DATAVAULTS APP: PERSONAL DATA PLATFORM IN THE SMART CITIES DOMAIN AND ITS POTENTIAL INTEGRATION INTO IDS ARCHITECTURES (INTERNATIONAL DATA SPACE)

María-José López-Osa, Idoia Murua-Belacortu, Urtza Iturraspe-Barturen, Enrique Areizaga-Sanchez, Valentin Sanchez-Pelaez y Ana-Isabel Torre-Bastida

TECNALIA, Basque Research and Technology Alliance (BRTA). Geldo Auzoa, 700 Edificio - 48160 Derio (Vizcaya). Tfno: +34 902 760000

DOI: https://doi.org/10.6036/10789 | Received: 13/nov/2022 • Reviewing: 14/nov/2022 • Accepted: 12/jan/2023

To cite this article: LÓPEZ-OSA, María-José; BELACORTU, Idoia-Murua; ITURRASPE-BARTUREN, Urtza; AREIZAGA-SANCHEZ, Enrique; SANCHEZ-PELAEZ, Valentin; TORRE-BASTIDA, Ana-Isabel. DATAVAULTS APP: PERSONAL DATA PLATFORM IN THE SMART CITIES DOMAIN AND ITS POTENTIAL INTEGRATION INTO IDS ARCHITECTURES (INTERNATIONAL DATA SPACE). DYNA. May-June 2023, vol. 98, n.3, p. 301-306. DOI: https://doi.org/10.6036/10789

ABSTRACT:

Current data sharing architectures, like International Data Space – IDS, still lack maturity and certain requirements or functionalities that prevent them from being complete solutions when it comes to implementing intelligent, exploitable, secure, and reliable data spaces on which to develop a healthy data economy. One of these main shortcomings is the treatment of personal data and it is that these reference models are highly oriented to guarantee the sovereignty of the data for industrial data, and therefore the access control and exploitation of personal data is out of their scope, forgetting the exceptional privacy and usage conditions that must be considered in this type of data. In this paper ***we present the DataVaults project****, and the platform implemented to allow the discovery of this kind of data and at the same time, control that only those who have been specified by the owner user have access and under the terms that he has previously established. Our approach allows the exchange of information not to be limited to industrial or business domains, where sharing is focused only on impersonal industrial data. The validation will be carried out in cases of use started in Smart Cities for the treatment of data and decision-making of citizens.*

Key Words: *Data sharing, International Data Spaces, personal data privacy, access and usage control.*

RESUMEN:

Las arquitecturas actuales de intercambio de datos, como International Data Space – IDS, carecen de madurez y de ciertos requisitos o funcionalidades que les impiden ser soluciones completas a la hora de implementar espacios de datos inteligentes, explotables, seguros y confiables sobre los cuales desarrollar una economía de datos saludable. Una de estas principales carencias es el tratamiento de los datos personales por estar muy orientados a garantizar la soberanía de los datos para ecosistemas industriales, y por tanto el control de acceso y explotación de los datos personales queda fuera de su alcance, olvidándose de las condiciones excepcionales de privacidad y uso que deben ser consideradas en este tipo de datos. En este artículo **presentamos el proyecto DataVaults**, y cómo en este se ha desarrollado una plataforma, que habilita mecanismos seguros, confiables y de preservación de la privacidad para permitir a las personas tomar posesión y control de sus datos e intercambiarlos según su voluntad. El objetivo ha sido permitir el descubrimiento de datos personales y a la vez, controlar que solo tengan acceso aquellos que hayan sido especificados por el propietario. Nuestro enfoque permite que el intercambio de información no se limite a dominios industriales o comerciales, donde el intercambio se centra solo en datos industriales impersonales. La validación se realizará en casos de uso del dominio de Ciudades Inteligentes.

Palabras Clave: Compartición de datos, International Data Spaces, privacidad de datos personales, control de acceso y uso.

1. INTRODUCTION

Organizations are becoming increasingly interested in the notion of "data as an asset" as they face increasing pressure to report a "single version of the truth" [1] and appreciate the opportunity to build business around that data.

