. This ability to componentize and to 'plug together' components is essential to build large, complex systems. Without interoperability, this becomes nearly impossible — as evidenced in the most famous myth of the Tower of Babel where the (in)ability to communicate (to interoperate) resulted in the complete breakdown of the tower-building effort.

A similar situation is related to data. The core of a "commons" of data (or code) is that one piece of "open" material contained therein can be freely intermixed with other "open" material. This interoperability is absolutely key to realizing the main practical benefits of "openness": the dramatically enhanced ability to combine different datasets together and thereby to develop more and better products and services (these benefits are discussed in more detail in the section on 'why' open data).

Providing a clear definition of openness ensures that when you get two open datasets from two different sources, you will be able to combine them together, and it ensures that we avoid our own 'tower of babel': lots of datasets but little or no ability to combine them together into the larger systems where the real value lies.

2.1 Concept and history of Open Data

Open Data are data accessible to everyone, made available by Public Administrations or private companies that can be reused and redistributed for different purposes.

The approach to the so-called governance of public affairs was established internationally in 2009 [2]. On the same day of his inauguration and as a first act, the President of the United States, Barack Obama published a memorandum on transparency and Open Government addressed to the executives of his administration.

The Memorandum was followed, just over a year later, by the Open Government Directive of 8 December 2009, focused on the principles of Transparency, Participation and Collaboration, which prescribe tasks, processes and organizational models that public bodies are called to follow, and the



Open Government Initiative which collects the American Federal Administration's Open Government initiatives in a site designed to encourage visitor participation and feedback [3].

Following the American example, many countries have approached the issue of Open Government, starting the process of changing the public scenario, towards greater transparency and closeness to citizens.

International organizations

In 2010, two international reports presented and recommended the Open Government model as a key for the development of the public sector and the productive fabric: the OECD report "Towards smarter and more transparent government" [4] and the UN Report on the state of eGovernment [5] in the world recommended the adoption of open administrative models.

European Union

In Europe, it is the Malmo Declaration on eGovernment policies of 2009 [6] that proposes a first path towards opening up European administrations in 5 years.

Italy

In 2011, the international Open Government Partnership initiative was launched for concrete commitment by governments to promote transparency, empower citizens, fight corruption and strengthen governance through technology. Italy joined the initiative in September 2011 and presented, in April 2012, a National Action Plan containing the main initiatives taken by the Government in the field of Open Government, which were later updated in 2014 with the Second Plan of action [7].

2.2 Reuse of open data (commercial reuse, accountability, Data Visualization)

The European Union estimated that the increasing availability of open data made available by the 27 member countries has led to an economic return of 75.7 billion euros in 2020, creating almost 25,000 new jobs related to the sector of data analysis by the end of the same year [8].

Businesses can benefit from the new knowledge that comes from open data by improving their business models or identifying new ones. There are already many innovative services based on open PA data that users use every day on their smartphones. Thanks to open data, developers, journalists, universities and research centers have at their disposal increasingly solid and reliable information tools to understand the extent of the phenomena in progress and to perform their work in the best possible way.

Furthermore, the free access to public information by all allows widespread transparency which improves the quality of the debate on public policies and - at the same time - makes the



administrative machinery more efficient. In this sense, we can speak of **Open data for Accountability**.

Thanks to the availability of open public data, individual citizens, trade associations, universities and research centers have new ways of verifying that the PA acts in full compliance with the rules.

There are many Italian examples on the subject:

1. Open Cohesion. Initiative [9] on cohesion policies in Italy that promotes the effectiveness of interventions through the publication of data on funded projects and widespread civic participation;
2. Open Balances. Platform [10] that makes the financial statements of all 8,100 Italian municipalities of the last twelve years navigable, comparable and downloadable;
3. Confiscated Assets. Participatory project [11] to promote transparency, reuse and enhancement of assets confiscated from mafia organizations, through the collection, analysis of data and monitoring of confiscated assets.

