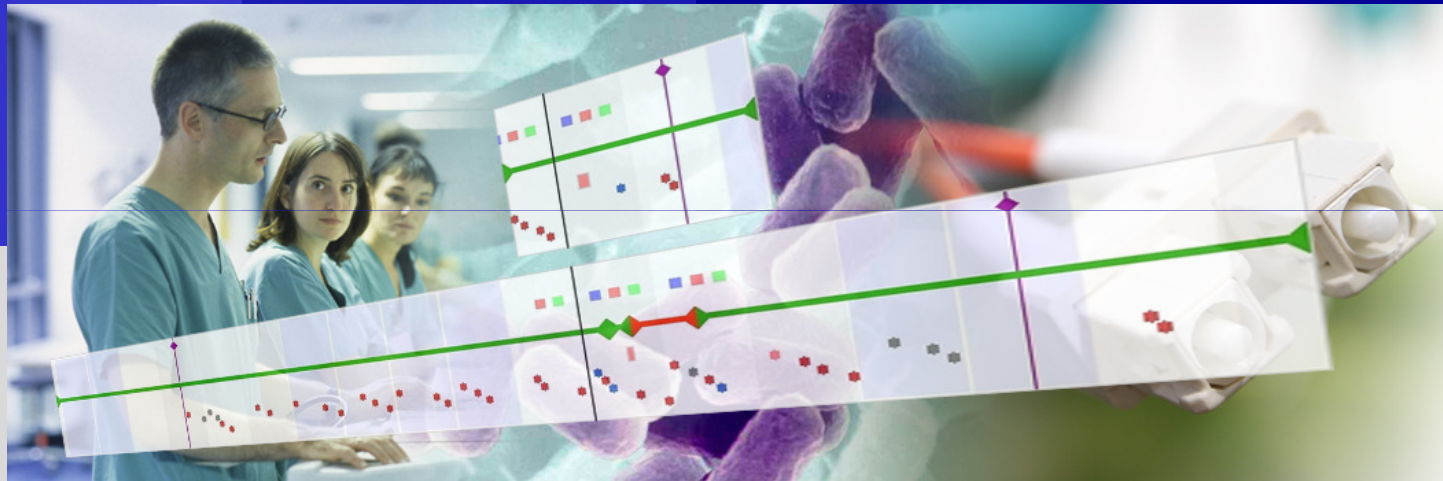


Implementing a Research Platform at Erlangen University Hospital

1st European i2b2 Workshop, 25-26.03.2013, Erlangen



Ganslandt T¹, Mate SN², Neue A², Engel I², Prokosch HU^{1,2}

¹ Medical Center for Information & Communication, Erlangen University Hospital

² Chair of Medical Informatics, Friedrich-Alexander-University Erlangen

Introduction

Erlangen University Hospital

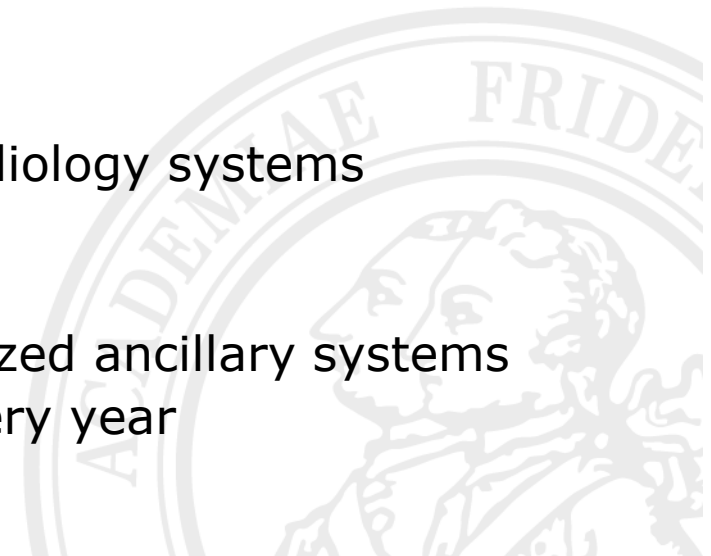


■ Academic hospital

- 1.350 bed maximum care facility
- 49 departments covering all areas of medicine
- 61.000+ inpatient and 304.000+ outpatient cases in 2012

■ Clinical IT Powerhouse

- 120+ IT staff
- enterprise-wide EHR, billing, lab and radiology systems
- 3 surgery and anesthesia systems
- 3 pharmacy and transfusion systems
- (unfortunately) many additional specialized ancillary systems
- => millions of datapoints generated every year



Introduction

Research Infrastructure Group

■ Mission

- provide platforms for the acquisition and management of study and biospecimen data
- integrate routine clinical data for secondary use
- provide access to combined data for research, administrative and strategic applications



Electronic Data Capture
(secuTrial)



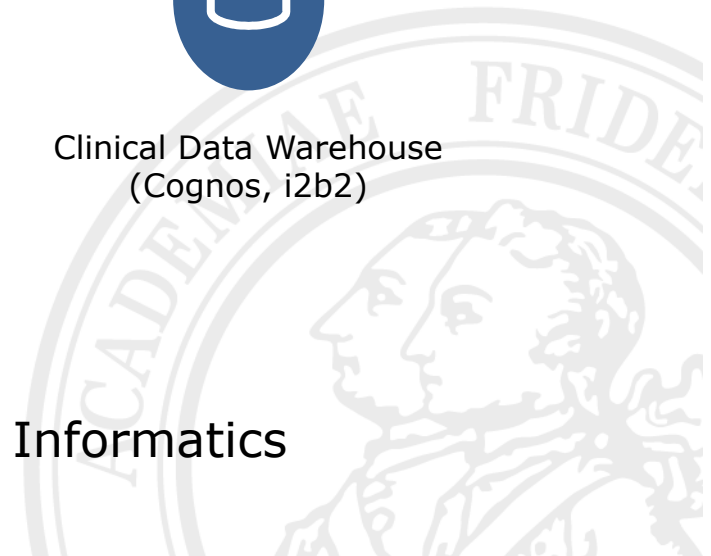
Biobank Management
(Starlims)



Clinical Data Warehouse
(Cognos, i2b2)

