A Framework to Identify Data Governance Requirements in Open Data Ecosystems

Armin Haberl, Christine Dagmar Malin & Stefan Thalmann

University of Graz, Business Analytics and Data Science-Center, Styria, Austria.
E-mail: armin.haberl@uni-graz.at, christine.malin@uni-graz.at, stefan.thalmann@uni-graz.at

Abstract Open data and open data ecosystems (ODEs) are important for stakeholders from science, businesses, and the broader society. However, concerns about data sharing and data handling are significant adoption barriers of ODEs that reduce stakeholder participation and thus the success of the initiative. Data governance (DG) is proposed as solution, but requirements of the three stakeholder groups combined are not clear and especially how they can be integrated in one DG concept. This paper develops a framework, supporting elicitation of DG requirements in ODEs. The framework builds on a series of stakeholder workshops and literature research resulting in DG requirements and DG mechanisms. The resulting framework includes five main dimensions: (1) data usability, (2) ethical and legal compliance, (3) data lineage, (4) data access and specified data use, and (5) organizational design.

Keywords: open data, ecosystem, data governance, requirements, framework.

University of Maribor Press

DOI https://doi.org/10.18690/um.fov.4.2022.22
ISBN 978-961-286-616-7

1 Introduction

In recent years open data has become a valuable resource in a digitized world. Data is considered open when it " […] can be freely used, modified, and shared by anyone for any purpose"(Open Knowledge Foundation, n.d.). Open data can enhance transparency and public engagement for public institutions, facilitate innovation and development of new products and services in businesses, or can enable research by providing combined open data sets (Janssen et al., 2012). Based on these expected benefits, governments worldwide increasingly started to propose and implement open data platforms (United Nations, 2020). Promising examples are the European Open Science Cloud (EOSC) and Gaia-X aiming to funnel data from national or regional initiatives into larger combined infrastructures (Bonfiglio, 2021; European Commission, 2019).

Despite the potential of open data and open data ecosystems (ODEs), there are several barriers that can reduce the willingness of stakeholders to participate and to share data. Primarily, these barriers are legal concerns, knowledge protection concerns or technical concerns (Zeiringer and Thalmann, 2021). Such barriers are especially important in co-opetitive settings between science and industry (Kaiser et al., 2020). In this regard, data governance (DG) is frequently proposed to mitigate these barriers in ODEs. So far research on DG in ODEs focussed on specific stakeholder groups such as science, society, or business. Requirements for DG involving all stakeholder groups and finding a suitable consensus is missing so far. The maximal benefit of ODEs, however, can only be realized if all three groups participate in ODEs. Thus, the paper aims to clarify the following research question:

How to identify data governance requirements for different stakeholder groups in ODEs?
Therefore, we conducted a study in a regional ODE focusing on data sharing and collaboration between science, business, and society.

2 Background

Data ecosystems can be defined as a "set of networks composed by autonomous actors that directly or indirectly consume, produce or provide data and other related resources" (Oliveira & Lóscio, 2018). These actors can be divided into platform owners and users. Platform users can be subdivided into a supply side (e.g., data

providers or data analysts) and a demand side (e.g., data consumers) (S. U. Lee et al., 2018a). The DG requirements of platform users are the focus of this paper, since the fulfilment of their requirements ultimately determines if a data platform will be used or not. One promising way to identify DG requirements of stakeholders is to investigate the barriers that prevent participation in ODEs. Beno et al. (2017) identified privacy and security issues, missing economic and strategic incentives, legal constraints, and technical difficulties as core barriers. To mitigate some of these barriers, DG can be implemented in ODEs. DG refers to the allocation of decision rights and accountabilities for data related decision-making in organizations (Khatri & Brown, 2010). The effects of well implemented DG include the reduction of data related risks and positive impacts on organizational performance (Gregory, 2011). Abraham et al. (2019) proposes structural mechanisms (e.g., roles and responsibilities), procedural mechanisms (e.g., standards, monitoring), and relational mechanisms (e.g., communication, training) as core elements of DG frameworks. For each of these dimensions concrete DG mechanisms matching the requirements of all stakeholder groups need to be defined. Several studies explored DG requirements in the context of ODEs or other data ecosystems that can serve as a benchmark for the development of DG (see e.g., Al-Ruithe & Benkhelifa, 2017; D. Lee, 2014; S. U. Lee et al., 2017; Van Den Broek & Van Veenstra, 2015; Welle Donker & van Loenen, 2017; Wende & Otto, 2007).