For some time, then, the concept of using data for administrative transparency alone has been abandoned, in favor of a commercial reuse of open data, i.e., the possibility that companies use the open data of the PA in order to generate products and services for creating social and economic value. Through the reuse of open data, the PA can contribute to the creation of new economies and at the same time strengthen the existing one. Although the possibilities of data reuse are infinite and depend only on the creativity of the developers, it is possible to draw a first distinction between the different forms of reuse:

- Development of Applications, which reuse public data to offer innovative services to their users (e.g., Moovit [12]);
- Data journalism or data-based journalism, where journalists reuse aggregate data from public sources to tell a story (e.g., InfoData from Il Sole 24 ore [13]);
- Enrichment, for subjects who already manage data-based services and who, thanks to public data, can strengthen their business (e.g., Immobiliare.it or Zillow.com [14]; Insurance companies, etc.).

3. Public data in the PA and Open Data legislation in Italy

The three pillars of Open Government are transparency, participation and collaboration [15].

Transparency

Institutions are required to provide citizens with data and information on the decisions they make and their actions. True transparency requires that this information must be usable, that is, easy to access, understandable and usable. The goal is to create a system of trust within the local community towards the work and choices made by the entities.

Participation

Decision-making processes must be open to the contribution of citizens and in general of collective intelligence generated from below. This is one of the central nodes of the model, whose purpose is to improve the quality of the political-administrative choices of public bodies, through the proposal of interventions that are effectively linked to the needs and requirements of citizens and the reduction of conflict.

Cooperation

In the open model, institutions are not intended as structures in their own right, but as subjects inserted within a collaborative and participatory network. Therefore, individual entities are called upon to use innovative tools and methods that aim to improve collaboration, both between the various levels of administration and between different entities.

Governance - conditioned by transparency, participation and collaboration - becomes a shared process that makes it possible to identify and define the real needs of citizens (as well as to give them answers), therefore strengthening their trust in institutions.

By focusing on these three concepts, an administration can put into practice a completely new form of relationship with its interlocutors (citizens and businesses). A relationship that changes the structure and the very nature of the administration from within, requiring a radical renewal of models, processes and tools used by the PA to establish a relationship with the citizens and with other public structures. A slow and complex process, that of Open Government, which requires a strong political endorsement.

The EU has no formal definition of open government. However, it stated in its vision on public services ([16]) that Open Government encompasses three core aspects (Figure 1):

Open assets: “government data, software, specifications and frameworks that are open so that anyone can freely access, use, modify, and redistribute its content with no or limited restrictions such as commercial-use or financial charges”, i.e., Open government data;

Open services: “digital public services that can be reused by other public administrations or by third parties in order to provide value-added services”, i.e., efficient e-government and

Open engagement: “opening up the processes for public policy making to the whole of society, including civil society, businesses, labor unions, or even individual citizens”.

