Supporting Knowledge Discovery in Medicine

Dominic GIRARDI[a], Klaus ARTHOFER[b]

[a]RISC Software GmbH – Research Unit Medical Informatics

[b]Upper Austria University of Applied Sciences – School of Management, Austria

dominic.girardi@risc.uni-linz.ac.at ● klaus.arthofer@fh-steyr.at

Introduction

- Knowledge discovery in medicine requires the deep involvement of medical domain experts [1]
- Data handling, processing, and analysis is known to be a major technical obstacle to scientific (bio-) medical research projects [2].
- Combination of IT knowledge and domain knowledge
- Limited funding → lack of IT support

[1]K. J. Cios, G. W. Moore. **Uniqueness of medical data mining**. Artificial intelligence in medicine, 26(1):1{24}, 2002.
[2]N. Anderson, E. Lee, J. Brockenbrough, M. Minie, S. Fuller, J. Brinkley und P. Tarczy-Hornoch, **Issues in biomedical research data management and analysis: Needs and barriers**," Journal of the American Medical Informatics Association, pp. 478 - 488, 2007

Method

CALUMMA [3][4] is an Ontology-based

- data definition-
- data acquisition-
- data preparation-
- data validation-
- data exploration-

infrastructure for data-intense, scientific research performed by domain experts with a minimum IT expert support.

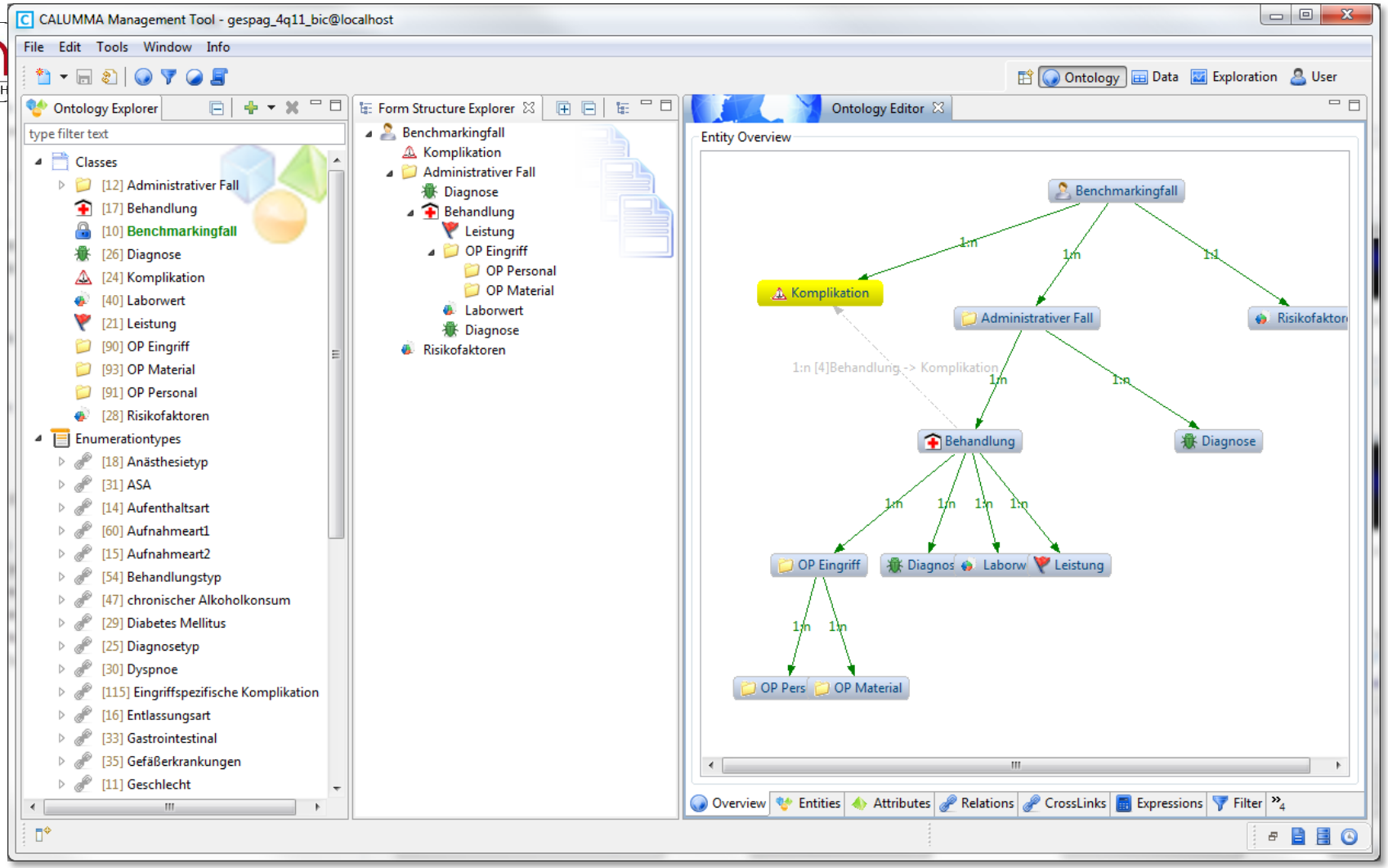
[3] Arthofer K., Girardi, D., Giretzlehner M. (2012): **Ein Ontologiebasiertes System zum Extrahieren, Transformieren und Laden von Daten in Krankenanstalten**. Proceedings of eHealth2012, May 2012, Vienna Austria

