

PP07 Experiences of Managing Quality Registry Data for Effective Exploration

Catharina Dahlbo, Capish, Malmö, Sweden
catharina.dahlbo@capish.com

INTRODUCTION

Swedish national quality registries are used to continually evaluate and improve the quality of Swedish health care. The data in these registries is also an important asset in research.

A collaborative project has been carried out with one of the most comprehensive registries in Sweden, CPUP – Cerebral Palsy Follow-Up Program, where data collection started already in the nineties. Representatives of the quality registry were looking for a tool that could give users access to the data and allow for a more dynamic work process.

Expectations of Exploration

- To find connections and correlations
- To explore dependencies between events and episodes
- To explore the development of one patient over time in relation to others
- To identify patient cohorts and compare them

Issues to Address

- Difficult to achieve data quality
- Data structure did not support the proposed analysis
- Limited ways of exploring data longitudinally
- Data only accessible to a limited number of users
- Multiple stakeholders with different interests

Components of an Effective Exploration

Data can be

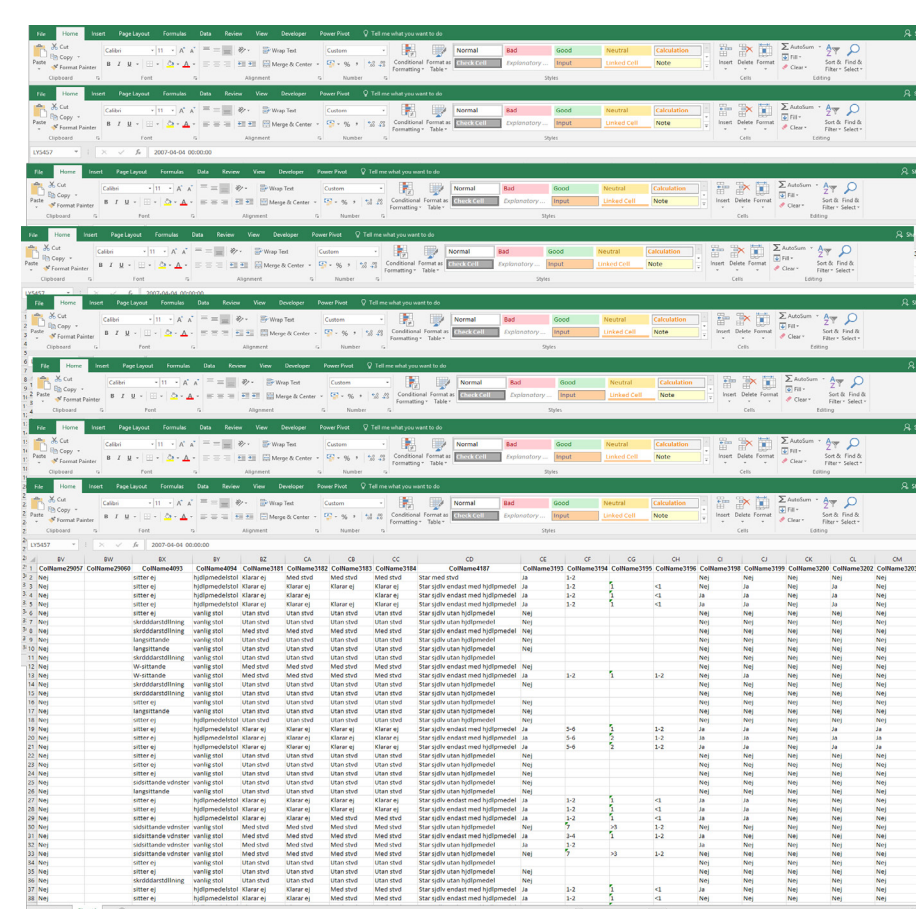
- accessible both cross-sectional and longitudinally
- searched freely using a free-text search
- related in real or relative time
- accessed by many users independent of location

Users can

- explore their own theories
- save and share sessions with others
- get a holistic view of each patient combined with aggregate views of the complete dataset
- easily identify, create and compare cohorts
- easily find and follow an individual patient

DATA COLLECTION

The used data was collected in the CPUP registry during the past 20 years.