In this context, the concept of data governance [2], which has existed for decades in the field of ICT [3], is framed, but due to the emergence of new paradigms such as IoT, *Big Data* or Artificial Intelligence, a redefinition of the concept of governance is necessary, including concepts such as data sharing and data sovereignty. This redefinition lies in new areas, and its implications with business models. Alhasam et al. [4] make a comprehensive analysis of the increasing popularity of this area of research, mentioning that they have found more than 110 activities related to data governance.

ISSN: 0012-7361 eISSN: 1989-1490 / DYNA. May-June 2023, vol. 98, n.3, p. 301-306. DOI: https://doi.org/10.6036/10789

The three objectives that can be associated with proper data governance are as follows, according to [5]:

- Ensure that the data meets the needs of the business.
- Protecting and managing data as a valuable business asset
- Reduce data management costs.

More information in Annex 1: Data Governance.

This paper presents a platform (app) for the control and exploitation of personal data composed of Cloud app and Personal app and focused on the Smart Cities domain. We also introduce the potential of deriving the concepts of this platform to an IDS (International Data Space) architecture. All this work has been carried out in the framework of the DataVaults project (see section 3).

The article is organized as follows: an introduction in which we present the problem, the methodology carried out to solve it and the main contributions of our work. Then, in section 2, we address the data sharing problem by presenting different technological approaches available in the current literature, as well as some references of works with personal data. In section 3, we explain the DataVaults project[1] describing the Personal app and the Cloud app, as well as the integration of the access policy editor in the IDS architecture. Then, the results of the use cases in which the operation and usability of the personal data platform has been validated are presented. The article ends with a series of conclusions and future steps.

1.1 Research problem and methodology

The most important concept to consider is security: is this data exchange secure? how is it ensured during data sharing that the owner of the data will always be the same and that the integrity of the data will not be breached?

From these questions we can deduce the three crucial concepts of data sharing and with very different meanings such as: data privacy, data integrity (quality) and data sovereignty.

The European Commission has proposed new rules to facilitate data sharing across the European Union and across sectors, with the aim of creating wealth for society, increasing citizens' and businesses' control and trust in relation to their data, and providing an alternative model to the data processing practices of major technology platforms. These measures will lay the foundation for a new European form of data governance that is in line with EU values and principles such as personal data protection (GDPR), consumer protection and competition rules (European Data Strategy, 2020).

Privacy refers to the implementation of policies and procedures to determine who can access the data. For example, specifying a temporary space of use, access rights (read, proprietary), etc.

Integrity refers to the policies and procedures that can be implemented to establish the quality and veracity of the data within a given enterprise and protect it from outside attacks. Policies that help ensure the confidentiality, quality and availability of data.

Finally, data **sovereignty** is the ability of individuals, corporations, or governments to decide with self-determination how, for what purpose and at what price, third parties can use their data. Data sovereignty is an important term for regulatory and data security purposes.

Considering the environment and problems previously presented regarding the sharing of personal data, without the loss of privacy, sovereignty and will over them of the owner user, in this section we state two objectives that represent the foundations of our study and the approach of the need for the personal data management platform that is the leitmotiv of the DataVaults project:

Objective 1 *Centralize all sources of user data and information in a single point of control, to make it easier for users to manage and set their access and usage policies.*

Objective 2 *Enable business generation through personal data, and enable ways to integrate personal data management into existing data sharing architectures, where collaborative models are already in place.*

Based on both objectives, the research methodology has been established as shown in the following figure.

