

A Federated Cloud Solution for Transnational Mobility Data Sharing

Extended Abstract

Emanuele Carlini¹, Thierry Chevallier², Patrizio Dazzi¹, Francesco Lettich¹, Raffaele Perego¹, Chiara Renso¹ and Salvatore Trani¹

¹*Institute of Information Science and Technologies (ISTI), National Research Council (CNR), Pisa, Italy*

²*AKKA Technologies, Toulouse, France*

Abstract

Nowadays, innovative digital services are massively spreading both in the public and private sectors. In this work we focus on the digital data regarding the mobility of persons and goods, which are experiencing exponential growth thanks to the significant diffusion of telecommunication infrastructures and inexpensive GPS-equipped devices. The volume, velocity, and heterogeneity of mobility data call for advanced and efficient services to collect and integrate various data sources from different data producers. The MobiDataLab H2020 project aims to deal with these challenges by introducing an efficient and highly interoperable digital framework for mobility data sharing. In particular, the project aims to propose to the mobility stakeholders (i.e., transport organising authorities, operators, industry, governments, and innovators) reproducible methodologies and sustainable tools that can foster the development of a data-sharing culture in Europe and beyond. This paper introduces the key concepts driving the design and definition of a cloud-based data-sharing federation we call the *Transport Cloud platform*, which represents one of the main pillars of the MobiDataLab project. Such platform aims to ensure transnational access to mobility data in a secure, efficient, and seamless way, and to ensure that FAIR principles (i.e., mobility data should be *findable, accessible, interoperable, and reusable*) are enforced.

Keywords

Data-sharing, Mobility Data, Cloud Platforms, Federated Platforms

1. Introduction

Over the recent years the European Union devoted many resources and efforts to promoting data-sharing initiatives, platforms, and policies across its member states. Indeed, in the *A European Strategy for Data* vision¹ the European commission recognizes how exploiting data allows the private sector to continuously innovate and create new types of thriving businesses. At the same time, public entities and institutions can leverage public data to better understand societal dynamics and make informed decisions. The data sharing vision driving the policies

SEBD 2022: The 30th Italian Symposium on Advanced Database Systems, June 19-22, 2022, Tirrenia (PI), Italy

✉ emanuele.carlini@isti.cnr.it (E. Carlini); thierry.chevallier@akka.eu (T. Chevallier); patrizio.dazzi@isti.cnr.it (P. Dazzi); francesco.lettich@isti.cnr.it (F. Lettich); raffaele.perego@isti.cnr.it (R. Perego); chiara.renso@isti.cnr.it (C. Renso); salvatore.trani@isti.cnr.it (S. Trani)

🆔 0000-0003-3643-5404 (E. Carlini); 0000-0001-8504-1503 (P. Dazzi); 0000-0001-6914-2961 (F. Lettich); 0000-0001-7189-4724 (R. Perego); 0000-0002-1763-2966 (C. Renso); 0000-0001-6541-9409 (S. Trani)



© 2022 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

¹https://ec.europa.eu/info/sites/info/files/communication-european-strategy-data-19feb2020_en.pdf

of the European Union is that of a “[...] *single European data space as a market for data, open to data from across the world, where personal as well as non-personal data, including sensitive business data, are secure and businesses also have easy access to an almost infinite amount of high-quality industrial data, boosting growth and creating value, while minimising the human carbon and environmental footprint*”². Accordingly, several European projects and initiatives have been heavily financed in the recent past – just to name a few, GAIA-X³ [1], SoBigData++⁴ [2], FENIX⁵, SUNFISH⁶ [3], MOBINET⁷, and the Data for Road Safety Initiative⁸.

Let us now focus our attention on *mobility data sharing*, which is a particular instantiation of the vision described above and the main topic of our work. In general terms, mobility data can be defined as data that provides information on mobility patterns. For instance, in the urban domain it can be encountered in the form of network description, timetable information, car traffic, public transportation or other mobility modes, parking data, and accessibility data [4]. It is then clear that a common mobility data space would enable different stakeholders to share their data into a single, possibly distributed, platform that facilitates access, pooling, and sharing of data from existing and future transport and mobility databases.

