A Universal Data Model for Data Sharing Under the European Data Strategy

Malte Hansen[1(✉)], Nils Gruschka[1], and Meiko Jensen[2]

[1] Department of Informatics, University of Oslo, Oslo, Norway
{maltehan,nilsgrus}@ifi.uio.no
[2] Karlstad University, Karlstad, Sweden
meiko.jensen@kau.se

Abstract. The current European data strategy foresees a novel ecosystem of data sharing and data trading among public and private sector organizations in the EU member states. The focus is on enabling and fostering data sharing among the stakeholders while maintaining compliance with existing EU and national data protection legislation, such as the European General Data Protection Regulation (GDPR).

However, managing data sharing in such a compliant manner requires additional metadata to be exchanged amongst the actors in this ecosystem. Therefore, this paper proposes a novel data model for managing data sharing activities. This model takes current and planned regulations (e.g., the Data Governance Act) and the resulting data ecosystem architectures (e.g. data intermediaries) into account and is applicable to different actions that are necessary for compliant data exchange, like data subject rights requests or intellectual property enforcement.

Keywords: data sharing · data model · European data strategy · Data Governance Act · GDPR

1 Introduction

The well-known quote "Data is the new gold!" is more than a decade old, but illustrates nowadays more than ever the value of data and data exchange for commercial enterprises as well as for the public sector. However, when not handled in compliance with legislation, especially data protection regulations, data sharing can violate the rights and freedom of individuals, which led to a second expression: "Data is the new uranium!". On the other hand, the fear of high fines for violation of data protection regulations can lead to a complete blockade of any data exchange. In light of this dilemma, the goal of the European data strategy is to enable data sharing while complying with EU and national legislation, such as the European General Data Protection Regulation (GDPR) [1].

Consequently, it becomes necessary for data controllers, data processors, data intermediaries, and other actors in this ecosystem to address these legal requirements in their data sharing agreements and platforms. More specifically, they

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024
K. Rannenberg et al. (Eds.): APF 2023, LNCS 13888, pp. 3–19, 2024.
https://doi.org/10.1007/978-3-031-61089-9_1

need to provide means for managing federated data sharing scenarios—involving multiple actors—in such a way that legal compliance is maintained, especially concerning GDPR obligations like purpose binding and data subject rights (cf. Art. 5, 15ff. GDPR [1]). This requires standardized interactions among data sharing actors in a decentralized, federated manner.

This paper proposes a novel data model for managing data sharing activities in such a pan-European data ecosystem. Based on the roles defined in the respective legislation, we analyze the needs for interaction and metadata exchange, and we derive a universal data model that can generically be utilized for managing data sharing interactions in a compositional, decentrally organized, legally compliant manner. The model can serve multiple purposes, such as data subject rights enforcement, data breach notification, intellectual property rights enforcement, and many more.

The paper is organized as follows. In the next section, we provide the legal and policy background for this work, based on the novel European legislation and data strategy. Section 3 then summarizes the state of the art in research on these topics. In Sect. 4, we define the proposed universal data model, providing its requirements and core technical aspects. The subsequent section illustrates the use of this data model in practice, based on an example scenario from the logistics domain. In Sect. 6, we discuss different areas of application of the data model, and the paper concludes with a discussion of relevant properties and open issues in Sects. 7 and 8, respectively.

2 Background

In this section, we briefly present the European data strategy [2] and its relevant regulations, which lead to a demand for an optimized data model for data sharing in the European market.

2.1 The European Data Strategy

Data has been identified as an essential resource by the European Commission, able to foster economic growth, research, and societal progress if used appropriately. This led to the commission's proposal of a common European data strategy [2]. The EU data strategy aims to create a single market for data, connecting public and private actors across multiple sectors. In this market, data is intended to flow freely for the benefit of all involved parties, facilitating access and re-use of data, hence optimizing data use. A set of European regulations were proposed to build the framework for this data ecosystem, fostering the elaboration of practical and clear rules for the access and use of data, while guaranteeing compliance with existing privacy and data protection legislation like the GDPR. These combined efforts shall strengthen the position of the EU as an attractive competitor in the global data economy, offering a fair, secure, and dynamic environment for data flows. To achieve these proclaimed goals, the

European Commission has proposed several new European regulations, including the Data Governance Act [3] (DGA), Data Act [4], Digital Services Act [5], and Digital Markets Act [6].

2.2 The Data Governance Act and Data Intermediaries

The European Data Governance Act (DGA) [3], which will come into force in September 2023, acts as a central cornerstone in the European data strategy. It aims to increase data availability and facilitate the reuse of data across the European market. To achieve this, common European data spaces are to be built, including actors from both public and private entities, that allow for sharing and reuse of data across multiple sectors. A key role in this new environment will fall upon the so-called Data Intermediaries (DI). DIs are designed to act as a mediator between different Data Controllers (DC), either storing or requesting data sets. Defined as a benevolent actor in the DGA, a DI should not have a commercial interest in using the data it obtains by itself. Rather, a DI's main duty is to fulfill requests for data issued by other actors, aggregating applicable data sets from their data sources, and distributing them to the requesting organizations in a secure and privacy-preserving way. In these data sharing scenarios, they will aid in enforcing data subject rights, as well as compliance with other relevant obligations of the GDPR and applicable European laws and regulations.