[1] DataVaults - Empowering Secure Data Storage, Sharing and Monetization

ISSN: 0012-7361 eISSN: 1989-1490 / DYNA. May-June 2023, vol. 98, n.3, p. 301-306. DOI: https://doi.org/10.6036/10789

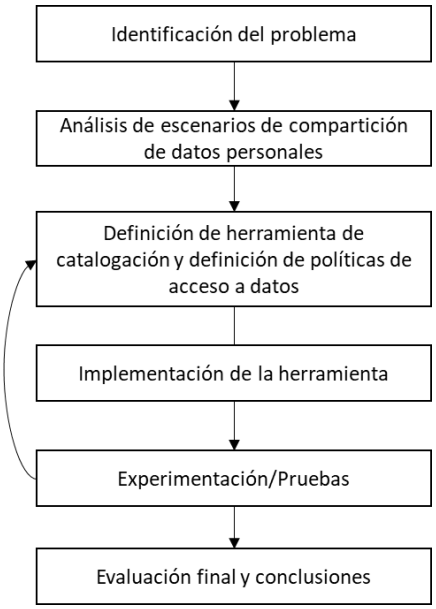

Illustration 1 - Research methodology

1.2 Major contributions

The main contribution achieved in this research project has been the implementation of the personal data platform that provides the user with secure, reliable and privacy preserving mechanisms and allows them to take ownership and control of their data and share it at will. The goal is to enable the personal data value chain to become a more open, multi-faceted and multi-tiered ecosystem governed by smart contracts to safeguard data ownership, privacy and usage and attribute value to all entities that generate value within this chain and, in particular, to the data owners.

To this end, in this article we present two of the contributions currently achieved at the technological level:

- A system for managing data sources related to the user, what we have called a personal data cataloging system, implemented within the ***DataVaults*** project, in the ***Personal App*** and the ***Cloud App*** explained in section 3. In this way, the discovery of this type of data is encouraged and, at the same time, it is possible to control that only those who have been specified by the user and under the terms previously established by the user have access to the data.
- Translator of user policies into models accepted by the main reference architecture for data sharing, IDS, which is a sub-module included as functionality for the ***DataVaults: Personal App*** part of the platform.

This solution opens the door to new possibilities in the exploitation of personal data, and even proposes a starting point for establishing policies to control the use of personal data.

In addition, we would like to highlight as a secondary contribution the analysis of data sharing scenarios carried out, as it allows us to analyze the approaches at a technological level, detecting problems and new challenges that may arise when sharing personal data, in search of promoting new business models and opportunities.

2. RELATED TASKS

The definition of data ownership is far from clear in these times of organized cyber-attacks, microtargeting, targeted advertising and technology giants such as Google, Apple, Facebook and Tencent. Both private users and companies using cloud services and external servers are often unaware that their externally stored data does not belong to them alone.

ISSN: 0012-7361 eISSN: 1989-1490 / DYNA. May-June 2023, vol. 98, n.3, p. 301-306. DOI: https://doi.org/10.6036/10789

In a context of data sharing, such as the new data spaces enacted by the European community[2] , ensuring the governance of personal data is of vital importance, in order to increase the control and trust of citizens and companies in relation to their data, and to offer business models that increase the benefit to society, citizens and companies.

The great avalanche of data and its growing importance is driving a new wave of large-scale data-rich intelligent environments. These ecosystems present new challenges and opportunities in the design of the architecture needed to implement them, as raised in the work presented by Curry et al [6].

In the aforementioned article [7] a very interesting taxonomy is presented when classifying this type of systems - ecosystems, as can be seen in the illustration below, and which is based on the conjunction of previous definitions by Koening [8] and Maier [9]:

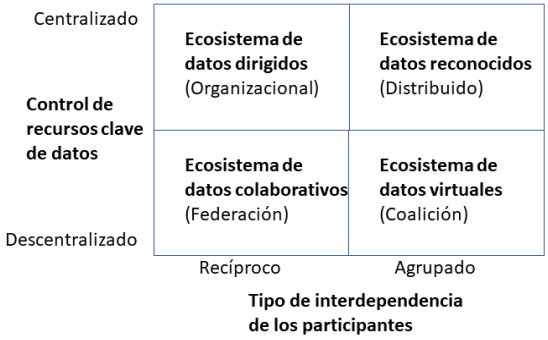

Illustration 2 - Data ecosystems

More information in Annex 2: Data Ecosystems.

