EXPLORING TECHNOLOGIES FOR DATA ALTRUISM ORGANIZATIONS

DATALOG

DATA FOR PLANETARY WELLBEING

A Product White Paper from DATALOG. v2

DATALOG is a Data Altruism Organization focused on data for planetary wellbeing. The goal of DATALOG is to enable data interoperability to serve altruistic purposes. This white paper focuses on the need to identify the technologies that can enable data governance for Data Altruism Organisations. First, we describe different requirements to enable and promote data governance and stewardship schemes. Second, we survey existing experiences in data stewardship to categorize available data models, technologies and architectures. Third, we define the needs of DATALOG and the potential architecture to be developed. Fourth, we explore the possibilities of integrating with the existing European Data Space initiative to foster interoperability in future ecosystems.

AUTHORS (IN ALPHABETICAL ORDER):

Giovanni Maccani (Ideas For Change)

Jim Ahtes (i2cat)

Manuel Portela (Universitat Pompeu Fabra)

Toni Rubio (DesideDatum)

Vladimir Estivill-Castro (Universitat Pompeu Fabra)

DATALOG

DATA FOR PLANETARY WELLBEING

EU Recognised Data Altruism Organisation

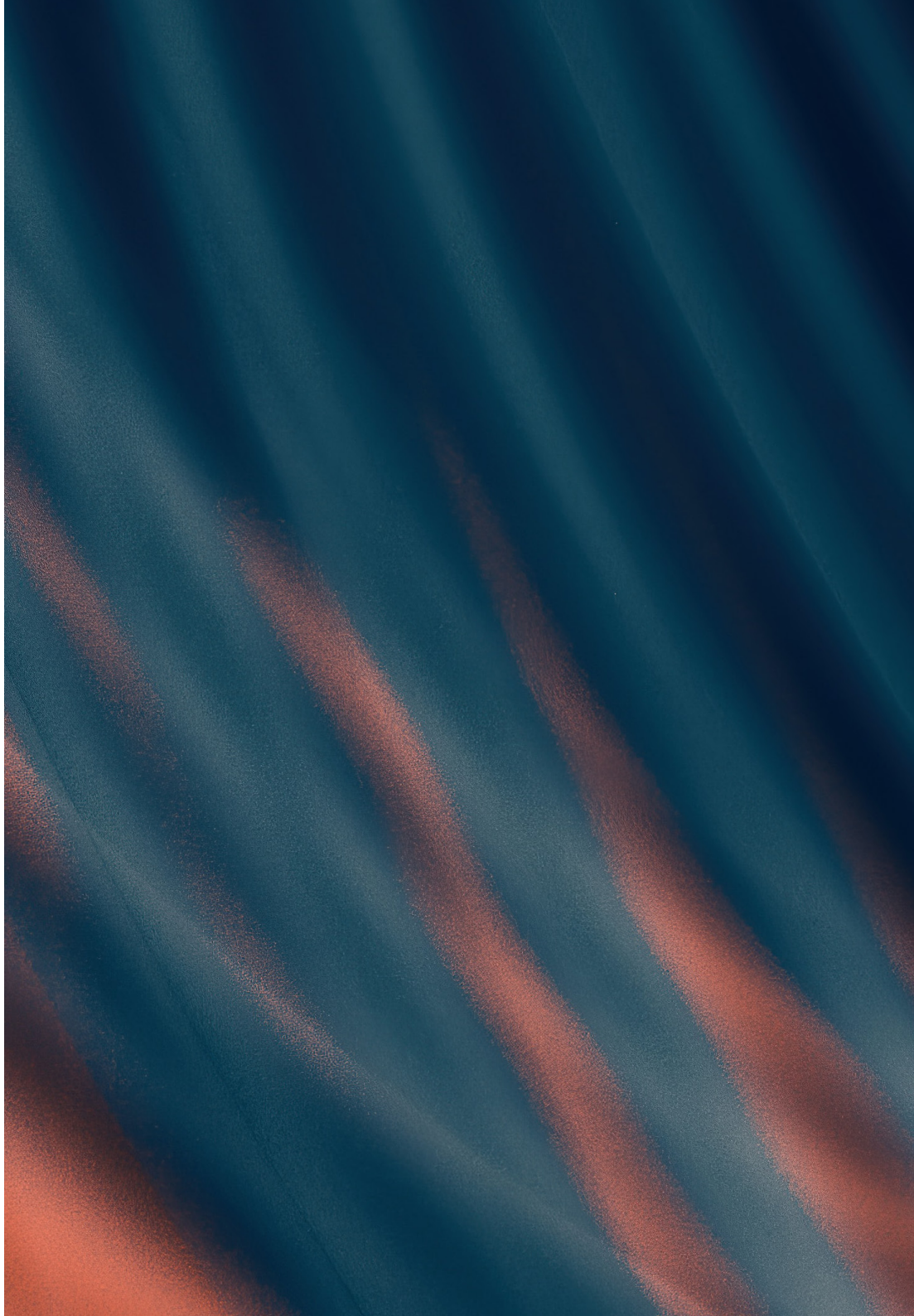
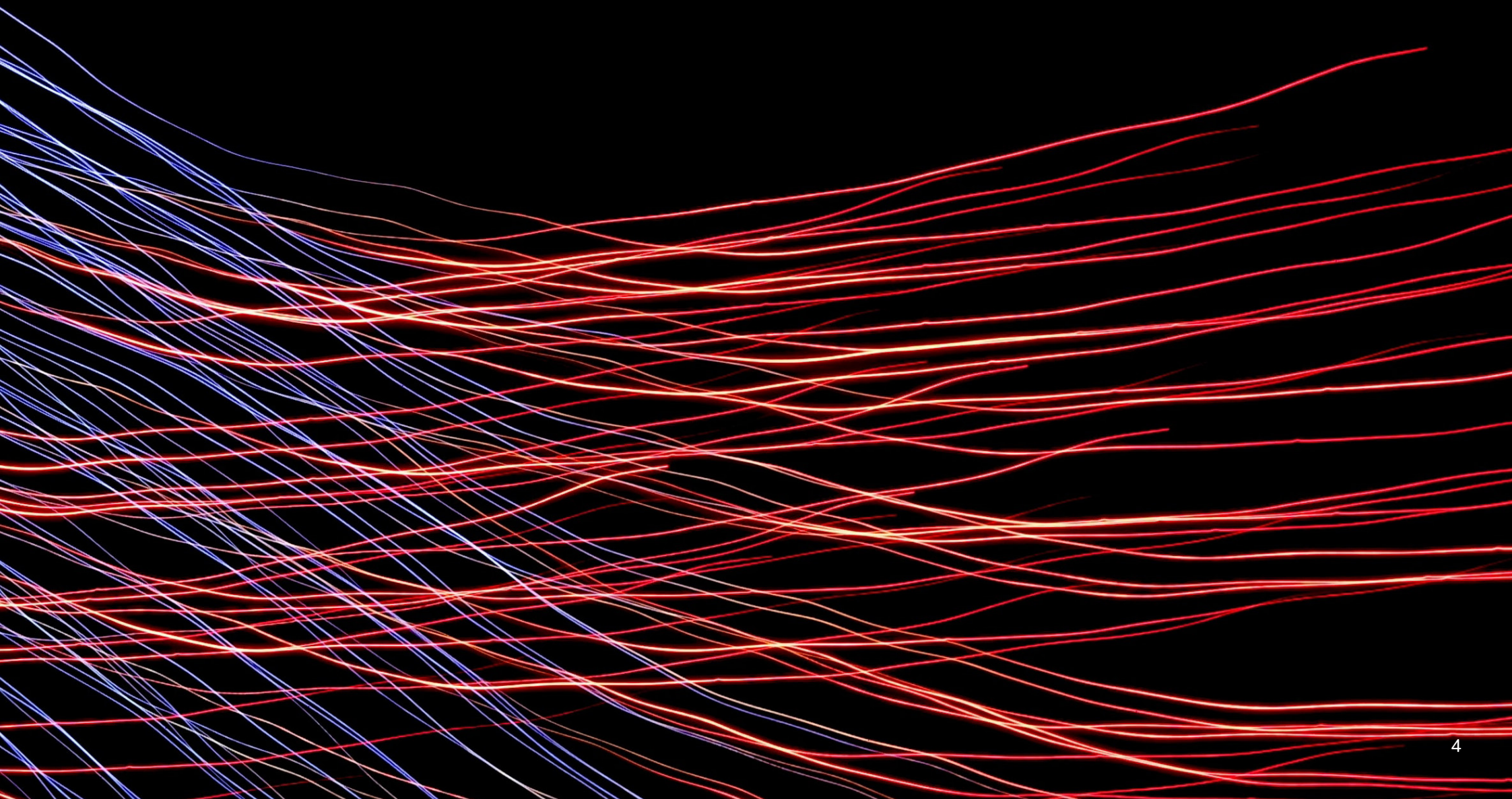