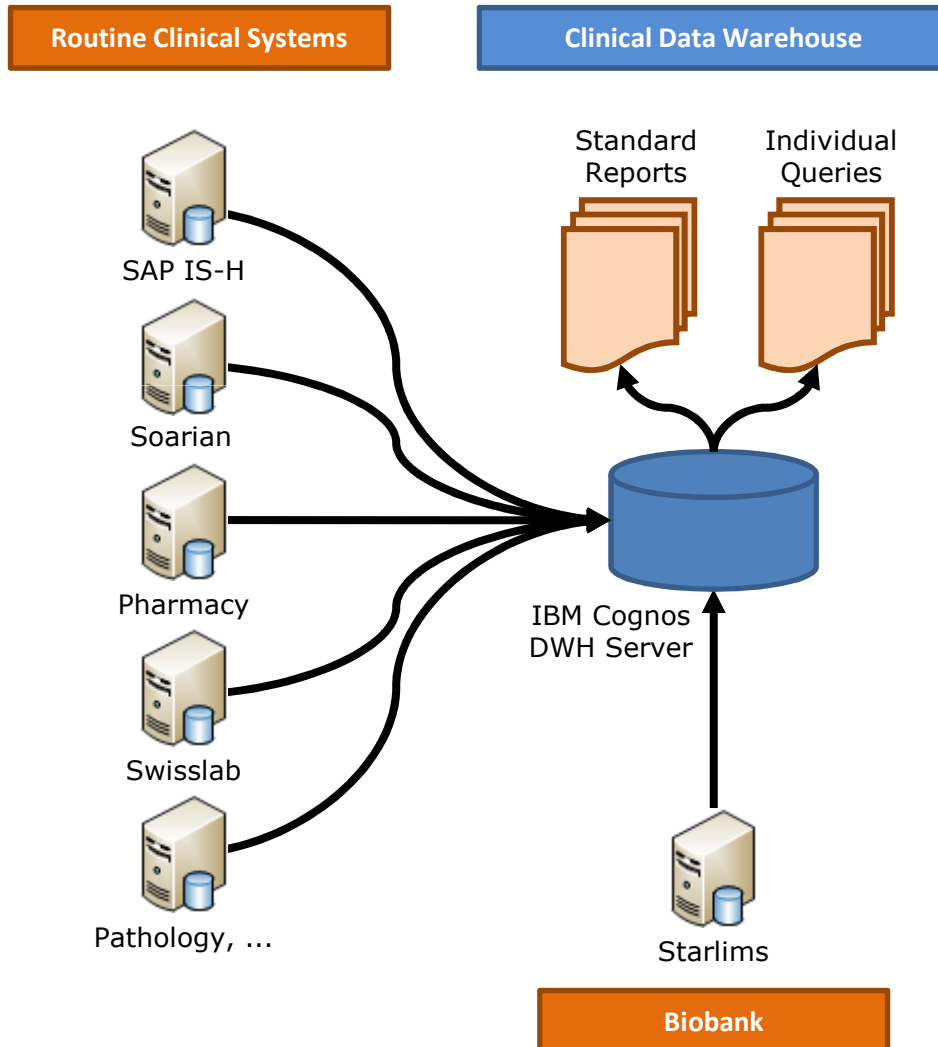
■ Organization

- based within central IT services
- close collaboration with Chair of Medical Informatics



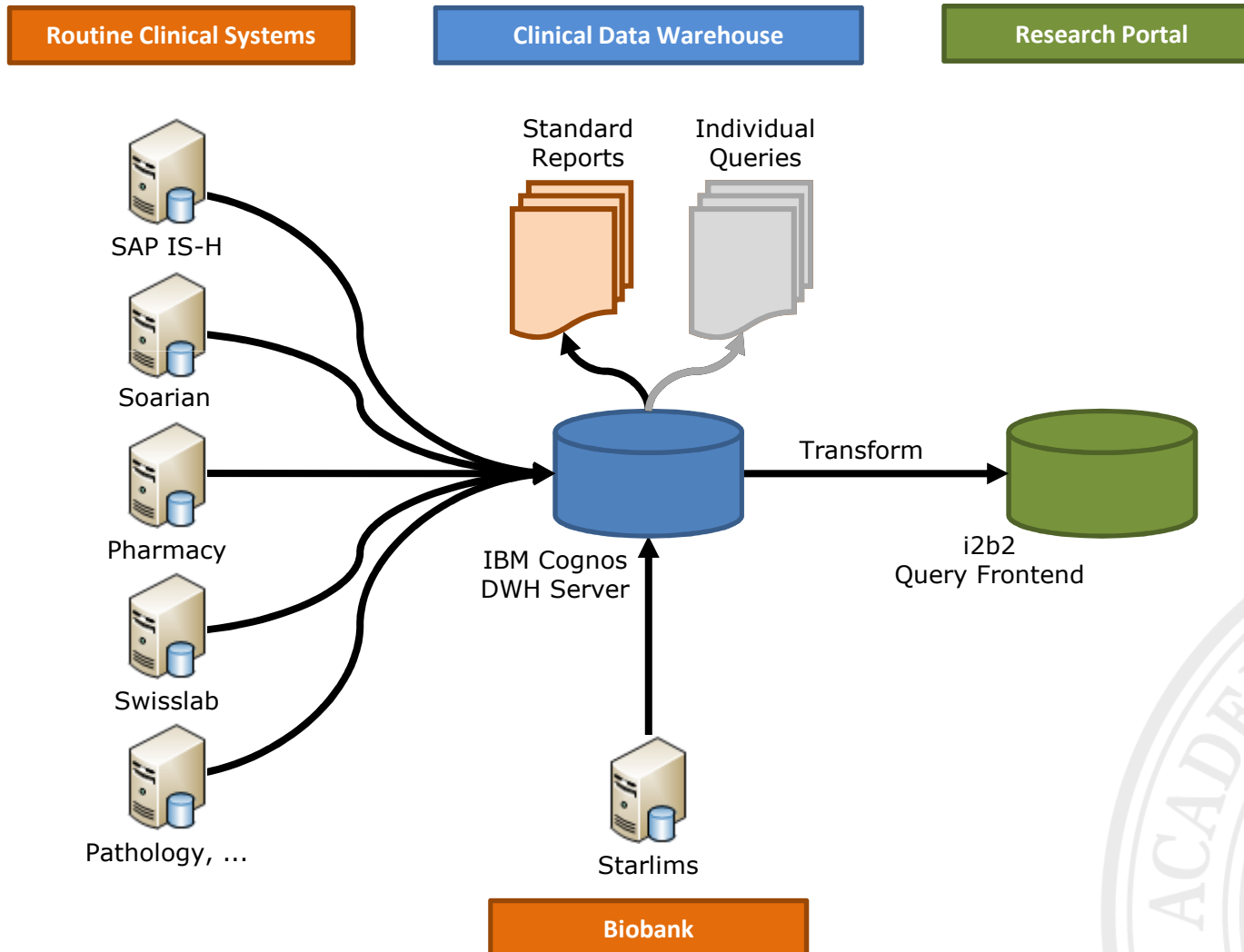
Introduction

Existing clinical Data Warehouse



Introduction

Goals: Research Portal



Introduction

Goals: Research Portal

- Empower
 - research users to query the clinical Data Warehouse themselves

- Leverage
 - existing cDWH database
 - i2b2 query frontend

- Design
 - a core dataset for research use
 - extensible with project-specific input