An open issue leading up to the official enforcement of the DGA is how the interaction between the DIs and the other actors, most importantly DCs and Data Subjects (DS), will be realized in detail. The pooling and distribution of data sets of personal data across different European countries, actors, and sectors, while also offering easy enforcement of data subject rights and transparency, requires a common baseline for all involved parties, such as a standardized collaboration protocol for mutual interaction. This leads to the demand for a universal data model for data sharing under the European data strategy. This model can then later be leveraged to design and optimize processes and help the DGA fulfill its role and obligations assigned in the DGA and reach the goals of the European data strategy.

2.3 Digital Markets Act and Market Fairness

A key element in the European Data Strategy is the empowerment of innovative small-size market actors. The data strategy addresses multiple current stumbling stones for data sharing, such as the tendency of device manufacturers and market platform operators to hoard data obtained from their customers for themselves, rather than making them accessible to other market actors. Here, the Data Act and the Digital Markets Act define several conditions under which the collection and sharing of data may even become mandatory by law.

As one example, the Digital Markets Act obliges huge digital market platform operators, the so-called *gatekeepers* (like social media giants, cloud service providers, and global scale service providers), to make the data they collect on their digital platforms accessible to other, less powerful market actors in a fair

an appropriate way (as long as several specific conditions are met). However, at the very same time, these gatekeepers remain the primary contact point for DSs trying to exercise their data subject rights. Hence, it becomes inevitable for these gatekeepers to keep track of all data sharing transactions they perform. Thus, gatekeepers require a standardized data sharing communication protocol with their data sharing partners, in order to implement this legal obligation. Here, the proposed universal data model for data sharing may play a key role.

3 Related Work

P3P [7] is a legacy framework for privacy preferences that has since seen many adaptions (e.g. [8–10]). Ulbricht and Pallas [11] developed a language to address consent requirements that came with the GDPR. Becher and Gerl [12] introduce a privacy preference language with the aim to provide privacy language compatibility via privacy interfaces. A prominent approach for applying privacy policies are sticky policies [13]. Iyilade and Vassileva introduce a policy language for data sharing [14]. An overview of additional privacy-related policy languages can be seen in [15].

For modeling data sharing agreements (DSA) specifically, some previous approaches exist. Swarup et al. present a language for DSAs that declares obligations and constraints in the form of "distributed temporal logic" that defines data flows and data storage [16,17]. CNL4DSA [18] is another example of a DSA language that focuses on increased user-friendliness. CNL4DSA later acts as a basis for a DSA lifecycle management framework [19]. However, all these contributions focus on the execution and management of privacy policies and DSAs, rather than their role in the data sharing process, and do not yet consider the obligations that came with the GDPR.

Additionally, there is research on the application of the GDPR in the context of various processes that impact data sharing. privacyTracker [20] introduces a framework that allows for a reliable construction of a data trail. Insynd [21] unifies privacy-preserving transparency and logging. Our previous work [22] discusses how to annotate data in face of the GDPR. LPL [23] is a privacy language that defines expressions for privacy properties in the context of the GDPR. TILT [24] is a language for transparency information in a machine-readable format based on requirements from the GDPR. In previous work [25] we presented a data model for the execution of right-of-access requests specifically. In the data model proposed below, we will combine these aspects and put them into the context of data sharing of the European data strategy.

Further, ENISA has released a report on how to engineer personal data sharing [26], specifying the interactions between DI and DS or DC, respectively, which serves as a guideline for the model.

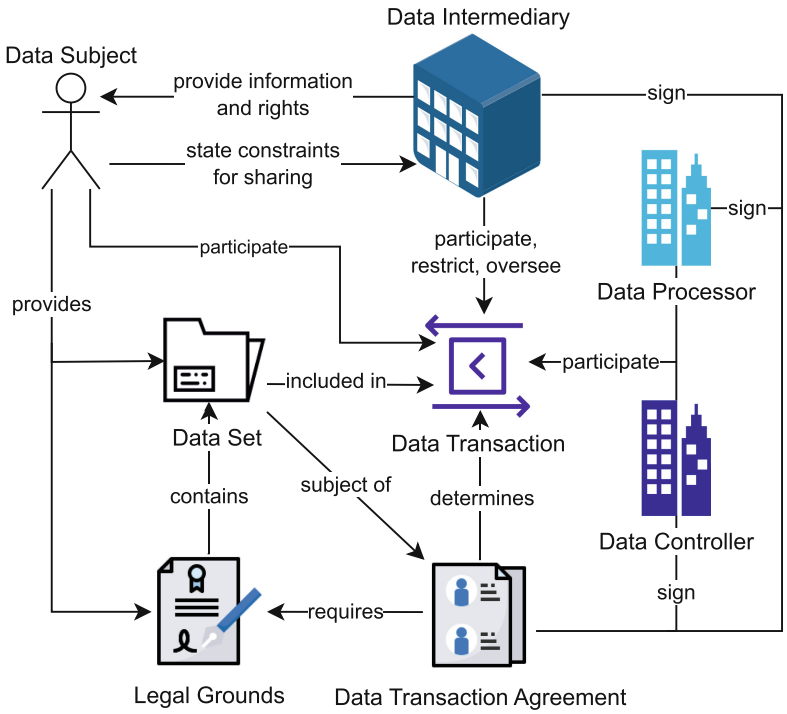

Fig. 1. Overview of interactions between relevant actors and resources in Data Model