Mobility data sharing – and enabling it appropriately – plays also a critical role in the decarbonization of the European Union. For instance, the European Green Deal⁹ and the European Data Space strategy¹⁰ strictly depend on the capacity of organizations to digitalize and share mobility data, since data-sharing and smart mobility solutions constitute an essential pillar toward the decarbonization of the European transportation sector. Indeed, over the recent years many systems enabling connected and automated *multi-modal mobility* took advantage of smart mobility solutions leveraging shared data and artificial intelligence. For instance, mobility data sharing can support several novel applications that improve intermodal connections in transport hubs – e.g., solutions that search for optimal levels of vehicle availability in car and bicycle sharing systems. All these considerations align with the goals pursued by the Global Roadmap of Action toward Sustainable Mobility [5], which states that mobility data-sharing programs and platforms can help the transition toward greener, safer, more accessible, and more efficient mobility systems.

Tackling the heterogeneity and peculiarities of mobility data, as well as the various constraints related to their safe and trusted sharing, is a core objective of the European H2020 *MobiDataLab project*¹¹. MobiDataLab envisions the usage of an open and federated cloud-based architecture to easily and practically enforce complex and often contrasting requirements coming from FAIR (Findability, Accessibility, Interoperability, and Reusability) and privacy principles. Indeed, a federated cloud can in principle support the sharing of arbitrary resources from arbitrary application domains, with arbitrary consumer groups across multiple administrative domains

²<https://eur-lex.europa.eu/legal-content/EN/TXT/>

³<https://www.data-infrastructure.eu/GALAX/Navigation/EN/Home/home.html>

⁴<https://plusplus.sobigdata.eu/>

⁵<https://fenix-network.eu/>

⁶<http://www.sunfishproject.eu/>

⁷<https://cordis.europa.eu/project/id/318485>

⁸<https://www.dataforroadsafety.eu/>

⁹https://ec.europa.eu/info/strategy/priorities-2019-2024/european-green-deal_en

¹⁰<https://digital-strategy.ec.europa.eu/en/policies/strategy-data>

¹¹<https://mobidatalab.eu/>

[6]. The MobiDataLab project primary objective is to foster data sharing amongst transport authorities, operators, and other mobility stakeholders operating in Europe, which in most cases want to maintain their data governance.

In the following we discuss the *Transport Cloud platform*, the cloud-based data-sharing federation the MobiDataLab project intends to propose.

2. A cloud-based architecture for FAIR mobility data sharing: the MobiDataLab Transport Cloud platform

The MobiDataLab Transport Cloud platform aims to facilitate access to mobility data in an open, interoperable, transnational, and privacy-preserving way. The platform also aims to adopt the FAIR principles [7] to model said access, i.e., mobility data available in a vast and diverse ecosystem that possibly encompasses many different sources should be *findable*, *accessible*, *interoperable*, and *reusable*.

The vision behind these goals stems from the needs and interests of the stakeholders behind the MobiDataLab project. Indeed, these are public and private institutions that have an interest to either act as mobility data *providers* (for instance, provide real-time public transportation data, road-network data, vehicle data), or are interested in *consuming* mobility data through the access and services provided by the transport cloud platform. Therefore, the platform has been designed according to federated cloud principles to offer solutions that strive to reduce (and, whenever possible, eliminate) current technical limitations that act as barriers to mobility data sharing and reuse.

In the following, we provide more details on the transport cloud platform that is being developed by the MobiDataLab project. First, we focus on the actors interacting with the transport cloud platform. Subsequently, we focus on the inner components of the architecture to detail how the platform enables, facilitates, and promotes mobility data sharing between data consumers and data producers.

2.1. Actors interacting with the transport cloud platform

Figure 1 introduces a first abstract perspective on the transport cloud platform architecture. The architecture is composed of a collection of *components* that perform key operations within the transport cloud. In the Figure it is also possible to identify four different types of actors that can interact with the transport cloud, namely:

- *Administrators* are individuals in charge of managing user accounts and the platform access, deploying applications within the platform, and configuring the components operating within the platform.
- *Developers* are individuals who deal with the deployment and integration of transport cloud components.
- *Data consumers* are entities that use the data and services available within the transport cloud. Relevant examples can be data scientists, researchers, domain experts, transport