- Respect
 - boundaries given by data protection and ownership



Methods

Extraction, Transformation & Loading

- Cognos DataManager platform
 - used for transformation of routine data

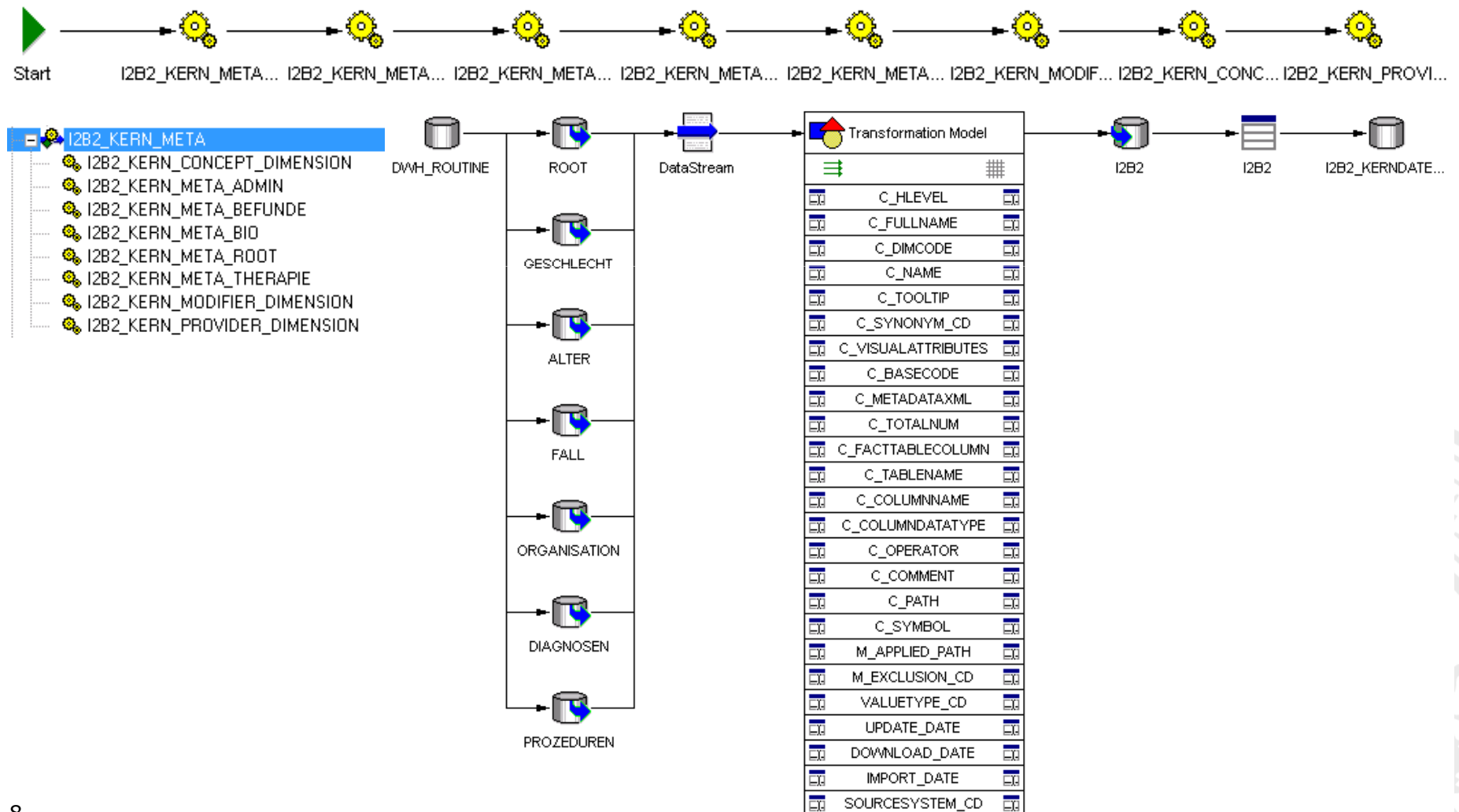
- IDRT platform
 - used for extraction of research-specific data (e.g. ODM)

- Ontology data
 - re-use of star-schema dimensions from routine DWH
 - custom dimensions for research-specific datasets