TABLE OF CONTENTS

INTRODUCTION



INTRODUCTION

This whitepaper describes DATALOG, an emerging Data Altruism Organisation created in Barcelona, Spain. This represents an innovative community which promotes responsible data sharing and reuse through new data governance mechanisms. DATALOG is an established legal entity developed with the goal of empowering citizens to take ownership and decisions on the consumption data from their utilities (water, electricity, gas) to promote responsible consumption, fight energy poverty and foster sustainability.

The use of data has been growing exponentially over the last decade. The continuous rise of open data is an important catalyst for data use across domains, actors and institutions. Driven by principles of transparency and service innovations, more and more governments and organizations make their data available to the public in different forms. This wave mobilized civil society towards generating new solutions to solve societal and environmental issues. Open data initiatives not only enabled the use of data for altruistic purposes but also allowed the creation of new services or the improvement of existing ones, or even the rise of new jobs and businesses. Based on the achievements so far, the Open Data Policy lab argues we are currently in the third wave of Open Data whereby we need "a much more purpose-directed approach than prior waves; it seeks not simply to open data, but to do so in a way that focuses on impactful reuse, especially through inter-sectoral collaborations and partnerships. The Third Wave pays at least as much attention to the demand as to the supply side of the data equation, and it is concerned not simply with data itself but with the broader technical, social, political and economic context within which data is produced and consumed."[1] Rethinking and shaping principles of data governance and governance of data ecosystems is therefore of paramount importance.

To foster more collaboration, organizations should look beyond open data. Open data has proven to be useful but limited. Not all data can be open; particularly those data sets with private or confidential information. Therefore these new collaborations should contemplate different ways of governing data. Thus, the urgent need for intermediation between the data owners, processors and data users. This intermediation role shall ensure security and trust, while empowering people to exercise their rights in terms of how their data is shared and under what conditions.

The Data Governance Act[2] is the regulation that most explicitly addresses data intermediaries and thus can have a key role in promoting and sustaining them. The DGA proposal was published in 2020, it entered into force in June 2022 and will be applicable from September 2023 following a 15-month grace period. This will be complemented by the Data Act[3], in negotiations between the European Parliament and Member States at the time of writing, which clarifies the access rights of such data, which entities can do so, and under what conditions.

Data intermediaries can be of different types, depending on the governance schemes, needs, interests and stakeholders' involvement: Data cooperatives, Data trusts, Data Unions, Data Marketplaces, Personal or Sectorial Data Spaces or Data Sharing Pools are some examples among others. Within these, the directive defines Data Altruism Organisations as the vehicles for individuals and companies to share their data voluntarily for the benefit of society, i.e. for the common good.

[1] https://opendatapolicylab.org/third-wave-of-open-data/
[2] Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European data governance and amending Regulation (EU) 2018/1724 (Data Governance Act) https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32022R0868
[3] https://ec.europa.eu/commission/presscorner/detail/en/ip_22_1113

DATA TRUST

Currently, Data Trusts represent the most common form of Data Altruism Organisations. According to Delacroix & Montgomery[4], a data trust is "a legal structure to govern data use. Inspired by the legal framework set by trust law, a data trust is a mechanism for individuals to pool their data rights (or data) into an organization – the trust – which is then tasked with governing its use according to conditions stipulated at its establishment.[...] By pooling their rights into a trust, individuals may be better placed to acquire a voice in the setting of terms and conditions around data use, owing to the power associated with data aggregation. By enabling access to aggregated data, with consent negotiated on a collective basis, data trusts could remove key obstacles to realizing the potential underlying large datasets."

Data trusts are acknowledged to be viable and effective data governance models to address planetary wellbeing through new forms of urban data management (e.g. Construction Data Trust, Place, Data Trusts Initiative, Data Collaboratives, Commonwealth Data Trust, OpenEnergy). The kind of data considered is typically not strictly personal, sensitive or confidential, but should be kept anonymized and aggregated to avoid individual identification and profiling. In addition, compared to other forms of data governance, this data is not normally consumed (analyzed, exploited, etc.) by the users that it belongs to. Consequently, users' interest in being heavily involved is rare. Involvement should be optional and adapted to different interest levels, discarding direct representation governance models like Data Cooperatives or Data Unions.

On the other hand, more peer-to-peer solutions like data Marketplaces or Data Spaces assume a clear value in data fostering sharing among peers. The full potential of peer-to-peer solutions have not been achieved yet. This is largely due to the unavailability of this data which prevents the growth of interest in using and sharing it, thus triggering virtuous cycles of data generation-sharing-use. The Marketplaces or Data Spaces models also assume that all the parties have the knowledge and infrastructure to contribute their part in the transaction, which is different to the context of Data Trusts where citizens lack the means to exploit their data in different ways. Balancing the inequalities of data transactions reveals the need for tools and infrastructure for smaller and unrepresented actors in the data market. This is a primary aspect addressed in DATALOG. In conclusion, we have chosen the data trust model because it allows a representative model that will ensure different levels of participation with legal mechanisms for protecting the rights of the data owners, together with tools and infrastructure that enable new forms of data sharing while finding new value on data to achieve social and environmental wellbeing.