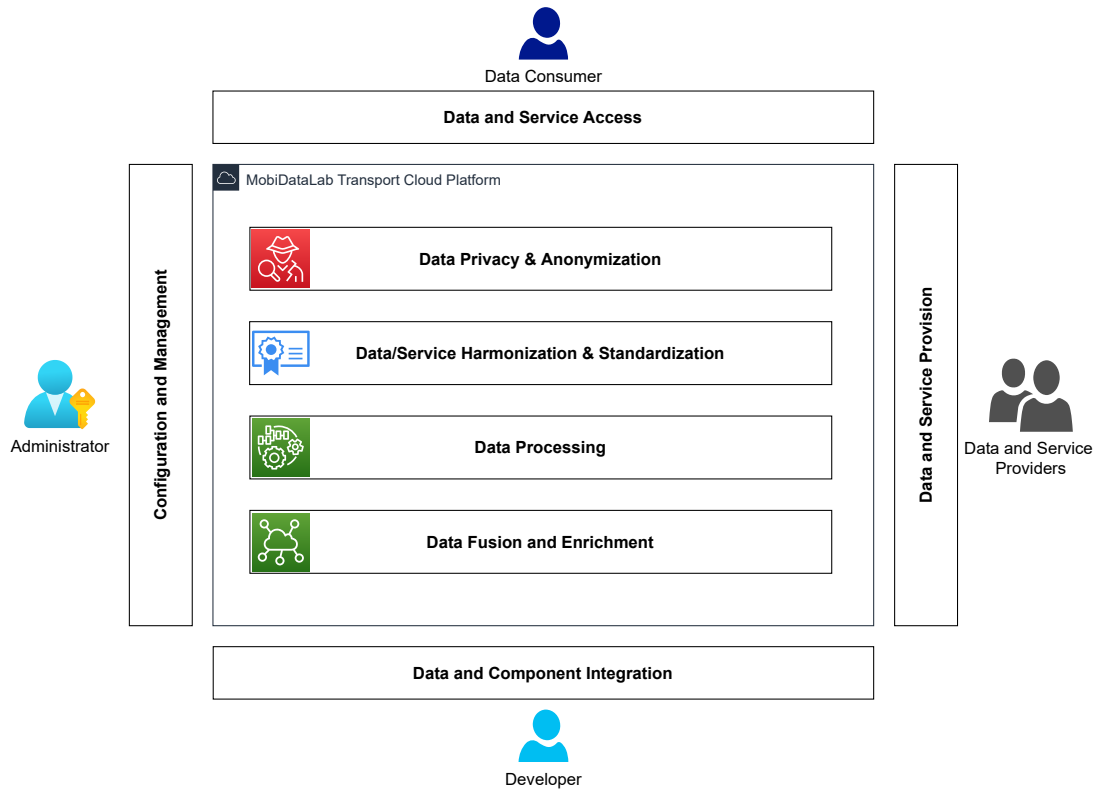


Figure 1: The MobiDataLab transport cloud platform architecture – simplified view.

customers, or external services that use the data and services offered by the transport cloud platform.

- *Data providers* are entities that provide, either passively or actively, data or services to the Transport Cloud. Some examples that were deemed relevant to the MobiDataLab project are Trip planners (e.g., Navitia¹², HERE¹³), MobiDataLab stakeholders (e.g., transport operators or public institutions that actively share their data and services for the good of the MobiDataLab project), and open data/services providers (e.g., OpenStreetMap¹⁴).

2.2. Information flow and key components within the transport cloud

The architecture of the transport cloud platform has been primarily designed to promote mobility data sharing between *data consumers* and *data providers*. Indeed, the components operating within the platform have the role of powering, sustaining, and adding value to the information flowing between these two types of actors. In the following, we focus on how the data consumers