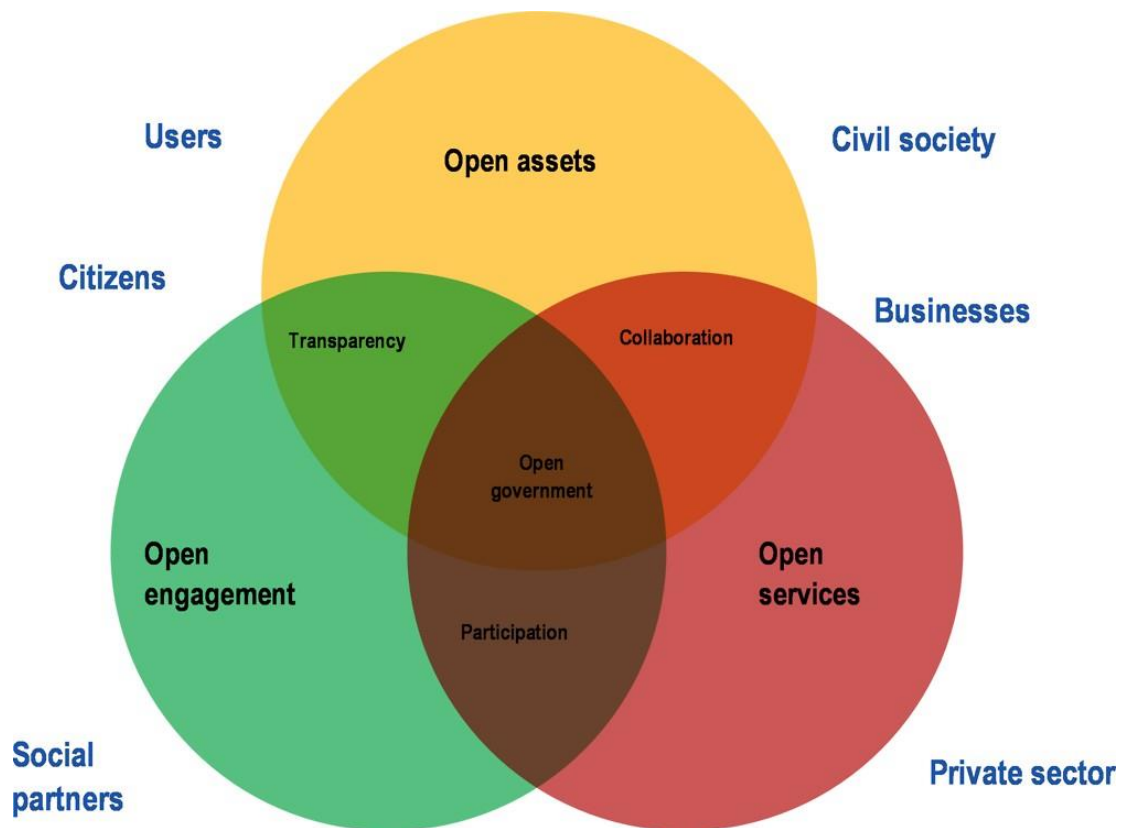


Fig. 1: Open Government: pillars and public services

3.1 Current legislation and full Open Data

The concept of open data goes far beyond transparency. Public data has enormous potential, which however cannot be fully expressed if the data is not truly open (full open data).

To do this, it is not enough to publish data, but any limitations (legal, financial or technological) to the reuse by others must also be eliminated. In Art. 68 of the Digital Administration Code (CAD), amended by Legislative Decree 217/2017, paragraph 1-ter, data is understood to be open if [17]:

- Data is available under the terms of a license that allows use by anyone, including for commercial purposes, in a disaggregated format;
- Data is accessible through information and communication technologies, [...] suitable for automatic use by computer programs and provided with the relative metadata (a series of instructions and descriptions that help to read and interpret the data correctly).

As a corollary to what has been said so far, the Manifesto for Open Government highlights that to be defined as such, open data must be¹:

- **Complete.** It must include all the components that allow it to be exported, used online and offline, integrated and aggregated with other resources and disseminated online;
- **Primary.** It must be presented in a sufficiently disrupted manner, in order to be used by users to integrate and aggregate it with other data and content in digital format;
- **Timely.** It must be made public as quickly as is necessary to preserve its value;
- **Accessible.** It must be transmissible and interchangeable between all users on the network directly through Internet protocols, without any signing of contract, payment, registration or official request;
- **Computer readable.** Machine-readable, i.e., processable automatically by the personal computer;
- **Not owners.** Users must be able to use and process data through non-proprietary programs, applications and interfaces;
- **Free from licenses that limit its use.** The data cannot be subject to copyright or intellectual rights, nor patents that may limit access and especially the use and reuse by users. Furthermore, the data is "open" if users are guaranteed any way of use, even for commercial purposes;
- **Reusable.** Users must be enabled to reuse and integrate data, to the point of creating new resources, applications, programs and services of public utility for the user community;
- **Searchable.** Users must be able to search for data and information with ease and immediacy using ad hoc search tools, such as databases, catalogs and search engines;
- **Permanent.** The peculiarities described up to now must characterize the data during its entire life cycle on the web.

When making data with these characteristics available to citizens, businesses and other users, the Public Administration (PA) lays the foundations for creating Open Government as it places transparency, collaboration and participation at the center of its work, while opening up to dialogue and in direct confrontation with private individuals.

¹ The Digital Administration Code (CAD) is a single text that brings together and organizes the rules concerning the computerization of the Public Administration in relations with citizens and businesses. Established with legislative decree n. 82 of 7 March 2005, it was subsequently modified and integrated first with legislative decree n. 179 of 22 August 2016 and then with legislative decree n. 217 of 13 December 2017 to promote and implement digital citizenship rights.

3.2. The regulatory path in Italy

The PA produces, manages and stores an enormous wealth of data. Part of the public data is produced as the ultimate purpose of the administrative action (e.g., ISTAT) whereas another part is a secondary product compared to the public services offered (e.g., Data on public transport).

Through the paradigm of open government, the PA can make these public information assets available to all external parties, whether businesses, citizens or trade associations contributing to the economic and social development of the country.