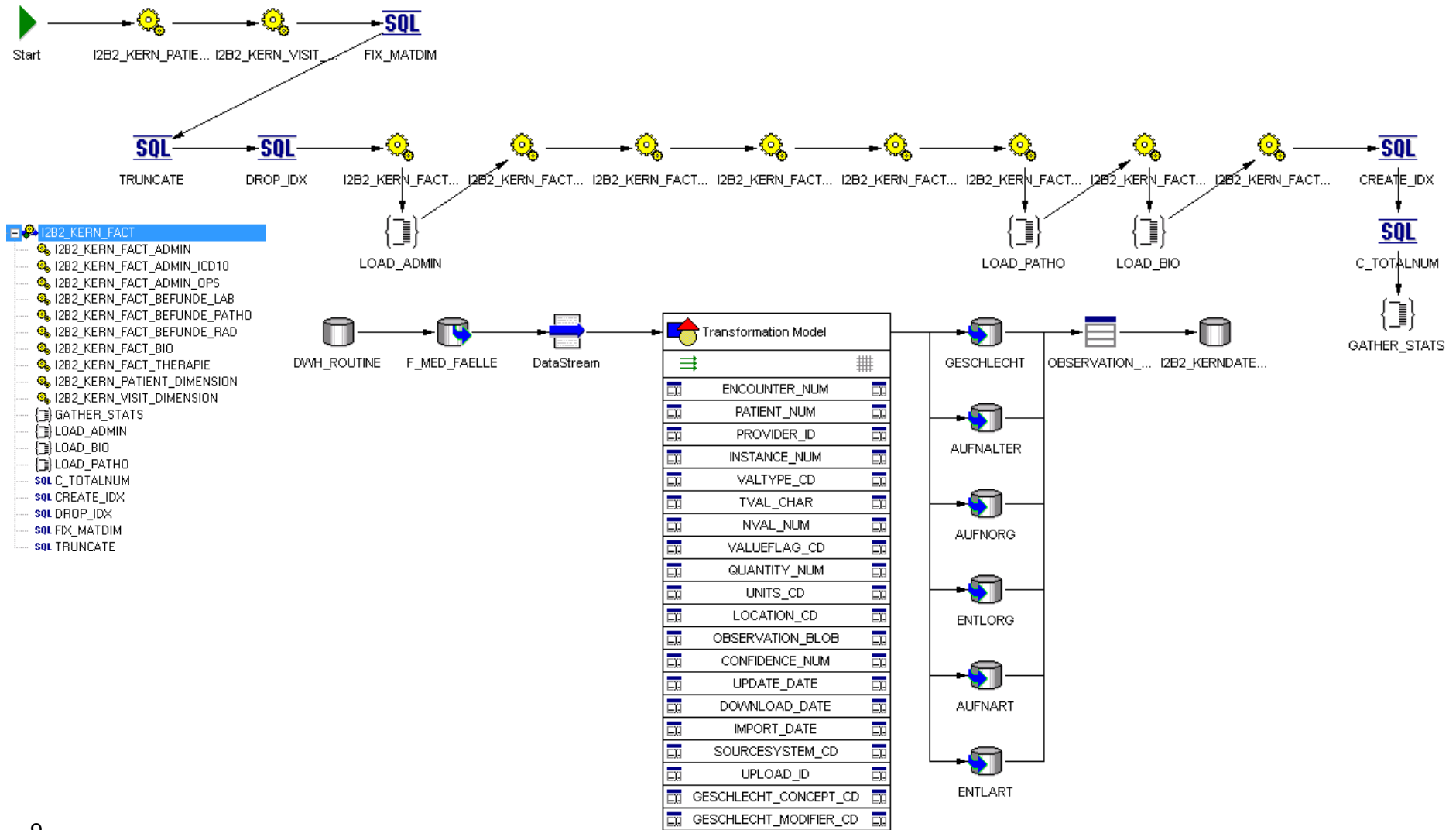
Methods

ETL: Ontology data



Methods

ETL: Observation Fact Data



Methods

Implementation details

■ Performance optimization

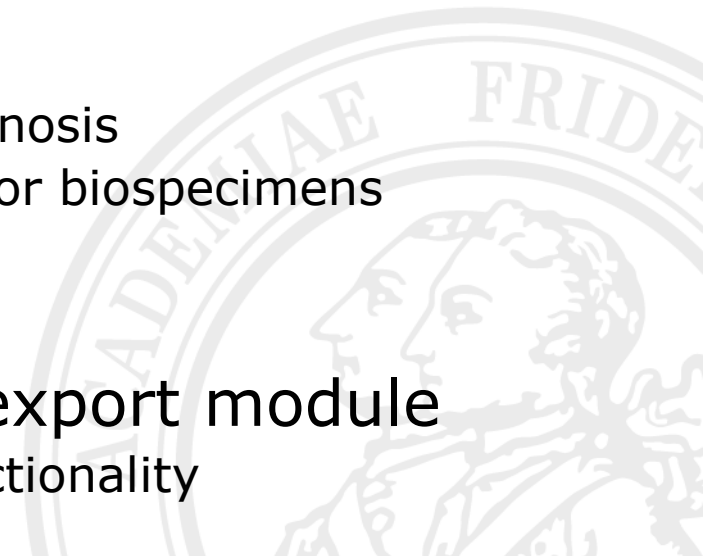
- use of Oracle SQL*Loader throughout ETL
- calculation of TOTAL_NUM data for i2b2 query optimization
- gathering of Oracle database statistics (post-ETL)
- partitioning of observation fact table and indexes based on schema key
- modification of indexes to support queries including modifiers

■ Extensive use of Modifier Codes

- to represent primary vs. secondary diagnosis
- to implement detailed sub-information for biospecimens and pathology reports

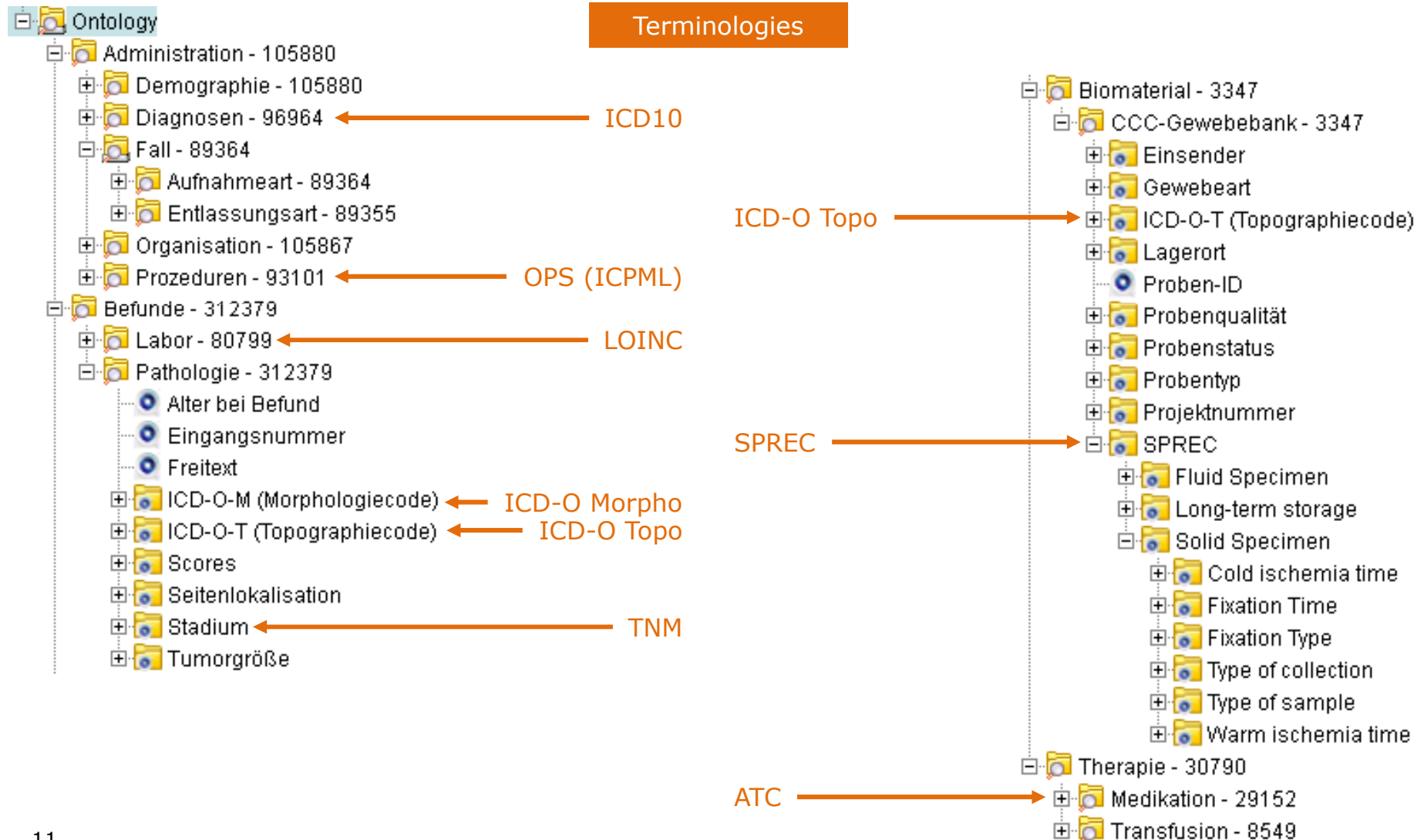
■ Redesign of i2b2 Webclient XLS export module