[4] Delacroix, Sylvie and Montgomery, Jessica, From Research Data Ethics Principles to Practice: Data Trusts as a Governance Tool (November 23, 2020). Forthcoming in: Pogrebna, G. & Hills, T. (2023) eds. Handbook of Behavioural Data Science, Cambridge University Press, Available at SSRN: https://ssrn.com/abstract=3736090 or http://dx.doi.org/10.2139/ssrn.3736090

DATALOG

Telecoms, utilities, mobility or food sectors are highly connected through sensors and smart systems. These data are complex to analyze, siloed, and, in many cases, they are inaccessible due to confidentiality and privacy concerns. This unavailability makes research and urban innovation hard to accomplish. DATALOG aims to unlock this value through increasing the knowledge of urban flows and developing better public policy by making granular data accessible and providing data analysis tools with a privacy-by-design approach. DATALOG is therefore a Data Altruism Organisation that promotes volunteer data collection and altruist data sharing for the common good and research purposes.

Good data governance and flexible data interoperability will help industries to develop innovative products and services, making the economy more efficient, sustainable and responsible. With more accessible data, the public sector can develop better public policies, more transparent governance, and more efficient public services. Thanks to the DATALOG tools, those citizens and businesses that adhere to the Data Altruism Organisation will be able to know their detailed consumption patterns, to compare themselves with other consumer profiles, and gain awareness about their impact on their city. This ultimately creates a community that promotes the collective fight against climate change. From here, new data sources and potential services will be continuously researched, thus exploring how these emerging data governance mechanisms can be applied for positive impact across domains.

DATALOG is a Data Altruism Organisation with participatory principles that offers Data governance mechanisms through anonymization and aggregation algorithms. Consequently, DATALOG is looking to develops visualization and prediction tools to help understand the urban impact and reduce greenhouse emissions. The main advantage of DATALOG is the fiduciary model that helps to protect the privacy and confidentiality of data while allowing to develop data sharing practices to foster a data-driven planetary well-being and better urban policies.

FINDING THE VALUE IN DATA THROUGH DATA STEWARDSHIP

FINDING THE VALUE IN DATA THROUGH DATA STEWARDSHIP

One of the critical challenges in the data re-use process is to find value in data while protecting the privacy of the data subjects and the interests of other parites[5]. There is a vast academic literature on the so called usability-privacy tradeoff. However, in practice, organizations often hold and process data for specific purposes, while lacking the knowledge and resources to find secondary uses of such data. In many cases, the potential value of data is not realized because of the need for more expertise or skills. The GovLab[6] calls for fostering Data Stewards in organizations to tackle this problem.

"Data Stewards are organizational leaders or teams empowered to create public value by re-using their organization's data (and data expertise); identifying opportunities for productive cross-sector collaboration and responding proactively to external requests for functional access to data, insights or expertise. They are active in both the public and private sector, promoting trust within and outside their organization."

Data stewardship is essential in private companies, public administrations, cooperatives or any organization that holds and processes data. Data stewardship is not an obligation but sets the path to a responsible data management[7-8]. Nevertheless, in order to find the right value of data, it is necessary to understand the organization's role and its place within the ecosystem. This role can change at any time due to the transformation process, adopting new roles in the market. Data re-use and data sharing practices come with a big responsibility. Organizations must coordinate their procedures and methods to protect and prevent risks when using, holding, and processing data. Data stewards should work closely with data protection officers (DPO) to avoid negative consequences.

DATALOG, as a Data Altruism Organisation (inspired by the history and experiences of data trusts), has a data steward in its heart and embedded in its structure. The foundational document of the organization establishes an advisory board that acts as the data steward. All data intermediary organizations should have similar figures in the ground of their governance structure to have enough authority to make decisions based on data and continuously define new value propositions that enrich and enhance their objectives.

Once data value is defined, the next step is to put in place a strategy for data governance. This involves defining the requirements and the deployment of technologies that allow management and control of all data flows inside and outside the organization. In the following sections, we explore current technologies used for data governance (both in internal and external settings) and the architectures that are being built around the emerging European Data Spaces infrastructure.

[5] O'Hara, K. (2019). Data Trusts: Ethics, Architecture and Governance for Trustworthy Data Stewardship. Web Science Institute.
[6] https://thegovlab.org/static/files/publications/wanted-data-stewards.pdf
[7] https://www.theodi.org/article/defining-responsible-data-stewardship/
[8] Verhulst, S. (2021). Reimagining data responsibility: 10 new approaches toward a culture of trust in re-using data to address critical public needs. Data & Policy, 3, E6. doi:10.1017/dap.2021.4

USE CASES OF A DATA ALTRUISM ORGANISATION

USE CASES OF A DATA ALTRUISM ORGANISATION

A research institution needs access to aggregated data for a research project

Multiple research institutions and universities are constantly looking for data to develop research projects and analysis. Sometimes, data accessed through open data portals is enough. In other cases, leveraging Article 89 of the GDPR, specific data sharing contracts can facilitate this process. Data sharing contracts can be, however, problematic. Typically, these can take longer than expected and the technical mechanisms for sharing the data are not (yet) standardized, making it difficult to effectively access the required data. Through the agreement with their beneficiaries, DATALOG can negotiate directly with multiple parties that request data for research purposes under a standardized procedure. This not only speeds up the process but also reduces the resources needed to process and use the data in different ways. Data can be exchanged through specific datasets or through Application Programming Interfaces (API). In exchange, DATALOG can participate in research projects and incorporate new data to its database as a result of the sharing agreement, or having a monetary income to finance its operations.

A local administration needs to collect, manage and access to citizen data for a policy prospect

Local administrations could benefit from data from different sources to design new public policies or to monitor existing ones. Nevertheless, local administrations typically lack resources and knowledge to manage personal data, raising an important barrier to accessing and reusing data. A service of intermediation and data management could be offered on-demand. As an expert on data government and management, DATALOG can help this match-making process, while ensuring full compliance with current regulations.

Citizens want to know more about their consumption and the carbon footprint from their neighborhood

In return for their contribution to DATALOG, users can have access to selected products including individual consumption reports and recommendations. For example, a dashboard that shows their consumption during the previous week where the user can set up alarms to be triggered when certain limits are reached. This service can be free, include advertising or enhanced with characteristics by paying a subscription. Also, general and collective reports based on statistics generated from data held by DATALOG will be published, where access conditions can be established depending on beneficiaries' decision. These reports can be developed on a regular basis or on-demand in order to contribute to different organization objectives, in particular, those that work towards planetary wellbeing.