The legislator - aware of the advantages and opportunities created by open data - has introduced open data as an administrative practice on several occasions and at different levels.

The main references in this sense are [18]:

PSI - Public Sector Information - the European directive on public sector information favors the reuse of data from European public administrations. In particular, the directive obliges administrations to make the data in their possession available for reuse for both commercial and non-commercial purposes, in compliance with the legislation on the protection of personal data. The directive also extends this provision to cultural institutions such as libraries, including universities, museums and archives, which were previously excluded. The Directive, which aims to ensure the most widespread reuse of open data in the Member States of the European Union, is part of the regulatory instruments for implementing the Digital Single Market. It is estimated that the total value deriving from the reuse of public sector information could grow to 194 billion in 2030. Numbers that have prompted the EU to start a process of revision of the relevant EU legislation in order to fully exploit the information potential of the public sector for the European economy and society, eliminating the existing barriers in the various national contexts and identifying datasets whose reuse would bring strong commercial value, as well as in terms of accountability (high value dataset). The review process was consolidated with the publication of Directive 2019/1024 in the Official Journal of the European Union;

CAD - Digital Administration Code - Provides the reference regulatory definition for open data format and open data, also listing the main features. The Decree-Law of May 19, 2020 (Relaunch Decree) intervened by modifying some parts of the Code following the effects of the global Covid-19 pandemic, in order to favor the creation of data exchange agreements between public offices;

Legislative Decree n. 101/2018 implementing the GDPR - General Data Protection Regulation, seen in this perspective not only as an initiative for the protection of European users, but above all as an interesting example of regulation of a specific economic sector based on data by a public entity (EU);

Three-year Plan for Information Technology in the Public Administration - Guidance document for the country's digital innovation, the 2020-2022 version dedicates an entire chapter to data, defining

the objectives in compliance with the European data governance strategy and with the framework outlined by European Directive on Open Data and Reuse of Public Sector Information (2019/1024).

The subject of data is also taken up by the Strategy for technological innovation and digitization of the country 2025 [19]. In the overall elaboration, the chapter takes into account, in particular:

- the “First Challenge: a digital society” with particular reference to objective 3 of the same;
- the “Second challenge: an innovative country”.

More specifically, the public information assets and the use and sharing of data by administrations and individuals is enhanced and encouraged, and in line with the action "A09_Data for the cities of the future".

3.3 Open Data Maturity Report

By the European Data Portal, the Open Data Maturity Report aims to annually monitor the maturity level of open data in all Member States, following four dimensions starting from 2018 [20]:

- **Open Data policy**, which analyzes the presence of public policies aimed at disseminating open data; whether or not there are offices and governance structures to promote stakeholder participation; the use of regulatory frameworks and regulations useful for making the data opening processes homogeneous and automated;
- **Open Data portal**, which focuses on the state of advancement of national portals, in terms of functionalities for interaction between users, and examines the coverage of open data in different domains, as well as the approach and measures in place to ensure the interoperability;
- **Open Data impact**, which examines the activities carried out to monitor and measure open reuse and the impact resulting from such reuse in four areas: political, social, environmental and economic;
- **Open Data Quality**, which focuses on the measures taken by Member States to ensure the systematic collection of metadata from local portals and, when possible, analyzes compliance with the European standard DCAT-AP metadata model.

4. Open Data: licenses, formats, Raw data and LOD

4.1 The licenses

Licenses need to be separated from policies, since they need to be machine-readable. In addition, a single policy can refer to or require many licenses.

Creative Commons licenses are preferred because [21]:

- There are human readable, legal, and machine readable versions of each license;
- The licenses have been verified for many jurisdictions.

Machine readability is important for two reasons:

- This allows automated determination of conditions applicable to download and visualisation of digital objects (if any), and allows search facilities to distinguish results on the basis of license conditions;
- Composite objects (mash-ups) and derived objects require automated computation of a resulting license.

Creative Commons licenses allow three types of qualifications [21], [22], [23]:

- **Permissions:** rights of the owner that are accorded to the end user;
- **Conditions:** performance expected from the end user in return for the permissions - for example the requirement to cite a work;
- **Limitations:** rights that are expressly not granted to the end user (for example disallowing commercial applications).

There are legitimate considerations for making licenses restrictive:

- release of the data can harm an individual, community, or natural environment. Such restrictions apply in the case of ethics and privacy concerns (personal data); indigenous community resources, and endangered species, for example;
- release of the data can impact the commercial and financial affairs or rights of an individual or organization.

