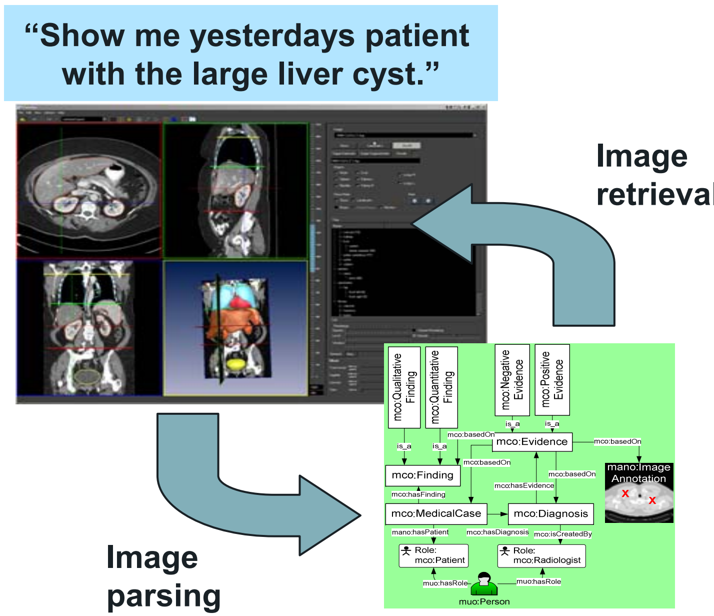
THESEUS
Neue Technologien für das Internet der Dienste

Medico

Scalable Semantic Image Search in Medicine



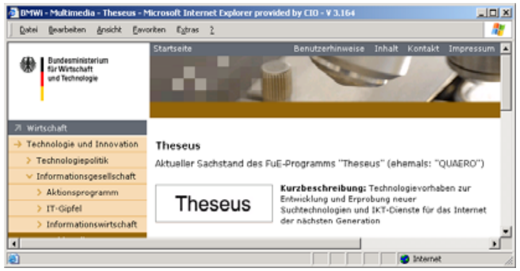
Gefördert durch das

Bundesministerium
für Wirtschaft
und Technologie

Semantic Imaging and Scalable Intelligent Image Search

– August 2006: **industry-led consortium** (empolis, Siemens, SAP) submits **Theseus** proposal
– October 2007: BMWi grants 90 Mio EUR to Theseus over 5Y
– Medico is one of 6 use cases in THESEUS
– Vision:

Provide Semantic Access to Medical Image Databases



Objectives:

- **Build the next generation of intelligent, scalable and robust search engines for the medical imaging domain**:
- Construct innovative, hierarchical information representations that will facilitate flexible image queries
- Formally address the intrinsic constraints of the medical imaging domain to define the space of queries
- Integrate higher level knowledge that will help explaining different semantic views in medical imaging: structure, function, and disease
- Create new synergies between semantics and image understanding

Medico Partners:

- **Siemens AG**
- **Ludwig Maximilians University**
- **Erlangen University Hospital**
- **Fraunhofer Institute for Computer Graphics**
- **German Center for Artificial Intelligence DFKI**

Limitations of Todays Hospital IT wrt Medical Images

- Medical images can only be searched using few parameters stored as meta data in so-called DICOM-headers (patient name, acquisition date, imaging modality etc) or indirectly by searching corresponding radiology reports
- ‘Content’ of the images can not be used for
  - quality control
  - data mining for clinical/epidemiological studies
  - decision making
  - workflow improvements

Theseus-Medico: Next Generation of Intelligent Medical Image Search Engines

THESEUS MEDICO

Clinical Literature

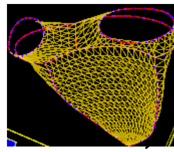
Domain Expert

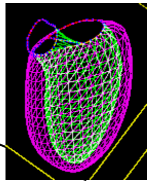
Define the set of typical queries for the domain (e.g., cardiology)

Derive minimal set of queries necessary for answering the set of all typical queries

- Semantics of the domain defined by the set of minimal queries
- Queries can be mined from literature or defined by advisory clinicians

Constrained Domain





HEART

Right Ventricle

Right Atrium

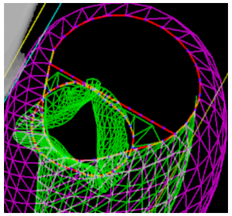
Left Ventricle

Cavity

Wall

Aortic Valve

Pulmonary Valve



Right Anterior Cusp

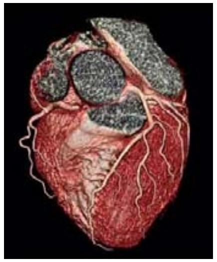
Epicardium

Endocardium

Left Anterior Cusp

- Knowledge is represented in a hierarchical ontology, e.g., Foundational Model of Anatomy