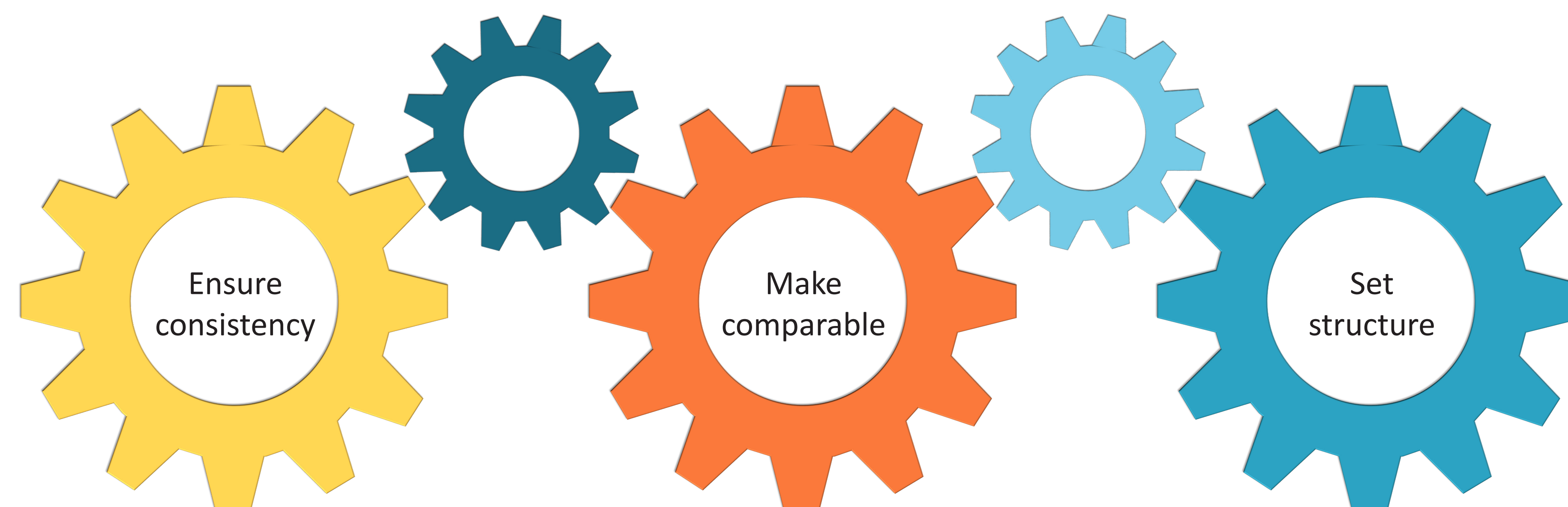


Source Data

Files contained up to 350 columns each.

DATA CURATION

Data curation is a complex process that has great impact on data exploration. Preparation, cleaning, translation, standardization, mapping, and modelling are examples of activities. If done well, the resulting structured data can be used for exploration and analysis of the data in many different contexts, without the need for additional data curation.



Data Quality Control

Corrections were performed on

- Duplicate entries
- Conflicting answers
- Comments in result fields
- Differently spelled values
- Dates

Standardization

- Data was
 - standardized to ISO, ICD10 etc.
 - translated to a common language
 - compared to references
- A uniform terminology was used
- Common measurement units were ensured
- Coded values were mapped to understandable terms