[4] Girardi, D.; Arthofer, K.; Giretzlehner, M. (2012) **An Ontology-Based Data Acquisition Infrastructure**. Proceedings of 4th International Conference on Knowledge Engineering and Ontology Development, pages 155-160, DOI: 10.5220/0004108101550160, October 2012, Barcelona Spain

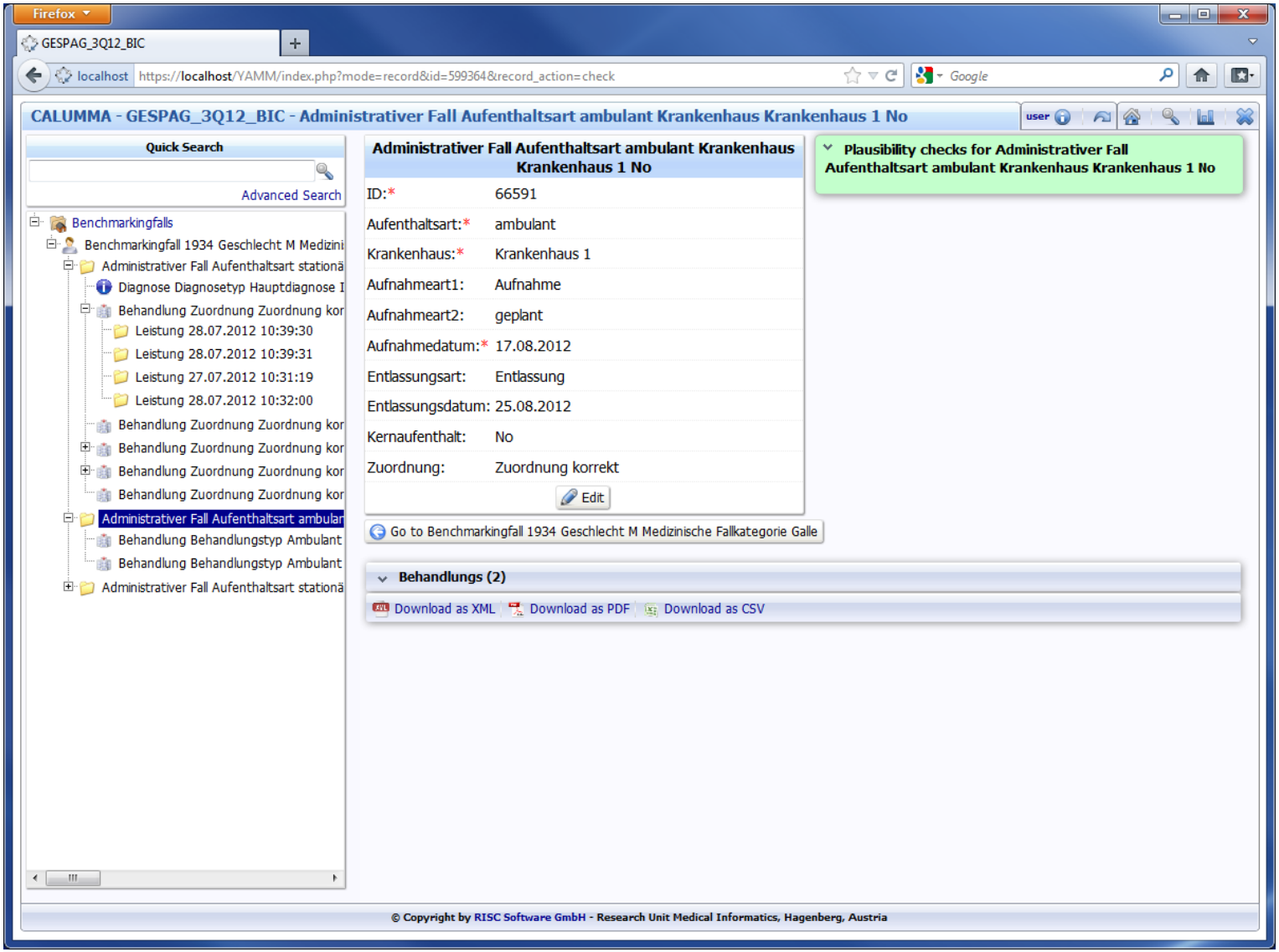
Method

- Ontology-centred research data infrastructure
- Central Ontology is defined by the domain expert, and contains:
  - Data elements of interest and their attributes
  - Data relations
  - Validity rules
- The remainder is **automatically created** at runtime
  - Web interface (input forms, overview tables, search masks,…)
  - ETL interface (import of electronically available data)
- Changes to the ontology have immediate effects the whole infrastructure immediately