These studies provide an insight into the requirements of ODEs but do not differentiate between ODEs with different stakeholder groups or only allow to indirectly assess the DG requirements through intermediate factors or configurations. The indirect assessment is problematic for two main reasons: (1) Contingency factors or configurations represent generalizations that are a useful starting point but can by design not depict the reality of every individual case. (2) Indirectly obtaining DG requirements excludes stakeholders from the development process of DG, which can lead to less trust, motivation, and participation of stakeholders compared to decentralized approaches (S. U. Lee et al., 2018a). A DG framework is therefore needed that allows developing ODEs to directly assess DG requirements from their individual stakeholder groups. Such a framework is to our best knowledge not yet available, although existing frameworks can serve as base for its development. Showing the strengths and shortcomings of those existing frameworks is not withing the scope of this paper.

3 Method and Procedure

A three-stage study (see table 1) was conducted in a regional ODE, that aims to develop a reference model for collaborative data use and to foster data cooperation between industry and science. Overall, 31 experts in data-related domains and had professional experience in Styrian universities and research facilities, public administrations, at global players in the automotive industry and pharma industry, insurance companies, and consulting agencies participated. Most of the stakeholders were in leading or top-level positions.

Table 1: Overview of the research process

	Stage 1	**Stage 2**	**Stage 3**
Method	Requirement workshops	Literature research	Validation workshop
Details	16 stakeholders from 3 domains	DG mechanisms from 26 sources	16 stakeholders from 2 domains
Results	DG requirements	DG framework	Validated & prioritized DG framework

Stage 1 - Requirement workshops: First, DG requirements were identified during three separate online workshops in March and April 2021 for stakeholders from science (n=6), public institutions (n=4), and business (n=6). Each workshop consisted of brainstorming, clustering, prioritization, and an in-depth discussion regarding the DG requirements. The workshops were audio-recorded, transcribed and analysed using the thematic analysis by Braun and Clarke (2006).

Stage 2 - Literature research: A literature research was conducted to extend the requirements from stage 1 into a DG framework. As a result, 26 papers were identified and analysed using the approach of Braun and Clarke (2006). Each sub-requirement was structured into three levels of requirement fulfilment. Thus, a DG framework was developed that combines the empirically identified DG requirements with corresponding DG mechanisms obtained in the literature.

Stage 3 - Validation workshop: To validate and prioritize the developed DG framework, an online workshop with 16 participants was conducted in December 2021. The identified main DG requirements were discussed and prioritized the by distributing

10 'priority points' to one or more requirements. Furthermore, they defined their minimum level of requirement fulfilment. Two final questions investigated the overall importance of DG in ODEs and the influence of the designed DG framework on their willingness to participate.

4 Results

The results of our study are five main stakeholder requirements for DG in ODEs: (1) data usability, (2) ethical and legal compliance, (3) data lineage, (4) data access and specified data use, and (5) organizational design (see table 2). Out of the sixteen stakeholders only fifteen participated in the prioritisation. All fifteen of them voted for the requirement 'data usability' and allocated 43 (28,7%) points to this requirement, more than to any other DG requirement. Also, all 15 voters voted for 'ethical and legal compliance' and assigned 35 (23,3%) priority points. The other requirements were not prioritized by all voters.

Table 2: Overview, definitions, and prioritisation results of the DG requirements (n=15)

	Requirement definition	**Priority** sum=150	**Unique voters**
Data usability	Users can use the data with the provided recourses for the allowed use cases.	**43** (28,7%)	15
Ethical & legal compliance	Legal regulations are followed, and data is used ethically, fairly, and transparently.	**35** (23,3%)	15
Data Lineage	Every transformation or alteration of data, from data origin to the current form, is traceable.	**31** (20,7%)	13
Data access & specified data use	Data providers specify, who can access their data under which conditions, for which purposes and what are allowed use cases.	**28** (18,7%)	14
Organizational design	Decision rights and responsibilities are clearly specified within an organizational structure.	**13** (8,7%)	10