In this section we therefore analyze the technological approaches and architectures that enable sharing. To learn more about the constraints and solutions that exist in the control of access to personal data and the current state of the art in terms of personal data platforms in the literature, *see Annexes 3: International Data Space and 4: Controlling Access to Personal Data*.

2.1 Sharing through secure access and usage policies - IDS Architecture

One of the main problems is to know what happens to the data once they leave the repository or platform of the owner provider, how to control, monitor or trace their use to facilitate subsequent profitability. Until now, the concern has been to protect access to the data (as described, for example, in articles [10] and [11]), but its use by potential consumers has not been controlled and, therefore, this is the greatest current challenge, especially at the technological level.

There are different attempts in the literature to define reference architectures around the concept of data sharing in distributed environments. The approaches vary significantly according to their characteristics such as: security conditions, purpose of sharing, or technological approach, and above all according to the application domain they address. The following are the most important ones, trying to cover different domains

- In Industry 4.0, there are several initiatives at the international level where attempts are being made to organize the paradigm into a reference architecture that is easily understandable and widely adopted. This is the case of the International Data Spaces (IDS)3 architecture promoted by the German-based **IDS Association (IDSA)**, the **Reference Architecture Model Industry 4.0 (RAMI4.0)** developed by the German Electrical Manufacturers Association.
- In the healthcare domain, data privacy is the focus of most studies. An example is given in the paper [12], where a system is presented that addresses the problem of sharing medical data between medical repositories in an untrusted environment.
- Referring to the new ***Smart Cities*** paradigm, the article [13] proposes a trust model for data sharing in smart cities.

[2] https://ec.europa.eu/commission/presscorner/detail/es/ip_20_2102
[3] https://internationaldataspaces.org/

ISSN: 0012-7361 eISSN: 1989-1490 / DYNA. May-June 2023, vol. 98, n.3, p. 301-306. DOI: https://doi.org/10.6036/10789

- Also, in the context of Smart Cities, and within the REPLICATE project[4] the objective is to increase the quality of life of the population of a number of cities, taking into account the impact that technologies have when creating services that these cities can offer to their citizens. And another example of the benefits of sharing personal data can be seen in the SmartEnCity project[5] , which aims to combat climate change by making use of a city's energy consumption data, data that is shared by enhancing the intelligent use of technology when sharing data.

3. DATAVAULTS: SECURE PLATFORM FOR THE EXCHANGE OF PERSONAL DATA

This section describes the overview of the platform proposed in the project, and details the two modules built to support the use of personal data in sharing ecosystems: the individual's personal data source cataloging system and the editor/translator of access policies to the semantics of the IDS reference architecture.

Tecnalia participates in the DataVaults project6 , funded by the European Union's Horizon 2020 research and innovation program under agreement no. 871755. The full name is "Persistent Personal Data Vaults Empowering a Secure and Privacy Preserving Data Storage, Analysis, Sharing and Monetisation Platform".

The main objective of this project is to offer a platform that puts personal data at its core, which can come from various sources (wearables, web APIs, smart home sensors, personal data records, etc.), and defines secure, reliable and privacy-preserving mechanisms that allow individuals to take ownership and control of their data and share it at will, through flexible data sharing and fair compensation schemes with other entities (companies or not).

DataVaults seeks to address the issues that make citizens skeptical about sharing their personal data by providing solutions related to privacy, ethics and intellectual property rights. The solution developed in DataVaults allows individuals to de facto own their data and share it according to their rules, through flexible data sharing schemes and with the possibility of receiving compensation.

To this end, DataVaults provides a framework in which to perform end-to-end management of personal data from various sources such as IoT sensors, wearable devices, social networks, health data, demographic profiles etc....and made available to third parties from a single point, all in a secure and reliable manner.

Data owners retain control of the data to decide with whom to share it and under what conditions, including the compensation they wish to receive for making it available to third parties. The role played by the companies or organizations that want access to this data is that of data requesters, or DataSeekers, who are asked to comply with a series of conditions in order to be able to use it. This system generates trust and allows collaboration and the knowledge that no unauthorized use will be made of the data.