Semantic Image Annotations



Creating Hierarchical Representations through Image Parsing

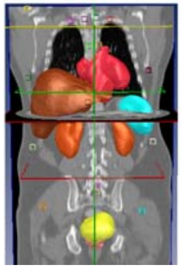
- Enrich hierarchical knowledge structure with imaging attributes

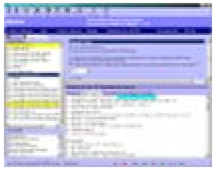
Combining Image Parsing and Semantic Knowledge

Generic Image Understanding is still a long-term agenda due to the high-complexity of the problem. However, the medical domain is constrained and hence it is possible to define an almost complete set of queries for capturing its semantics.

Semantic Image Annotations

THESEUS MEDICO





- Automated image parsing
- (Multi-modal) manual annotation of images
- Extraction from DICOM headers and DICOM structured reports
- Byproduct of reporting
- Automated extraction from radiology reports

Semantic Image Annotations

(RDF triple store)

Medical thesauri and ontologies
RadLex, FMA, ICD-10 etc

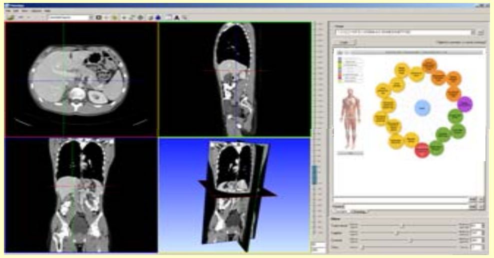
Semantic Image Annotations

THESEUS MEDICO

Semantic search (supporting e.g. synonyms (esp. multilingual) and hyperonyms)
"Show me yesterdays patient with the large multifocal spleen lesions."

Context based volume retrieval /

"Semantic navigation" of 3D data sets



Clinical decision support, workflow improvements, quality control
e.g. linking ICD codes or biopsy outcomes to medical images

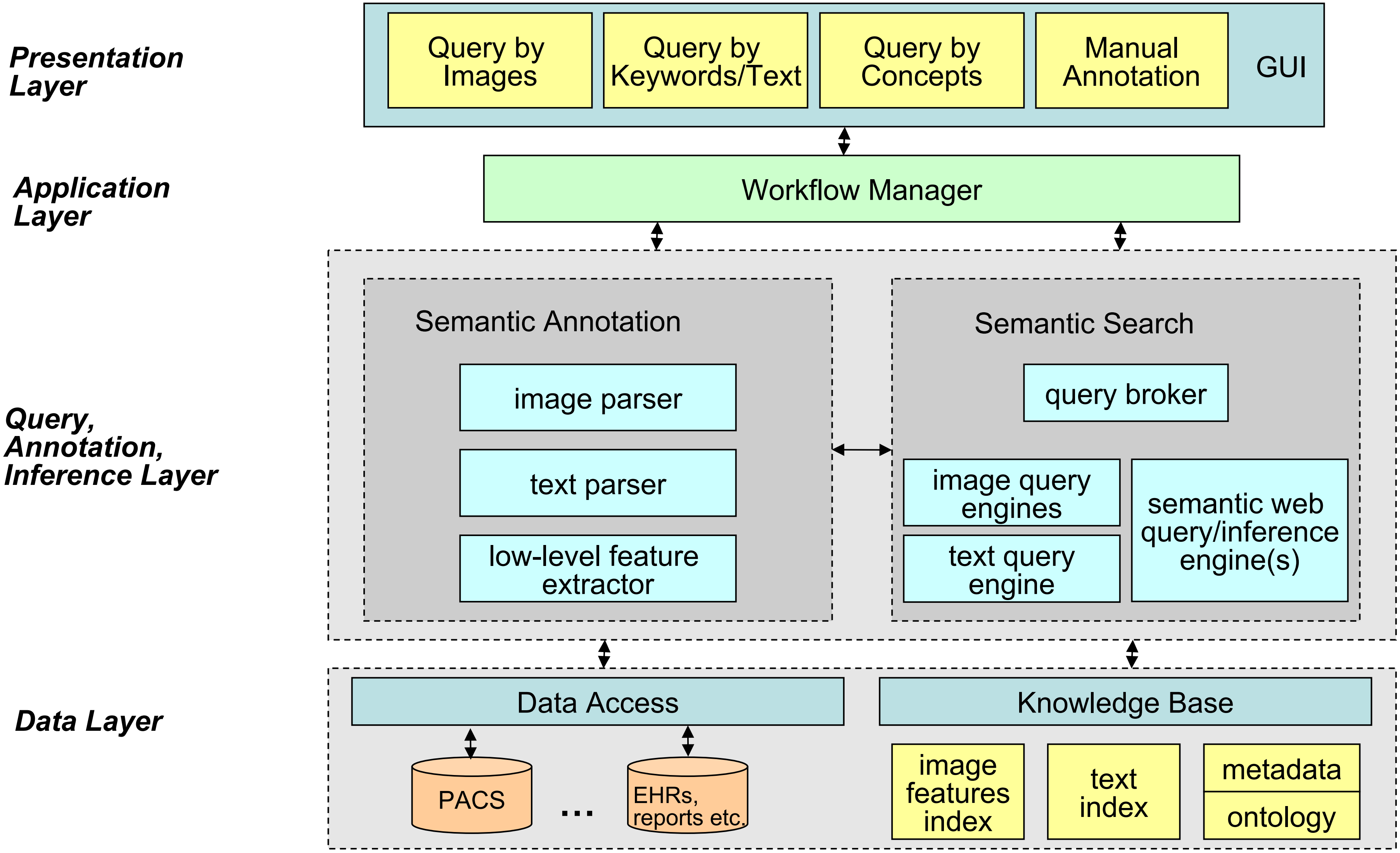
Translational research (clinical trials, patient recruiting, epidemiological studies)