4.1 Data usability

The usability of the provided data is key feature of an ODE, as one stakeholder expressed "[…] it is important, that there is a process in place to ensure, that the provided data will enter the system with controlled quality, regarding completeness, plausibility and in the end also data quality. This is a crucial starting point for all the data processing and the conclusions drawn from the provided data [E1]". The interviewee highlights the relationship between quality of input data and the outcome of analytic projects and that "this is a crucial starting point". In addition to the data quality of the data set itself, interviewees mentioned suitable data formats as very important from a technical perspective as well as metadata to understand the data set. Building on this, other stakeholders highlighted that this is not only true for technical data quality measures, but also for more qualitative measures such as relevance and up-to-dateness. See table 3 for details of the main requirement 'data usability'.

Table 3: DG mechanisms for 'data usability' according to literature research

	Level 1	**Level 2**	**Level 3**
Data quality	Quality standards	Level 1 + data cleaning	Level 2 + stewards and data enrichment
Metadata	Non-standardized	Standardized in the data platform	Standardized for different domains
Data formats	One format chosen by data providers	formats chosen by data providers	formats according to user needs
Data updates	No updates	Regular updates + versioning	Near-time updates + versioning

Data Quality - Level 1: Data quality can be ensured by continuously measuring and assessing the uploaded data according to data quality standards and metrics (DAMA, 2010; Otto et al., 2007). Level 2: The data platform can offer additional services to further improve data quality. A data cleaning service can be implemented to correct errors, standardize information, and validate uploaded data (Comerio et al., 2010). Level 3: The introduction of data stewards that are responsible for the quality and use of data can further improve data quality (DAMA, 2010). An additional data enrichment service can add value to existing data sets by incorporating data from other data sources (Comerio et al., 2010).

Metadata - Level 1: Metadata can be implemented without standardized meta data elements such as free texts (Welle Donker & van Loenen, 2017). Level 2: Preferably, metadata and data documentation is standardized throughout the data platform (Welle Donker & van Loenen, 2017). Level 3: Commonly agreed upon metadata standards are used that are valid for certain domains such as e.g., healthcare (Welle Donker & van Loenen, 2017; Zuiderwijk et al., 2012).
Data formats - Level 1: A basic policy might only require data providers to provide data in a single data format of their choice (Welle Donker & van Loenen, 2017). Level 2: To further increase the usability of data, data providers can be required to offer data in different formats (Welle Donker & van Loenen, 2017). Level 3: An advanced policy can even require data providers to offer specific formats that are requested by data users (Welle Donker & van Loenen, 2017).

Data updates - Level 1: In a rudimentary implementation data providers upload data to the platform and do not update their data at any point (Welle Donker & van Loenen, 2017). Level 2: Data providers have to update data regularly (e.g., every year) and offer versioning (Welle Donker & van Loenen, 2017). Level 3: Updating data near-time and offering versioning can additionally improve data usability for data users (Welle Donker & van Loenen, 2017).

4.2 Ethical and legal compliance

General Data Protection Regulation (GDPR) was mentioned by several stakeholders, as one said: "I think it is an essential area that generally must be considered, since it won't be possible to avoid dealing with personal data [E1]". This statement highlights, that the data platform needs mechanisms to ensure the correct use and processing of personal data and to ensure compliance with the GDPR. Furthermore, the user´s ethical responsibilities while using data was highlighted. As solution the stakeholders proposed monitoring the operations in the data platform and the certification of the DG program. Thereby, it was requested that the overall technological implementation of the data platform should ensure data security and compliant use and processing of data. See table 4 for the main requirement 'ethical and legal compliance'.