The architecture that supports the data sharing system between data owners and organizations or companies consists of two distinct parts:

- ***Personal Platform:*** the application that data owners manage as a configurator of conditions under which to share data. The different parts of this configurator allow each data owner:
  - Add a data source to your space
  - Manage the different sources of added data
  - Select a data set to share
  - Anonymize this data or part of it
  - Select the data access control policy to apply or create a new one.
  - Specify a price
  - Establish a license
  - Encrypt data
  - Allowing them to be part of a "person" entity
  - Using TPM (Trusted Platform Module) as authentication method

[4] https://replicate-project.eu/
[5] https://smartencity.eu/
[6] https://www.datavaults.eu/about/

ISSN: 0012-7361 eISSN: 1989-1490 / DYNA. May-June 2023, vol. 98, n.3, p. 301-306. DOI: https://doi.org/10.6036/10789

- Obtain an estimate of exposure risk
- View and manage transactions

- ***Cloud Platform***: the entry point to the platform in which data seekers select the characteristics they want the data they are looking for to meet, filling in their profile with relevant information for the system to provide the data they are looking for and that are in accordance with the conditions offered by the "seekers".

The image below provides a conceptual view of both parts and their sub-modules:

Illustration 3 - DataVaults Global Architecture

We have highlighted in blue the module on which the cataloging system is based within the Personal App and in red the main parts in which the editor/translator of the access policies to the semantics of the IDS architecture has been implemented, this is explained in detail within the subsection "Integration of the DataVaults policy editor with IDS architecture."

More information in Annex 5: Detailed description of the DataVaults app.

INTEGRATION OF DATAVAULTS POLICY EDITOR WITH IDS ARCHITECTURE

The access policy system mentioned in the previous point is not associated with a standard format that can be understood by other systems. The platform developed in DataVaults is considered a prototype to demonstrate the benefit of sharing personal data from the perspective of proprietary control of the same, and from that perspective, interoperability with other types of systems has not been prioritized. However, being a system conceptually transferable to other protocols and formats, an additional functionality has been developed, not included in the initial requirements of the solution and therefore not part of the use cases. This functionality allows transforming the policy created from the Policy Editor in the Sharing Configurator, to a format more similar to the model developed within IDSA, an association whose main objective is the creation of a global standard for the management of data spaces in an international way. *This standard is explained in Annex 3: International Data Space.*

This translation is an approximation to the IDS format, based on the ODRL standard7 . The process by which this transformation could be obtained is facilitated, but in order to perform it in detail, the implementation, and the information to be transformed would have to be improved. This is a demonstration of the possibilities that this type of policy-based control systems can offer.

[7] https://www.w3.org/TR/odrl-model/

ISSN: 0012-7361 eISSN: 1989-1490 / DYNA. May-June 2023, vol. 98, n.3, p. 301-306. DOI: https://doi.org/10.6036/10789