PACS
(image archive)

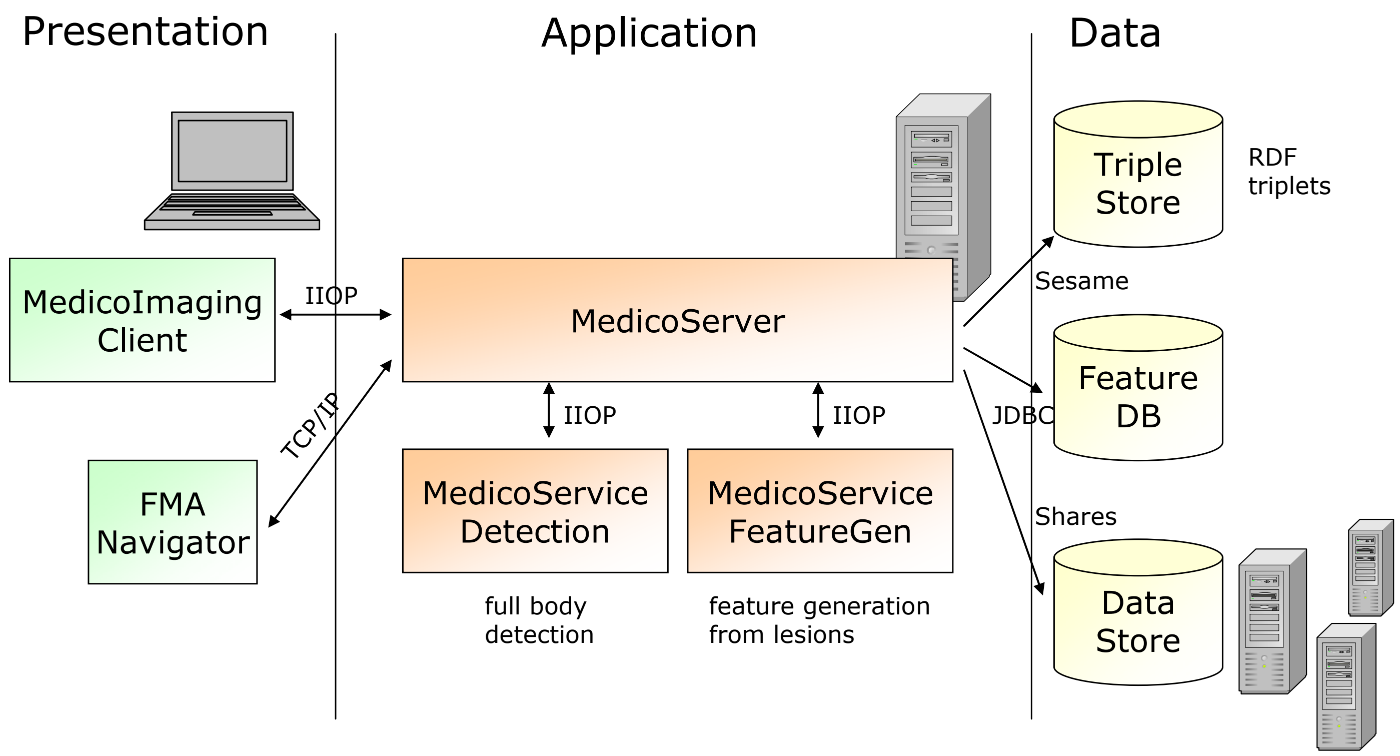
Semantic Image
Annotations
**("spleen", "splenomegaly",
"spleen lesion", "hypodense", "multifocal")**