4 Data Model

The goal of the proposed data model is to provide a format for data sharing in compliance with the EU data strategy and EU data regulations, potentially serving as a standardization. A participant in a data sharing scenario is obliged to prove compliance with these regulations, fulfill requests for DS rights, and report privacy breach notifications. Currently, the information that an actor can provide is restricted to their own involvement, which makes it also very difficult to audit these processes from the outside. Especially for the intended role of DIs and gatekeepers, this is very impractical. This data model aims to offset these deficiencies by providing the required metadata. While the model will focus on personal data transactions and the role of the DI, it will also be applicable to non-personal data use cases, such as intellectual property rights. As depicted in Fig. 1 the model will revolve around data transactions. A data transaction is defined as an exchange of a data set between two actors that is regulated by a data transactions agreement (DTA) We define a DTA as a generalization for agreements for the exchange or collection of data between two parties, such as data sharing or data-use agreements. An actor participating in a data transaction can be a DS, DC, data processor, or DI. The DI can additionally serve as a controlling instance for these transactions and aid the DS in the execution of

their rights. As the scope of all relevant actors and applications in the EU data strategy is quite wide, we have to focus on the most relevant factors in this work. Hence, the model will not address the conditions, obligations, and execution of the underlying DTA. Rather, we will introduce a wrapper for data transfers that provides the necessary information to respect compliance with EU regulations, promotes transparency, and facilitates the execution of DS rights.

4.1 Requirements

This leads us to the requirements for the data model. First, it needs to depict a traceable data flow, allowing us to see the sources and destinations of data. The data sharing scenario demands a class for the data transaction and the underlying agreement, describing the conditions for the data transaction. For the transaction, agreement, and data trail the involved actors and their respective roles, as senders or recipients, as well as the data sets that are being exchanged, have to be included. Additionally, the model must be able to display several constraints given by the GDPR. Collection and processing of personal data under the GDPR require a purpose [1, Art. 5(2)], as well as a legal ground [1, Art. 6], e.g. consent. Further, special categories of personal data, e.g. health data, underlie stricter rules for processing [1, Art. 9].

4.2 Classes and Attributes

An overview of all classes and attributes can be seen in Fig. 2. One attribute all classes share is the identifier. Generally, the ids depend on a known context. This means, that two actors who engage in a data transaction know or negotiate the shared local id, e.g. for the DTA. For the legal ground, purpose, and category classes standardization across the EU data market is recommendable, which will be discussed in Sect. 7. The attributes in all the classes are limited to the strictly necessary ones. If the context of a scenario requires additional attributes they can be added to the existing model on demand.

The data set class describes the properties of the underlying data that it is wrapping. The exact method of wrapping is not covered in this work, as it is not essential in the data sharing process. Different solutions are possible, e.g. as a reference or embedded in the document. The best option depends on the scenario. The data set class contains references to the categories, legal grounds, and DSs of the underlying data. A data set can contain more than one of each of these fields, e.g. aggregation of data can lead to multiple DSs or categories of data in a single data set instance. Also, the data subject field can be empty to consider the scenario of non-personal data. The creation and retention date fields exist to serve as control instances for compliance with the time frame of legal grounds and DTAs. Lastly, the source field refers to the data transformation or transaction that was used to obtain the data set. Over this reference, the sources and destinations of a data set can be depicted, which is required for a Right of Access request [1, Art. 15(1)(c)+(g)].

