

Secondary Use of Claims Data from the Austrian Health Insurance System with i2b2: A Pilot Study

Florian Endel¹ Georg Duftschmid²

¹Vienna University of Technology, ASC (florian@endel.at)

²Medical University of Vienna, CeMSIIS

eHealth Vienna, 2016-05-24



Outline

- 1 Background: claims data
- 2 i2b2
- 3 Evaluation
- 4 Conclusion & outlook

Objective

Can claims data research with GAP-DRG
be raised to a new level using i2b2?

background GAP-DRG, a database providing *claims data* from Austrian health care system


i2b2 data warehouse framework for clinical data and research, integrated with...


docker software containers, custom ontologies and test data from GAP-DRG

user feedback by re-enacting cohort extraction on test data

Outline

- 1 Background: claims data
- 2 i2b2
- 3 Evaluation
- 4 Conclusion & outlook

 General Approach for Patient-oriented Ambulant DRGs

 Grundlagenforschung für ambulante
patientenbezogene DRGs

GAP-DRG became a proper name

[gapdrg.endel.at/dokuwiki/doku.php/gapdrg:about]

... inherited from one of the first larger projects

- claims data
- from the Austrian public health- and social insurance system
- governed by:
Main Association of Austrian Social Security Institution (HVB)
- accounting & administrative information
- routinely collected data
- linked data
- secondary data use
- meta-data and registries
- pseudonomized, encrypted, indirectly personalized

2006 & 2007 whole Austrian population

- main source of this project

2008 - 2011 Lower Austria: "GAP-DRG2"

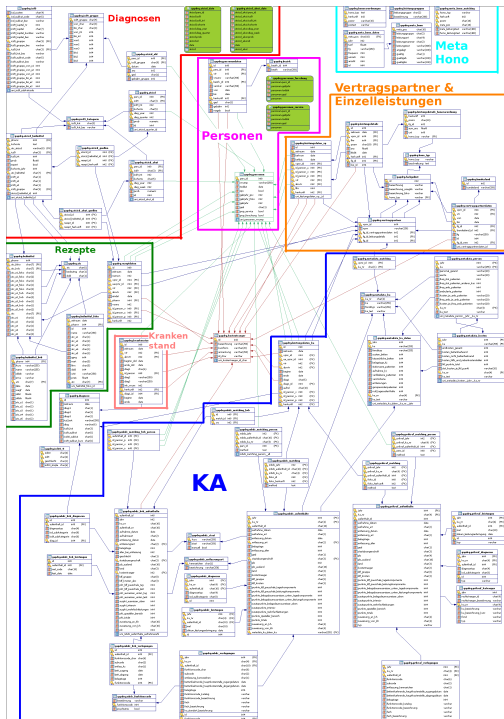
- rich data of good quality
- vastly advanced data model, loading procedures, ...
- *regional restriction*

Outlook GAP-DRG 3

- ongoing effort

- relational data structure
- foreign keys enforcing validity
- normalization
- optimized indexes

- historization guaranteed by constraints (GAP-DRG2)



Claims data in GAP-DRG

Advantages

- maximum flexibility for developers
- manual optimization of database queries
- full data access
- automation, reproducibility, ...

Disadvantages

- good understanding of databases and SQL strictly required
- utilization complicated and error-prone
- much experience necessary
- no granular user rights management
- no abstraction of data model, might change with new version

Outline

- 1 Background: claims data
- 2 i2b2
- 3 Evaluation
- 4 Conclusion & outlook

- Informatics for Integrating **B**iology and the **B**edside
 - National Center for Biomedical Computing in Boston, MA
- "clinical data informatics framework", data warehousing application
- for clinical data, reuse of EHR records and patient centered data
- open source, extendable, modularized
- various user interfaces, API (SOAP, REST)
- role based security model
- SHRINE (Data Sharing Network)

- ⇒ hard to deploy, complex data integration & ontology

Implementation & deployment with docker

- automatic, self-contained build using software containers^a
- completely reproducible process
- container-based deployment: no prerequisites, evaluated successfully
- various versions, simple upgrade, regression tests
- still only available in development environment...

^aall scripts available on GitHub: github.com/FlorianEndel/i2b2-Docker

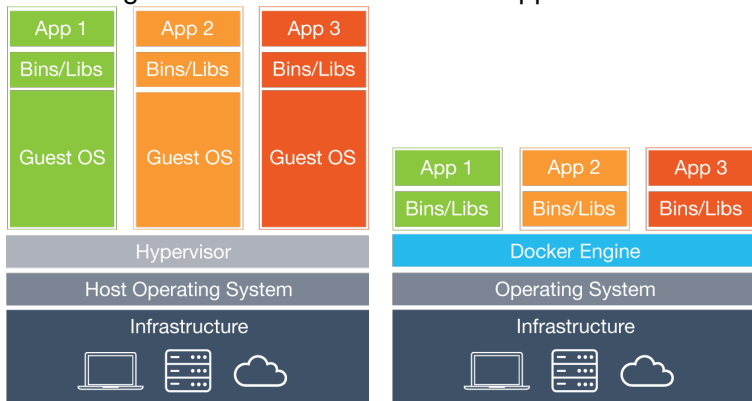
Data & ontology

- randomly distorted test data from GAP-DRG
- custom ontology (> 23.000 items)