HIS: further structured
clinical and lab data
(EPR)

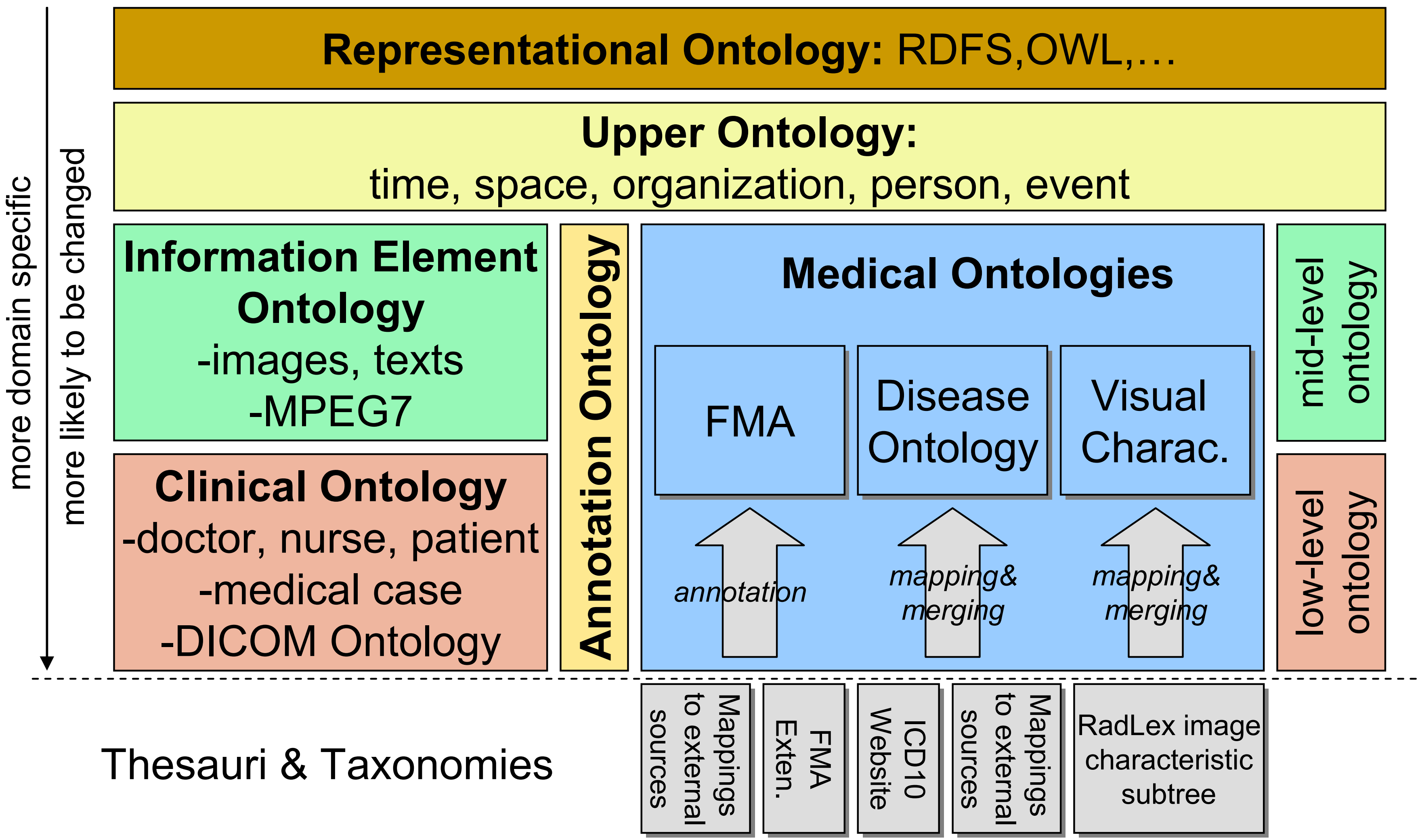
System Overview



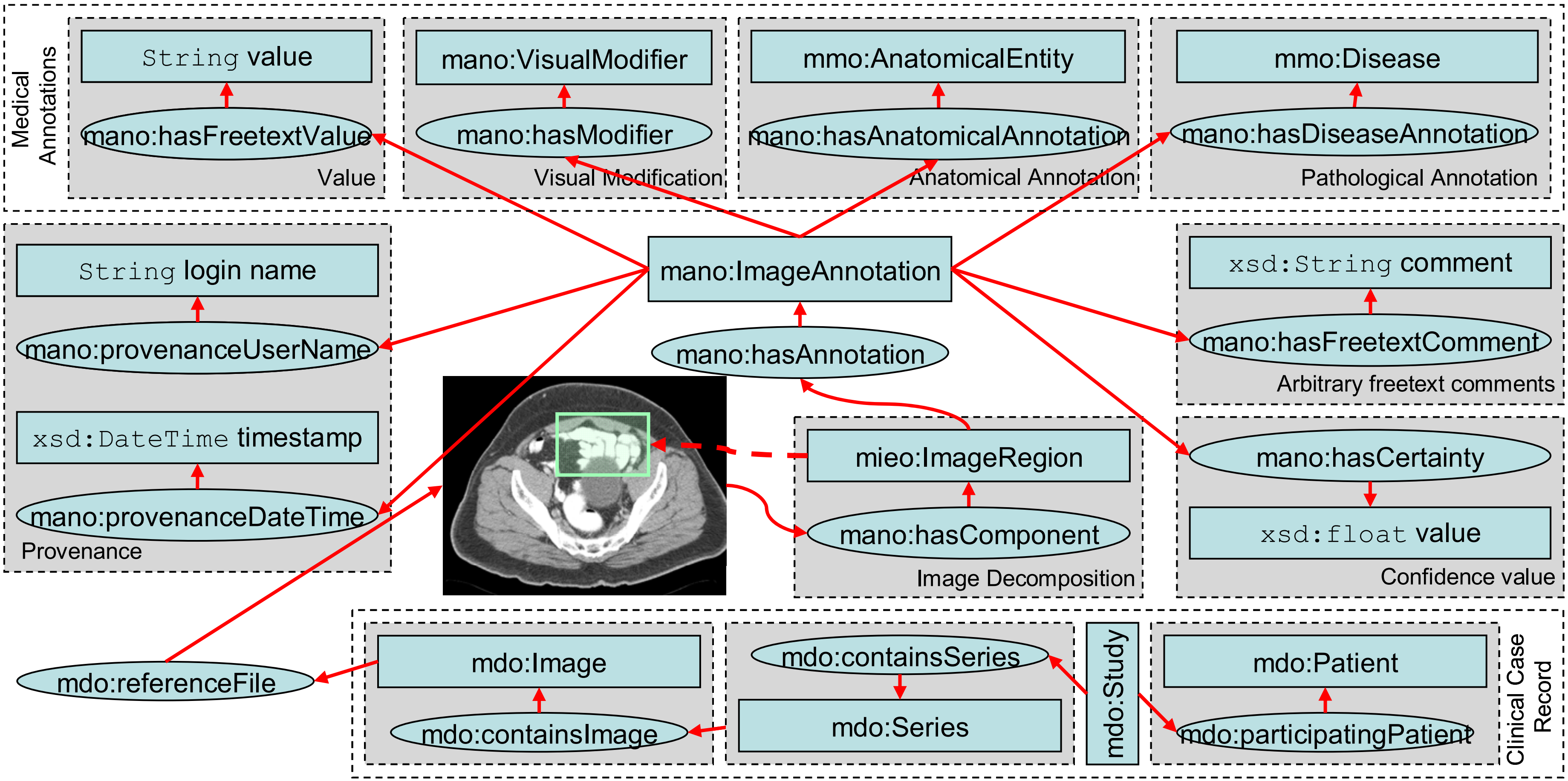
Architecture Status



MEDICO Ontology Hierarchy



Schema for Semantic Image Annotations



Ontology Alignment

- Why medical ontology alignment?
  - Medical image contents contain knowledge along 3 related domains: anatomy, radiology, diseases
  - This machine processable knowledge is available in domain ontologies
  - Integration is necessary
    - for obtaining a coherent view of the domain
    - for image annotation and for subsequent search
  - The integration is not straightforward
    - Conceptual heterogeneity: different conceptualizations (views), same knowledge
    - Terminological heterogeneity: different naming conventions, same concepts
  - Medical ontology aligment explicitly states the heterogeneity via mappings