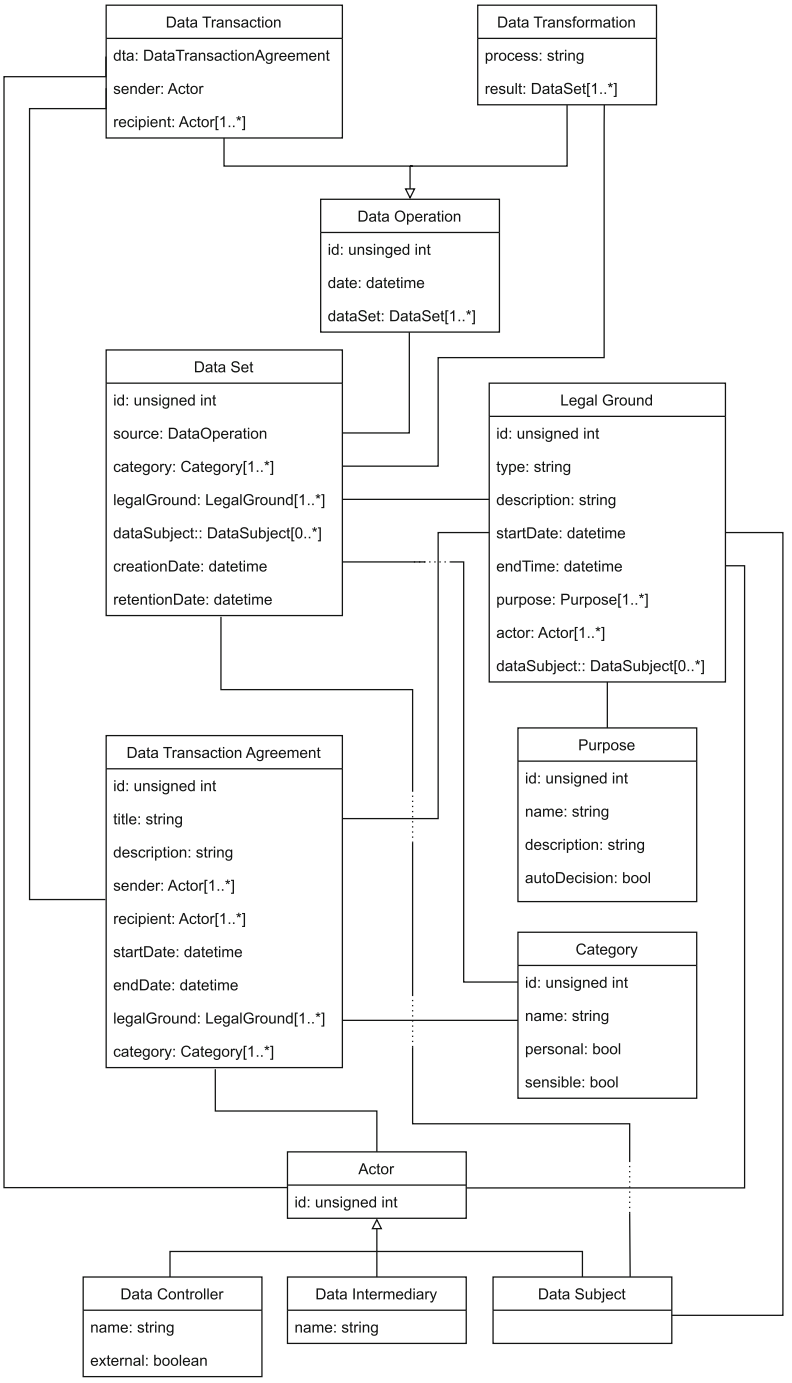

Fig. 2. Classes and attributes in the data sharing data model

The data operation class is a generalization of data transactions and data transformations. Each operation contains a timestamp and a list of the data sets included in the operation. A data transformation is any processing of a data set, where you take a data set and change its properties to generate a new data set. While a data transformation has to be considered as a possible source for a data set, it is not relevant to the data sharing scenario. Therefore, the field describing the process is kept general as a string. Instances of the data transaction class are the realization of any data sharing process. They include the underlying DTA, as well as the involved actors. A transformation always contains one actor that sends the transaction and one or more actors who receive it. This can be leveraged to create a data trail. An actor is the generalization of either a DS, DC, or DI. A DC instance can describe a data processor as well. If a distinction is required at a later point, the model can be expanded. The external attribute of the DC can be used to identify data transactions that leave the EU.

DTAs include a title and a description. The sender and recipient attributes define which actors are allowed to act as the sending or receiving party in any data transaction under this DTA respectively. The start and end dates define the time frame in which the DTA can be used as a basis for a data transaction. Additionally, it is defined which legal grounds and categories can be leveraged in the transaction. These can then be compared with the corresponding fields in the data sets, which are attached to the transaction, to verify compliance.

The legal ground contains a description and time frame as well. Further, the type can be defined, e.g. consent, contract, or legal obligations. It also covers which actors are allowed to act upon this legal ground and the data of which DSs they are allowed to process in this case. Additionally, a legal ground must contain at least one purpose for the processing of data. The purposes are modeled in their own class, including a field for marking the existence of automated decision-making in it. Lastly, the category class describes the category of data that is contained in a data set. It can be distinguished between non-personal, personal, and sensible personal data.

5 Illustrating Example

In order to give a better overview of the application of the proposed data model, this section describes an example of a common data exchange scenario between DCs and DIs, as is shown in Fig. 3.

Our DC, *City C Bus*, is a private bus company operating several bus lines in City C, State Z. *City C Bus* considers introducing new bus lines inside of City C, giving commuters a direct connection between the central station and the new business parks popping up on the outskirts of City C. To get an overview of potential customers, *City C Bus* wants to analyze mobility data sets for train, bus, and car commuters in the region. For this purpose, they contact *State Z Commute*, a DI responsible for mobility in State Z. *City C Bus* requests the following data sets: Mobility data for individuals for workdays between 06:00 and 18:00, including passenger ID, starting location, destination, vehicle, start

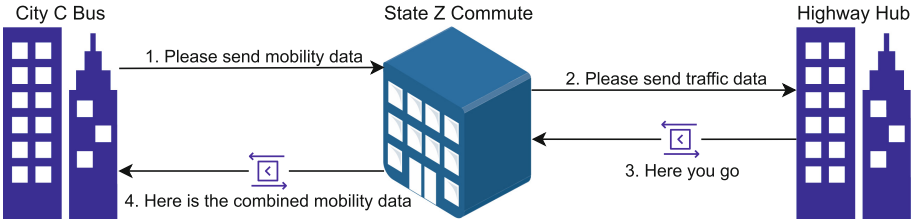

Fig. 3. Overview of the scenario for a data request of City C Bus to the DI State Z Commute

time, end time, and information about ticket subscription for public transport. While the exact content of a data set is not relevant to the data sharing process, it is important to note, that it contains personal data in the form of personal identifiers and location data. *State Z Commute* has access to the public transport data in State Z itself. For the highway traffic data, they depend on *Highway Hub*, an independent DC, as a data source.