Creative Commons licenses cannot be made more restrictive, and hence cannot be applied for the conditions of use identified above.

Table 1 provides a summary of licenses that should be available within a typical policy. Some of the more restrictive licenses may not be required in an open data policy [24].

License URL	Short Description
Public Domain: CC0	Use this universal tool if you are a holder of copyright or database rights, and you wish to waive all your interests in your work worldwide.
Attribution 4.0 International (CC BY 4.0)	Free to share, adapt, and apply the work, even for commercial purposes, provided that you give appropriate credit.
Attribution-NoDerivatives 4.0 International (CC BY-ND 4.0)	Free to share, even for commercial purposes, provided that you give appropriate credit.
Attribution-ShareAlike 4.0 International (CC BY-SA 4.0)	Free to share, adapt, and apply the work, even for commercial purposes, provided that you give appropriate credit and distribute the adaptations under the same license.
Attribution-NonCommercial 4.0 International (CC BY-NC 4.0)	Free to share, adapt, and apply the work, for non-commercial purposes, provided that you give appropriate credit.
Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0)	Free to share, adapt, and apply the work, for non-commercial purposes, provided that you give appropriate credit and distribute the adaptations under the same license.
Attribution-NonCommercial-NoDerivs : (CC BY-NC-ND)	Free to share, provided that you give appropriate credit, apply the work for non-commercial purposes only, and do not distribute modifications or derivatives. Note that this is not a free culture license.
Restricted: Ethics and Privacy ("No Harm")	The work is restricted in its use in one of the following ways: the end user must be known to the provider of the data, and/ or can only access the data under controlled conditions. Publication of derivatives to be agreed with the data provider.
Restricted: Commercial Value Non-standard - URL provided by owner	The work is available only on payment of an access fee or license fee. The owner of the work is free to determine any further restrictions on use by way of a license for the work.

Restricted: Classified Government Data Non-standard - URL provided by owner Access to the work is restricted due to government classification.	Access to the work is determined on individual merit and the owner of the work has full discretion in this regard.
--	--

4.2 The formats

In 2010, the inventor of the World Wide Web and the creator and advocate of the Semantic Web and Linked Data, Sir Tim Berners-Lee, suggested a 5-star deployment scheme for Linked Open Data [25]. The rating begins at one star and data gets stars when proprietary formats are removed and links are added.

To ensure that the data is read and interpreted by computer applications, the information must be represented through structured formats, that is, in tables organized in rows and columns and aggregated together according to a logical scheme.

With respect to formats for open government data, it is common practice to refer to the five-star scale of Tim Berners-Lee which evaluates the quality of the formats used (Figure 2):

- One star - The data is available on the web (in any format such as Word or PDF), but released with an open license so that it can be considered open;
- Two stars - The data is available in a structured format that can be interpreted by software (for example a Microsoft Excel spreadsheet);
- Three stars - The data is in a structured format, as for the two-star level, however this format is not proprietary, but is itself free and standardized (for example a file in CSV format);
- Four stars - In addition to respecting all the previous criteria, the data makes use of open standards (such as RDF and SPARQL) to identify information that can be identified in a timely or aggregated manner through a URL, i.e., a unique link that points directly to the resource;
- Five stars - The data meets all other criteria and also contains links to other data in order to provide context to your information. In this case we can actually talk about Linked Open Data.