- complete refactoring with extended functionality



Results

Core Dataset Ontology



Results

Core Dataset Observation Facts

■ Population

- 310,000 Patients with 1.2 million episodes of care

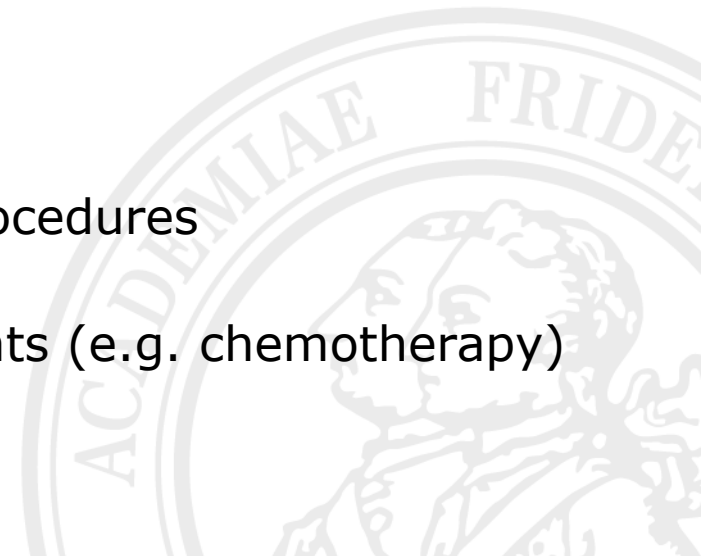
■ Pathology and biospecimen data

- 584,000 free text pathology reports
- 4,300 annotated reports (ICD-O, TNM, UICC, ...)
- 10,000 annotated cryospecimens

■ Routine clinical data

- 2.1 million demographic elements
- 2.4 million diagnoses and 1.8 million procedures
- 34 million lab findings
- 462,800 pharmacy and transfusion events (e.g. chemotherapy)

■ Total: 48 million observations



Results: Sample Query

Colorectal carcinoma (ICD C18 - C20)

i2b2 Query & Analysis Tool Project: CCC Biobank User: Thomas Ganslandt Find Patients | Analysis Tools | Message Log | Help | Logout

Navigate Terms **Find Terms**

- Ontology
 - Administration - 105880
 - Demographie - 105880
 - Diagnosen - 96964
 - 01. Bestimmte infektiöse und parasitäre Krankheiten - 17524
 - 02. Neubildungen - 49444
 - Hauptdiagnose
 - Nebendiagnose
 - C00-C14: Bösartige Neubildungen der Lippe, der Mundhöhle und des Pharynx - 41
 - C15-C26: Bösartige Neubildungen der Verdauungsorgane - 6951
 - Hauptdiagnose
 - Nebendiagnose
 - C15: Bösartige Neubildung des Ösophagus - 740
 - C16: Bösartige Neubildung des Magens - 974
 - C17: Bösartige Neubildung des Dünndarmes - 214
 - C18: Bösartige Neubildung des Kolons - 1940
 - C19: Bösartige Neubildung am Rektosigmoid, Übergang - 312
 - C20: Bösartige Neubildung des Rektums - 1946
 - C21: Bösartige Neubildung des Anus und des Analkanals - 203
 - C22: Bösartige Neubildung der Leber und der intrahepatischen Gallengänge - 8
 - C23: Bösartige Neubildung der Gallenblase - 85
 - C24: Bösartige Neubildung sonstiger und nicht näher bezeichneter Teile der Ga
 - C25: Bösartige Neubildung des Pankreas - 966
 - C26: Bösartige Neubildung sonstiger und ungenau bezeichneter Verdauungsor
 - C30-C39: Bösartige Neubildungen der Atmungsorgane und sonstiger intrathoraka
 - C40-C41: Bösartige Neubildungen des Knochens und des Gelenkknorpels - 487
 - C43-C44: Melanom und sonstige bösartige Neubildungen der Haut - 4053
 - C45-C49: Bösartige Neubildungen des mesothelialen Gewebes und des Weichteil
 - C50-C50: Bösartige Neubildungen der Brustdrüse [Mamma] - 6126
 - C51-C58: Bösartige Neubildungen der weiblichen Genitalorgane - 2401
 - C60-C63: Bösartige Neubildungen der männlichen Genitalorgane - 800
 - C64-C68: Bösartige Neubildungen der Harnorgane - 800
 - C69-C72: Bösartige Neubildungen des Auges, des Gehirns und sonstiger Teile de
 - C73-C75: Bösartige Neubildungen der Schilddrüse und sonstiger endokriner Drüs
 - C76-C80: Bösartige Neubildungen ungenau bezeichneter, sekundärer und nicht nä

Query Tool

Query Name: C18: Bösartige @10:23:54

Temporal Constraint: Treat all groups independently

Group 1			Group 2			Group 3		
Dates	Occurs > 0x	Exclude	Dates	Occurs > 0x	Exclude	Dates	Occurs > 0x	Exclude
Treat Independently			Treat Independently			Treat Independently		
C18: Bösartige Neubildung								
C19: Bösartige Neubildung								
C20: Bösartige Neubildung								

one or more of these **AND** drop a term on here

Run Query Clear Print Query 1 Group New Group

Query Status

Finished Query: "C18: Bösartige @10:23:54" [7.3 secs]
 Compute Time: 5 secs

Number of patients for "C18: Bösartige @10:23:54"
 patient_count: 3466

Results: Sample Query

Add Staging from Pathology report: pT2/3

i2b2 Query & Analysis Tool Project: CCC Biobank User: Thomas Ganslandt Find Patients | Analysis Tools | Message Log | Help | Logout

Navigate Terms **Find Terms**

- Ontology
 - Administration - 105880
 - Befunde - 312379
 - Labor - 80799
 - Pathologie - 312379
 - Alter bei Befund
 - Eingangsnummer
 - Freitext
 - ICD-O-M (Morphologiecode)
 - ICD-O-T (Topographiecode)
 - Scores
 - Seitenlokalisierung
 - Stadium
 - Regression
 - TNM
 - 0. Präfix TNM
 - 1. T
 - T0
 - T0a
 - T1
 - T1a
 - T1b
 - T1b1
 - T1b2
 - T1c
 - T2
 - T2a
 - T2b
 - T2c
 - T3
 - T3a
 - T3b
 - T3c
 - T3d
 - T4