In a general flow of a data request to a DI by a DC, depicted in Fig. 4, the DC, in this case, *City C Bus*, sends the request to the DI, *State Z Commute*. The DI then checks if the requested data can be sent from their own data store or if they require additional data from one of their data sources. A data source can be another DI, a DC, or a DS. For our scenario, this results in the following possible sequence of events:

First, the DI *State Z Commute* receives the request from *City C Bus*. *State Z Commute* checks its own data store for the requested data. To get a complete data set, they send a request for the highway traffic data to their data source *Highway Hub*. The response from *Highway Hub* now contains the first data transaction, as seen in Listing 1.1. The format of the data sets in this scenario is JSON. Additionally, ids have been replaced with clear names in most instances for the sake of readability. In the data transaction instance, we can see the timestamp, attached data sets, sender, and recipient of the transaction. Further, the basis for the data transaction, the DTA (see Listing 1.2) is included. It shows us who is allowed to send and receive data sets between the involved actors and defines the properties of the data sets that can be sent, namely the time frame, legal ground, and categories of data. We can now use this information to confirm the legitimacy of the data transaction by comparing the information in the DTA with the values in the data transaction instance and the attached data sets. To assert the correct usage of the DTA, you can compare the date, sender, and recipient fields in the data transaction and DTA instances. For compliance of the sent data sets with the DTAs a comparison between the date, legal ground, and category fields in the involved DTA and data set instances is possible. By doing so we can see that everything is legitimate in this transaction.

After receiving the transaction with the highway traffic data from its data source, *Highway Hub*, the DI, *State Z Commute*, aggregates the data set with the mobility data from their own data store. This can be modeled as a data

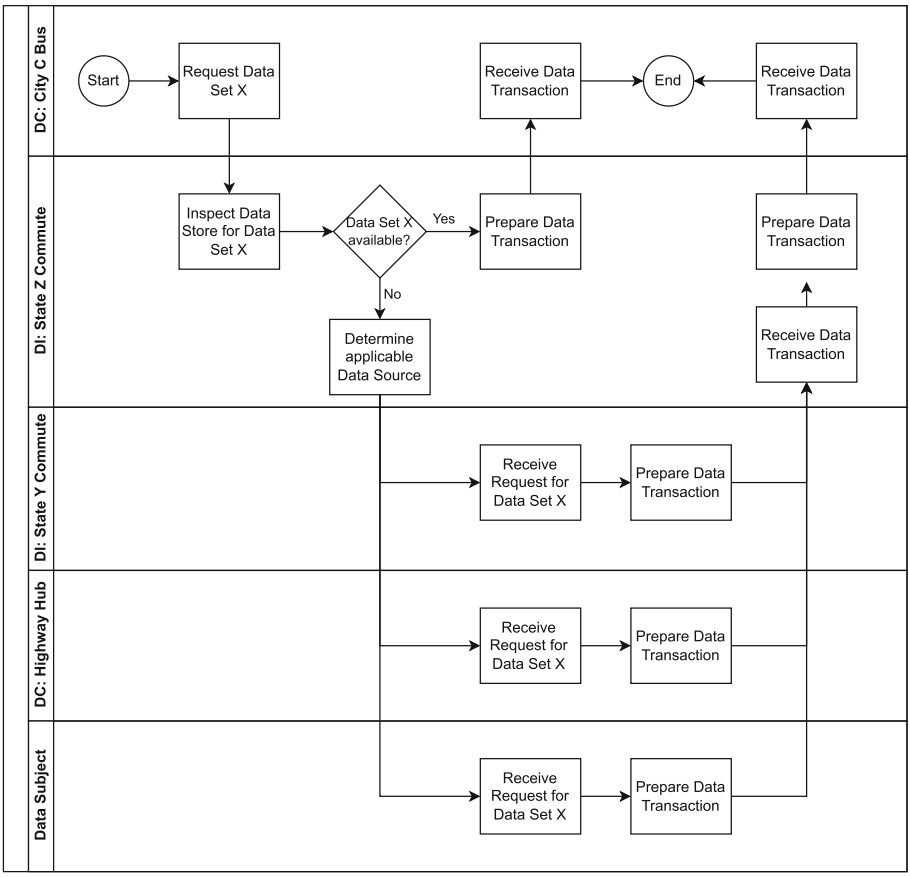

Fig. 4. Simplified flow of events for a data request to Data Intermediary State Z Commute by Data Controller City C Bus

transformation. They then send the complete data to *City C Bus* and complete the request. The resulting data transaction, see Listing 1.3, uses the same DTA, as seen in Listing 1.2, since the agreement covers all three involved actors. However, if we audit the legitimacy of this data transaction now, we can see that *State Z Commute* committed a mistake. The sender and recipients defined in the agreement only allow for a data flow from the DCs, *City C Bus* and *Highway Hub*, to the DI, *State Z Commute*. Therefore, the selection of this DTA for this data transaction was not valid and compliance was breached.