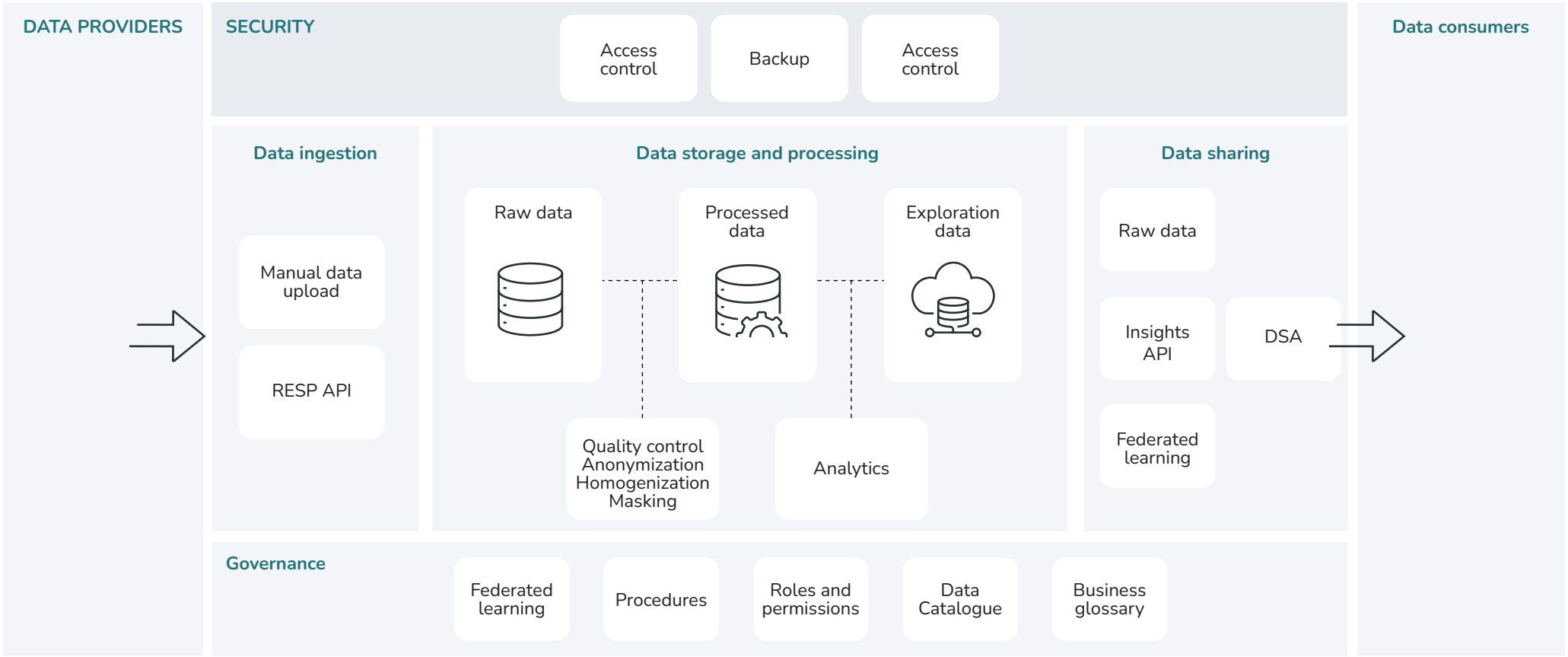
A company needs to effectively share controlled data to a third party

Sometimes companies and institutions hold data that are complex or that require substantial cleansing efforts before being shared. Through the experience acquired for data processing and data governance, DATALOG can facilitate tools and knowledge to help organizations to make data available through standardization, cleaning and anonymization. DATALOG can also help organizations to develop its data stewardship strategy and find value of data through different methodologies across multiple sectors.

DATA MANAGEMENT AND GOVERNANCE

DATA MANAGEMENT AND GOVERNANCE

From a technological perspective, there are many considerations and challenges to define the infrastructure supporting DATALOG. These include the following interrelated layers: data collection, data modeling and storage, data quality, security, governance, and data sharing.



Data Ingestion

In DATALOG, citizens provide data voluntarily, either directly uploading their data or authorizing the platform to collect the data from other sources such as the energy or water-providing company on their behalf. For this reason, the infrastructure enables direct data upload into the platform rather than acting merely as an interface accessing external sources.
There are be two mechanisms to upload the data into the platform. The ideal option is to directly obtain the data from the databases from the service providers, with a previous authorisation of the citizens. In this scenario, automatic processes to connect with external sources are triggered. The effectiveness of this process depends to a great extent on the level of maturity of the utility organization that holds the data. In this point, experiences like data spaces will be particularly useful, as it may simplify the process to integrate with different sources without the need to develop specific mechanisms.

Alternatively, there will be a module for the citizens to directly upload the data into the platform. However, in this scenario, ensuring the quality of the data may be particularly challenging and, therefore, the applicability of this data for analysis can be potentially affected.

Data storage and processing

As for any data-driven project, raw historical data is necessary to make valuable decisions. Usually, the more granular this data is, the more value can potentially be achieved. However, in this case, we must take into account several considerations about privacy and control over the data.

While data should be anonymized once it enters the platform, we must preserve individual registers in order to comply with the right of withdrawal as well as to enable personalized services (i.e. the individual consumption reports). This way, citizens should be able to remove their data whenever they want.

To ensure an additional security measure we shall divide the structure of the data storage into three layers: Layer one will contain data as it is, Layer two contain refined data, and Layer three will store data processed specifically for specific use cases and models.

Data Quality

When working with citizen data, data quality and completeness is one of the main challenges.

We therefore need to ensure that the data reaching the platform is of good quality. If we are integrating with an external source via API or similar, the process may be easier. But still, specific controls to check for data integrity, timeliness, accuracy, completeness, consistency, validity and uniqueness are needed.

When working with data uploaded manually by individual citizens, other factors have to be considered. The interface to upload the data has to be designed to be as clear and easy to use as possible, and controls to avoid mistakes should be implemented.

Also, double checking mechanisms should be implemented, such as requiring the citizen to upload the data and a picture of the document of the bill, or automatically checking for outliers according to other similar data points.

Once the data has reached the platform and we have ensured the proper quality requirements, there are other challenges to be taken into account. DATALOG may receive much data for a specific urban area, while only some from other regions. Also, data might be biased whether in the collection process or in the information provided by citizens that share their data with DATALOG. All these factors may pose challenges in the generalizability of the insights derived from the data, thus potentially affecting the quality of the final service.

A data quality process is in place to ensure data accuracy, consistency, completeness, and integrity across the whole data lifecycle. In this way, DATALOG also establishes data quality standards and mechanisms for data validation, cleansing, and remediation.

Data Sharing

DATALOG encompasses three different mechanisms for sharing data and insights for decision-making. First, sharing raw data with third parties only happen under strict control of privacy and data protection, avoiding sharing personal information and implementing aggregation and masking methods when necessary, according to GPDR requirements.