    - help reduce semantic ambiguity
    - information loss

Query Pattern Mining

THESEUS MEDICO

What do clinicians and radiologists search for?

Goal: Predict meta information relevant for annotating medical data

Domain sources

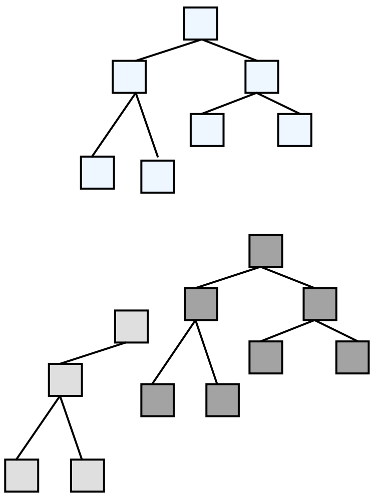
- Proprietary
- Publicly available
- Untructured
- Semi structured

PubMed

RIS Reports

Semantic sources

- FMA Ontology-Human Anatomy
- RadLex Terminology- Radiology
- NCI Cancer Thesaurus
- Features from medical images



Predicting Clinical Queries

Query Patterns =
- most frequent concepts in domain corpora
- +their interrelations

<Ontology Concept>	relation	<Ontology Concept>
<Anatomical Structure>	relation	<Anatomical Structure>
LymphNode	**located-in**	**Neck**

How big is the enlargement of the **lymph node** **in** the **neck**?