6 Application Areas

As seen in the scenario described above, the model can be applied to the exchange of data between different DCs and DIs that have an agreement to properly define

Listing 1.1. Data Transaction between *Highway Hub* and *State Z Commute*

```
{
  "id": 30091,
  "date": "2023-02-01 13:27:53",
  "dataSet": [15454, 15455],
  "dta": [Z State traffic route utilization agreement],
  "sender": Highway Hub,
  "recipient": [State Z Commute]
}
```

Listing 1.2. Data transaction agreement between the actors in State Z

```
{
  "id": 7,
  "title": "Z State traffic route utilization agreement",
  "description": "Formal agreement on the exchange of
      route utilization data between operators of public
      traffic routes in Z State and Z State Commute for
      the fulfillment of legal obligations",
  "sender": [City C Bus, Highway Hub],
  "recipient": [State Z Commute],
  "startDate": "2019-01-01 00:00:00",
  "endDate": "9999-12-31 23:59:59",
  "legalGround": [Requirement to share route utilization
      with competent authority],
  "category": [location, public transport id]
}
```

their data transactions. Alternatively, the model can also be applied to data transactions with DSs. For example, the collection of personal data from a DS can be realized in the same way as the transaction in the scenario. Here, the DTA could be a terms of service document instead. In this way, the intended roles of the DIs and gatekeepers in the EU data strategy can be fulfilled with this data model.

As shown in the scenario, validating the legitimacy of data transactions can be done automatically, for example with pre-defined queries. This opens the path for preventive and detective measures against compliance breaches as well. Scanning incoming data sets against said queries can prevent potential breaches. Consequently, compliance with European regulations should improve across the whole European data market. Additionally, for DIs in particular this is a very potent tool to facilitate the distribution and re-use of data sets.

Another use case for the proposed data model is the improved ease of implementation for DS rights and transparency. DS rights are rights to enact control about one's personal data granted to European citizens by the GDPR [1, Chapter

Listing 1.3. Data Transaction between *State Z Commute* and *City C Bus*

```
{
  "id": 30158,
  "date": "2023-02-02 21:43:18",
  "dataSet": [15454, 15455],
  "dta": [Z State traffic route utilization agreement],
  "sender": State Z Commute,
  "recipient": [City C Bus]
}
```

3]. Examples include the Right of Access, Right to Erasure, or Right to Data Portability. As an example, a DI could create a data trail for a DS by processing the different sources and destinations of their data transactions, potentially supplementing them by requesting additional information about data transactions for the same data sets from their data sources. This data trail has a lot of possible applications. For example, it can be leveraged to fulfill a request for the deletion of personal data by a DS. In the context of data altruism, this can also be applied to the withdrawal of consent. It can further be used for privacy breach notifications. As the implementation of these DS rights in companies is often lackluster, the concept of Data Subject Rights as a Service (DSRaaS) has been introduced (cf. e.g. [27]). The goal of DSRaaS is to introduce a service provider for DS rights that acts as a bridge between the DSs and DCs, removing obstacles in both the implementation and execution of DS rights. With their goal of aiding in the enforcement of DS rights, DIs are a natural fit for the role of a DSRaaS provider. For the execution of this role, an appropriate data model is required. As the proposed data model fulfills the requirements for this purpose, it builds an important cornerstone in the realization of the DSRaaS architecture.

While the depicted scenario handles a transaction of personal data, it can be applied to non-personal data as well, by leaving the DS field in a data set instance empty or pointing to the rights holder in case of intellectual property rights scenarios instead. A prime example of a use case for non-personal data would be the enforcement of intellectual property rights [28].

7 Discussion and Open Issues

The data model assumes that data sets of non-personal data contain a category, legal ground, and therefore purpose as well. While adding this information to non-personal data is not required by law, it has its use cases, as can be seen in the previously discussed case of intellectual property rights. Further, in cases of mixed personal and non-personal data, or cases where data is incorrectly classified as non-personal data, the addition of this information can serve as a tool to guarantee or correct compliance. To fill out this information default values can be applied. For example, the framework for free flow of non-personal data

[29] serves as a prime example of possible legal grounds. Adding this classification to data would additionally facilitate possible measurements of the success of this framework in the future. This does not imply that the data model must be used for every data set, regardless of conditions.

Another important aspect is how the model manages changes to any of its elements. In the proposed model, changes to an instance of any of its classes are restricted, meaning that each change is a new instance of said class, meaning that e.g. changing the actors that are allowed to act as a sender of data in a DTA, would create a new document of this DTA. This is done to be able to verify under which conditions a data sharing scenario occurred. If, for example, a DTA is adjusted at a later point in time, it is important to still be able to see the version under which the transaction happened to investigate any potential compliance breaches. While this greatly improves the long-term verification of compliance, it creates additional data, which contradicts the principle of data minimization [1, Art. 5(1)(c)]. Therefore, to keep the additional data to a minimum, changes that do not alter the nature of the data sharing process, such as a rephrasing of a description, should not generate a new instance of an element.