Query Tool

Query Name: C18: B6-T2@10:25:28

Temporal Constraint: Treat all groups independently

Group 1			Group 2			Group 3		
Dates	Occurs > 0x	Exclude	Dates	Occurs > 0x	Exclude	Dates	Occurs > 0x	Exclude
Treat Independently			Treat Independently			Treat Independently		
C18: Bösartige Neubildung			1. T [T2]					
C19: Bösartige Neubildung			1. T [T2a]					
C20: Bösartige Neubildung			1. T [T2b]					
			1. T [T2c]					
			1. T [T3]					
			1. T [T3a]					
			1. T [T3b]					
			1. T [T3c]					
			1. T [T3d]					

one or more of these AND one or more of these AND drop a term on here

Run Query Clear Print Query 2 Groups New Group

Query Status

Finished Query: "C18: B6-T2@10:25:28" [20.8 secs]
 Compute Time: 17 secs

Number of patients for "C18: B6-T2@10:25:28"
 patient_count: 220

Results: Sample Query

Add CEA lab test

i2b2 Query & Analysis Tool Project: CCC Biobank User: Thomas Ganslandt Find Patients | Analysis Tools | Message Log | Help | Logout

Navigate Terms | **Find Terms**

- Mineral; bone; joint; connective tissue - 64637
- Neural; Muscular - 5396
- Nucleotides/sides and derivatives - 42050
- Porphyryns - 143
- Protein - 67958
- Renal function - 76012
- Small molecules - 33135
- Specimen Physical properties; artifacts - 8295
- Sugars/Sugar metabolism - 68443
- Tumor markers - 14962
 - Cancer Ag 125 [Units/volume] in Serum or Plasma - 4810
 - Cancer Ag 15-3 [Units/volume] in Serum or Plasma - 4465
 - Cancer Ag 19-9 [Units/volume] in Body fluid - 8
 - Cancer Ag 19-9 [Units/volume] in Serum or Plasma - 6153
 - Cancer Ag 72-4 [Units/volume] in Serum or Plasma - 725
 - Carcinoembryonic Ag [Mass/volume] in Body fluid - 10
 - Carcinoembryonic Ag [Mass/volume] in Serum or Plasma - 11365
 - Chromogranin A [Mass/volume] in Serum or Plasma - 662
 - Fibroblast growth factor 23.intact [Mass/volume] in Plasma - 17
 - Fibroblast growth factor.C-terminal [Units/volume] in Plasma - 294
 - Prostate Specific Ag Free [Mass/volume] in Serum or Plasma - 995
 - Prostate Specific Ag Free/Prostate specific Ag.total in Serum or Plasma - 977
 - Prostate specific Ag [Mass/volume] in Serum or Plasma - 2602
 - Sialate [Presence] in Urine - 1
 - Squamous cell carcinoma Ag [Mass/volume] in Serum or Plasma - 741
- Vitamins - 6892
- Chemistry; challenge - 3172
- Coagulation - 74023
- Cytology - 209
- [Document ontology - 0]
- Drug/Tox - 10599
- Fertility testing - 3493
- Hematology - 75536
- Human Leukocyte Antigens - 2098
- Microbiology - 47186

Query Tool

Query Name: C18: -T2-Carci@10:27:52

Temporal Constraint: Treat all groups independently

Group 1			Group 2			Group 3		
Dates	Occurs > 0x	Exclude	Dates	Occurs > 0x	Exclude	Dates	Occurs > 0x	Exclude
Treat Independently			Treat Independently			Treat Independently		
C18: Bösartige Neubildun			1. T [T2]			Carcinoembryonic Ag [Mas		
C19: Bösartige Neubildun			1. T [T2a]					
C20: Bösartige Neubildun			1. T [T2b]					
			1. T [T2c]					
			1. T [T3]					
			1. T [T3a]					
			1. T [T3b]					
			1. T [T3c]					
			1. T [T3d]					

one or more of these
AND
one or more of these
AND
one or more of these

3 Groups

Query Status

Finished Query: "C18: -T2-Carci@10:27:52" [21.9 secs]
 Compute Time: 19 secs

Number of patients for "C18: -T2-Carci@10:27:52"
 patient_count: 184

Results: Sample Query

Visualization of Demographics

i2b2 Query & Analysis Tool Project: CCC Biobank User: Thomas Ganslandt Find Patients | Analysis Tools | Message Log | Help | Logout

Navigate Terms Find Terms

- Ontology
 - Administration - 105880
 - Befunde - 312379
 - Biomaterial - 3347
 - CCC-Gewebebank - 3347
 - Einsender
 - Gewebeart
 - ICD-O-T (Topographiecode)
 - Lagerort
 - Proben-ID
 - Probenqualität
 - Probenstatus

Demographics

Specify Data View Results Plugin Help

Patient Count: 184

Age in Years

20-30	2
30-40	1
40-50	5
50-60	31
60-70	52
70-80	69
80-90	20
90-100	4

Sex

F	79
M	105

Vital Status (Deceased)

N	6
Y	178

Workplace

- ganslats

Previous Queries

- C18-T2-Car-Pro@10:28:55 [1-14-2013] [ganslats]
- Results of C18-T2-Car-Pro@10:28:55 [1-14-2013] [ganslats]
- Patient Set for "C18-T2-Car-Pro@10:28:55" - FINI
- Number of patients for "C18-T2-Car-Pro@10:28:55" - FINI