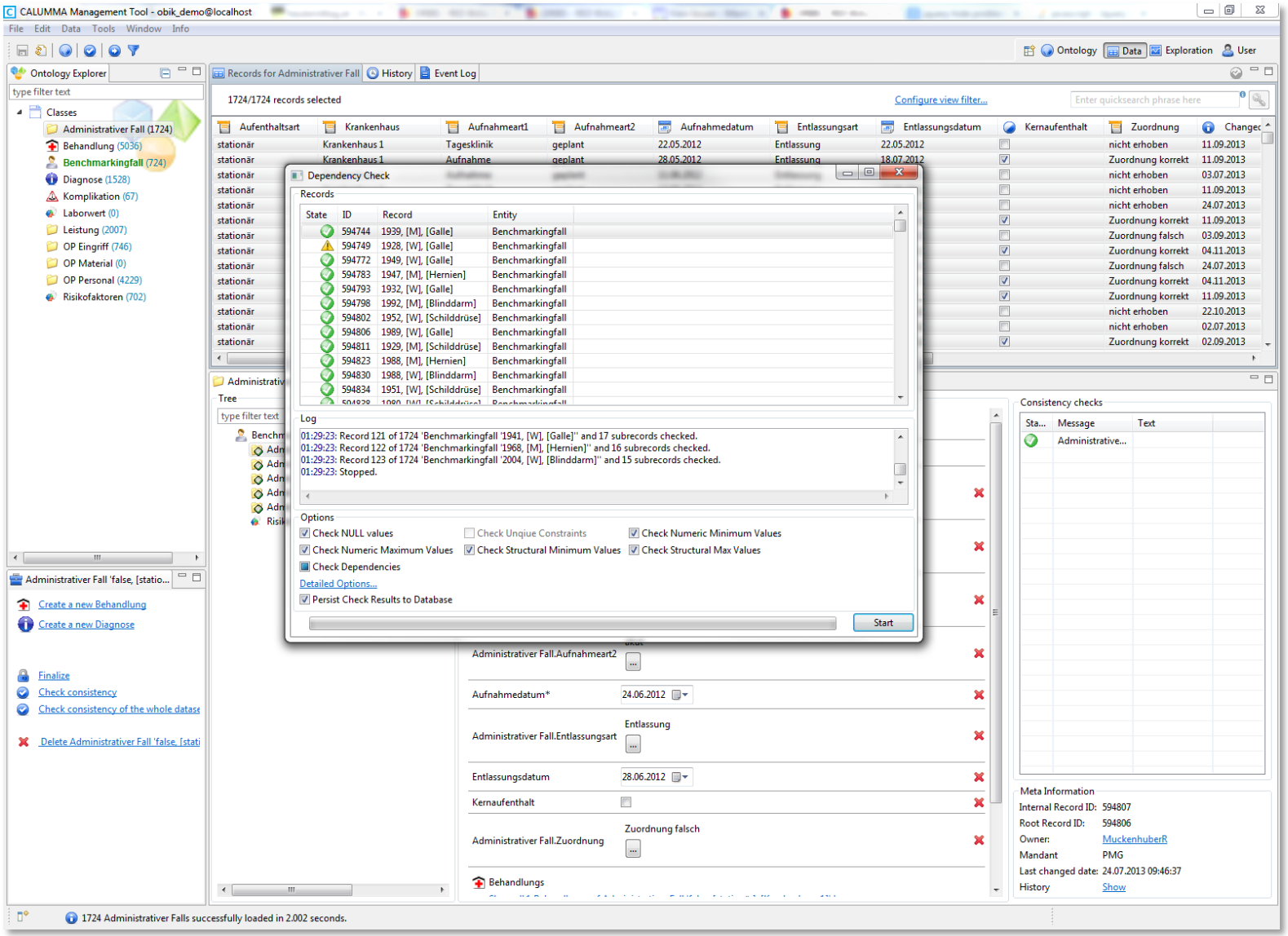
Ontology Editor



Web Interface



Data Validation

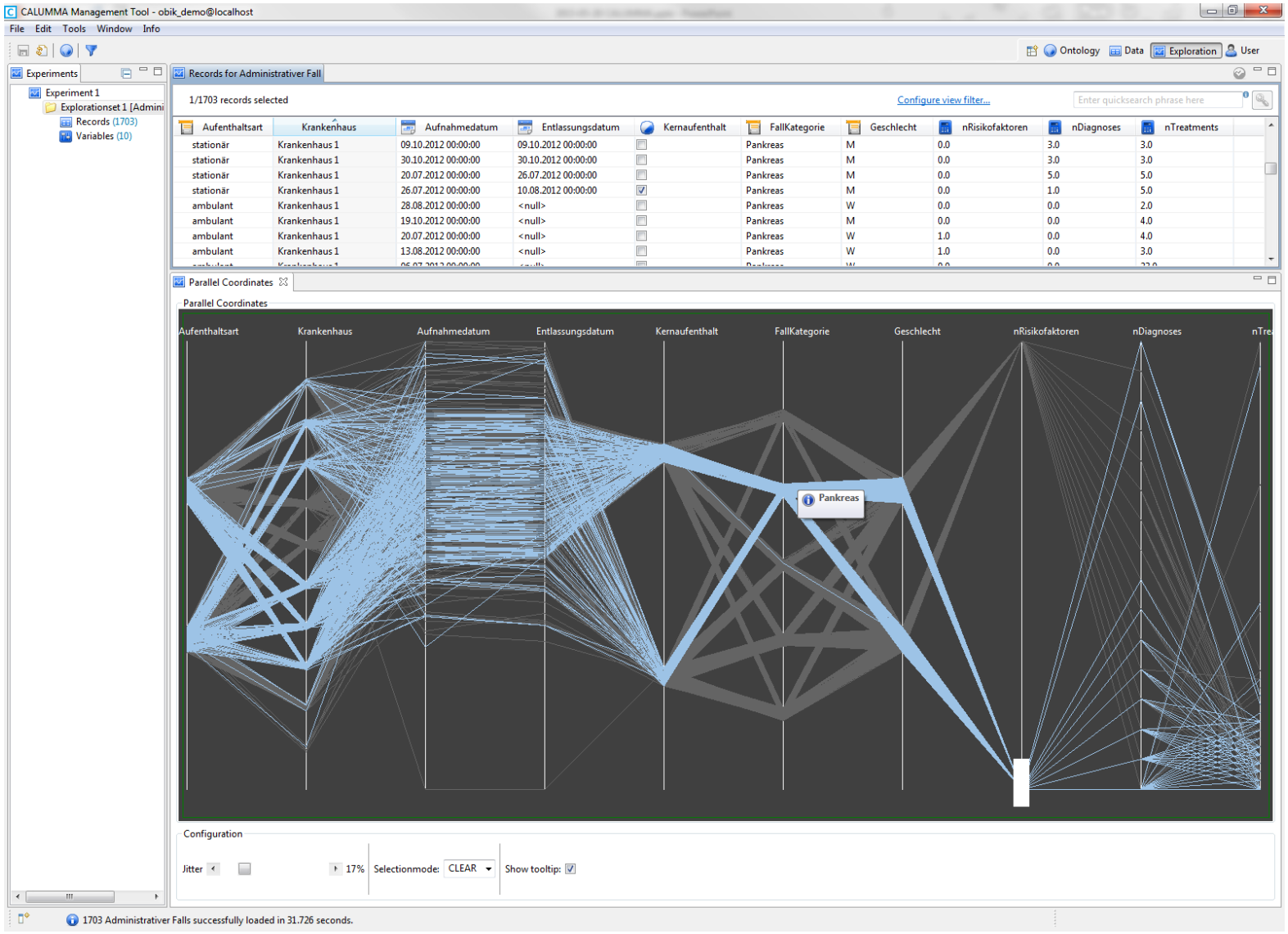


Ontology guided Knowledge Discovery