Table 4: DG mechanisms for 'ethical and legal compliance'

	Level 1	Level 2	Level 3
Monitoring and audits	Done by the platform owner	Done by the platform owner and platform users	Level 2 + external certification
Compliance tools	DG is techno-logically integrated	Level 1 + automated adaptations	Level 2 + measurements and indicators

Monitoring and audits - Level 1: A basic implementation can consist of regular audits and monitoring of the DG program by the owner of the data platform (Abraham et al., 2019; DAMA, 2010; S. U. Lee et al., 2018b). Level 2: In a more decentralized setup, also the platform users audit and monitor the DG program (DAMA, 2010; S. U. Lee et al., 2018b) and thus provide a more objective and unbiased assessments (DAMA, 2010). Level 3: Certifications according to international standards like ISO/IEC 38505-1 that are assured by auditors offer the highest standard to ensure compliance (ISO, 2017; Johannsen et al., 2020).

Compliance tools - Level 1: DG controls can be technologically integrated into a data platform (Al-Ruithe et al., 2019; Gheorghe et al., 2009). These controls can ensure compliance, but are not necessarily linked to specific compliance objectives (Gheorghe et al., 2009). Level 2: Once all integrated controls are linked to explicit compliance objectives, they can automatically adapt when objectives change (Gheorghe et al., 2009). The compliance objectives can be centrally stored in a policy repository and should be adjusted according to legal and ethical requirements (Gheorghe et al., 2009). Level 3: In addition to these automated controls, compliance indicators can be implemented to automatically measure the degree of compliance of DG processes (Al-Ruithe et al., 2019; Gheorghe et al., 2009). An example for these indicators is the number of instances where personal data was not correctly anonymized (Gheorghe et al., 2009).

4.3 Data lineage

A way to trace the data and its transformations in an ODE was requested. One interviewee highlighted the importance of data lineage for the overall trust into the data platform and the provided data: "[…] it is important for those, that use the data,

to build a form of trusted environment and to make clear where data comes from [E2]". Other stakeholders stated that not only the origin of the data, but also the applied transformations are important to ensure reproducibility of research. Especially for this reproducibility the granularity of the lineage data matters since a greater level of detail enables more accurate reproduction of research results. Interviewees pointed out that this lineage information needs to be communicated very clearly to the platform users, making the access to this information an important requirement. See table 5 for details of 'data lineage'.

Table 5: DG mechanisms for 'data lineage according' to literature research

	Level 1	**Level 2**	**Level 3**
Lineage type	Input data	Level 1 + transformations	Level 2 + update lineage
Granularity	Information about data sets	Information about data sets or tuples	Information about tuples
Access	Visualisation	Level 1 + queries	Level 2 + API

Lineage type - Level 1: The lineage information clarifies from which input data given output data was derived, but does not specify which transformations were applied during the derivation process (Ikeda & Widom, 2009). Level 2: Additional lineage information can contain the transformations that were applied during the derivation process ranging from simple aggregations or algebraic operations to complex procedures using custom code (Cui & Widom, 2003; Ikeda & Widom, 2009). Level 3: An extension to level 2 can be a data lineage system, that combines lineage information of derived data (input data and transformations) with the update history of input data (Das Sarma et al., 2008). This allows data users to view different versions of derived data depending on the version of input data they select (Das Sarma et al., 2008).

Granularity - Level 1: Coarsely-grained data lineage contains information about entire data sets (Ikeda & Widom, 2009). An example can be information about the input data set or transformations that were used to produce an output data set (Ikeda & Widom, 2009). Level 2: An extension is to provide lineage by offering information about data sets and individual data-tuples (Simmhan et al., 2005). This might require the use of dataset abstractions to track data in more general forms than datasets or

tuples (Foster, 2003; Simmhan et al., 2005). Level 3: In a finely-grained setup lineage information is available about every given data-tuple (Zafar et al., 2017).

Access - Level 1: A fundamental approach can be to visualize lineage information using a derivation graph (Simmhan et al., 2005). Level 2: Additional queries on the lineage data can be offered, e.g. selecting data with specific transformations (Simmhan et al., 2005). Level 3: To complement the access through visualization and queries, application programming interfaces (APIs) can allow users to implement their own data lineage services (Simmhan et al., 2005).

4.4 Data access and specified data use

The ability to limit data access and data use is a key feature of ODEs as one managing director expressed "[…] data security is important, that it is clear how data is provided and that sensitive data stays secure inside the platform and access rights, and roles are ensured [E3]". This statement highlights how important data access control mechanisms are to ensure, that sensitive data is only available to the intended data users. Further, it was stated that the available data needs to be distributed under appropriate licenses that must be communicated clearly to the data users. Finally, data should only be used for the intended use cases since the context can have substantial influence on the possible interpretation of data. See table 6 for details of 'data access and specified data use'.