Plugins

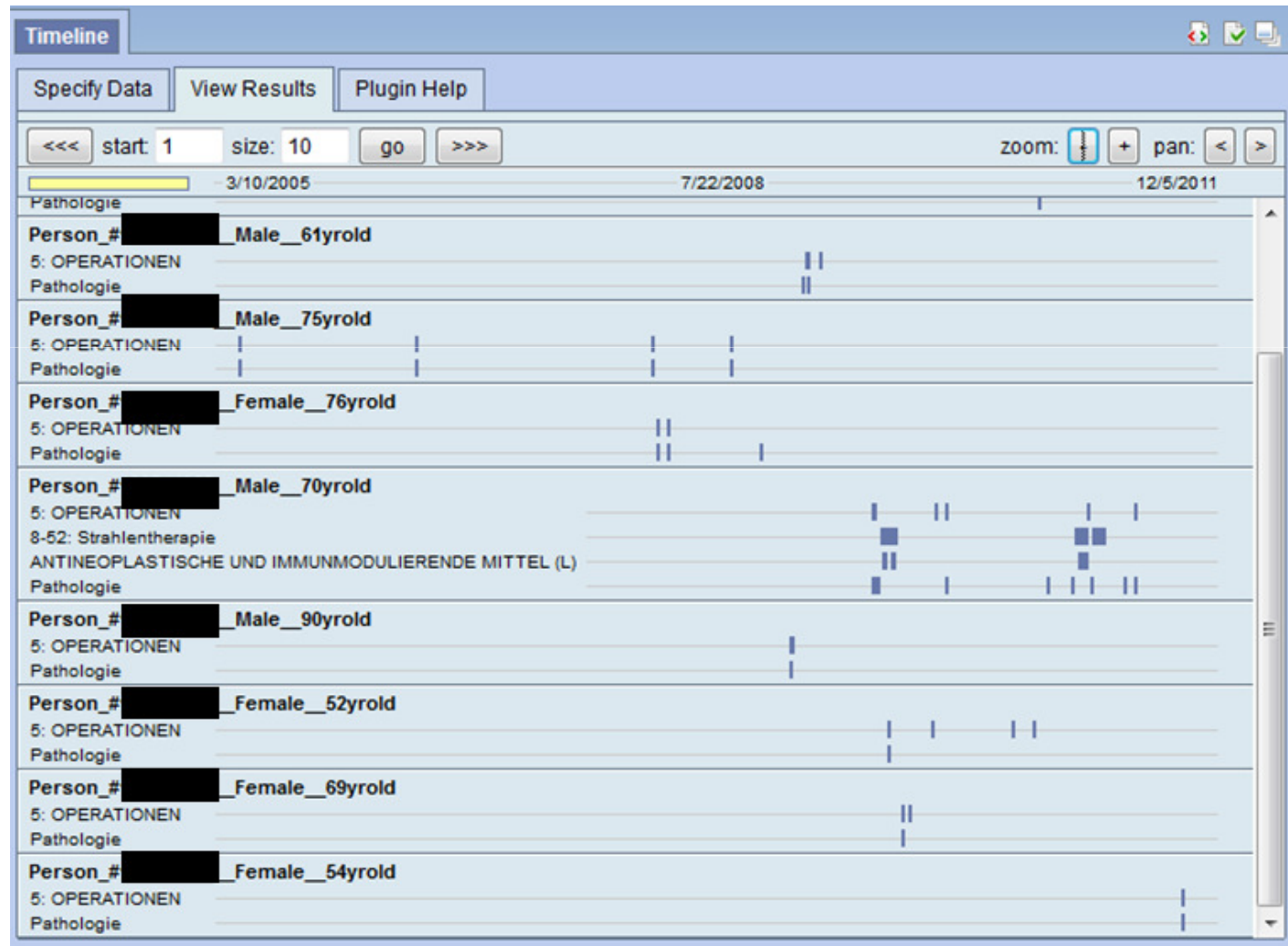
Detailed List View Category: ALL

ExportXLS
This plugin tabulates unidentified patient data, and applicable diagnoses from specified concepts, of a Patient Set; as well as provides convenient exportation of these data to an Excel spreadsheet.

Demographics (1 Patient Set) - Simple Counts

Results: Sample Query

Visualization of Timeline

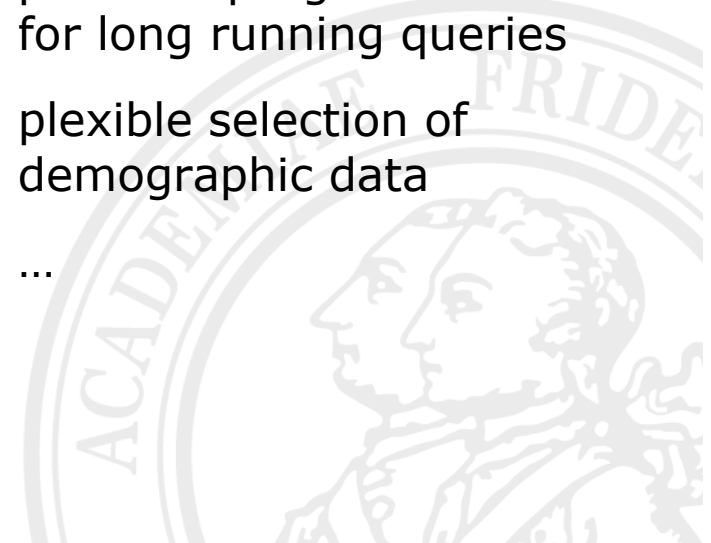


Results

Redesign of i2b2 Webclient XLS export module

The screenshot shows the 'ExportXLS' webclient interface. It has a title bar with standard window controls and a menu bar with 'Specify Data', 'View Results', 'Plugin Help', and 'Settings'. Below the menu bar, there is instructional text: 'Drop a patient set and at least one concept (ontology term) onto the appropriate input boxes below, then click the "View Results" tab to retrieve the respective observations in the selected patient set. For more info refer to the "Plugin Help" tab.' There are two input boxes: 'Patient Set: Drop a Patient Set here' and 'Concept(s): Drop one or more Concepts here'. Below these are 'Output Options:' including a 'Formatting:' dropdown set to '1 row per observation (detailed, 1 column per observation detail)', a checkbox for 'Replace Patient IDs with Ascending Numbers starting at 1', and a section 'Include Patient Demographic Data:' with checkboxes for Sex, Age, Birth Date, Birth Year, Vital Status, Language, and Race. At the bottom, there is a section 'Options that can cause long running time:' with checkboxes for 'Resolve Concept/Modifier Codes' and 'Include Ontology Path of Concepts'.

- exports **all** fact instances of a patient
- supports modifier-based fact data
- supports paged queries, automatically iterates over queries paged by server
- provides progress feedback for long running queries
- plexible selection of demographic data
- ...



Results

Redesign of i2b2 Webclient XLS export module

ExportXLS

Specify Data | View Results | Plugin Help | Settings

Click the one of the buttons on the right to download the following table in the appropriate format. [CSV Export](#) [HTML/XLS Export](#)

Patient Information
for Patient Set '18 years old@10:43:26 [1-10-2013] [demo] [PATIENTSET_91]'

	Patient ID	Timestamp Start	Timestamp End	Observation Set	Observation Concept Code	Observation Concept Denotation	Observation Value	Observation Unit	Observation Modifier
1	1000000041	2003-11-17	2003-11-17	Circulatory system	ICD9:427.89	Other specified cardiac dysrhythmias			
2	1000000041	1998-11-27	1998-11-27	Venous PCO2 (Group:VPCO2)	LOINC:2021-4	VPCO2 (LOINC:2021-4)	36	mm/hg	
3	1000000041	2006-11-09	2006-11-09	Venous PCO2 (Group:VPCO2)	LOINC:2021-4	VPCO2 (LOINC:2021-4)	33	mm/hg	
4	1000000044	2006-04-28	2006-04-28	Circulatory system	ICD9:447.6	Arteritis, unspecified			

- several row- and column-based output formats
- exports as CSV and HTML-coded XLS
- resolves concept & modifier codes

Patient Information
for Patient Set '18 years old@10:43:26 [1-10-2013] [demo] [PATIENTSET_91]'