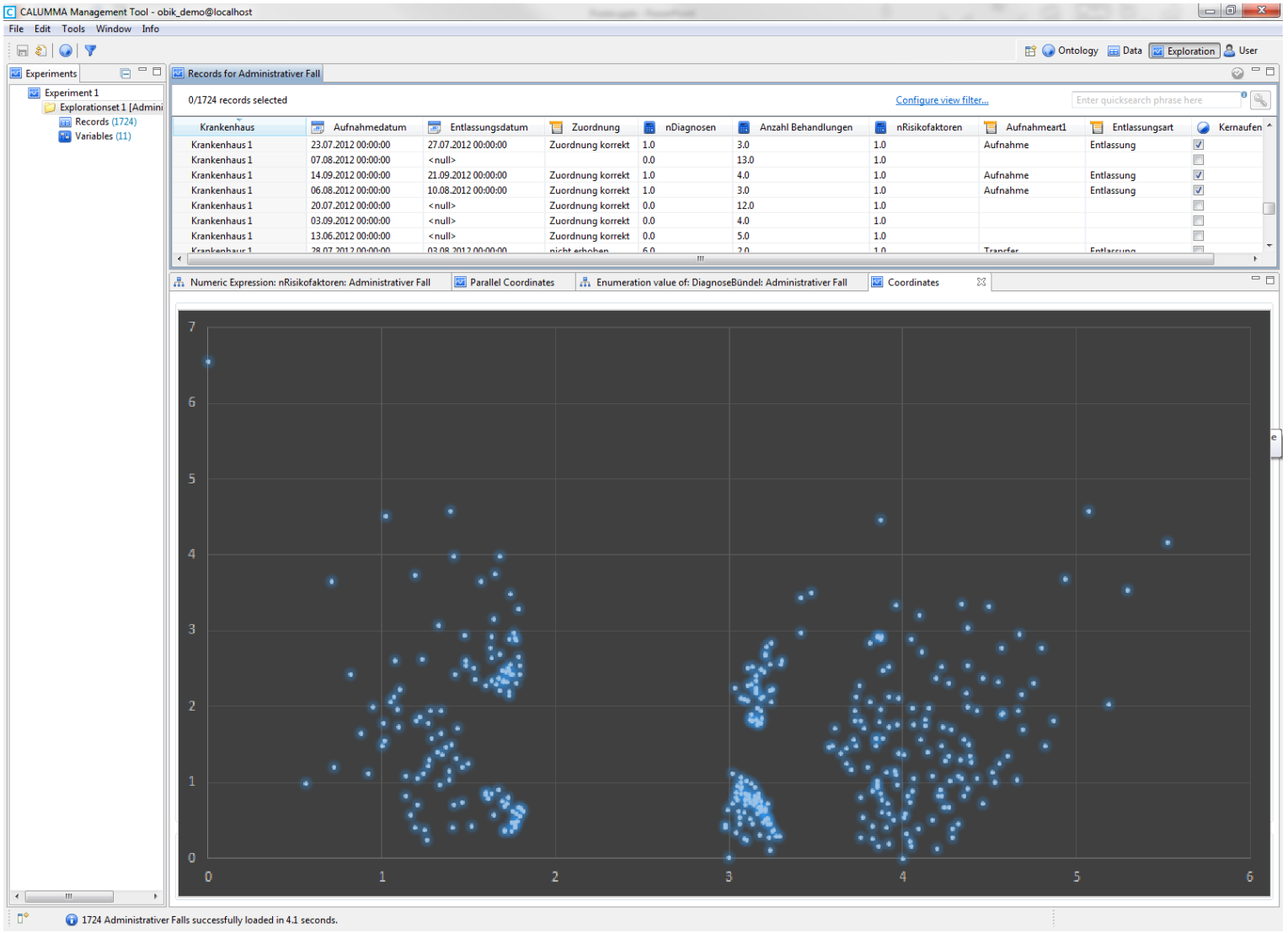
- Technical barriers:
  - Data pre-processing,
  - transformation,
  - normalization
- → automatize with ontology meta-information
- Integrate well-selected set of analysis and visualization methods for exploratory data analysis
- One-click from data to mining