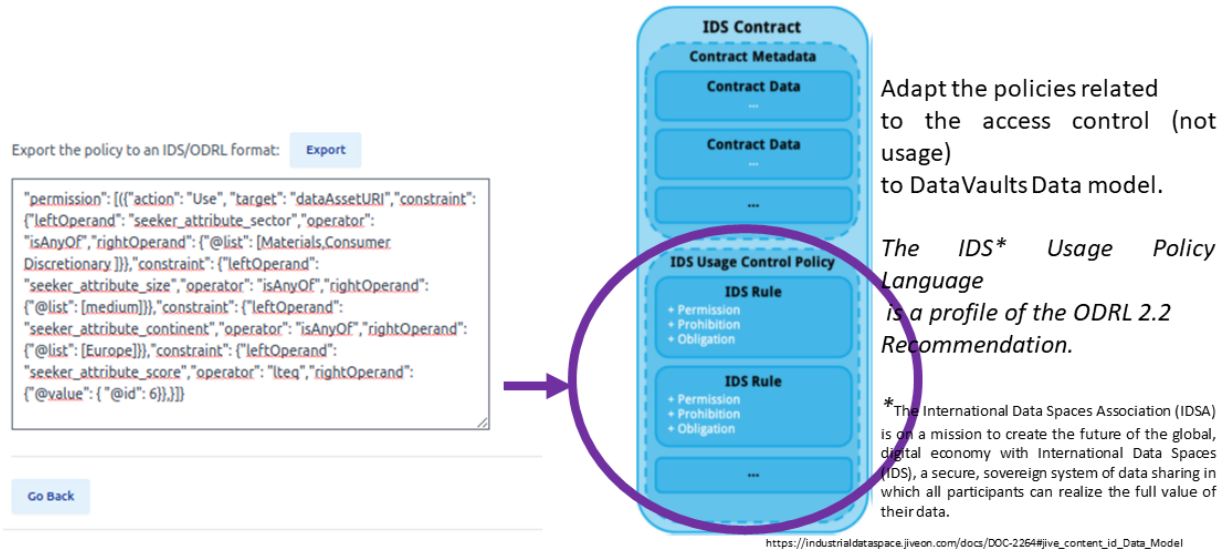

Illustration 4 - Transforming policies to IDS format

RESULTS OF THE USE CASES

The use of personal data in smart cities is conditioned by the availability of these data to build services that allow citizens to benefit from the services offered by these cities, based on the data available to them. There is still reluctance to share data and that is why many initiatives are currently working on this aspect, in order to offer greater confidence and peace of mind to citizens when making their data available to cities.

As an outreach outcome of the DataVaults project, the consortium has published a book [14] that addresses the problems a Smart City faces in making use of currently available data and how this can be improved through the use of emerging technology that allows a citizen to share their personal data, adding value.

The following table shows a summary of indicators by use case, providing an overview of the real usefulness of the tool.

Table 1 - Results

Use Case	Utility raised	Perceived result	Indicator / Value
Scenario 1-Energy trader	Provide permissions to the marketing company on users' personal data and allow them to monitor the use of their data.	Users can track the use of their data and even request and receive financial compensation.	No. of users who use the Personal App to share their data (10-20) No. of transactions generated around the data economy (>100 transaction with a value of less than 30,000 euros)
Scenario 2 - Italian City	Users will be able to provide their data for the improvement of urban mobility services, and each month they will receive a discount on various museum and hall tickets managed by the city council.	Users will receive better citizen services and leisure benefits from the city, without losing control over their data.	Number of users who use the Personal App to share their data (10 -100) Number of shared data (200 shared data sets)
Scenario 3 - Greek city	Users will be able to provide their data to analyze how to encourage tourism to the local market of Piraeus, and each	Users will receive increased participation in local market activities and will also receive financial	Number of users who use the Personal App to share their data (<20)

ISSN: 0012-7361 eISSN: 1989-1490 / DYNA. May-June 2023, vol. 98, n.3, p. 301-306. DOI: https://doi.org/10.6036/10789

	month they will be compensated with a discount at different locations in the market.	compensation in the form of discounts.	Increase in the number of customers (10%)

Based on the results, we can say that the acceptance of the Personal App has been good, with an average of 20 users, and the general perception of the individuals can be described in two dimensions: regarding the use of their data, the feeling has been of control and security, and regarding the perceived rewards, the general opinion is of satisfaction due to being able to measure and manage the rewards obtained by this sharing.

For a more detailed description of the scenarios, see Annex 6: Description of usage scenarios.