Second, there is also be the possibility to access insights. Offered through an authenticated REST API: we plan to implement a set of predefined questions in the platform and make these accessible upon request. For example, insights such as predicting gas or electricity consumption in a particular area will be available through the API.

Finally, DATALOG continues to explore the use of federated learning strategies to train specific models for third parties without the data leaving the platform.

All use of data or services from the platform is done under the establishment of a Data Sharing Agreement. These are agreements that define the purpose of the data sharing, cover what happens to the data at each stage and help all the parties involved to be clear about their roles and responsibilities.

Data Governance

Data governance processes are critical in many different types of organizations, and it is particularly sensitive and relevant in a data altruism organisation. In this area, the first step is to define roles, policies and procedures that guarantee a proper data governance implementation. The technological solution implemented has to be consistent with these.
Also, it is necessary to classify data assets within the platform based on sensitivity, value, and usage. Determine data ownership and define the rights and responsibilities associated with data access and usage.

Finally, effective auditing on the use of data and transparenting it is essential to sustain credibility and confidence of all stakeholders.

Security

In terms of security, given the sensitivity of the information DATALOG deals with, an approach of security by design needs to be applied from the beginning. This includes aspects such as establishing regular data backup procedures and a disaster recovery plan to ensure data can be restored in case of data loss, system failure, or other disasters. Data storage should be protected to avoid server downtime or intrusion to prevent any loss in the information. Data used in machine learning models should be taken care to avoid potential threats, like adversarial attacks or data leakage.

INTEGRATION WITH DATA SPACES

INTEGRATION WITH DATA SPACES

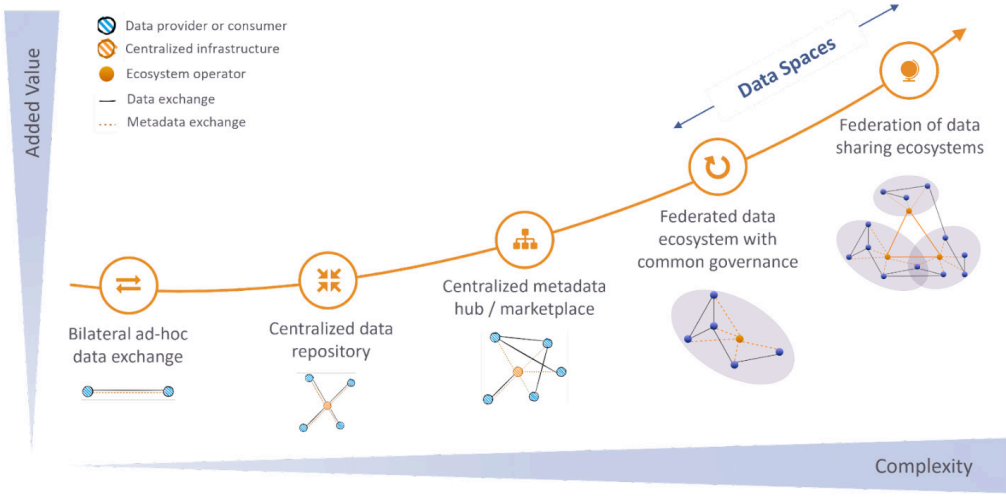
As defined above, a Data Altruism Organisation is meant to centralize decision-making and data. This concentration is useful for the case of looking for fiduciary duties and controlled data practices. Nevertheless, a different approach is needed when coordinating multiple actors with a purpose in common but with higher complexity. Data Spaces are one of the approaches adopted under the European Data Strategy to gather and share data from specific industrial sectors, particularly between organizations (B2B, B2G), by setting up a common governance over distributed data ecosystems. A data altruism organisation that uses and exchanges data from multiple sectors needs to be ready to be interoperable with different data sources. For this reason, we explain the nature and the technical aspects of the current approaches to data spaces to foresee how DATALOG can and will interact with data spaces in the future.

INTRODUCTION TO DATA SPACES

A Data Space is a federated ecosystem where voluntary, sovereign and secure data sharing materializes from a decentralized infrastructure following shared governance, organizational, regulatory and technical mechanisms. A Data Space facilitates the discoverability, access and re-use of the data provided by its participants and is complemented by data-driven services that are meant to generate added value for everyone involved[9] [10]. A fundamental principle is that of data sovereignty, where data providers have full control and traceability of their data across the value chain.

From a technical point of view, the data space model is the evolution of traditional data sharing and integrated solutions. Unlike traditional models, based on a centralized node of data or metadata, a data space has a highly decentralized and federated infrastructure, where participants communicate directly with each other, without the shared data being distributed by a central node outside the data provider.

This direct communication is possible thanks to standardized techniques and solutions, where the different participants in the data space can be connected through interoperable components and a common governance framework.



Evolution of data spaces as a federated data sharing architecture (source: Oficina del Dato, SEDIA, Spanish Government)[11]

[9] Nagel, Lars, & Lycklama, Douwe. (2021). Open DEI: Design Principles for Data Spaces - Position Paper (1.0). https://doi.org/10.5281/zenodo.5105744
[10] https://datos.gob.es/es/blog/la-importancia-de-desplegar-espacios-europeos-de-datos
[11] Alberto Palomo, Chief Data Officer - State Secretariat for Digitalization and Artificial Intelligence. Gaia-X Workshop on Tourism Data Space, June 9nd, 2022. https://gaia-x.eu/event/how-can-dataspaces-contribute-to-tourism-development-in-europe-through-citizen-centered-offerings/

KEY CHARACTERISTICS OF DATA SPACES

A data space is based on four fundamental design principles[12]:

Security and data sovereignty: security in data sharing is ensured through certification processes and approval of participants. Likewise, each participant maintains sovereignty over the data it makes available to the different participants through the definition of access, usage and re-use policies.

Equality of conditions in the data economy: a data space facilitates the entry of new actors to a given sector, making available the data offered by the different participants, always respecting their respective conditions and policies of use. This availability, in turn, generates incentives for new participants, allowing them to focus their efforts on the development and innovation of new digital solutions based on available data, instead of a market defined in terms of data exclusivity by monopolistic players.

Decentralized soft infrastructure: The infrastructure of a Data Space is not a monolithic infrastructure, but a collection of standards, reference models and software solutions that meet the predefined technical, legal and economic requirements to establish secure data exchange.

Participatory governance: the essence of a data space is the representation of the needs and conditions of all the actors involved, from both the private and public sectors, for the creation of value. These governance mechanisms foster the liquidity of the ecosystem and the development of an ecosystem around the data space. and environmental wellbeing.