- Feasibility test: Parallel Coordinate Visualisation

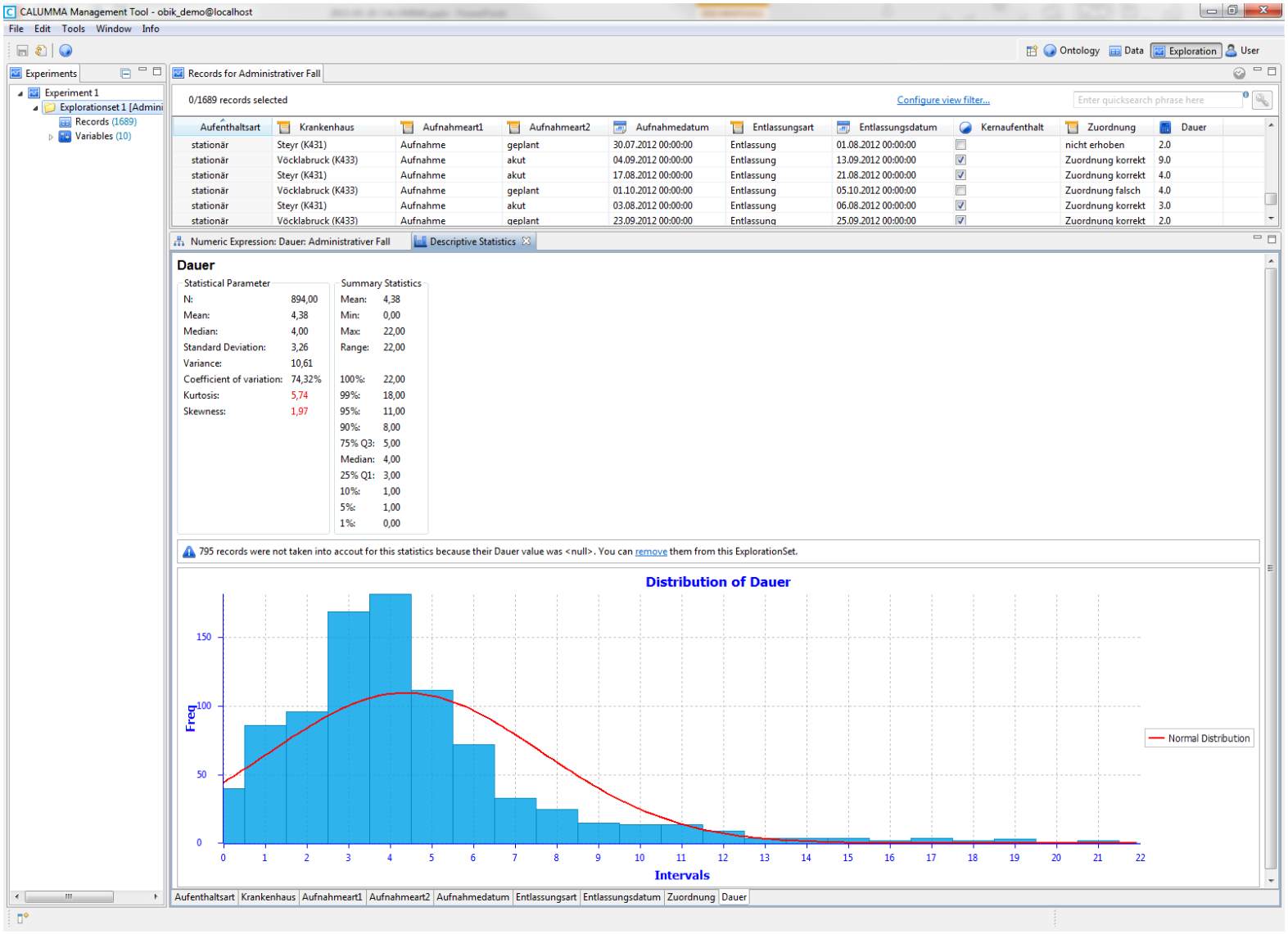
Results


CALUMMA Management Tool - obik_demo@localhost
File Edit Tools Window Info
Ontology Data Exploration User
Experiments
Experiment 1
Explorationset 1 [Admini
Records (1703)
Variables (10)
Records for Administrativer Fall
1/1703 records selected
Configure view filter...
Enter quicksearch phrase here
Aufenthaltsart Krankenhaus Aufnahmedatum Entlassungsdatum Kernaufenthalt FallKategorie Geschlecht nRisikofaktoren nDiagnoses nTreatments
stationär Krankenhaus 1 09.10.2012 00:00:00 09.10.2012 00:00:00 Pankreas M 0.0 3.0 3.0
stationär Krankenhaus 1 30.10.2012 00:00:00 30.10.2012 00:00:00 Pankreas M 0.0 3.0 3.0
stationär Krankenhaus 1 20.07.2012 00:00:00 26.07.2012 00:00:00 Pankreas M 0.0 5.0 5.0
stationär Krankenhaus 1 26.07.2012 00:00:00 10.08.2012 00:00:00 Pankreas M 0.0 1.0 5.0
ambulant Krankenhaus 1 28.08.2012 00:00:00 <null> Pankreas W 0.0 0.0 2.0
ambulant Krankenhaus 1 19.10.2012 00:00:00 <null> Pankreas M 0.0 0.0 4.0
ambulant Krankenhaus 1 20.07.2012 00:00:00 <null> Pankreas W 1.0 0.0 4.0
ambulant Krankenhaus 1 13.08.2012 00:00:00 <null> Pankreas W 1.0 0.0 3.0
Parallel Coordinates
Parallel Coordinates
Aufenthaltsart Krankenhaus Aufnahmedatum Entlassungsdatum Kernaufenthalt FallKategorie Geschlecht nRisikofaktoren nDiagnoses
Pankreas
Configuration
Jitter 17% Selectionmode: CLEAR Show tooltip:
1703 Administrativer Falls successfully loaded in 31.726 seconds.