Access control - Level 1: A basic method of access control can be based on the identity or role of users (Majumder et al., 2014). In such a setup the platform owner could allow access to data depending on the role of a given user (S. U. Lee et al., 2018b; Majumder et al., 2014). Level 2: To give data providers more control about their data, they can specify access rules to their data depending on the role of a given user (Majumder et al., 2014). Level 3: A policy-based access control method can allow data providers to define individual access policies beyond identity or roles by encrypting data before data upload (Majumder et al., 2014).

Table 6: DG mechanisms for 'data access and specified data use'

	Level 1	Level 2	Level 3
Access control	By platform owner – based on user roles	By data provider – based on user roles	By data provider – based on policies
Licenses	Standard license	Licensing options	Level 2 + custom licenses
Use cases	Platform standards prohibit certain purposes of use	Data providers prohibit certain purposes of use	Level 2 + certain contexts of use are prohibited

Licenses - Level 1: The simplest solution is to define a standard license that applies to all data of the data platform and allowing users to use and combine different data sets (Martin et al., 2013). Level 2: The data platform can allow data providers to choose from a defined set of licenses that should apply to their published data (Immonen et al., 2018). Selection guidelines can be offered to guide data providers in their selection process (Alamoudi et al., 2020; Immonen et al., 2018). Level 3: In the advanced case data providers can choose from a defined set of licenses and customize those licenses (Alamoudi et al., 2020; Immonen et al., 2018).

Use cases - Level 1: The implementation of data access standards and corresponding service level agreements are a simple way to restrict data usage (Abraham et al., 2019; Al-Ruithe et al., 2019; Khatri & Brown, 2010). Level 2: In a more advanced case data providers could specify access requirements, that limit the use of their data to certain purposes (Custers & Uršič, 2016). These purposes may include social use (non-profit), professional use (for-profit) or academic use (Abella et al., 2019). Level 3: Data providers could additionally restrict the data usage to certain contexts and prohibit data recontextualization, e.g. health data can be used for diagnostics but not for health insurances (Custers & Uršič, 2016).

4.5 Organizational design

A clear definition and distribution of DG roles and responsibilities in an ODE are of enormous importance for the organizational design, as one interviewee highlighted "[…] it is simply important to know, what kind of role distribution is present in such a platform. […] It is no simple organizational task to determine, who is going to take on which role [E3]". In addition, he expressed the requirement of

platform neutrality, that should ensure that the (decision) rights of the platform users are not restricted by the platform owner. See table 7 for details of 'organizational design'.

Table 7: DG mechanisms for 'organizational design' according to literature research

	Level 1	Level 2	Level 3
Decision rights	Mostly held by platform owner	Shared by platform owner and users	Mostly held by platform users
Roles and responsibilities	Roles are implemented	Level 1+ Roles are regularly adapted	Level 2 + Training for role owners

Decision rights - Level 1: In a data platform with a centralized DG design, the data related decision rights and control are mostly held by the platform owner (Abraham et al., 2019; S. U. Lee et al., 2018b; Lis & Otto, 2021). This centralized form of DG can be simpler, but also lacks transparency and stakeholder participation (S. U. Lee et al., 2018b). Level 2: A more decentralized approach can allow data users to share some of the data related decision rights with the platform owner (Abraham et al., 2019; S. U. Lee et al., 2018b; Lis & Otto, 2021). Level 3: A fully decentralized or self-organized approach can locate most of the decision rights to the platform users and only keep core decision to the platform owner (Abraham et al., 2019; S. U. Lee et al., 2018b; Lis & Otto, 2021). This approach can lead to more transparency and stakeholder participation compared to the more central approaches of level 1 + 2 (S. U. Lee et al., 2018b). However, this approach is also complex and difficult to execute (S. U. Lee et al., 2018b).