4. CONCLUSIONS AND FUTURE WORK

A key aspect of a good data management and data governance strategy is managing access control. Access control is the ability to control at all times who can access the personal data stored in your systems. But there are many types of data, in business and industry organizations are becoming aware of this need because they assumed that their data is an asset and provides additional value. But for example, in the domain of a city, the data produced by citizens are often not properly protected nor are they aware of the value they can achieve. It is said today that all the personal data of an individual can reach a value close to 1,000 euros, and yet most citizens do not make use of them or protect them properly. And one of the main reasons is because technologically no one provides them with the means and tools to do so. In this article we have tried to highlight the most important concepts related to sharing control systems when we talk about personal data, especially in a context such as smart cities for the reasons we have already explained. For this we have explained our main contribution which is the implementation of an access control system in the framework of the DataVaults project, especially the original and interesting point is the policy editor, which focuses on the characteristics of personal data and also provides an alignment or translation to the model understood by the popular IDS data sharing architecture.

In the future and as next steps we have two lines of work in perspective, a greater integration of personal data in the IDS architecture, deepening our work on access policies for sharing compatible with the model and flow of this architecture, and implementing it in the system presented here. And as a second line we seek greater validation by citizens, providing them with applications and visual tools with which they can autonomously validate their permissions and how they want to manage them, being able to introduce changes according to their needs.

REFERENCES

[1] Khatri, V., & Brown, C. V. (2010). Designing data governance. Communications of the ACM, 53(1), 148-152.

[2]Pisani, E., & AbouZahr, C. (2010). Sharing health data: good intentions are not enough. Bulletin of the World Health Organization, 88, 462-466.

[3]McSweeney, A. (2010). Data Governance: Keystone of Information Management Initiatives. Dostupno na: https://www. hashdoc. com/documents/998/data-governance-keystone-ofinformation-management-initiatives. [30.05. 2017].

[4]Alhassan, I., Sammon, D., & Daly, M. (2018). Data governance activities: A comparison between scientific and practice-oriented literature. Journal of Enterprise Information Management.

[5]Panian, Z. (2010). Some practical experiences in data governance. World Academy of Science, Engineering and Technology, 62(1), 939-946.

[6]Curry, E., & Sheth, A. (2018). Next-generation smart environments: from system of systems to data ecosystems. IEEE Intelligent Systems, 33(3), 69-76.

[7]G. Koenig, "Business Ecosystems Revisited," Management, vol. 15, 2012, pp. 208–224.

[8]M. W. Maier, "Architecting Principles for Systems-of-Systems," Systems Engineering,

Wiley, 1998.

[9]Jarke, M., Otto, B., & Ram, S. (2019). Data Sovereignty and Data Space Ecosystems.

[10]Suhendra, V. (2011, December). A survey on access control deployment. In International conference on security technology (pp. 11-20). Springer, Berlin, Heidelberg.

[11]Tourani, R., Misra, S., Mick, T., & Panwar, G. (2017). Security, privacy, and access control in information-centric networking: A survey. IEEE communications surveys & tutorials, 20(1), 566-600.

[12]Xia, Q. I., Sifah, E. B., Asamoah, K. O., Gao, J., Du, X., & Guizani, M. (2017). MeDShare: Trust-less medical data sharing among cloud service providers via Blockchain. IEEE Access, 5, 14757-14767.

[13]Cao, Q. H., Khan, I., Farahbakhsh, R., Madhusudan, G., Lee, G. M., & Crespi, N. (2016, May). A trust model for data sharing in smart cities. In 2016 IEEE International Conference on Communications (ICC) (pp. 1-7). IEEE.

ISSN: 0012-7361 eISSN: 1989-1490 / DYNA. May-June 2023, vol. 98, n.3, p. 301-306. DOI: https://doi.org/10.6036/10789

[14]Tophan, S., Boscolo, P. & Mulquin, M. (2022). Personal Data-Smart Cities: How cities can utilise their citizen's personal data to help them become climate neutral. River publishers, Denmark.

ACKNOWLEDGMENTS

The work described in this article is partially funded by the DataVaults project, co-funded by the Horizon2020 program of the European Union (H2020-ICT-2019-2) under Grant Agreement No. 871755.

ISSN: 0012-7361 eISSN: 1989-1490 / DYNA. May-June 2023, vol. 98, n.3, p. 301-306. DOI: https://doi.org/10.6036/10789