Results


CALUMMA Management Tool - obik_demo@localhost
File Edit Tools Window Info
Ontology Data Exploration User
Experiments
Experiment 1
Explorationset 1 [Admini
Records (1724)
Variables (11)
Records for Administrativer Fall
0/1724 records selected
Configure view filter...
Enter quicksearch phrase here
Krankenhaus | Aufnahmedatum | Entlassungsdatum | Zuordnung | nDiagnosen | Anzahl Behandlungen | nRisikofaktoren | Aufnahmeart1 | Entlassungsart | Kernaufen
Krankenhaus 1 | 23.07.2012 00:00:00 | 27.07.2012 00:00:00 | Zuordnung korrekt | 1.0 | 3.0 | 1.0 | Aufnahme | Entlassung
Krankenhaus 1 | 07.08.2012 00:00:00 | <null> | | 0.0 | 13.0 | 1.0
Krankenhaus 1 | 14.09.2012 00:00:00 | 21.09.2012 00:00:00 | Zuordnung korrekt | 1.0 | 4.0 | 1.0 | Aufnahme | Entlassung
Krankenhaus 1 | 06.08.2012 00:00:00 | 10.08.2012 00:00:00 | Zuordnung korrekt | 1.0 | 3.0 | 1.0 | Aufnahme | Entlassung
Krankenhaus 1 | 20.07.2012 00:00:00 | <null> | Zuordnung korrekt | 0.0 | 12.0 | 1.0
Krankenhaus 1 | 03.09.2012 00:00:00 | <null> | Zuordnung korrekt | 0.0 | 4.0 | 1.0
Krankenhaus 1 | 13.06.2012 00:00:00 | <null> | Zuordnung korrekt | 0.0 | 5.0 | 1.0
Numeric Expression: nRisikofaktoren: Administrativer Fall
Parallel Coordinates
Enumeration value of: DiagnoseBündel: Administrativer Fall
Coordinates
0 1 2 3 4 5 6 7
1724 Administrativer Falls successfully loaded in 4.1 seconds.

Results


CALUMMA Management Tool - obik_demo@localhost
File Edit Tools Window Info
Ontology Data Exploration User
Experiments
Experiment 1
Explorationset 1 [Admini
Records (1689)
Variables (10)
Records for Administrativer Fall
0/1689 records selected
Configure view filter...
Enter quicksearch phrase here
Aufenthaltsart Krankenhaus Aufnahmeart1 Aufnahmeart2 Aufnahmedatum Entlassungsart Entlassungsdatum Kernaufenthalt Zuordnung Dauer
stationär Steyr (K431) Aufnahme geplant 30.07.2012 00:00:00 Entlassung 01.08.2012 00:00:00 nicht erhoben 2.0
stationär Vöcklabruck (K433) Aufnahme akut 04.09.2012 00:00:00 Entlassung 13.09.2012 00:00:00 Zuordnung korrekt 9.0
stationär Steyr (K431) Aufnahme akut 17.08.2012 00:00:00 Entlassung 21.08.2012 00:00:00 Zuordnung korrekt 4.0
stationär Vöcklabruck (K433) Aufnahme geplant 01.10.2012 00:00:00 Entlassung 05.10.2012 00:00:00 Zuordnung falsch 4.0
stationär Steyr (K431) Aufnahme akut 03.08.2012 00:00:00 Entlassung 06.08.2012 00:00:00 Zuordnung korrekt 3.0
stationär Vöcklabruck (K433) Aufnahme geplant 23.09.2012 00:00:00 Entlassung 25.09.2012 00:00:00 Zuordnung korrekt 2.0
Numeric Expression: Dauer: Administrativer Fall
Descriptive Statistics
Dauer
Statistical Parameter
N: 894,00
Mean: 4,38
Median: 4,00
Standard Deviation: 3,26
Variance: 10,61
Coefficient of variation: 74,32%
Kurtosis: 5,74
Skewness: 1,97
Summary Statistics
Mean: 4,38
Min: 0,00
Max: 22,00
Range: 22,00
100%: 22,00
99%: 18,00
95%: 11,00
90%: 8,00
75% Q3: 5,00
Median: 4,00
25% Q1: 3,00
10%: 1,00
5%: 1,00
1%: 0,00
795 records were not taken into accout for this statistics because their Dauer value was <null>. You can remove them from this ExplorationSet.
Distribution of Dauer
Freq
Intervals
Normal Distribution
Aufenthaltsart Krankenhaus Aufnahmeart1 Aufnahmeart2 Aufnahmedatum Entlassungsart Entlassungsdatum Zuordnung Dauer

Conclusion & Outlook

- Technically feasible
- Little awareness in medical community about non-standard visualization and analysis methods
- Reduce reservations with usability

- Further Research & Challenges
  - Distance based methods
  - Distance calculation
  - Parameterization of algorithms
  - Performance with bigger data volumes (> 100000)