Roles and responsibilities - Level 1: A basic implementation can define DG roles and responsibilities of the data platform (Al-Ruithe & Benkhelifa, 2017). Level 2: Additionally, the roles and responsibilities can be regularly reviewed and adapted to meet the changing requirements (Al-Ruithe & Benkhelifa, 2017). Level 3: Further, the role owners can also receive role-specific training to reduce user errors, increase productivity and increase compliance (Otto et al., 2007).

5 Discussion and Outlook

We developed a framework for DG requirement identification validated by stakeholders from science, public institutions and business. Feedback from stakeholders underlined the importance of DG in ODEs and confirmed the positive effects of well-designed DG on stakeholder participation.

The main **theoretical contribution** is the clarification of DG requirements for ODEs with the three mentioned stakeholder groups. In addition, we also provide corresponding DG mechanisms that match the identified DG requirements. This complements existing frameworks that either do not consider varying DG requirements of different stakeholder groups and matching DG mechanisms (Abraham et al., 2019; Al-Ruithe & Benkhelifa, 2017; Welle Donker & van Loenen, 2017) or only indirectly identify DG requirements through intermediary factors or configurations (S. U. Lee et al., 2017; Van Den Broek & Van Veenstra, 2015; Wende & Otto, 2007).

As the main **managerial contribution**, the newly developed framework can be used in ODEs to directly obtain DG requirements from corresponding stakeholder groups. This inclusion of stakeholders in the development process of DG can lead to more satisfaction, trust, motivation and participation of these stakeholders (S. U. Lee et al., 2018b). Fulfilling the DG requirements of different stakeholder groups can help to overcome the open data adoption barriers as described by Beno et al. (2017).

The main **limitation** of this paper is, that the developed DG framework is based on requirements of a regional ODE. Even though the framework allows to differentiate fulfilment levels, the overall DG requirements and sub-requirements might differ in other regional contexts and with different stakeholder groups. A generalization of our results was not intended but can serve as a foundation for regional or contextual adaptations. The scope of this research did not include the implementation of DG and it is therefore not possible to make claims about the practical feasibility of the proposed DG framework.

Future research should explore, how the identified DG framework and the suggested DG mechanisms can be implemented in ODEs. In this regard the perspective of Open Educational Recources in the context of internationalization (Pirkkalainen et. al, 2010) should be taken into account. Additional research is also needed to adapt the framework for other regional contexts and different stakeholder groups.

References

Abella, A., Ortiz-de-Urbina-Criado, M., & De-Pablos-Heredero, C. (2019). The process of open data publication and reuse. Journal of the Association for Information Science and Technology, 70(3), 296–300. https://doi.org/10.1002/asi.24116

Abraham, R., Schneider, J., & vom Brocke, J. (2019). Data governance: A conceptual framework, structured review, and research agenda. International Journal of Information Management, 49, 424–438. https://doi.org/10.1016/j.ijinfomgt.2019.07.008

Alamoudi, E., Mehmood, R., Aljudaibi, W., Albeshri, A., & Hasan, S. H. (2020). Open Source and Open Data Licenses in the Smart Infrastructure Era: Review and License Selection Frameworks. In R. Mehmood, S. See, I. Katib, & I. Chlamtac (Eds.), Smart Infrastructure and Applications (pp. 537–559). Springer International Publishing. https://doi.org/10.1007/978-3-030-13705-2_22

Al-Ruithe, M., & Benkhelifa, E. (2017). Cloud data governance maturity model. Proceedings of the Second International Conference on Internet of Things, Data and Cloud Computing, 1–10. https://doi.org/10.1145/3018896.3036394

Al-Ruithe, M., Benkhelifa, E., & Hameed, K. (2019). A systematic literature review of data governance and cloud data governance. Personal and Ubiquitous Computing, 23(5), 839–859. https://doi.org/10.1007/s00779-017-1104-3

Beno, M., Figl, K., Umbrich, J., & Polleres, A. (2017). Open Data Hopes and Fears: Determining the Barriers of Open Data. 2017 Conference for E-Democracy and Open Government (CeDEM), 69–81. https://doi.org/10.1109/CeDEM.2017.22

Bonfiglio, F. (2021). Gaia-X: Vision & Strategy. https://gaia-x.eu/sites/default/files/2021-12/Vision%20%26%20Strategy.pdf