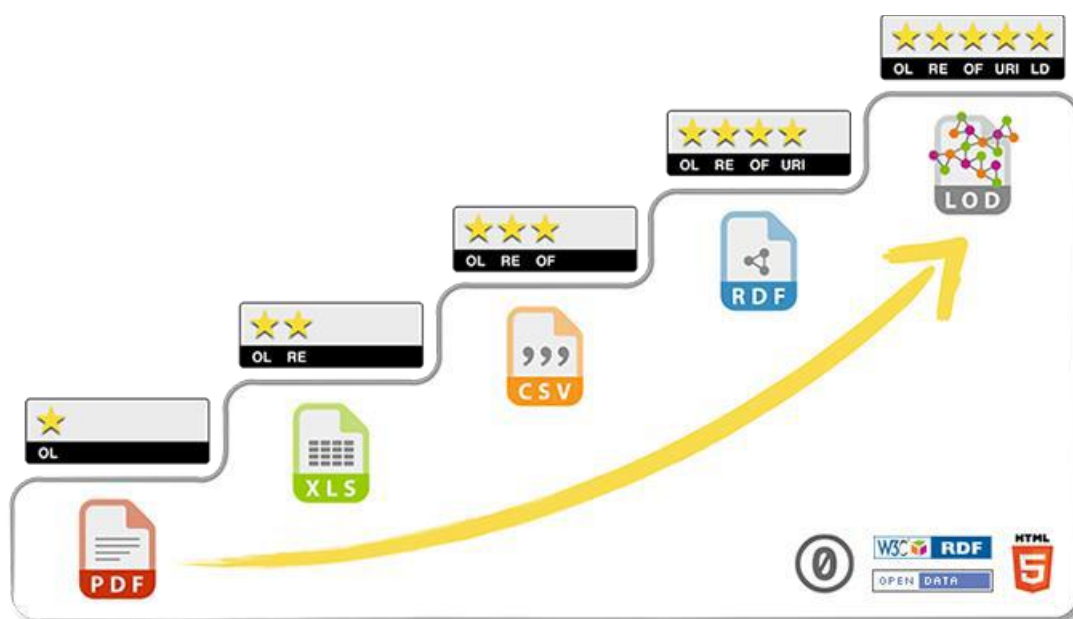


Fig. 2: A 5-star deployment scheme for Linked Open Data

5. SEBASTIEN Datasets

Datasets in the Sebastien project consist of (i) weather and climate information, (ii) territorial information and (iii) animal welfare information. Weather and climate data is based on meteorological variables and their combination. It is used to evaluate the effect of climatic conditions on animal welfare, behavior, production, reproduction etc. The territorial data considers the territorial characteristics, especially land surface, that can interfere with bioclimatic conditions and animal welfare. Animal-based data refers to metabolic, behavioral, productive and reproductive performance traits of animals that could be influenced by climatic variables.

The datasets have been divided into *background* and *foreground* datasets.

The *background datasets* can be further divided into datasets generated by a project partner prior to the project or under other funding and datasets provided by third-party organizations.

The *foreground datasets* are generated by the partners during the Sebastien project and under Sebastien funding.

The *background datasets* include the following datasets, containing weather and climate information (proposed in D2.2):

- E-OBS Gridded Observations;

- ERA5 Reanalysis;
- ERA5-Land Reanalysis;
- UERRA (Regional Reanalysis for Europe from 1961 to 2019);
- CERRA (Sub-daily Regional Reanalysis data for Europe from 1984 to present);
- C3S Seasonal forecast data;
- EURO-CORDEX data;
- VHR-REA_IT Dataset;
- ITALY8km-CM data;
- VHR-PRO_IT Dataset;
- COSMO-2I: COSMO at 2.2km – Italy area.

Access to these datasets is open, free and unrestricted.

The *foreground datasets* are:

- IoT data produced by sensors;
- Indicators produced as output of four services implemented in Sebastien.

In particular, IoT data is produced by a device named AnimalTalker. The AnimalTalker is composed of an AnimalCollar that receives data via Bluetooth (BLE) from different sensors, called AnimalButtons, and the AnimalButton(s). Each button is composed of a sensor, a battery and a Bluetooth microchip with its antenna. It sends the data collected in a string format using a BLE (Bluetooth Low Energy) service. The AnimalTalker gathers all the strings with their IDs and makes them available over the internet. The AnimalTalker has also a GPS module, an environmental temperature and humidity sensor, an accelerometer, and a NB-IoT/4G/5G module to allow internet connection.

The data collected by the AnimalCollar are:

1. the neck movement, using a 3 axis accelerometer, to register rumination and breathing;
2. the location (GPS signal);

3. the environmental data (temperature and relative humidity).

They will be released with CC BY 4.0 license and will be freely accessible through dedicated and standardized API of the catalogue.

Selected indices and indicators proposed to stakeholder are: (i) bioclimatic indicators; (ii) territorial indices and (iii) animal-based indicators.

The indicators and indices will be produced by the SEBASTIEN services, as described in the *M3 Report on Data Sources* document. Four services will be developed in SEBASTIEN:

- Service 1: Coping with environmental stressors for breeds;
- Service 2: Intensive farming risk management under climate extremes;
- Service 3: Extensive farming management and feed availability;
- Service 4: Livestock farming under risks from combined abiotic and biotic factors.