¹²<https://navitia.io/>

¹³<https://www.here.com/>

¹⁴<https://www.openstreetmap.org/>

MobiDataLab Architecture

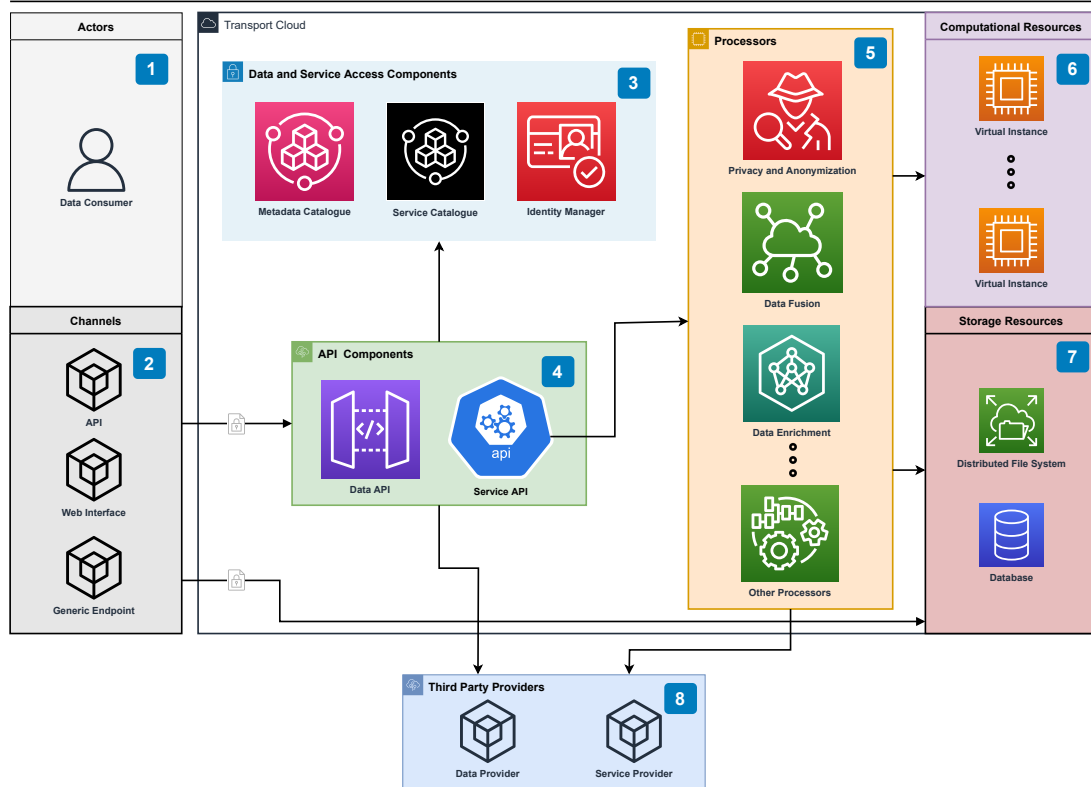


Figure 2: The MobidataLab transport cloud platform architecture – detailed view.

and producers interface and interact with the transport cloud platform, and elaborate on the functionalities that key components provide to sustain the underlying information flow. To this end, we introduce in Figure 2 a more detailed overview of the platform architecture.

On the one side of the transport cloud we have the *Data consumers* (box 1 in Figure 2), which can interact with the platform through the several *transport cloud channels* (box 2 in Figure 2, *Data and Service Access* box in Figure 1). Said channels are (1) *API endpoints* (mainly dedicated to REST API services), (2) *web interface endpoints*, i.e., endpoints dedicated to services which need interaction with the end user (for example scenarios involving data analysis and visualisation tasks), and (3) *generic endpoints* – for instance, a SPARQL endpoint may enable a data consumer to access some knowledge base via Resource Description Framework (RDF) queries.

Whichever the channel of choice, data consumers first interact with the transport cloud by authenticating themselves via the **identity manager** (box 3 Figure 2, *Data and Service Access* box in Figure 1). Once authenticated, data consumers proceed to submit their requests to the transport cloud platform via the API Components (box 4 in Figure 2), which in turn process the requests by querying the **metadata** and **service catalogues** (box 3 Figure 2, *Data and Service Access* box in Figure 1) to find out the appropriate data sources and services (either internal or

external to the transport cloud) the platform must use to satisfy the requests. It is worth noting that the technologies currently under examination for the implementation of said catalogues are being conducted according to the FAIR principles.

On the other end of the transport cloud we have the **third-party providers**, i.e., the *data providers* previously mentioned (box 8 in Figure 2). The third-party providers always represent the information *entry point* of the platform as they are *responsible* for the *provision of information* in the form of datasets and services. The access mechanisms to these information sources are identified and implemented according to the operations to be performed and the types of data and services that need to be accessed. More precisely, information retrieved from third-party providers can either be *imported* within the Transport Cloud, thus requiring appropriate storage solutions (box 7 in Figure 2)¹⁵, or accessed *on the fly* through the use of specific data and service endpoints exposed by the providers. We report that the transport cloud will handle the latter type of access using the *data and service APIs components* (box 4 Figure 2), which may employ appropriate caching mechanisms to improve access efficiency.