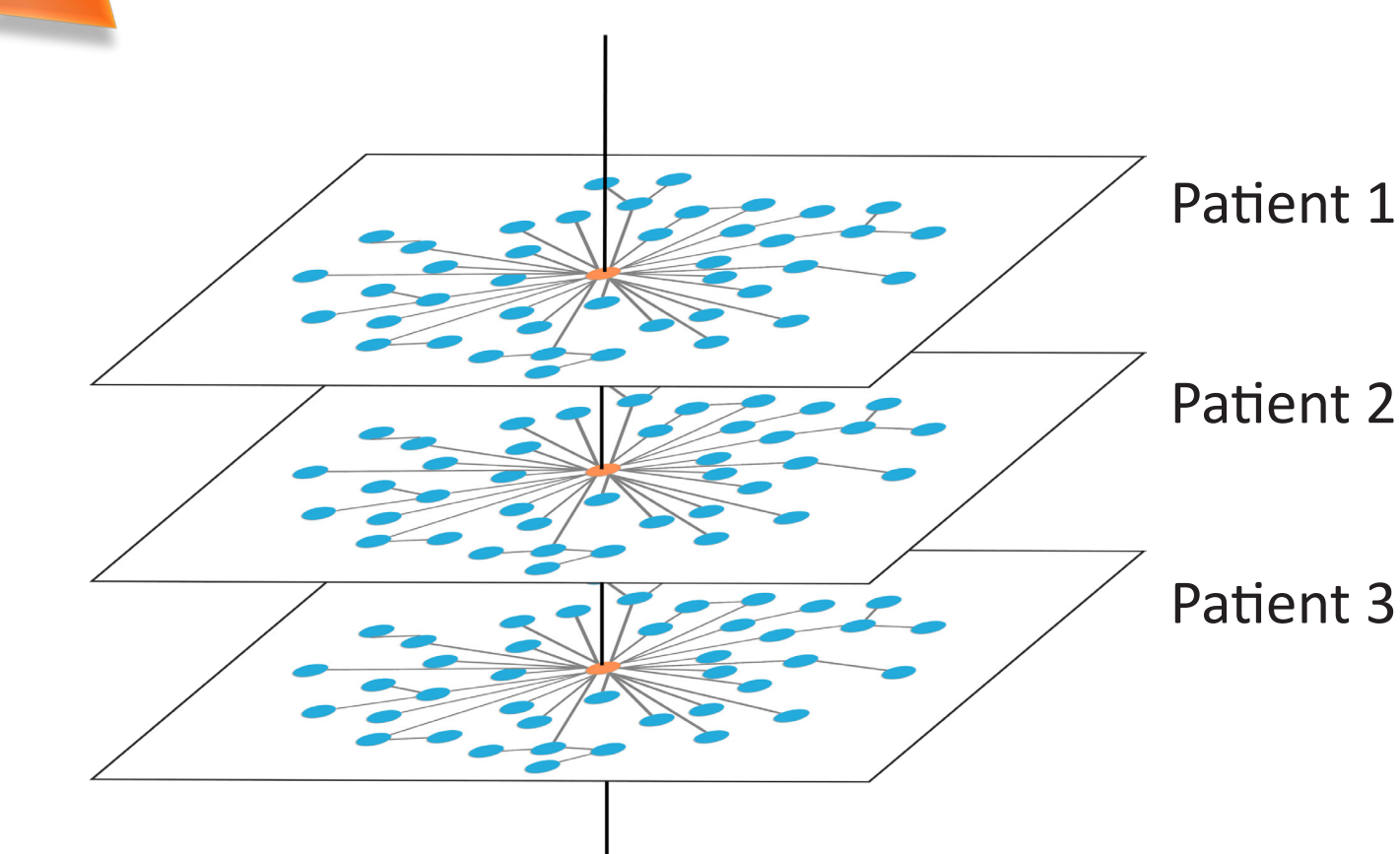
Effective Modeling

- Data was
 - grouped into well-known concepts
 - appended with metadata
- Time was related to events for individual patients
- Episodes were created
- Summaries were created