These services produce as output several data and maps that will be considered as datasets.

The foreground datasets – i.e., IoT real-time data for monitoring animal welfare and indicators built by the four Services - will be managed in SEBASTIEN by setting up a Data Lake that will store them and provide a unified API to facilitate access and download in different formats (netcdf, csv, maps, geoJSON, ...). The SEBASTIEN Platform Architecture and the Data Lake are described in detail in Milestone 5 – Report on Data Lake.

The Sebastien services and the related input and output datasets that will be under a license are described as follows.

5.1 Service 1: Coping with environmental stressors for breeds

This service is aimed at supporting livestock farming by matching breeds and identifying potential improvements for the different breeds, towards their acclimation, acclimatization and adaptation to environmental conditions and production needs. Acclimation means the coordinated phenotypic response developed by the animal to a specific stressor in the environment, whereas acclimatization refers to a coordinated response to several simultaneous stressors. Adaptation instead involves genetic changes as adverse environments persist over several generations.

This service will use climate data (from monitoring and modelling) for characterizing the farms environments in the recent and future periods, plus -omics data, phenotypic data (from monitoring

sensors and projects - e.g., LEO - by the partner AIA) and other datasets owned by farms/associations.

The huge amount of data will be combined through simple analytical/statistical to more complex AI/ML² approaches to find relationships between overall farming conditions, management options and observed adaptation potential/resilience.

Input data: The user will select:

- a given geographical area with increasing extension (e.g., province, region);
- a species/breeds;
- environmental settings (seasonal and long term 2050).

Output data: An assessment of production performance (production decline, reproductive decline, milk quality, welfare) based on long-term (2050) or seasonal climate change. The output will be single values, maps, tables that will be released with CC BY 4.0 license.

5.2 Service 2: Intensive farming risk management under climate extremes

This service aims to alert about the approaching (in short-term) or projected (in the long-term) dangerous environmental circumstances for cattle in the farms, due to indoor and outdoor conditions. These causes of discomfort, loss of reproduction and/or production (milk yield and milk quality) performances, including change in animal behaviour, are all assessed by integrating already archived and/or newly collected phenotypic and -omics data, thanks to the synergy with the HIGHLANDER project and other projects from partner AIA.

Dangerous conditions will not only refer to heat stress due to high temperatures, but also to the combination with other parameters (humidity, wind, solar radiation/shading when outdoors) to support the regulation of ambient conditions indoors and schedule outdoor periods. Several indicators/indices combining these parameters will be considered, with classification and thresholding systems fine-tuned for stakeholders according to species/breeds.

Input data: The user will select:

- farm location and other information;
- weather variables: weather forecasts at 48 hours (outdoor stable) and historical time series of temperature and relative humidity for both indoor and outdoor farm stables.

Output data: The forecast of indoor stable temperature and the prediction of the stress index for animals will be given. The output data composed of single values and maps will be released under CC BY 4.0 license.

² AI / ML – Artificial Intelligence / Machine Learning

5.3 Service 3: Extensive farming management and feed availability

Service 3 is aimed at the evaluation of the use of pastures, by means of information extracted from satellite data elaboration.

Input data: The user will select:

- the grazing area;
- the type of animals (bovine, ovine);
- the number of heads or days (if available).

Output data: The expected product will be indicators that allow obtaining a warning on pasture availability (e.g., evaluation of the number of days that the pasture is available, based on animal load) and the animal load (e.g., evaluating the maximum number of animals per number of days for a pasture). The output data composed of single values and maps will be released under CC BY 4.0 license.

5.4 Service 4: Livestock farming under risks from combined abiotic and biotic factors

This service aims to add further datasets about the livestock phenotypes that are usually associated with the presence of parasites and the outbreak of diseases, jointly with literature information about the environmental conditions that can potentially favor or trigger them.

Specifically, this service will allow the creation of disease/vector risk maps using long-term (2050) and seasonal climate data. A particular model will be implemented based on the weather/climatic conditions, which calculates the risk of disease on the basis of the vectors, for example. These maps will not be available for the whole Italy as localized data refer to the regions of Sardinia and/or Tuscany.

Input data: The user will select:

- a geographical area;
- the species;
- animal density;
- vaccination state;
- the seasonal forecast and long-term projections up to 2050.