Now that it is clear how data consumers and third-party providers (i.e., data producers) interact with the transport cloud, we focus on a key component that enables to *process* information flowing within the platform, i.e., the **processor** (box 5 in Figure 2, which encompasses the *Data Privacy & Anonymization*, *Data/Service Harmonization & Standardization*, *Data Processing*, and *Data Fusion and Enrichment* boxes in Figure 1). In generic terms, we define a processor as a component that models some function that inputs some data and produces an output according to a well-defined logic. This definition then allows to instantiate the notion of processor in many different ways, thus allowing the transport cloud to provide via multiple processors a potentially unlimited number of operations and services. For instance, in the context of mobility data sharing a processor may be used to perform *semantic enrichment* based on common vocabularies, *geographical enrichment* based on common geometries, *data format translation* when some data format must be reconducted to another one, *data fusion* when multiple datasets must be combined, *data anonymisation* to increase trust in the platform via privacy-preserving techniques, injection of license specification, and in general any data processing task that is relevant to the goals of the project. In general, it is clear how processors add further value to the transport cloud platform, as they give the ability to create novel data and services. Finally, we report that processors will lean on computational resources internal to the transport cloud platform (box 6 Figure 2).

3. Conclusion and future work

This paper introduces the concepts at the basis of the Transport Cloud platform, which is the cloud and data-sharing federation proposed by the MobiDataLab Project. The project kicked-off in February 2021 and aims to promote the transport data sharing culture in Europe. The MobiDataLab Transport Cloud platform is therefore intended to be an open and inter-operable platform that eases the access and integration of distributed and heterogeneous mobility data owned by distinct organizations. The platform will be available to public and private institutions

¹⁵Storage solutions that the platform will include are relational databases, spatial databases, and knowledge graph databases.

that have an interest to either act as mobility data providers or consume mobility data through the access and services provided by the transport cloud platform. The platform has been designed to be compliant with the FAIR principles. In the near future we plan to further advance the definition and the design of the Transport Cloud architecture, which will be then evaluated according to the needs of the MobiDataLab stakeholders.

Acknowledgments

This work has been partially supported by the European Union's Horizon 2020 Research and Innovation program under the projects ACCORDION (Grant agreement ID: 871793) and MOBIDATALAB (Grant agreement ID: 101006879).

References

- [1] A. Braud, G. Fromentoux, B. Radier, O. Le Grand, The road to european digital sovereignty with gaia-x and idsa, *IEEE Network* 35 (2021) 4–5. doi:10.1109/MNET.2021.9387709.
- [2] V. Grossi, B. Rapisarda, F. Giannotti, D. Pedreschi, Data science at sobigdata: the european research infrastructure for social mining and big data analytics, *International Journal of Data Science and Analytics* 6 (2018) 205–216.
- [3] F. P. Schiavo, V. Sassone, L. Nicoletti, A. Margheri, FaaS: Federation-as-a-Service, *arXiv e-prints* (2016) arXiv:1612.03937. arXiv:1612.03937.
- [4] E. Carlini, P. Dazzi, F. Lettich, R. Perego, C. Renso, Cloud and data federation in mobidatalab, in: M. Cafaro, L. Ferrucci, H. Kavalionak, A. Makris (Eds.), *FRAME@HPDC 2021: Proceedings of the 1st Workshop on Flexible Resource and Application Management on the Edge*, Virtual Event, Sweden, 25 June, 2021, ACM, 2021, pp. 39–40. URL: <https://doi.org/10.1145/3452369.3463819>. doi:10.1145/3452369.3463819.
- [5] S. M. for All (SuM4AllTM) initiative, Sustainable mobility: Policy making for data sharing, 2021. URL: <https://www.wbcds.org/Programs/Cities-and-Mobility/Transforming-Mobility/Digitalization-and-Data-in-Urban-Mobility/Policy-to-Enable-Data-Sharing/Resources/Sustainable-mobility-Policy-making-for-data-sharing>.
- [6] R. B. Bohn, C. A. Lee, M. Michel, The NIST cloud federation reference architecture, 2020. NIST Special Publication <https://doi.org/10.6028/NIST.SP.500-332>.
- [7] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. da Silva Santos, P. E. Bourne, et al., The fair guiding principles for scientific data management and stewardship, *Scientific data* 3 (2016) 1–9.