

UDAPTOR

Data Portability as a Service: whitepaper

I. EXECUTIVE SUMMARY

We are a team of 2 that has been performing academic and marketing research and developing UDAPTOR - a Platform for end-to-end Data Portability, for the past year. The mission of UDAPTOR is to give people more control over their personal data by providing them with the ability to switch between service providers with just a few clicks. We see UDAPTOR as one of the main drivers of building a secure space for seamless data transfers within the European Union.

Currently, we are focusing on the following streams of work:

- (1) **Data Interoperability & Compatibility.** We have designed a Data Portability System (DPS), which is able to convert user's data from one vendor format to another in an automatic way.
- (2) **Data Access.** We propose a Data Access Assistant tool that supports users during the Data Request process and automates most steps. It will be integrated with the DPS as the main channel for retrieving the data from various providers.
- (3) **Data Awareness.** We propose the popularization of Data Protection related topics through content creation, marketing, and promotion.

Overall, our proposal for UDAPTOR is the following (Fig.1): use the blog and campaigns to build a Data Awareness foundation and develop the Data Access Assistant and the Data Portability System, which altogether will make the Platform for end-to-end Data Portability. The platform will make sure that each step of switching providers, i.e. getting aware of the opportunity, requesting and receiving the data, converting it to a usable format, and transferring the data; will happen in a seamless way by means of a single product. Consequently, it will bring the community towards breaking data silos, and start-ups/SMEs will get the chance to access users' data to provide their best service from day-0.

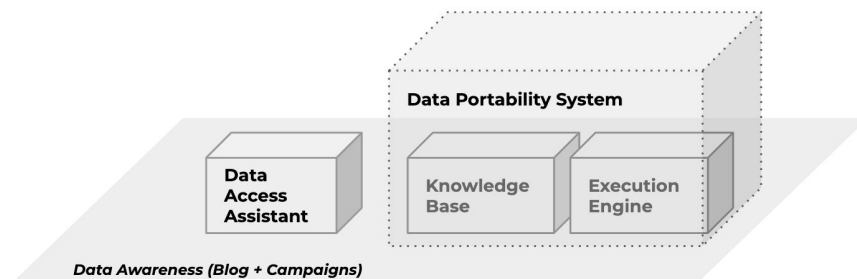


Figure 1 - the platform for end-to-end Data Portability.

II. DEEP DIVE

The main part of the proposed solution is the Data Portability System (DPS) built according to a blueprint designed by our team in our academic research. The core module of the DPS system is a Knowledge Base – a 3-layer Knowledge Graph (namely physical, vendor and canonical layers) that uses RDFS and OWL notations to represent schemas of the data (see Fig. 2), each layer serves certain needs, and layers are connected between each other.

The physical model layer represents the schema that shows how the data is physically stored in files and is used as an I/O tool. The vendor model layer shows the relation between concepts and artifacts, it does not reveal any physical structure of the data and it is used as a connection between physical and canonical models. The canonical model layer is an abstract model that connects all vendors in the Knowledge Base using 2 kinds of mappings – Local-as-view and Global-as-view. This allows us to leverage from mapping composition for constructing direct mappings from one data source to another.

