

The logo for entimo, featuring the word "entimo" in white lowercase letters on a solid green rectangular background.

Behind the scene: Technical aspects and challenges of metadata extraction and clinical data management

Ronald Steinhau
Entimo AG - Berlin/Germany

Content

- Motivation | Industry Challenges
- Use Cases
- Technical Challenges
- Status
- Outlook

Motivation

- Industry Challenges:
 - Mapping of legacy studies
 - Wave of mergers and acquisitions
 - Evolution of standards
- CDISC Share Pilot Report, Conclusions:
 - Manual terminology harmonization possible for short lists
 - Technological support required for search, viewing and comparison of numerous, long lists

Project Setup

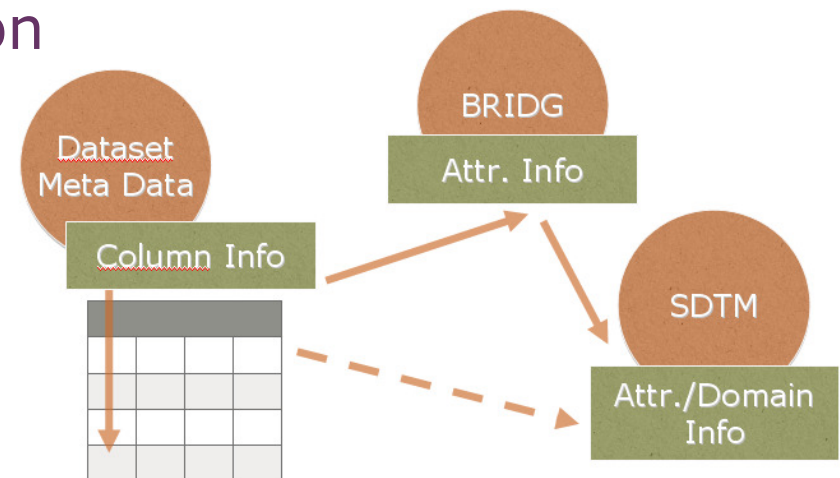
- Research Project
 - Funded by the German Federal Ministry of Economy and Technology
- High Level Goals:
 - Investigate mechanisms for recognition and extraction of metadata
 - Store and explore metadata in contexts
 - Re-use information from comparable studies
 - Include into analysis different metadata **and** data
 - Provide support in mapping creation and issue resolution (full automation is illusionary)

Use Cases (1)

- Harmonize codelists
 - Compare lists, codes **and** their parts
 - Visualize differences | matching degree
 - Provide help for codelist merge
- Reuse mapping sequences
 - Search for existing mapping algorithms based on metadata, links and data profiling results
 - Display matching degree and issues
 - Suggest issue resolution
 - Improve matching with growing information base (adaptive learning)

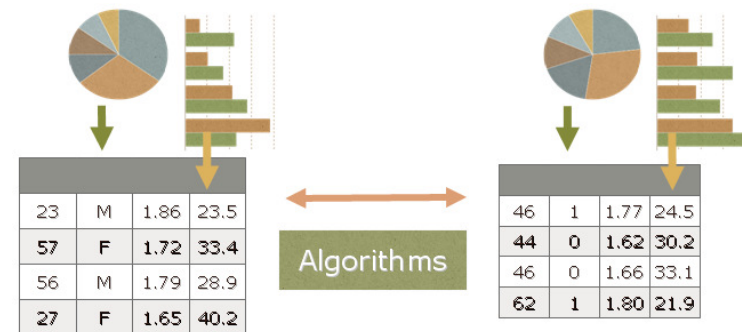
Use Cases (2)

- Explore interconnected metadata systems:
 - Support definition and storage of relations between metadata systems
 - Allow easy navigation between contexts
 - BRIDG, SDTM, company standards,...
 - (Re-)use link information for mapping sequence generation



Use Cases (3)

- Compare datasets
 - Use statistical/profiling information (data approach)
 - Find similar datasets (metadata driven approach)
 - Possibility to exclude certain data from comparison
 - Detect suitable mappings
 - Adjust matching levels



Technical Challenges: Metadata storage and cross linking

- Flexible importers/exporters required:
 - Different input formats (XMI,EAP,XML,Excel,SAS...)
 - Multiple metadata standards with heterog. structures (BRIDG, HL7, SDTM, ADaM...)
 - Different datatype systems
- Unified storage helpful (based on OMG):
 - Easy links between different models
 - Unified search and navigation
 - Simpler reuse of metadata definitions
 - Easier mappings between datatype system