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative Research in Psychology, 3(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Comerio, M., Truong, H.-L., Batini, C., & Dustdar, S. (2010). Service-oriented data quality engineering and data publishing in the cloud. 2010 IEEE International Conference on Service-Oriented Computing and Applications (SOCA), 1–6. https://doi.org/10.1109/SOCA.2010.5707184

Cui, Y., & Widom, J. (2003). Lineage tracing for general data warehouse transformations. The VLDB Journal The International Journal on Very Large Data Bases, 12(1), 41–58. https://doi.org/10.1007/s00778-002-0083-8

Custers, B., & Uršič, H. (2016). Big data and data reuse: A taxonomy of data reuse for balancing big data benefits and personal data protection. International Data Privacy Law, ipv028. https://doi.org/10.1093/idpl/ipv028

DAMA. (2010). The DAMA guide to the data management body of knowledge: DAMA-DMBOK guide (M. Mosley, M. Brackett, S. Earley, & D. Henderson, Eds.; First edition). Technics Publications, LLC.

Das Sarma, A., Theobald, M., & Widom, J. (2008). Data Modifications and Versioning in Trio (Technical Report ILPUBS-849; p. 14). Stanford University.

European Commission. (2019). European Open Science Cloud (EOSC) strategic implementation plan (S. Jones & J. Abramatic, Eds.). Publications Office. https://doi.org/10.2777/202370

Foster, I. (2003). The virtual data grid: A new model and architecture for data-intensive collaboration. 15th International Conference on Scientific and Statistical Database Management, 2003., 11. https://doi.org/10.1109/SSDM.2003.1214945

Gheorghe, G., Massacci, F., Neuhaus, S., & Pretschner, A. (2009). GoCoMM: A governance and compliance maturity model. Proceedings of the First ACM Workshop on Information Security Governance - WISG '09, 33. https://doi.org/10.1145/1655168.1655175

Gregory, A. (2011). Data governance — Protecting and unleashing the value of your customer data assets: Stage 1: Understanding data governance and your current data management capability. Journal of Direct, Data and Digital Marketing Practice, 12(3), 230–248. https://doi.org/10.1057/dddmp.2010.41

Ikeda, R., & Widom, J. (2009). Data Lineage: A Survey [Technical Report]. Stanford InfoLab. http://ilpubs.stanford.edu:8090/918/

Immonen, A., Ovaska, E., & Paaso, T. (2018). Towards certified open data in digital service ecosystems. Software Quality Journal, 26(4), 1257–1297. https://doi.org/10.1007/s11219-017-9378-2

ISO. (2017). ISO/IEC 38505-1:2017. ISO. https://www.iso.org/cms/render/live/en/sites/isoorg/contents/data/standard/05/66/56639.html

Janssen, M., Charalabidis, Y., & Zuiderwijk, A. (2012). Benefits, Adoption Barriers and Myths of Open Data and Open Government. Information Systems Management, 29(4), 258–268. https://doi.org/10.1080/10580530.2012.716740

Johannsen, A., Kant, D., & Creutzburg, R. (2020). Measuring IT security, compliance and data governance within small and medium-sized IT enterprises. Electronic Imaging, 2020(3), 252-1-252–11. https://doi.org/10.2352/ISSN.2470-1173.2020.3.MOBMU-252

Kaiser, Rene, Stefan Thalmann, Viktoria Pammer-Schindler, and Angela Fessl. "Collaborating in a research and development project: knowledge protection practices applied in a co-opetitive setting." WM 2019-Wissensmanagement in digitalen Arbeitswelten: Aktuelle Ansätze und Perspektiven-Knowledge Management in Digital Workplace Environments: State of the Art and Outlook (2020).

Khatri, V., & Brown, C. V. (2010). Designing data governance. Communications of the ACM, 53(1), 148–152. https://doi.org/10.1145/1629175.1629210

Lee, D. (2014). Building an open data ecosystem: An Irish experience. Proceedings of the 8th International Conference on Theory and Practice of Electronic Governance, 351–360. https://doi.org/10.1145/2691195.2691258

Lee, S. U., Zhu, L., & Jeffery, R. (2017). Design Choices for Data Governance in Platform Ecosystems: A Contingency Model. ArXiv:1706.07560 [Cs]. http://arxiv.org/abs/1706.07560

Lee, S. U., Zhu, L., & Jeffery, R. (2018a). A Data Governance Framework for Platform Ecosystem Process Management. In M. Weske, M. Montali, I. Weber, & J. vom Brocke (Eds.), Business Process Management Forum (pp. 211–227). Springer International Publishing.

