Problem

- All about personalization
- All about bundling of services/purposes
- Whitepaper: negotiation of roles as consumers and data subjects
- But also:
  - citizens,
  - participants in scientific endeavours,
  - workers,
  - etc

Ultimately: as individuals associating with each other through data

PersonalData.IO wants to explore this, fast.
Content consumers

- Where do a lot of disinformation problems come from?
- Adtech mess
Erosion of purpose limitation

"Personalisation"

"The collection and processing of information about your use of this site to subsequently personalize advertising for you in other contexts, i.e. on other sites or apps, over time. Typically, the content of the site or app is used to make inferences about your interests which inform future selections."

Non-monetary value:

- trust
- sovereignty

● content consumers?
● consumers?
● citizens?
“Lumascape”
Citizens

- How do we study disinformation on Facebook?
- Continue negating our agency as data subjects...
- ... our right to balance privacy concerns...
- ... and our right to participate in science

Facebook’s legal team, policy team, and communications team—all risk-averse—had to sign off. “We had like 80 or 90 drafts,” says Stamos. The main thing that was taken out were examples of such posts, which the legal team feared would violate various privacy laws or decrees. Weirdly, he says, the Russians—even though they were lying about their identities—were considered customers of Facebook Ireland, and covered by strong privacy laws.

Also (paraphrase): “this is too hard” “would cost us too much”
Facebook internals

http://tiny.cc/FB-data-model
Value: assess fairness, mobility, VAT, social security
Workers + city population

Aren’t cities best placed to assess this?
How?


http://tiny.cc/Uber_data_model
Not that crazy...
Algebraic Property Graphs
Joshua Shinavier
Uber Technologies
Ryan Wisnesky
Conexus AI

Abstract
In this paper, we use algebraic data types to define a formal basis for the property graph data models supported by popular open source and commercial graph databases. Developed as a kind of inter-lingua for enterprise data integration, algebraic property graphs encode the binary edges and key-value pairs typical of property graphs, and also provide a well-defined notion of schema and support straightforward mappings to and from non-graph datasets, including relational, streaming, and micro-service data commonly encountered in enterprise environments. We propose algebraic property graphs as a simple but mathematically rigorous bridge between graph and non-graph data models, broadening the scope of graph computing by removing obstacles to the construction of virtual graphs.

2012 ACM Subject Classification Data modeling, Data exchange, Graph-based database models

Keywords and phrases Property Graphs, Algebraic Data Types, Algebraic Databases
Algebraic Property Graphs

Joshua Shinavier
Uber Technologies

Ryan Wisnesky
Conexus AI

Abstract
In this paper, we use algebraic data types to define a formal basis for the property graph data models supported by popular open source and commercial graph databases. Developed as a kind of inter-lingua for enterprise data integration, algebraic property graphs encode the binary edges and

Metadata graph

- Hundreds of thousands of structured datasets at Uber
- Data protections and user trust
  - GDPR and other regulations, Uber’s own data policies
  - What kind of user data? Where is it?
- Heroic numbers of manual annotations
  - Limited expressivity, limited guarantees
  - Inference is required
- Two birds: in annotating datasets, standardize and compose schemas
  - Now we have a true global knowledge graph
  - Investigating efficient reasoning and “No ETL” solutions
PersonalData.IO

Virtual “tiers-lieu”: hackerspace, coworking, hackathon,...

Put skills together, activate individuals

- SAR crowdsourcing
- ecosystem visualizations
- stakeholder engagement (rinse and repeat)
- personal data analysis
- regulator complaints
- code generation (?)
Focus areas (?)

- different types of agency
- profiling data: access, not portability
- key “hinges”
  - consents: freely provided → portability
  - identity

- make data mobile within one purpose
  - Instagram ← → WhatsApp integration a win for me
  - better cut: identity/consent/resource
  - I will seek myself (and others will help) portability across purposes
  - DTP denies possibility of the latter
Scientifically understanding problematic internet use

Project with hospitals in Vaud and Geneva (early)
data transfer project
Bundling

- semantic + syntactic commodification
- security
- governance

(Libra: semantic + syntactic commodification, financialization, governance)

Facebook, Google, etc decide on what’s ok
API keys

Acquire API keys

The Hosting Entity must acquire API keys for all the providers it wants to be able to transfer data to or from. This allows the Hosting Entity to decide which providers it wishes to interact with. This step also allows each provider to have control, if they wish, over which other providers or Hosting Entities they give keys to. This step may also require the Hosting Entity to agree to the terms and conditions required by providers to use their APIs.
Reciprocity

A healthy data portability ecosystem has providers that allow equivalent import and export functionality. Providers that import data but don’t allow a similar level of export may pose a risk to users by trapping their data into a service. There are several possible ways to promote reciprocity in the DTP ecosystem. We have identified several methods, which we list below, and will work with Partners to further explore these and other options.

Data Portability Provider Pledge

Providers can work together to create a data portability pledge that requires them to follow best practices on portability. Host Platforms can seek to support providers that commit to the pledge, user interfaces can display providers that commit to the pledge to users, and reports can be published on the state of the ecosystem with regards to reciprocity.

In source code

Contributions into the main Source Code Repository, hosted on GitHub, are encouraged to contain an exporter coupled with each importer. This is to ensure at least an attempt at reciprocity.
Dear developers,

my name is Paul and I’m one of the founders of CamperBoys, a German camper rental company. To complement our rental service, we are developing a mobile travel assistant that provides our customers with detailed information about campsites, restaurants and other POIs.

Together with Stefan Mager, PhD candidate at the Ludwig Maximilian University of Munich, who is researching data portability, we plan to participate in the DTP by adding an integration to our app, which allows our POIs to import their data from different services like Google and Facebook.

To keep up with the latest developments and evaluate the feasibility of our project, it would be great if you could invite us to the dtp-discuss group and to the Slack channel:

- pizzininipaul@gmail.com
- stefan.mager91@goolemail.com

If there are any questions or concerns, please let us know.

Best
Paul
Dear developers,

my name is Paul and I’m one of the founders of CamperBoys, a German camper rental company. To complement our rental service, we are developing a mobile travel assistant that provides our customers with detailed information about campsites, restaurants and other POIs.

Together with Stefan Mager, PhD candidate at the Ludwig Maximilian University of Munich, who is researching data portability, we plan to participate in the DTP by adding an integration to our app, which allows our POIs to import their...