Another observation of relevance must be made with respect to the relation of the proposed universal data model to the data minimization principle of the GDPR. Obviously, introducing a detailed tracking mechanism of data sharing interactions as proposed here introduces a large amount of additional data, metadata, to be precise, with respect to ongoing data processing. This may be perceived as being in contrast to the data minimization principle of the GDPR. For instance, Recital 57 states that it should not be mandatory for data controllers to store additional personal information (like the data transactions proposed here) for the sole purpose of complying with the data subject rights obligations. However, it is necessary to understand that this recital does not prohibit the collection of such metadata either. Further, data subject rights and data breach notifications are a mandatory part of the GDPR's legal obligations, hence justifying the storage of metadata of data processing activities in general. When it comes to anonymization, where all links to any data subjects are removed from the data, the resulting anonymous data obviously no longer requires tracking of its processing for implementing GDPR obligations, hence the proposed data transactions are no longer—and should no longer be—created for such anonymized data. For all other cases, where the link to any DS is still explicitly and intentionally part of the personal data, keeping track of its processing entities remains required and mandated by the GDPR. The same holds true for cases where other legal grounds apply.

An important question is how the data model would perform in partial dissemination, as it is not realistic that every DI and DC immediately adopt the proposed model. If we reflect upon the exemplary scenario shown above, we can now imagine that one DC, e.g. Highway Hub, did not adopt the model. This means, that the data the DI, *State Z Commute*, receives from the DC does not contain the information attached to the normally referenced data transaction agreement class. However, as the data transaction agreement between the two

parties does exist anyways, the DI can still adapt this agreement according to the model and add the information on their own. This means that for the DI the use cases of the model still exist. However, the ability to build a data trail would be limited in this scenario. As the data flow can only reliably be traced between actors that have adopted the data model, the data trail would stop after reaching the DC here.

As mentioned in Sect. 4.2, the categories of data, as well as the legal grounds and purposes for data processing, would greatly profit from standardization across the European data market, and possibly further beyond. This would greatly increase the comprehensibility of these elements, especially for non-specialists. Consent, as a legal ground, specifically profits from this, as a common baseline would facilitate consent management, especially in distributed systems. Further, standardized processes for the handling of data with specific properties, e.g. data raised for the purpose of advertising, could be introduced, promising a strong tool for auditing and compliance as well.

The question of how to create identifiers for DSs has to be examined closely as well. While a local id works great in a specific context, another solution has to be found when the problem is applied to a larger scope. If a DS, for example, withdraws its consent for the sharing of its personal data and wants all recipients of this data to delete it, the data might have traveled between multiple contexts. The id must remain traceable throughout these contexts to fulfill this request for deletion. This id then also must be designed in a way that allows for a maximum degree of anonymity and does not reveal any information to a third party that would have otherwise not been able to retract this information.

8 Conclusion

Data exchange and sharing is and will be an important cornerstone of both the European and global data market. Currently, these data exchanges are very hard to track reliably, which makes it difficult to verify compliance and exert DS rights. The data model we have developed aims to resolve this by introducing an approach centered around data transactions and data transaction agreements to attach the relevant metadata in data sharing scenarios. A data transaction references the metadata of data sets that are subject of the data sharing scenario, the actors that act as sender and recipients, and a reference to the underlying data transaction agreement between the involved actors. The data transaction agreement then further defines the conditions of the data transaction.

We demonstrated an exemplary application of this model in an illustrating scenario, featuring a common data sharing scenario under the DGA, a request for data to a DI by a DC. As shown in the example, by comparing the data in the data transaction, data transaction agreement, and the attached data sets, compliance can easily be verified, possibly in an automated process. Additionally, a data trail can be built based on the model that can be leveraged for further use cases, such as data breach notifications, DS right enforcement, and consent management. While the model focuses on the application for personal data, it

can also be used in non-personal data scenarios, e.g. for managing intellectual property rights.

The proposed model can serve a useful role in the development of other open issues, such as the standardization of legal grounds, purposes and categorization of personal data, and the development of privacy-preserving identifiers for DSs. While a complete adoption in the European data market would help, not only in the progression of these issues, but also in improving compliance and DS right enforcement in data sharing scenarios within all of Europe, the model can fulfill most of its use cases in a partial dissemination as well. To achieve these improvements, a standardization of the data sharing process based on this model would therefore be a valuable consideration.

Acknowledgements. The contribution of M. Jensen was partly funded by the Swedish Knowledge Foundation (KK-Stiftelsen) as part of the TRUEdig project.

References

1. European Parliament and Council. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance), 4 May 2016. http://data.europa.eu/eli/reg/2016/679/oj/eng. Accessed 24 Apr 2018
2. European Commission. European data strategy – Making the EU a role model for a society empowered by data (2022). https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy_en
3. Proposal for a Regulation of the European Parliament and of the Council on European data governance (Data Governance Act). COM/2020/767 final
4. Proposal for a Regulation of the European Parliament and of the Council on harmonised rules on fair access to and use of data (Data Act). SEC(2022) 81 final - SWD(2022) 34 final - SWD(2022) 35 final
5. Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act)
6. Regulation (EU) 2022/1925 of the European Parliament and of the Council on contestable and fair markets in the digital sector and amending Directives (EU) 2019/1937 and (EU) 2020/1828 (Digital Markets Act)
7. Cranor, L.F.: P3P: making privacy policies more useful. IEEE Secur. Priv. **1**(6), 50–55 (2003). https://doi.org/10.1109/MSECP.2003.1253568
8. Agrawal, R., et al.: An XPath-based preference language for P3P. In: Proceedings of the 12th International Conference on World Wide Web, pp. 629–639 (2003)
9. Yu, T., Li, N., Antón, A.I.: A formal semantics for P3P. In: Proceedings of the 2004 Workshop on Secure Web Service, pp. 1–8 (2004)
10. Li, N., Yu, T., Anton, A.: A semantics based approach to privacy languages. Comput. Syst. Sci. Eng. **21**(5), 339 (2006)