Query Pattern Derivation

[Recent work: German radiology reports]

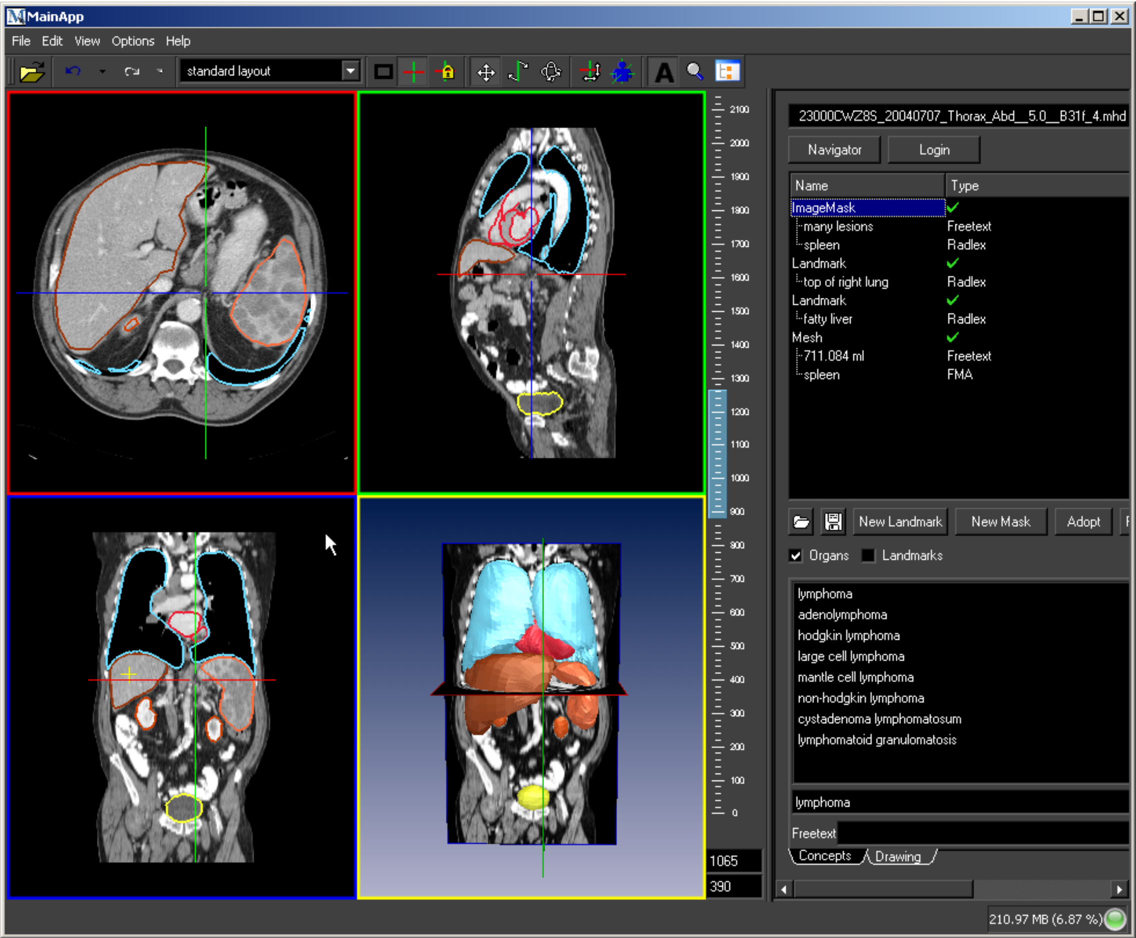
- Some particularities of the radiology reports
  - no well-formed sentences, lacking for examples verbs, and few punctuation signs
  - a lot of abbreviations and specific patterns (temporal information, spatial information used describing the position and the dimension of parts of the contents of the X-Ray, etc).
- Current WIP: extraction of
  - relevant terms and relations within the document
  - negative findings
  - spatial information that describes the content of the X-Ray
  - temporal information that indicates a development of the disease, also with links to former examinations of the patient (anonymised)
  - as a side effect: possibility to extend available terminological and semantic resources in the field (for example the German version of Radlex)

