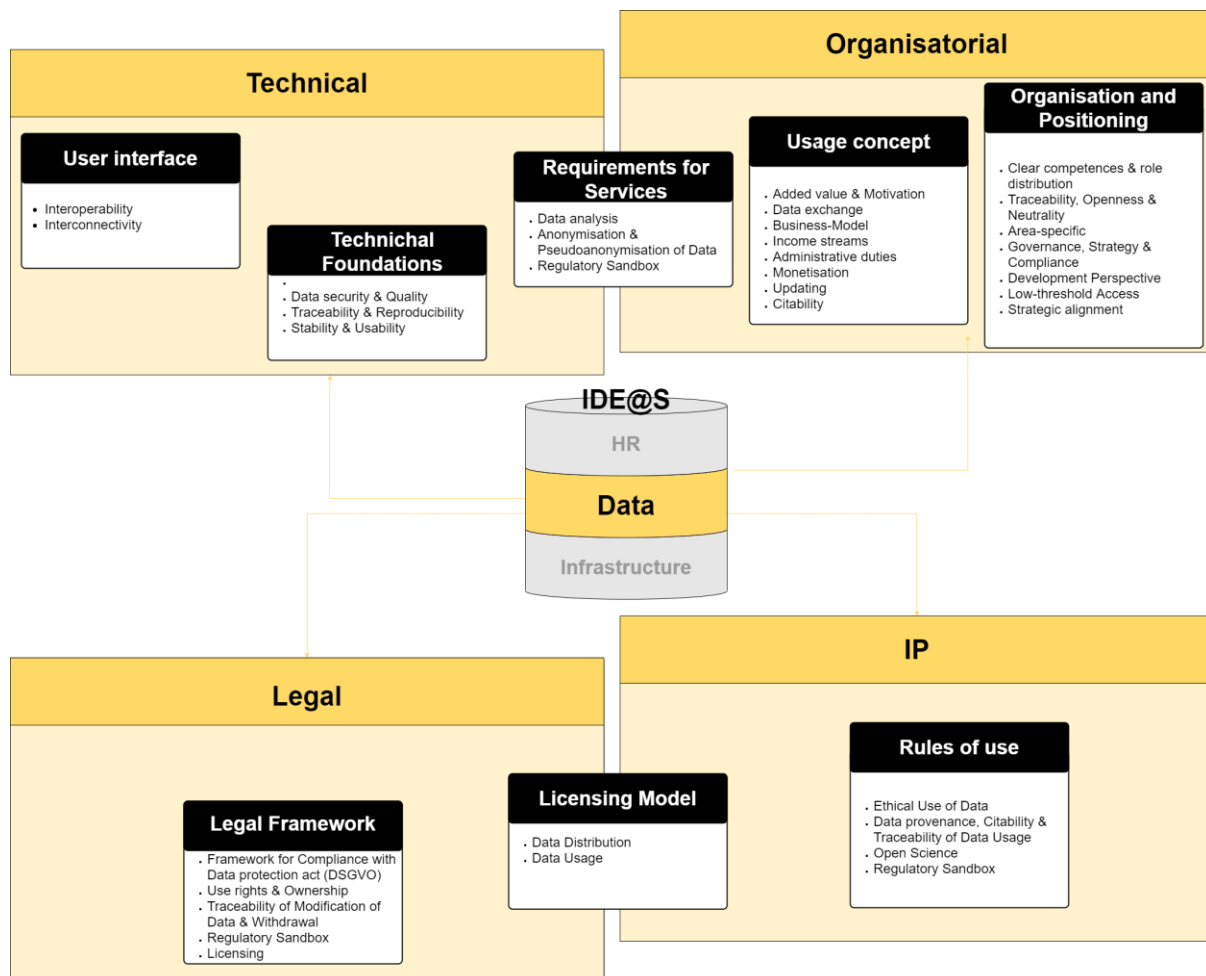


Requirements for a data exchange platform – Workshop series

The University of Graz hosted three workshops between 31st of March and 16th of April with 16 representatives from research organisations, public administration and businesses. The goal was to know the individual requests from each of the stakeholder groups for a collaborative data space in Styria. The interested parties discussed lively answers to the central question, *Which criteria should be met by a data platform in order for your company or organisation to use it?* The results from the workshops are reported here, and are graphically summarized in the diagram below.