Lee, S. U., Zhu, L., & Jeffery, R. (2018b). Designing Data Governance in Platform Ecosystems. Hawaii International Conference on System Sciences. https://doi.org/10.24251/HICSS.2018.626

Lis, D., & Otto, B. (2021). Towards a Taxonomy of Ecosystem Data Governance. Hawaii International Conference on System Sciences. https://doi.org/10.24251/HICSS.2021.733

Majumder, A., Namasudra, S., & Nath, S. (2014). Taxonomy and Classification of Access Control Models for Cloud Environments. In Z. Mahmood (Ed.), Continued Rise of the Cloud (pp. 23–53). Springer London. https://doi.org/10.1007/978-1-4471-6452-4_2

Martin, S., Foulonneau, M., Turki, S., & Ihadjadene, M. (2013). Risk Analysis to Overcome Barriers to Open Data: EJEG. Electronic Journal of E-Government, 11(1), 348–359. Publicly Available Content Database.

Oliveira, M. I. S., & Lóscio, B. F. (2018). What is a data ecosystem? Proceedings of the 19th Annual International Conference on Digital Government Research: Governance in the Data Age, 1–9. https://doi.org/10.1145/3209281.3209335

Open Knowledge Foundation. (n.d.). Open Definition—Defining Open in Open Data, Open Content and Open Knowledge. Retrieved 9 February 2022, from http://opendefinition.org/

Otto, B., Wende, K., Schmidt, A., & Osl, P. (2007). Towards a Framework for Corporate Data Quality Management (M. Toleman, A. Cater-Steel, & D. Roberts, Eds.; pp. 916–926). The University of Southern Queensland. https://aisel.aisnet.org/acis2007/109

Pirkkalainen, H., Thalmann, S., Pawlowski, J., Bick, M., Holtkamp, P., & Ha, K. H. (2010). Internationalization processes for open educational resources. In Workshop on Competencies for the Globalization of Information Systems in Knowledge-Intensive Settings, ICSOB.

Simmhan, Y. L., Plale, B., & Gannon, D. (2005). A survey of data provenance in e-science. ACM SIGMOD Record, 34(3), 31–36. https://doi.org/10.1145/1084805.1084812

United Nations (Ed.). (2020). Digitial government in the decade of action for sustainable development. United Nations.

Van Den Broek, T., & Van Veenstra, A. F. (2015). Modes of Governance in Inter-Organizational Data Collaborations. ECIS 2015 Completed Research Papers, Paper 188. https://doi.org/10.18151/7217509

Welle Donker, F., & van Loenen, B. (2017). How to assess the success of the open data ecosystem? International Journal of Digital Earth, 10(3), 284–306. https://doi.org/10.1080/17538947.2016.1224938

Wende, K., & Otto, B. (2007). A Contingency Approach to Data Governance (M. A. Robert, R. O'Hare, M. L. Markus, & B. Klein, Eds.; pp. 163–176). https://www.alexandria.unisg.ch/213308/

Zafar, F., Khan, A., Suhail, S., Ahmed, I., Hameed, K., Khan, H. M., Jabeen, F., & Anjum, A. (2017). Trustworthy data: A survey, taxonomy and future trends of secure provenance schemes. Journal of Network and Computer Applications, 94, 50–68. https://doi.org/10.1016/j.jnca.2017.06.003

Zeiringer, Johannes P., and Stefan Thalmann. "Knowledge sharing and protection in data-centric collaborations: An exploratory study." Knowledge Management Research & Practice (2021): 1-13.

Zuiderwijk, A., Jeffery, K., & Janssen, M. (2012). The Potential of Metadata for Linked Open Data and its Value for Users and Publishers. JeDEM - EJournal of EDemocracy and Open Government, 4(2), 222–244. https://doi.org/10.29379/jedem.v4i2.138