To ensure compliance with these principles, the architecture of a data space must be formed by technical and governance building blocks. These blocks address four different directions relating to design principles: sovereignty and security, interoperability and standardization, creating added value through data, and governance of the data space and its operations.[13]

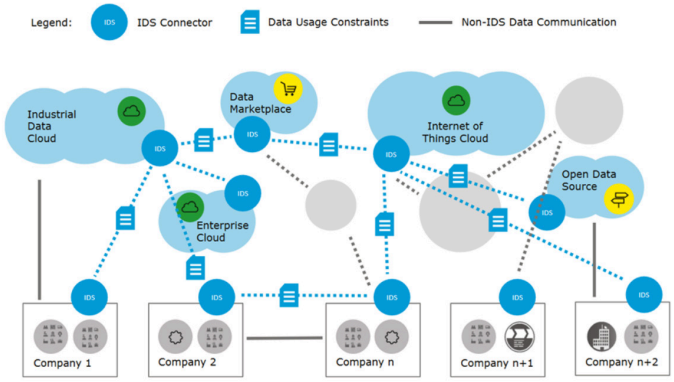
[12] Nagel, Lars, & Lycklama, Douwe. (2021). Open DEI: Design Principles for Data Spaces - Position Paper (1.0). https://doi.org/10.5281/zenodo.5105744

[13] Nagel, Lars, & Lycklama, Douwe. (2021). Open DEI: Design Principles for Data Spaces - Position Paper (1.0). https://doi.org/10.5281/zenodo.5105744

DATA SPACE ARCHITECTURES AND APPROACHES

The rise of data spaces and the different European initiatives and strategies that contemplate them, have created multiple solutions, approaches and reference architectures for their implementation. These solutions and initiatives contain design-level specifications, technological architecture and technical components needed to create data spaces themselves.

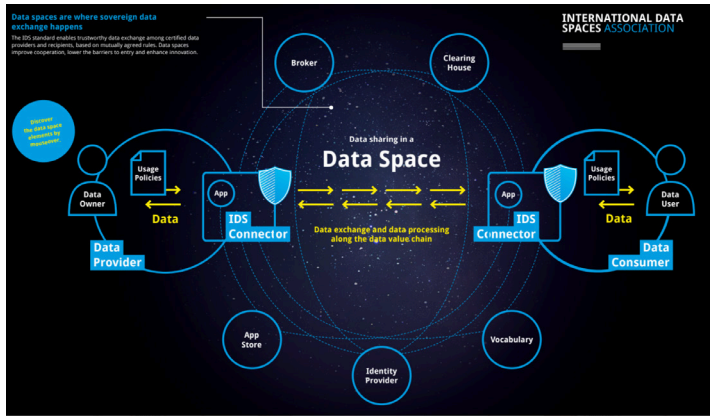
One of these alternatives is the reference architecture model created by the International Data Spaces Association (IDS-RAM), which defines all the necessary and complementary components required to implement a data space, respecting the principles of interoperability, sovereignty, trust and intermediation. A complementary architecture is that of Gaia-X[14], which focuses on the creation of federated services to provide an interoperability layer between different infrastructure providers and allow the generation of data spaces on top.



Peer-to-peer data space according to the IDSA association[15]

There is also an architecture centered around the FIWARE context broker[16], which acts as an interoperability layer to collect data from heterogeneous sources and offers it to a service layer. Finally, the iSHARE architecture model is a European standard for sharing international business data in a sovereign manner, governed by the requirements and guarantees of the iSHARE[17] foundation.

Although these solutions have been created independently, work has already begun on converging the different models to ensure greater interoperability between them through collaborations, such as the Data Space Business Alliance (DSBA)[18] and Data Space Support Centre[19].



Main components contemplated in the IDS-RAM reference architecture (IDSA)

[14] https://gaia-x.eu
[15] https://internationaldataspaces.org
[16] https://www.fiware.org/catalogue
[17] https://ishare.eu
[18] https://data-spaces-business-alliance.eu/dsba-releases-technical-convergence-discussion-document
[19] https://dssc.eu

EXAMPLES OF DATA SPACES IN THEIR PILOT PHASE

Using these solutions, several operational data spaces have already been launched in different sectors and across countries in Europe. Some examples are proposed below.

In the context of the mobility sector:

Mobility and Data Space (Germany)[20] which focuses on sustainable, multimodal and safe urban mobility. More than 200 stakeholders from the mobility sector in Germany from private companies, research centers and public administration are involved. It includes infrastructure, public transportation, traffic, road safety and environmental data.

Catena-X (Germany)[21] which focuses on connecting the supply chains of the automotive industry. It presents use cases to address several current issues in the sector, such as sustainability, simulation and traceability.

Eona-X (France)[22] which focuses on mobility, interurban travel and tourism, integrating industries like car companies, airlines and train operators for a federated catalog of multimodal data.

In the context of the energy sector:

SYNERGIES[23] which integrates energy grid systems & services, business innovation, big data management and citizen engagement to foster better energy services.

Enershare[24] which enables an electricity-centered data-driven energy ecosystem where energy (electricity, heat, natural gas), non-energy (e-mobility, water, security, health) and ICT/data stakeholders and operators can securely share, exchange and reuse energy data assets.

In the context of smart cities:

Data Space for Smart and Sustainable Cities and Communities (DS4SSCC)[25] engages different stakeholders, including local administrations, to enable green and digital transition. This is the first phase of a smart cities data space for Europe, focusing on the roadmap needed for future implementation.

Urban Data Space (Bulgaria)[26] which focuses on air quality prediction and social facilities coverage, it is held by the London Metropolitan University and the GATE Institute in Sofia.

USAGE (Urban Data Space for Green Deal)[27] involves the Open Geospatial Consortium, local administrations and technology companies that will mix earth observation and Internet of things to deliver the EU Green Deal to the cities.

In the context of manufacturing and logistics:

A prominent example is the Industrie 4.0[28], an industrial data sharing network in Germany.

Smart ConnectedSupplier Network (SCSN)[29] defines a communication and data sharing standard in the logistics sector in the Netherlands.