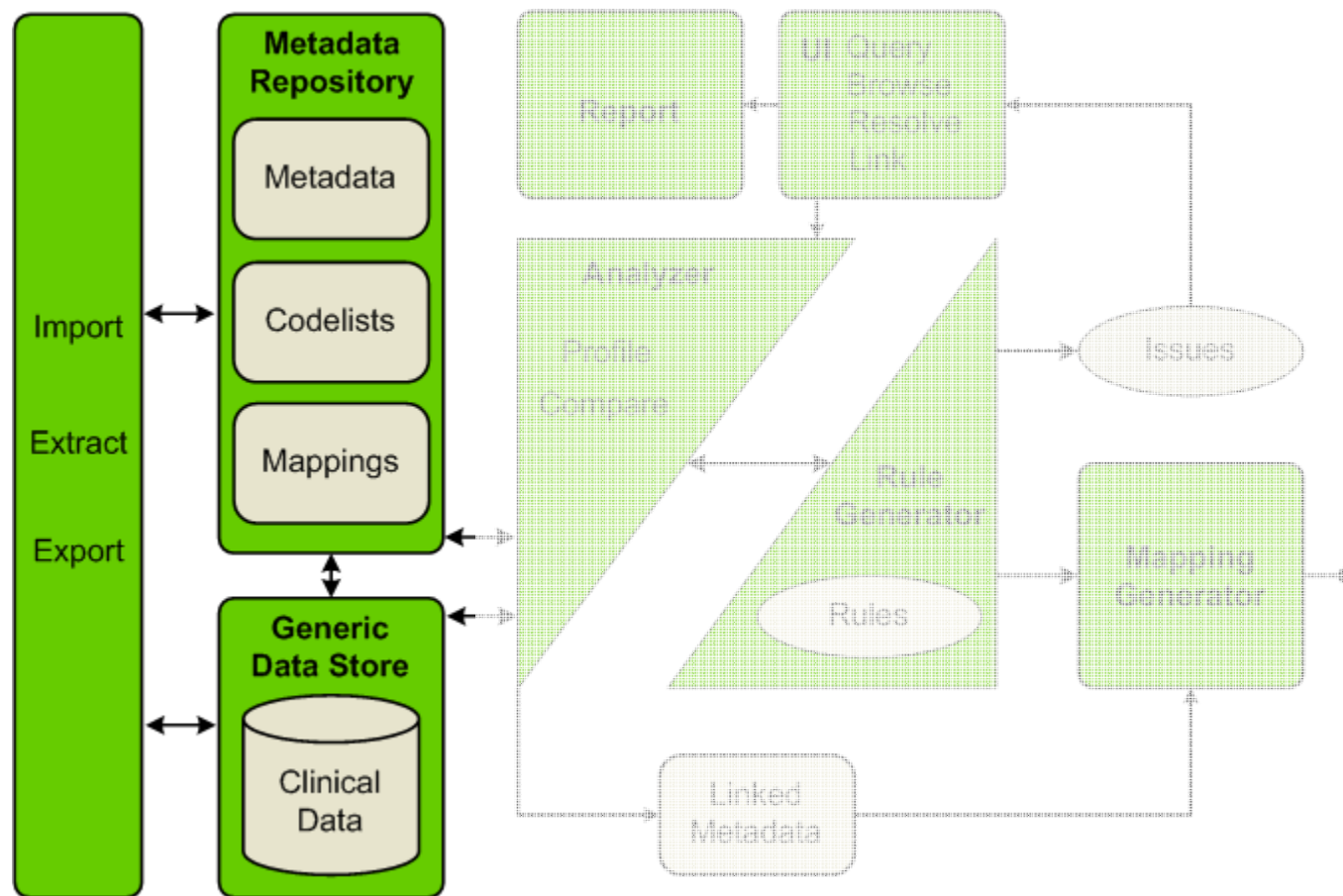
Technical Challenges: Clinical Data Repository

- Generic storage necessary:
 - Any dataset shall fit!
 - No redundancy
 - Links to metadata repository/contexts
- Performance requirements:
 - Provide fast import and access
 - Support large data volumes
- Pharma specific requirements:
 - Fulfill 21 CFR Part 11 (e.g. audit trail)

Technical Challenges: Exploration & Comparison Reports

- Metadata Navigation
 - Metadata of a selected data entity represented in one or more metadata contexts
 - Relation of different metadata types (code lists, domain definitions, mappings...)
 - Access to data based on metadata and vice versa
- Comparison & Reports
 - Interactive report (showing data and/or metadata)
 - Active report elements (drill-down and navigation)

Workflow | Architecture



Status

- Technical infrastructure in place:
 - High-performance clinical data repository (data warehouse)
 - Unified metadata repository
- Implemented modules:
 - Metadata and raw data importers/extractors
 - Profiling/statistical analysis module
 - Flexible domain specific language (DSL)
 - Metadata navigator
 - Interactive reports with drill-down/-up capability and hyperlink based navigation

Outlook & Conclusion

- Working on improvements:
 - UI user-friendliness
 - End integration
 - Large-scale “field tests” with customer test data
- Conclusion
 - Generic, fast repository for linked metadata and data combined with exploration, analysis and visualization means are able to reduce complexity and improve productivity

entimo

Many thanks for your attention!

Questions?

For more information or cooperation,
contact st@entimo.de