Example of Radiology Report

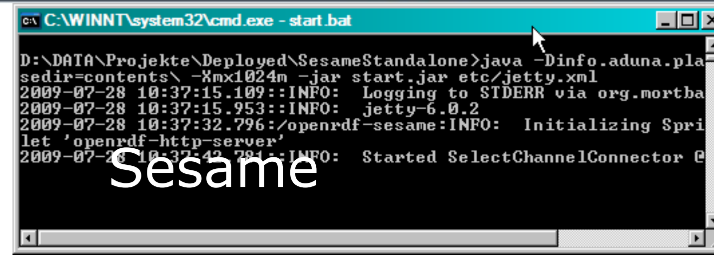
THESEUS MEDICO

Diagnostic Findings (German „Befund“)	Assessment (German „Beurteilung“)
Normal liver **size** with two sufficiently indicative lesions **in segment 7** with a **diameter** of app. **1,5 cm** (image 5) and **1 cm** (image 11)….**No former** records. …Normal wide heart shadows…aorta sclerosis…Trachea is **not** compressed. ….**On the left** many small round densifications….	In comparison to the pre-MRT on **03.03.2009 no change** in the **number** and **dimension** of liver lesions in segment 7….**No** lymphoma manifestations **in the region** of investatigation…**no** inlfiltrate… Executions in left Hilus….it is recommended to compare the relevance on the basis of **previous** records.
<u>in German (original)</u> Leber **normal groß** mit zwei signalreichen Läsionen im **Segment 7**, die einen **Durchmesser** von ca. **1,5 cm** (Bild 5) und **1 cm** (Bild 11) haben…. **Keine** Voraufnahmen. ….**Normal breiter** Herzschatten. Aortensklerose. Trachea **nicht** eingeengt …. **linksseitig** mehrere kleine rundliche Verdichtungen…	**<u>in German (original)</u>** **Im Vergleich zur** Vor-MRT vom **03.03.2003 unveränderte Zahl** und **Größe** der Leberläsionen **im** Segment 7, …….**Keine** Lymphommanifestationen im Untersuchungsgebiet….**Kein** Infiltrat, …Pneumothorax. Verrichtungen am Hilus **links** …..Relevanz Vergleich mit **Voraufnahmen empfohlen**.