11. Ulbricht, M.R., Pallas, F.: YaPPL - a lightweight privacy preference language for legally sufficient and automated consent provision in IoT scenarios. In: Garcia-Alfaro, J., Herrera-Joancomartí, J., Livraga, G., Rios, R. (eds.) DPM CBT 2018. LNCS, vol. 11025, pp. 329–344. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00305-0_23
12. Becher, S., Gerl, A.: ConTra preference language: privacy preference unification via privacy interfaces. Sensors **22**(14), 5428 (2022)
13. Pearson, S., Casassa-Mont, M.: Sticky policies: an approach for managing privacy across multiple parties. Computer **44**(9), 60–68 (2011)
14. Iyilade, J., Vassileva, J.: P2U: a privacy policy specification language for secondary data sharing and usage. In: 2014 IEEE Security and Privacy Workshops, pp. 18–22, May 2014. https://doi.org/10.1109/SPW.2014.12
15. Kasem-Madani, S., Meier, M.: Security and privacy policy languages: a survey, categorization and gap identification. arXiv preprint arXiv:1512.00201 (2015)
16. Swarup, V., Seligman, L., Rosenthal, A.: A data sharing agreement framework. In: Bagchi, A., Atluri, V. (eds.) ICISS 2006. LNCS, vol. 4332, pp. 22–36. Springer, Heidelberg (2006). https://doi.org/10.1007/11961635_2
17. Swamp, V., Seligman, L., Rosenthal, A.: Specifying data sharing agreements. In: Seventh IEEE International Workshop on Policies for Distributed Systems and Networks (POLICY 2006), pp. 4-pp. IEEE (2006)
18. Matteucci, I., Petrocchi, M., Sbodio, M.L.: CNL4DSA: a controlled natural language for data sharing agreements. In: Proceedings of the 2010 ACM Symposium on Applied Computing, pp. 616–620. Sierre Switzerland: ACM (2010). ISBN: 978-1-60558-639-7. https://doi.org/10.1145/1774088.1774218. https://dl.acm.org/doi/10.1145/1774088.1774218. Accessed 16 Jan 2023
19. Ruiz, J.F., et al.: A lifecycle for data sharing agreements: how it works out. In: Schiffner, S., et al. (eds.) APF 2016. LNCS, vol. 9857, pp. 3–20. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-44760-5_1
20. Gjermundrød, H., Dionysiou, I., Costa, K.: privacyTracker: a privacy-by-design GDPR-compliant framework with verifiable data traceability controls. In: Casteleyn, S., Dolog, P., Pautasso, C. (eds.) ICWE 2016. LNCS, vol. 9881, pp. 3–15. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46963-8_1
21. Peeters, R., Pulls, T.: Insynd: improved privacy-preserving transparency logging. In: Askoxylakis, I., et al. (eds.) ESORICS 2016. LNCS, vol. 9879, pp. 121–139. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-45741-3_7
22. Jensen, M., Kapila, S., Gruschka, N.: Towards aligning GDPR compliance with software development: a research agenda. In: ICISSP, pp. 389–396 (2019)
23. Gerl, A., et al.: LPL, towards a GDPR-compliant privacy language: formal definition and usage. In: Hameurlain, A., Wagner, R. (eds.) Transactions on Large-Scale Data and Knowledge-Centered Systems XXXVII. LNCS, vol. 10940, pp. 41–80. Springer, Heidelberg (2018). https://doi.org/10.1007/978-3-662-57932-9_2
24. Grünewald, E., Pallas, F.: TILT: a GDPR-aligned transparency information language and toolkit for practical privacy engineering. In: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency. FAccT 2021, pp. 636–646. Association for Computing Machinery, New York (2021). ISBN: 978-1-4503-8309-7. https://doi.org/10.1145/3442188.3445925. Accessed 30 Nov 2022
25. Hansen, M., Jensen, M.: A generic data model for implementing right of access requests. In: Gryszczyńska, A., et al. (eds.) APF 2022. LNCS, vol. 13279, pp. 3–22. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-07315-1_1
26. Engineering Personal Data Sharing. ENISA. https://www.enisa.europa.eu/publications/engineering-personal-data-sharing. Accessed 16 Feb 2023

27. Hansen, M., Gruschka, N., Jensen, M.: Introducing the concept of data subject rights as a service under the GDPR. In: Schiffner, S., Ziegler, S., Jensen, M. (eds.) DPLICIT 2023, pp. 17–31. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-44939-0_2
28. European Commission. Enforcement of intellectual property rights. https://single-market-economy.ec.europa.eu/industry/strategy/intellectual-property/enforcement-intellectual-property-rights_en. Accessed 08 Feb 2023
29. Regulation (EU) 2018/1807 of the European Parliament and of the Council of 14 November 2018 on a framework for the free flow of non-personal data in the European Union (Text with EEA relevance.) Legislative Body: EP, CONSIL (2018). http://data.europa.eu/eli/reg/2018/1807/oj/eng. Accessed 14 Feb 2023