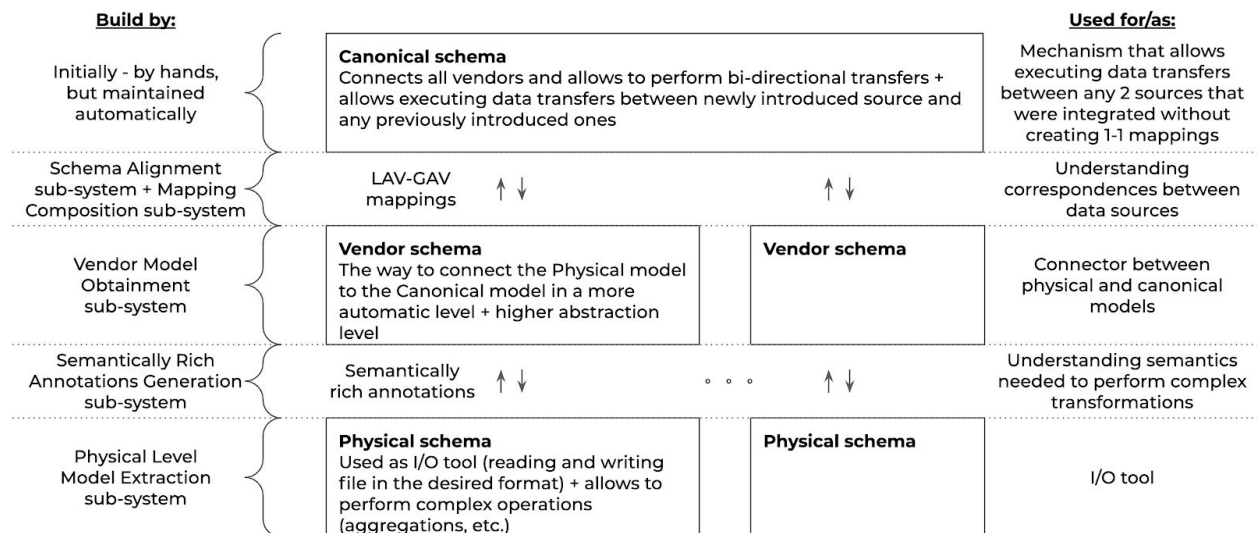


Fig.2 - The Knowledge Base, a high-level diagram

In Fig. 3 you can see the BPMN diagram that defines the main use cases and shows the general flow. There are 2 main streams: “registering a data source” aims at supplementing the Knowledge Base with new service providers (vendors) in a highly automated fashion, with human checks done by an operator of the system; “executing the conversion” aims at the conversion of data collections from one format to the other, in a completely automated manner.

In Fig. 2 and Fig. 3 we can see the other elements of the DPS: the Execution Engine, and the subsystems for Physical Level Model Extraction, Vendor Model Obtainment, Schema Alignment, Semantically Rich Annotations Generation and Mapping Composition. Both Figures describe what the modules are responsible for and at which point of the use cases they are contributing. To the date of application, our team already has the Physical Level Model Extraction and a beta version of the Knowledge Base and has been working following parts of the DPS: the sub-systems for and for Vendor Model Obtainment, the Execution

Engine module and the annotations for in-between layers. The rest of the elements of the DPS have been designed, proof of concepts have been made, and are ready to be implemented.