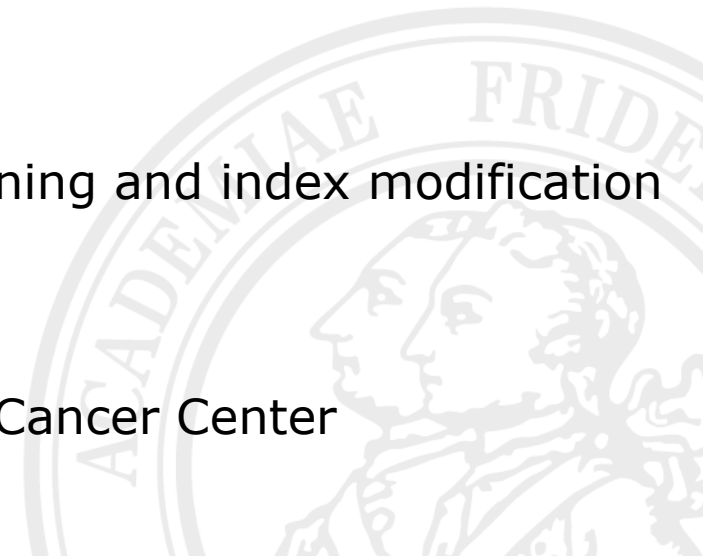
Patient ID	Arterial HCO3 (Group:ABASEX)	Circulatory system	Venous PCO2 (Group:VPCO2)
1		Other specified cardiac dysrhythmias [ICD9:427.89]	
2			VPCO2 (LOINC:2021-4) [LOINC:2021-4] = 36 mm/hg
3			VPCO2 (LOINC:2021-4) [LOINC:2021-4] = 33 mm/hg
4		Arteritis, unspecified [ICD9:447.6]	
5		Other specified cardiac dysrhythmias [ICD9:427.89]	
6		Other specified cardiac dysrhythmias [ICD9:427.89]	



Discussion

Achievements

- Integrated core dataset
 - covering most currently available data
 - research use permitted by board of directors
- ETL process
 - re-using existing content and platforms
 - ca. 4h additional run time (after regular warehouse load)
- Performance optimization
 - up to 10x improvement through partitioning and index modification
- Current use cases
 - biospecimen search for Comprehensive Cancer Center
 - EHR4CR local endpoint



Discussion

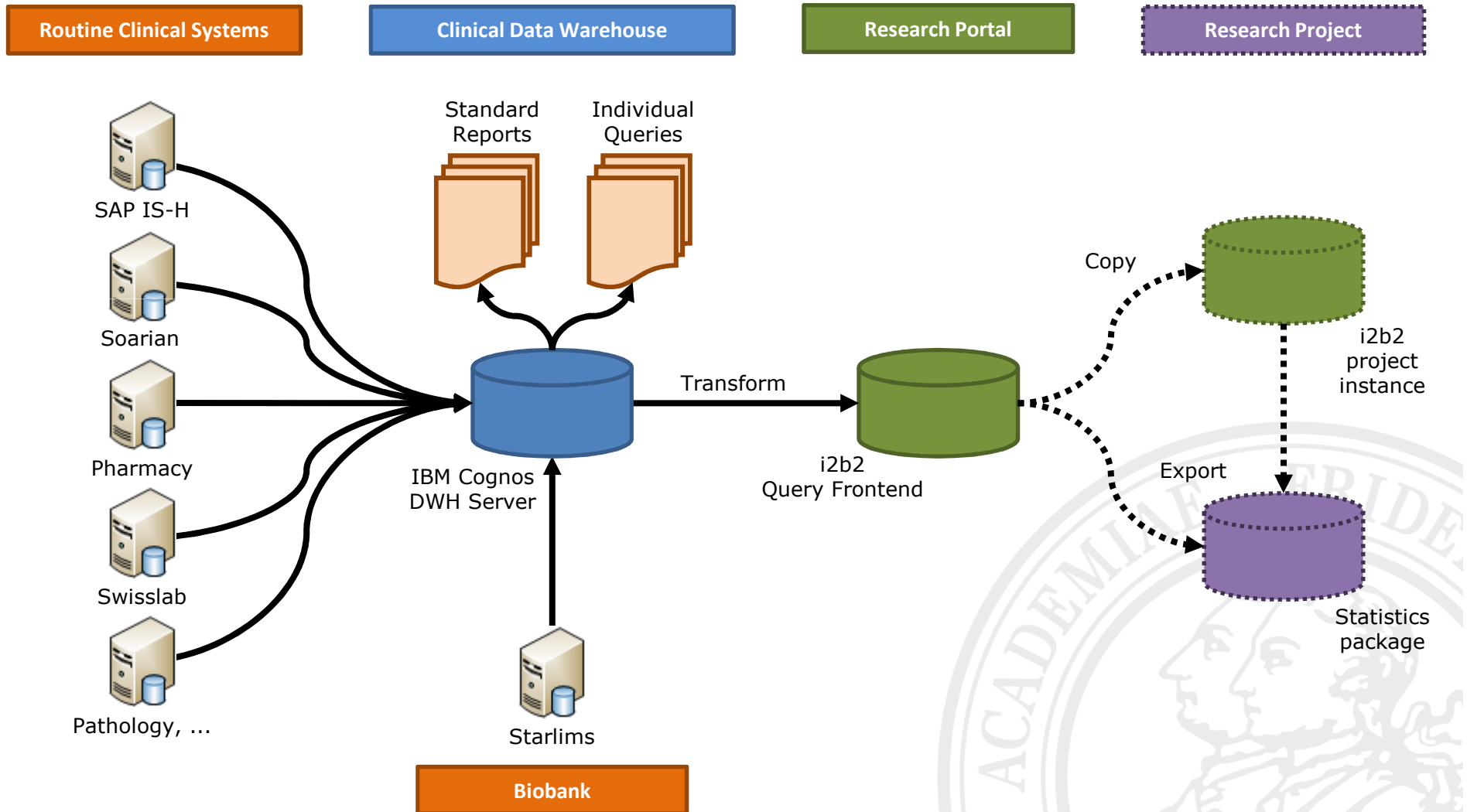
Remaining Challenges

- Permissions within i2b2
 - currently we need to load separate projects
- Automation of pseudonymization
 - currently needs manual configuration (pre ETL)
- Broad rollout to clinical users
 - including German language documentation & training
 - including monitoring of actual use
- Support for an additional "Research Project level"
 - request for project instance with specified dataset
 - integration with ETL, including follow-up data



Discussion

Extension of Research Project Level



Outlook

- Next up: broad clinical rollout
- Address the Research Project Level
 - investigate "project request" feature in Webclient
 - extend or re-implement
 - => looking for current users and/or collaborators! <=
- Integrate Omics data
 - including a look at the transMART platform
- Looking forward to i2b2 v1.7 :-)



Thank You for Your Attention!

