The Person Data Repository evolved from practical needs of the projects at the Berlin-Brandenburg Academy of Sciences and Humanities. It concerned the handling of data about persons that were either original prosopographical data or a byproduct of research. It was agreed upon that this data should be published and made retrievable, not only academy wide. The projects had the main period covered in common, the 19th century, but the material was very divers and the person data differed in depth and type. Therefore, it was thought of how to manage such heterogeneous conditions - the quintessence was to invent a tool or infrastructure that would enable various research projects to enter, store and retrieve their personal data as easily, independent and stable as possible and as well to make this data comparable and retrievable overall projects.

And that is why a few visionaries applied for and where funded since 2009 by the German Research Foundation (DFG). This support will last until 2015. We are integrated into the department TELETA (The Electronic Life of the Academy) and consist of three scholars and currently three students.

Before I explain the actual work of the PDR, let me position the PDR within the field of “prosopography” with the help of two quotes both from this source:


“As such [prosopography] is a system for organizing mostly scarce data in such a way that they acquire additional significance by revealing connections and patterns influencing historical process.” (S. 37, Keats-Rohan).

To achieve such “significance” the data sets can be thought of being “factoids”.

“The factoid was to refer to an assertion about an individual in a source, presented in the source as true. Rather than struggling over terminology and whether these assertions were true […] and therefore ‘facts’, this allowed us to record factoids from the source without establishing their positivist truth. Their connection to reality may be in doubt, but they are ‘historical facts’ by virtue of being contained within a primary source.” (S. 130, Smythe)

For us, this meets the essential idea of our project but we extend and interpret further: To gain insight into “connections and patterns” it requires the possibility to capture data fragmentarily and to relate these fragments to each other. That is the factoid approach and translated in our context as “aspects”. The capturing and relating should be supplemented by the use of a classification to enhance semantic meaning. Plus, naming the sources is mandatory. That is a way to not loose context and semantics. In the end, the researcher decides about trust, relevance and interpretation of the data contained in the repository.

What is it then that we do?

As I already indicated, we are doing research on an appropriate, flexible and alternative way of structuring, storing, and retrieving personal data. This results in an infrastructure that enables us to finally weave a network of prosopographical repositories and build a possibly interdisciplinary research platform.

At this point, I would like to stress, the PDR attempts to be an instrument for work with and on personal data in general. The offered services enable projects to handle their data in their specialized focus, questionnaire and context. Therefore, we do not differentiate between biographical, prosopographical or however focused project.
Let us turn back to the components the PDR builds on: data model, repository and application.

That is how our data model is structured:
Although the focus is theoretically on the person, the “person-object” does not contain any content information but identifiers. Any statement made about the person is written down in so-called “aspects”. The information originates from a source that makes a statement about the person. The source can be qualified in priority. The content of the statement is related to education, biography, family, occupation or any other kind of category. This category semantically classifies the aspect and through this the statement. Within an aspect you can markup fragment parts of the statement like (birth-)place or date, occupation, institution or any other type.
The classification used for categorizing will finally decide about comparability and connectivity. But in general, the projects can use their own classification and don’t have to follow default categories.

Hence, you have a person-object, aspects that point to it and sources proofing scientific trustworthiness. Both aspects and sources can reference to various persons, aspects or sources. By this, redundancy is reduced and a net of semantic information created.
I would like to add, the model is encoded in XML.

But for sure, not only the very detailed approach of structuring pieces of information as “aspects” is central to our work at the PDR. Also, only using the infrastructure of a repository allows the storing and retrieving of data under the same conditions and prerequisites for divers projects.
We are able to ingest a variety of materials by converting them into our data model; we are able to output and exchange data in different formats; we offer publishing services for e. g. web presentation or visualization. Plus, a client software can be used.

The repository, roughly explained, is structured like this:
The PDR-Allies provide connectivity for external applications of any kind and synchronize the access to the database and the SOLR index. Thereby, because of the modular parts of the repository, it enables software developers to control every single aspect of or develop own solutions for the actual processes of data storage and queries.

One example of a connector to our repository is the PDR Interface Tool. Build on php, you can use this access point for extracting or retrieving data from the database. We use it as the connector for the webfrontend. It provides search masks for the different object types and tools for visualization. Here, you see a screenshot of our frontend, that is for sure not yet finished. It is a crucial and difficult topic to develop an adequate website for this very ambitious and complex system.

Finally, we developed, and are in continuous improvement of the client software “Archive-Editor”, that helps you to enter and retrieve your data within your own workspace.
The workflow and the functionalities reflect the data model as they allow you to create person-objects, aspects and reference objects and to markup further down within the aspect's content. There are tools integrated, like a connector to the German National Library to atomically suggest and add the German Authority File (GND) to your data. This tool is called “Concord Service”. Furthermore, two webservices automatically recognize place and time data within a statement and is integrated into the workflow in the Archive-Editor. They are called “PDR Places” and “PDR Dates”.

Needless to say, that all components of the PDR are open source.

That brings me to the conclusion of my presentation: Let me briefly summarize what I have reported to you.

Theory: We aim at being an adequate infrastructure for the handling of person data. We extracted a theory from the prosopographical research but we don’t act to no extend exclusive to any other kind of science dealing with person data.

Flexible Data model: The overall goal of the model is to enable data to be contradictory, fragmentary, arguable or true while still remaining within the research context and, by this, being context-sensitive interpretable. The person-, aspect- and reference-object can reference to each other, the person-object
Modular repository architecture: The repository serves as the central access to the data and can be adapted by any other developer that needs even only parts of the whole infrastructure.

Applications: The Archive-Editor and the other services rank as very innovative and unique applications within the field of prosopographical research. Especially the Archive-Editor gained acknowledgment by e.g. european infrastructure projects as DARIAH or in practical use as in the case of MUSICI, a project focusing on musicians in Italy at the time between 1650 to 1750.

But there is still a lot to be done:
- extending the data model and implementing it: The object type will become generic, therefore, can include various types as event, artifact or place. The essential structure of the XML will not be changed but adapted.
- improving the website: Search options, filtering, presenting and visualizing must be adjusted to the underlying data model and data structure.
- finish connectors for GND and RDF: As getting connected is the overall goal for a repository, we will focus on connectors working for German Authority File (GND) and Resource Description Framework (RDF).
- substantiate cooperation: Already established cooperation will be substantiated and new partners chosen.

Let me finish here now: Thank you for inviting us to your workshop and we are glad to hear, that the PDR arouses interest around the world. We believe that the services and infrastructure that we already set up is meant to be used by other projects even if just partly.

Thank you for listening.