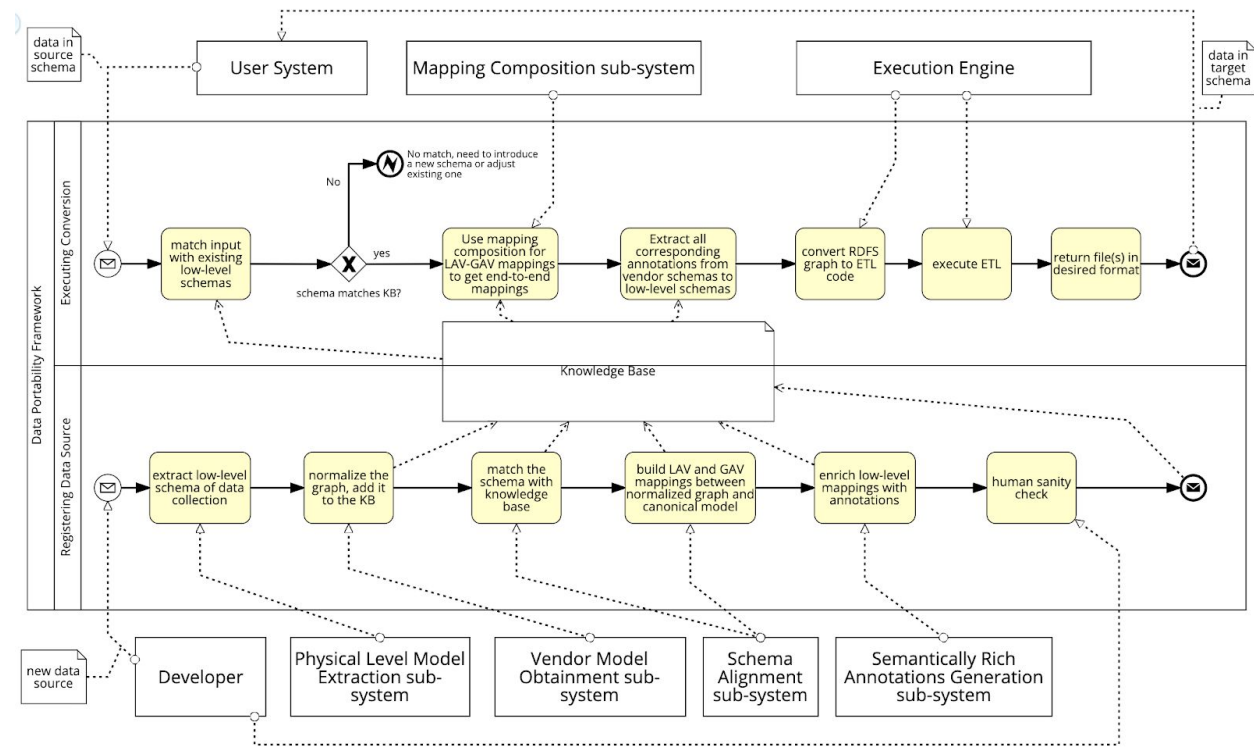


Fig. 3 - The BPMN diagram for the main use-cases for the DPS.

The DPS has the following principles that facilitate switching between providers in a seamless way:

- **automated registering of sources:** few hours to integrate a new vendor as data provider/receiver;
- **integrate once - convert anything:** introduction of a new data source allows conversions between it and any previously introduced data sources without any human interaction;
- **easy adjustments to changes:** in case of schema change, we only adjust the physical model layer;
- **transparency:** open-source code for file transformations, and built on top of open-source libraries;
- **performing syntactic and semantic conversions:** the DPS encompasses rich operations such as merging, aggregating, conversion, filters, etc. easing Data Portability for companies and users;
- **domain-agnostic approach:** Ontologies let us perform transformations within any domain (e.g. Music, Fitness, Finances) as well as across multiple domains;

We designed a whole ecosystem in order to introduce the DPS to people and businesses and maximize its utility. The ecosystem consists of a set of technological and business ad-hoc solutions that facilitate switching between service providers.

First of all, we are going to address the Data Access problem: the biggest pain point that our team identified during the launch of our first MVP was the complexity of data retrieval by users - it's a very varied and mostly manual process. We are developing a Data Access Assistant - a tool that automates the process of requesting and receiving the data, showing hints, and notifying people at important milestones. Currently, the Data Access Assistant supports only a few services to request the data from. We will scale it to a wider range of services, get it to a mature state and publish its code as open-source - heading towards building a community of Data specialists and enthusiasts around it.

The other very important point to take action upon is Data Awareness: our goal here is to empower individuals to exercise their Data Rights. The people who are aware of the topic will pioneer and popularise Data Portability opportunities. Thus, we launched a blog where we explain complex things in a simple way to people (Brexit and Data Protection, CCPA, etc.). Within this project our plan is to continue working in content creation, introducing new formats (e.g. series of posts on topics like "Data Protections in different countries", educational content to increase Data Literacy among the community, etc). We aim to significantly increase the number of people in this warmed up channel so that it can be used as a source of early adopters.