Output data: The expected product will be (i) seasonal risk maps for vectors/pathologies (average value deviation of the same period). The maps will be monthly updated; (ii) long term risk maps for vectors/pathologies (the map will be produced only once).

These maps will be released under CC BY 4.0 license.

6. Conclusions

This Milestone describes the open data policy of datasets included and produced in the SEBASTIEN project. Large amounts of datasets are managed by other projects and used in SEBASTIEN for obtaining different outputs. Raw data in the project is acquired by IoT sensors. Moreover, several indicators are produced in the four project services. The data will be open, free and released under license CC BY 4.0. Registration to the portal will be required to access the services. These datasets will be stored and managed in the SEBASTIEN Data Lake and will be available in the SEBASTIEN Data Portal for free access, visualization and analysis.

7. References

1. *Open Data Handbook*. Web Site: <https://opendatahandbook.org/guide/en/what-is-open-data/>
2. *Open Government*. Web Site: <https://obamawhitehouse.archives.gov/open/about>
3. B. Ubaldi (2013), *Open Government Data: Towards Empirical Analysis of Open Government Data Initiatives*, OECD Working Papers on Public Governance, No. 22, OECD Publishing.
4. OECD (2010), *Towards Smarter and More Transparent Government, E-government Project*, Web Site: <https://www.oecd.org/gov/digital-government/44934153.pdf>;
5. United Nations (2010), *United Nations E-Government Survey 2010*, Leveraging E-Government at a Time of Financial and Economic Crisis, DOI: <https://doi.org/10.18356/0e749d15-en>
6. Ministerial Declaration on eGovernment approved in Malmö, Sweden, on 18 November 2009. Web Site: <https://www.aoc.cat/wp-content/uploads/2014/09/declaracio-malmo-1.pdf>.
7. Open Government Partnership initiative (2011). Web Site: <https://open.gov.it/en/open-government/ogp-italy>
8. *Creating Value through Open Data*. Web Site: https://data.europa.eu/sites/default/files/edp_creating_value_through_open_data_0.pdf
9. *Open Coesione*. Web Site: <https://opencoesione.gov.it/>
10. *Open Bilanci*. Web Site: <https://openbilanci.it/>
11. *Confiscati Bene*. Web Site: <http://www.confiscatibene.it/>
12. *Moovit*. Web Site: <https://moovit.com/>
13. *Infodata*. Web Site: <http://www.infodata.ilsole24ore.com/>
14. *Zillow*. Web Site: <http://www.zillow.com/>
15. J. Millard (2017), *European Strategies for e-Governance to 2020 and Beyond*, in *Public Administration and Information Technology, Government 3.0 – Next Generation Government Technology Infrastructure and Services*, Springer International Publishing, Cham, http://dx.doi.org/10.1007/978-3-319-63743-3_1.
16. EU (2016), *Towards faster implementation and uptake of open government*, prepared by PWC.
17. Carullo G. (2017). *Gestione, fruizione e diffusione dei dati dell'amministrazione digitale e funzione amministrativa*. Torino, Giappichelli.
18. European Commission (2022), *Digital Public Administration factsheet 2022 Italy*. Web site: https://joinup.ec.europa.eu/sites/default/files/inline-files/DPA_Factsheets_2022_Italy_vFinal.pdf
19. Minister for Technological Innovation and digitization (2020), *The strategy for technological innovation and the plan 'Obiettivo 2025'*. Web Site: <https://assets.innovazione.gov.it/1610546390-midbook2025.pdf>
20. *Open Data Maturity Report 2021*. Web Site: https://data.europa.eu/sites/default/files/landscaping_insight_report_n7_2021.pdf.
21. Creative Commons: License Features, <https://creativecommons.org/choose>
22. Various: Choose a License, <https://choosealicense.com/>
23. Creative Commons: Considerations for licensors and licensees, https://wiki.creativecommons.org/wiki/Considerations_for_licensors_and_licensees



24. Hugo, W and Pauw, J (2013): Notes on Data Sharing Policy in South Africa - prepared for DST.
Web Site:
https://docs.google.com/document/d/1KnMAqwRbvfnkn0PeLIXTSjANB_DHgT4qbaEjm-HsY6k/edit?usp=sharing
25. 5 Stars Open Data. Web site: <https://5stardata.info/en/>