Software containers: docker



- something between virtual machine & "app"



- flexible, low resource consumption, independent
- open source, cutting edge, broad adoption

[www.docker.com/what-docker]

Outline

- 1 Background: claims data
- 2 i2b2
- 3 Evaluation**
- 4 Conclusion & outlook

Concept

- **get early feedback by observing users**
- to assess utility, potential usage scenarios, focus group(s)
- reproduce cohorts similar to a real-life project¹
- with the i2b2 web-interface using randomized test-data
- test with various user types
- very few have access and skills
- compliance of users and available resources

Procedure

- a short introduction is given
- short tasks have to be carried out while details are recorded
- feedback is gathered with a concluding interview

¹Endel, F. et al. Reusing claims data to assess parenthood as risk factor for myocardial infarction. *Studies in Health Technology and Informatics* 210, 979-979 (2015).

Implementation

- protocol and questionnaire
 - personal information
 - especially knowledge about GAP-DRG, computer systems
- introduction to the system and tasks
 - only rough demonstration of user interface
 - without explanation of content, security system, ...
- information extracted from observation & interview
 - evaluation of completeness, correctness
 - number of situations stuck / help needed for progression
 - does the situation improve with the second challenge?
 - what are the obstacles?

Extracted results

- tested with 6 people: 2 of each group
- general positive feedback

- even untrained create cohorts intuitively
- system experts dive right into exploration / idea generation
 - new ideas vs. compliance with protocol
- database experts save time but miss flexibility

- feedback for i2b2's user interface (and stability)

Outline

- 1 Background: claims data
- 2 i2b2
- 3 Evaluation
- 4 Conclusion & outlook**

Conclusion & outlook

Conclusion

- i2b2: a viable solution for claims data research with GAP-DRG
- install & deploy with docker: advantages & overcoming obstacles
- targeted users: system experts

Outlook

- deployment in GAP-DRG infrastructure
 - 2016, summer: new server including *docker*
- i2b2 images based on i2b2's git-repository

Ambition

- limited access for cohort exploration
- integrate claims data and EHR with SHRINE

Secondary Use of Claims Data from the Austrian Health Insurance System with i2b2: A Pilot Study

Florian Endel¹ Georg Duftschmid²

¹Vienna University of Technology, ASC (florian@endel.at)

²Medical University of Vienna, CeMSIIS

eHealth Vienna, 2016-05-24



Appendix: GAP-DRG Restrictions

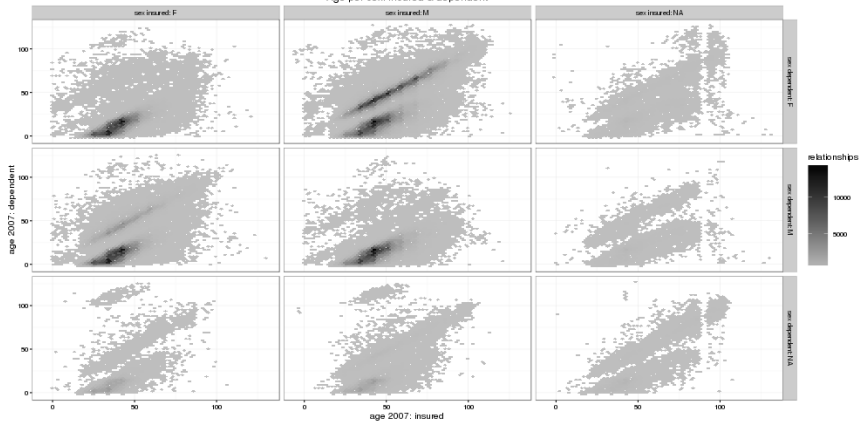
not included

- "clinical data", *omics, health status
- payment details (costs in €)
- personalized information, identifiable natural/legal entities
- social services (e.g. unemployment)
- most information not relevant for payment, e.g.
 - ambulatory unit of hospitals
 - "private" payments
- diagnoses from ambulatory contacts
- **genealogical** information

Appendix: Genealogical information

- relationship between patients
 - spouses (traditionally married couples)
 - parents and children
- information is not (currently) coded directly
- retrieved from **co-insurance** in patients' master-data
- complex algorithm
 - with disclaimers, assumptions, imprecision
 - based on specifics of the Austrian insurance system

Age per sex: insured & dependent



Appendix: Evaluation background

myocardial infarctions in young parents

- does young parenthood alter the risk for myocardial infarction?
Endel, F., Sauter, S., Koller, L., Niessner, A. & Duftschmid, G. Reusing claims data to assess parenthood as risk factor for myocardial infarction. *Studies in Health Technology and Informatics* 210, 979-979 (2015).
- cohorts retrieved by manually developed database queries

Genealogical information

- relationship between patients (spouses, child/parent)
- information is not (currently) coded directly
- retrieved from co-insurance in patients' master-data
- integrated into i2b2's ontology manually