Technical requirements

The technical requirements at the base of the IDE@S platform include data security and quality, data traceability and reproducibility, system stability and usability. Data security requires the secure exchange of the data and the clear definition of access rights. Data quality issues include plausibility, completeness and actuality of the data. Data traceability and provenance requires the possibility to track access to data, determining who accessed it, and if and how data were modified or analysed. The stability of the platform should guarantee it can meet current and future user requirements. Great importance is attached to good usability of the platform.

The IDE@S platform should show a clear concept for the user interface, which must enable interoperability and interconnectivity. IDE@S should in no way become an isolated solution, but instead be seamlessly integrated in existing processes, systems and initiatives. The interface should

also enable ETL processes (Extract, Transform, Load) as well as support widely established standards, like e.g. REST (Representational State Transfer) architecture.

From the technical perspective, IDE@S should comprise a definite service portfolio. For this it essential the exchange of resources, including data exchange, expertise and know-how. Besides access to the data, the system must provide data analysis tools. To meet the legal requirements from the technical side, it is desirable to have data anonymization and depersonalisation procedures as well as regulatory sandboxes in place.

Organisational requirements

The requirements for the organisational features of the data platform are of high priority to stakeholders. The governance structure is decisive for the data platform to be used at all. A specific request is the clear definition of responsibilities and duties. It is important to establish who leads the whole project and who keeps it alive. Moreover, the roles of the different stakeholders must be properly distributed. Besides the responsibilities of the organisation in the driving seat, those of users and providers must also be considered. Here, organizational principles such as traceability, openness and neutrality are important. Due to the differences among stakeholders, IDE@S must meet area-specific as well as general requirements. Furthermore, a governance strategy must be put in place and regulatory compliance must be established. Governance must also find an appropriate legal structure in which to embed the data platform. Regarding the strategic orientation, a roadmap for further development should be adopted and communicated to stakeholders from the very beginning. IDE@S must proactively approach many organisations to establish partnerships, for which a low-threshold access model would be desirable, since in many cases organisations still lack clear demands for data services. A low-threshold access would enable organisations to both benefit from and contribute to the data platform. The (possibly complex) strategic orientation must be properly formulated and remain traceable. Together with IDE@S' inner organisation, connection to and integration with existing open-data initiatives are important requirements, e.g. the European Open Science Cloud.

The discussion about the „usage concept“ of IDE@S placed further requests upon data use. Specifically, it was pointed that the data must provide an added value to the platform and to the participant organisations. This added value can also stem from available data analysis tools and should contribute to motivate data sharing. The benefit that can be obtained from the data exchange is the highest priority for some stakeholders and the basis for further considerations. Related to this, the business model of the platform and the income streams must also be clearly identified. Also important here is the cost of administrative tasks, which should not be higher than the expected income. Together with monetisation, non-financial aspects of data provision must be taken into account. In order for data to deliver added value, it must be kept up-to-date and of high quality. This requires a plan to update the platform at regular intervals. The updating tasks should not be the responsibility of individual users but be ascribed to the participant organisations. Citability of data can also provide added value to stakeholders and should also be ensured.

Legal requirements

From the legal point of view, IDE@S' overarching framework should comply with data protection regulations (DSGVO). In particular, user rights, data ownership, traceability of data modification and

withdrawal are important components of this framework. These should also hold for the use of available data services. The use of sandboxes requires to develop a usage concept that allows to choose which regulations apply in a particular case, so that their influence can be determined. Furthermore, data licensing policies compatible with the legal framework must also be designed.

IP-requirements

Requirements for IP rights were considered as part of the “rules of use”. Here, the discussion focussed in the requirements for the interactions within the platform ecosystem. Ethical usage of personal data and the establishment of corresponding regulations is of primary importance. The connection of data to their context must be kept at all times to avoid incorrect interpretation of the data. For the users it should be clear where do the data come from (data provenance) and what can they be used for. The above-mentioned citability is therefore crucial. A significant requirement is that data be used according to the FAIR principles. Individual usage and IP rights must be granted both for data upload as well as for data use. In this respect are regulations for data use decisive and will determine whether the platform is used at all. Finally, the platform should support and encourage the adoption of open science practices, at least for organisations already engaged in them. For this, open-access and open-data are the most relevant requirements. Regulatory sandboxes that allow to experiment with data would be an asset in this regard.

*Bericht erstellt von Christine Malin (Uni Graz)
Englische Bearbeitung: Miguel Rey Mazón (TU Graz)*