DATA ANALYSIS

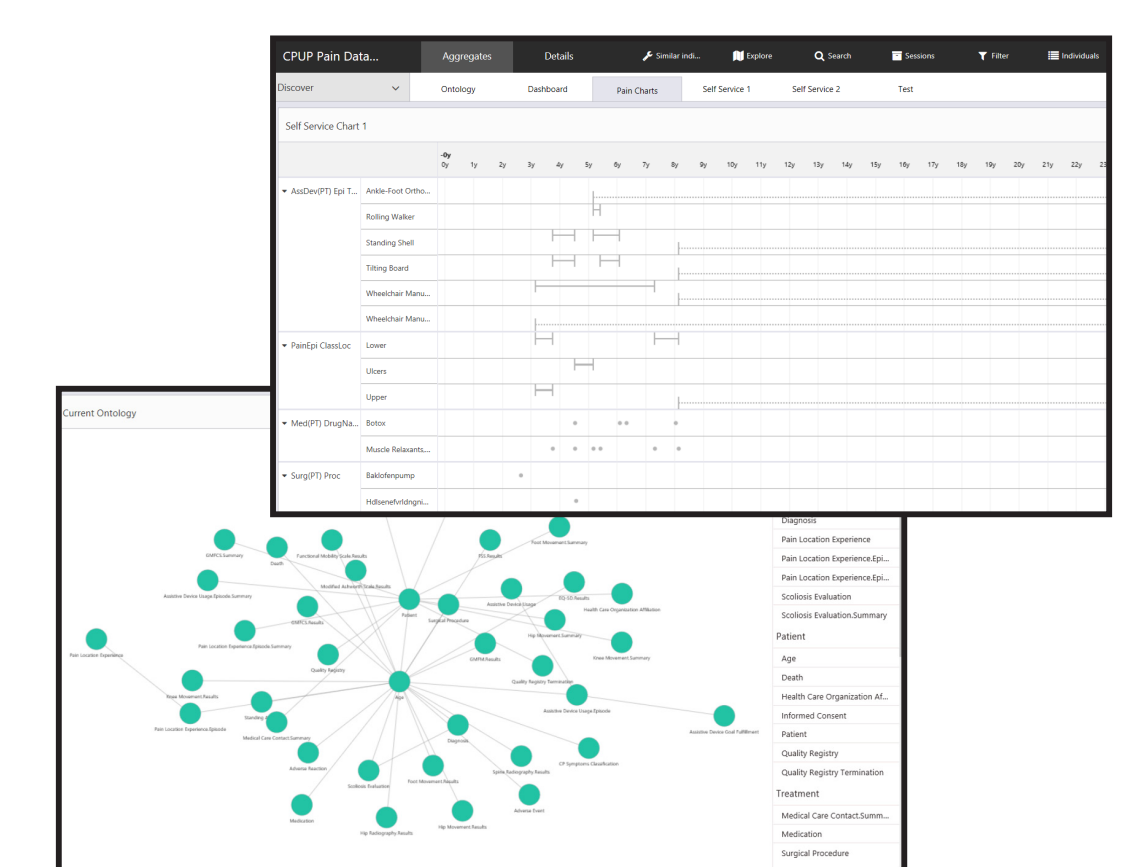
Capish Ontology

Curated data with its metadata was made available in so called 'holons', which were expressed as XML files. The holons and their relations were further described by the Capish Ontology.



Application for Exploration

The curated data was explored in tailor made and intuitive applications, that allow users to move from the entire population to detailed patient information, and vice versa.



CONCLUSION

The source data was not available in a way that facilitated exploration and optimal use of the data. The approach we took was to clean and model the data in a way that enabled a more objective analysis of the data which also improved data quality and consistency. This allowed us to make the most of the data and demonstrate the potential in data curation and the Capish technology. For Capish it was a confirmation that the concept will fill a gap for the health care industry.

Take Control of Data before Exploration!