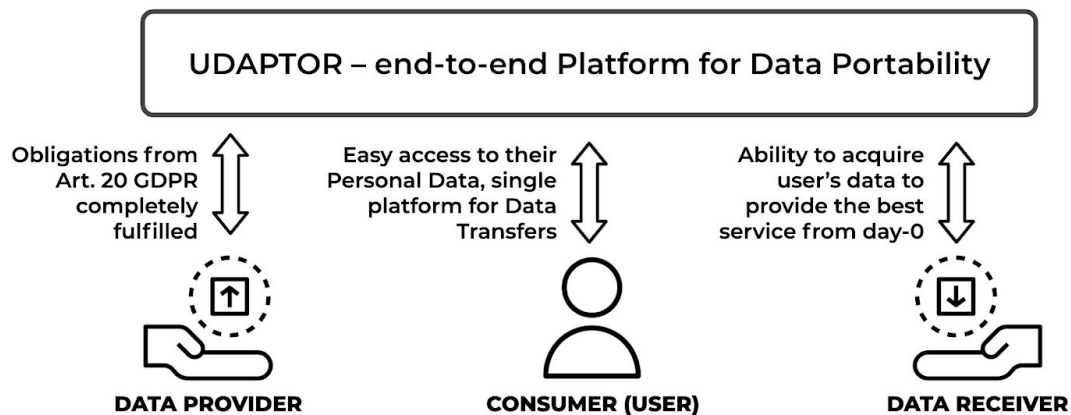


Fig. 4 - the three-way market for UDAPTOR.

We see our solution as the mediator of a 3-way market (Figure 4): we connect Data Provider (DP), Consumer (C), and Data Receiver (DR). UDAPTOR helps a Consumer to transfer data in a seamless manner with minimum actions required, it allows a Data Receiver to get the data in a usable format (outsourcing Data Engineering), and even more, it allows the Data Provider to fulfill their obligations regarding GDPR Art.20 in full scale. We see several examples of use-cases:

- 1) **B2B data transfer between direct competitors:** a user (C) switch music provider from Apple Music (DP) to Spotify (DR)
- 2) **B2B data transfer between complimentary services:** a user (C) transfers data from Grammarly (DP) to Duolingo (DR) to get more tailored lessons
- 3) **B2B data transfer for breaking Data Silos:** A house owner (C) transfers data from a current Energy Provider (DR) company to Zolar

- 4) **B2G data transfer for Public Good:** a citizen (C) transfers their anonymized Google's location data (DP) to the city transport department (DR) to help to conduct analysis on city planning
- 5) **G2G data transfers for Public Good:** transferring data from hospitals (DP) to a Federal Research Center (DR) for research purposes

The main challenges of the project will lie in making a people-oriented implementation of our theoretical research. That includes extending the formalized knowledge basis to new online services, accommodating new use cases of data transformation in the Execution Engine, and integrating all the involved parts of the DPS as a usable product that is transparent and trusted by people. This entails taking into consideration the classical variables of data-related problems such as Volume, Velocity, and Variety.

We also would like to elaborate on core values the solution that we believe would make a difference from the social-economic perspective:

- **Data Liberation:** Our solution gives people a tool that will increase control over personal data. As users should be the main beneficiaries, our service is free for people forever.
- **Market Democratization:** The DPS will increase the users' mobility among service providers, endorsing start-ups/SMEs, and questioning current data monopolies. It has great potential for breaking data silos and giving opportunities to young start-ups that struggle all across the EU.
- **Transfer-only Privacy:** UDAPTOR is a transfer tool built in a way that the data is deleted immediately upon completion to avoid any potential harm, the data is never stored or analyzed.
- **Data for Public Good:** Besides B2B data transfers the community can significantly improve from G2B, B2G, and G2G transfers, facilitating initiatives like Personalized medicine, The European Green Deal, etc.

In the near future, the solution will become a transparent people-oriented Data Trustee (the term from "A New Competition Framework for the Digital Economy" by the 'Competition Law 4.0' Commission) or Data Intermediary (from "A European strategy for data"). We want to work closely with regulatory and research institutions towards building a secure, homogeneous, and people-oriented data transfer environment within the European Union and across all sectors.