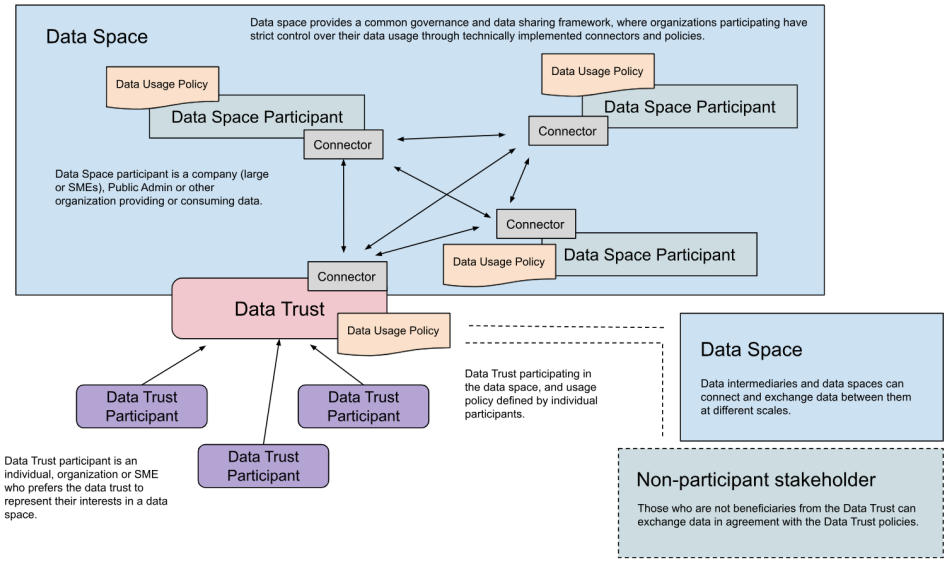
[20] https://mobility-dataspace.eu
[21] https://catena-x.net/en
[22] https://eona-x.eu
[23] https://energydataspaces.eu
[24] https://enershare.eu
[25] https://www.ds4sscc.eu
[26] https://idsa-bulgaria.gate-ai.eu/urban-data-space
[27] https://www.usage-project.eu/home
[28] https://www.plattform-i40.de/IP/Navigation/EN/Home/home.html
[29] https://smart-connected.nl/en

DATA ALTRUISM ORGANIZATIONS AND FUTURE EUROPEAN DATA SPACES

We briefly introduced the approach on Data Spaces to understand how different European stakeholders are working together to approach data interoperability. We believe Data Altruism Organisations have an important role within the European Data Strategy. One such role could be for these to provide common representation of its participants in larger Data Space deployments, to overcome the complexity of participating directly in the Data Space for individuals and even smaller businesses without the infrastructure to do so.

The DATALOG approach has at its core the ability to share and exchange multi sectoral data, and we foresee the capacity to engage with the Minimum Interoperable Mechanisms that are being developed for urban Data Spaces. However, these are the key differences, as a Data Altruism Organisation:

- Is a centralized entity, that acts independently,
- Has a fiduciary duty towards their beneficiaries,
- Can give representativeness and help to balance the power for those actors that lack of time, resources and knowledge to act in the data market,
- Has a direct relation towards their beneficiaries (in particular, those non-institutionalized),
- Can offer direct solutions to individuals, collectives and companies that are not in position to participate in a data space (e.g. lack infrastructure, data stewardship, or resources in general),
- Can negotiate and exchange data with data spaces on behalf of beneficiaries but also to other entities that are not part of any data space.
- By definition, focus solely on the common good, whereas data spaces not necessarily.



How Data Trusts could interact with Data Spaces

DEVELOPMENT PLAN

DEVELOPMENT PLAN

Objective and Deliverables

The main interest of the founders of DATALOG is to promote its adoption and growth. As a consequence, our development roadmap includes not only the implementation of the base infrastructure of DATALOG, but also to provide open source reference implementations of all the software components to be used. By releasing all these software components under open source licenses, we will encourage other cities and actors to develop and integrate their own infrastructures to build data altruism organisations in the measure of their specific needs.

We are also co-creating these tools with companies, social organizations, local administrations and citizens to make sure that our solution fits the needs of the urban ecosystem.

The DATALOG Stack under development , includes the following elements:

- Specification of the DATALOG Data Governance protocol
- DATALOG Data collection platform
- DATALOG Visualization platform
- DATALOG Individual Data Explorer

Subsequently, in order to promote adoption, we will also release implementations of other components that sit atop the aforementioned core components of DATALOG:

- Data Sharing Consent Dashboard to allow beneficiaries to lend data to third-parties
- Data Explorer Credits for recognizing expert contributors
- Ability to Request Data to add new sectors to DATALOG
- Participatory Funding with DATALOG to empower new initiatives

Short to Mid Term Roadmap

Establish the legal entity. Target: completed. A first step is to constitute the data altruism organisation as a legal entity to operate and perform their activities.

Survey about data donation and product development. Target: completed. To understand the needs from stakeholders and users, we released a general survey that questions the interests and worries about data sharing and privacy. This survey allowed us to set our goals and develop our pilot.

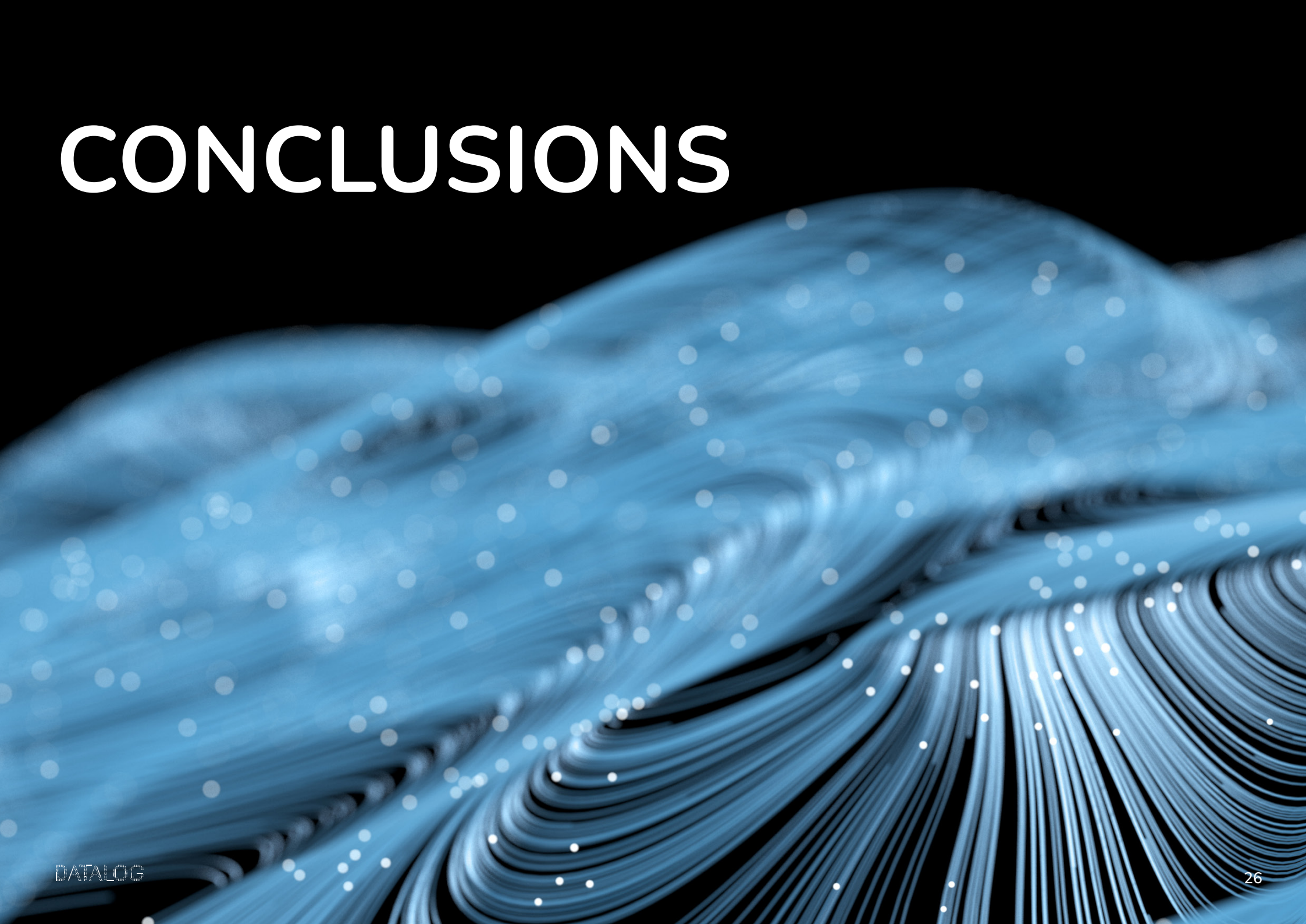
Data collection campaign and Pilot development. Target: Summer 2024. A first pilot will be set to test each of the core pieces of the DATALOG stack independently. We will start to collect, and process data in order to have the first models and visualizations.

Pilot Validation. Target: Q4 2024. A participatory validation of the pilot will be carried out to evaluate the first results and to iterate towards the first release of the MVP.

Release. Target: end of 2024. We will release our Minimum Viable Product (MVP) in Barcelona, where we will release the core pieces of the DATALOG stack and the protocol for data governance.

Beyond. The scenario is very nourishing. We see many potential applications, the opportunity to replicate and scale the model of DATALOG towards different horizons. We also foresee testing new applications to automate, enhance the data sharing and integrate new technologies to secure the privacy and confidentiality of data.

CONCLUSIONS



CONCLUSIONS

Data interoperability is a challenge that requires the alignment and collaboration of multiple actors. We show that there are many opportunities, models and architectures that can be adopted depending on several needs and requirements. To achieve a successful interoperability for planetary wellbeing we defined a combination of legal, social, technical and economic requirements that take the form of a Data Altruism Organisation called DATALOG.

A first pilot in Barcelona is taking place during 2023 and 2024 thanks to the support of Barcelona City Council[30] to test the proof-of-concept and develop the tools and gather the resources needed. With the experience, learning, and results obtained we expect to scale and replicate the model across Europe.

ORGANIZE:

upf. Universitat Pompeu Fabra Barcelona

IDEAS FOR CHANGE

SUPPORTED BY:

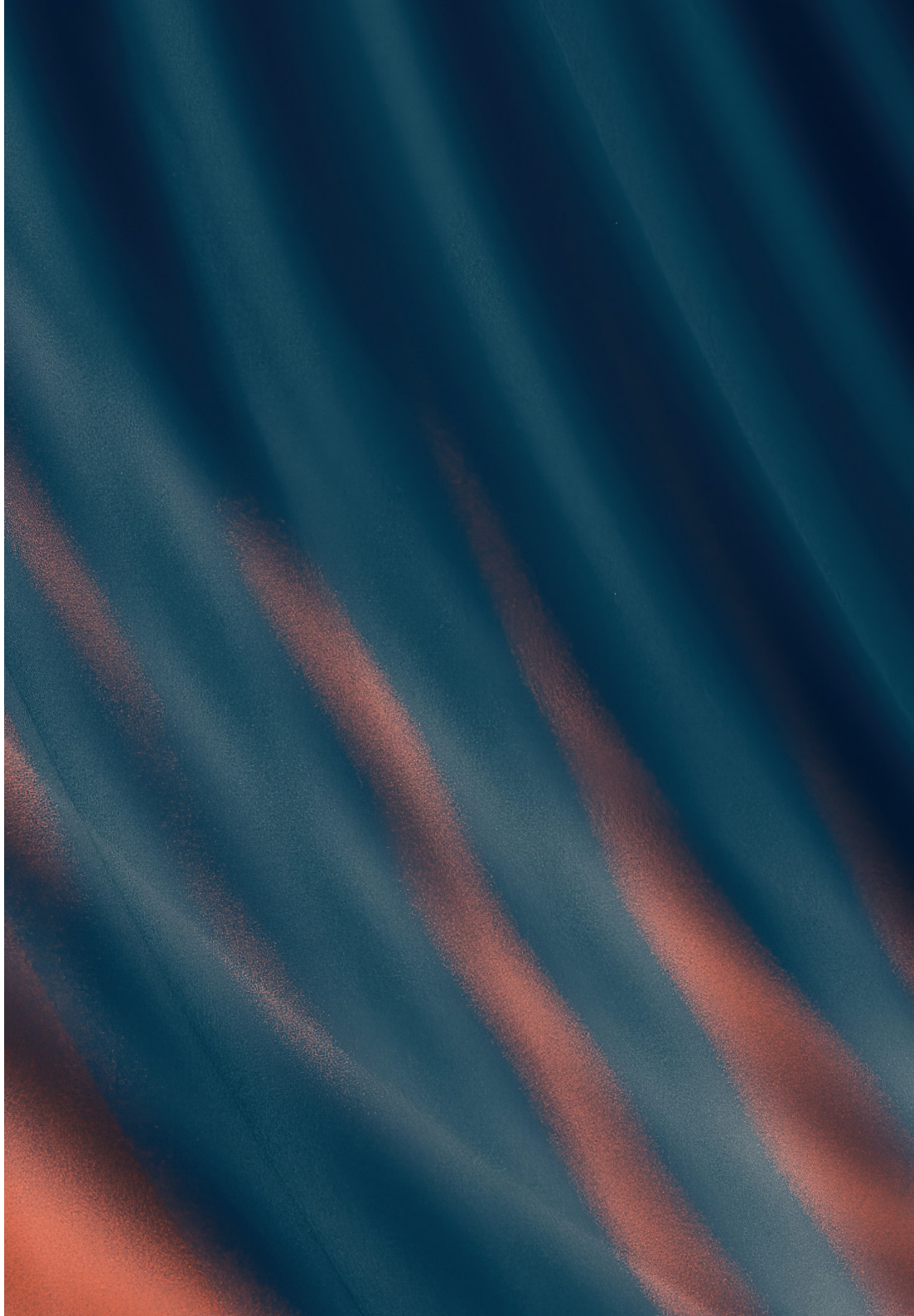
IN BARCELONA OVA

Ajuntament de Barcelona

[30] The DATALOG project was granted under the “La ciutat proactiva 2022” call from the Fundació BitHabitat from the Barcelona City Council.

This work is part of Maria de Maeztu Units of Excellence Programme CEX2021-001195-M, funded by MCIN/AEI /10.13039/501100011033.

DATALOG




DATALOG
DATA FOR PLANETARY WELLBEING
DATALOG.ES