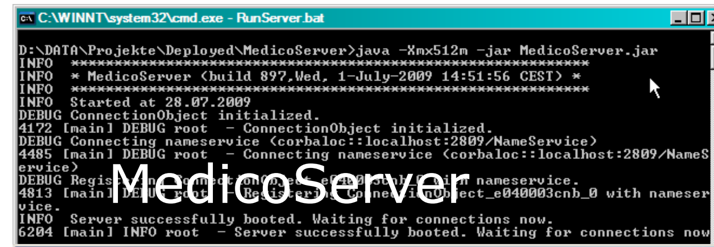
Screenshots



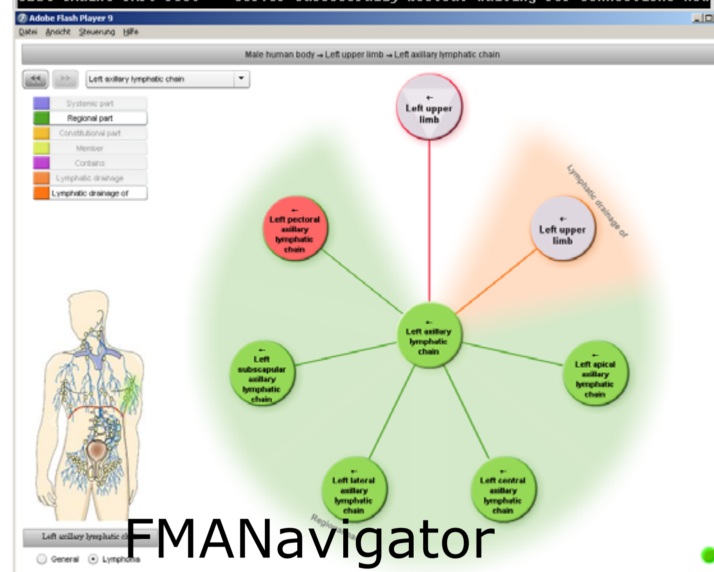
MedicoImagingClient



Sesame



MedicoServer



FMANavigator

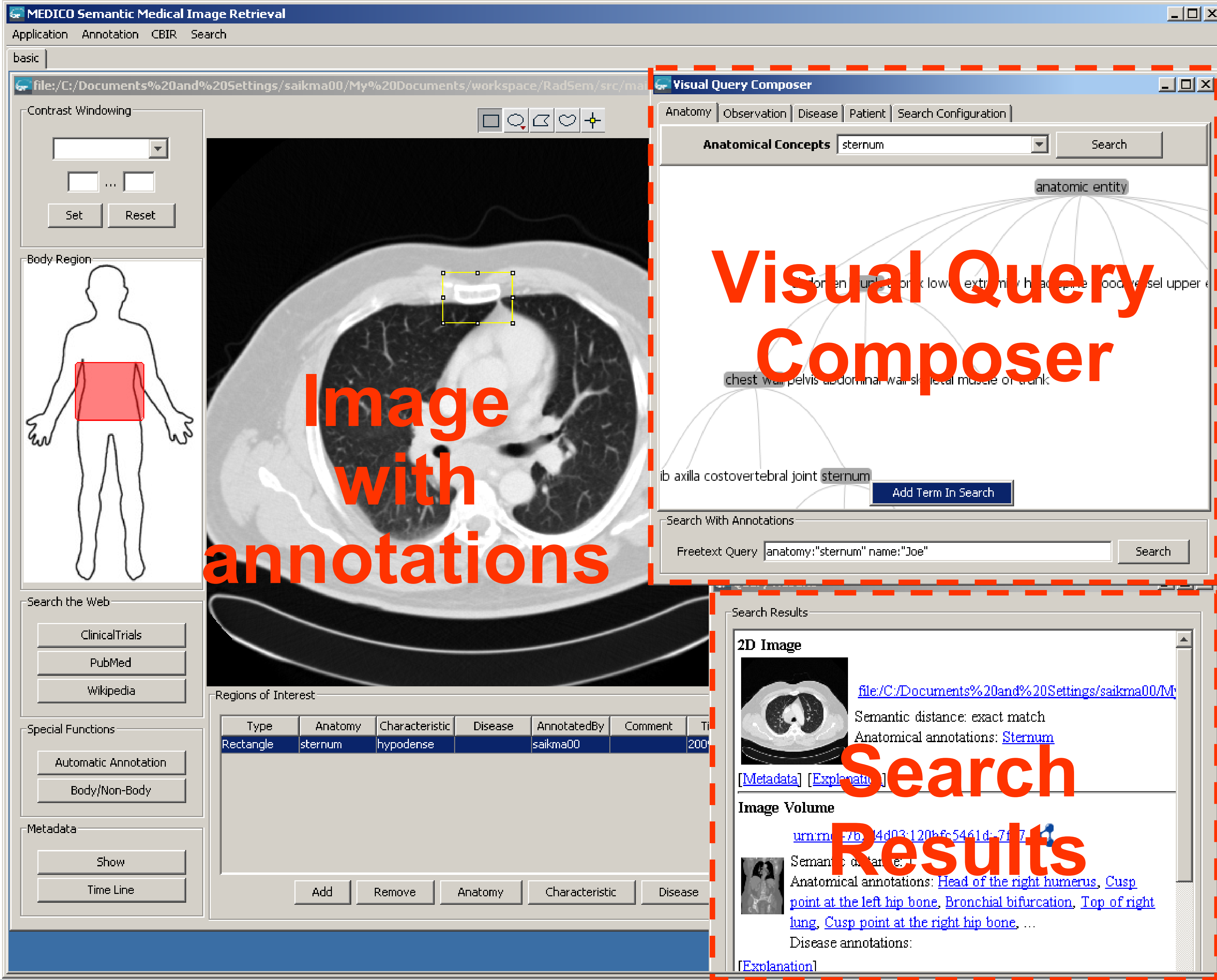

MEDICO Semantic Medical Image Retrieval
Application
Annotation
CBIR
Search
basic
Contrast Windowing
Set
Reset
Body Region
Search the Web
ClinicalTrials
PubMed
Wikipedia
Special Functions
Automatic Annotation
Body/Non-Body
Metadata
Show
Time Line
Image with annotations
Regions of Interest
Type
Anatomy
Characteristic
Disease
AnnotatedBy
Comment
Rectangle
sternum
hypodense
saikma00
Add
Remove
Anatomy
Characteristic
Disease
Visual Query Composer
Anatomy
Observation
Disease
Patient
Search Configuration
Anatomical Concepts
sternum
Search
anatomic entity
Visual Query Composer
chest wall
pelvis
ib axilla costovertebral joint
sternum
Add Term In Search
Search With Annotations
Freetext Query
anatomy:"sternum" name:"Joe"
Search
Search Results
2D Image
Semantic distance: exact match
Anatomical annotations: Sternum
[Metadata]
Search Results
Image Volume
Anatomical annotations: Head of the right humerus, Cusp point at the left hip bone, Bronchial bifurcation, Top of right lung, Cusp point at the right hip bone, ...
Disease annotations:
[Explanation]

Image Parsing

Image Parsing System

THESEUS MEDICO

challenges

many anatomical structures

many modalities

large data volume

generic
extensible

scalable

efficient

modular design

machine learning

pyramid multi-tiered

solutions

Detection Technology

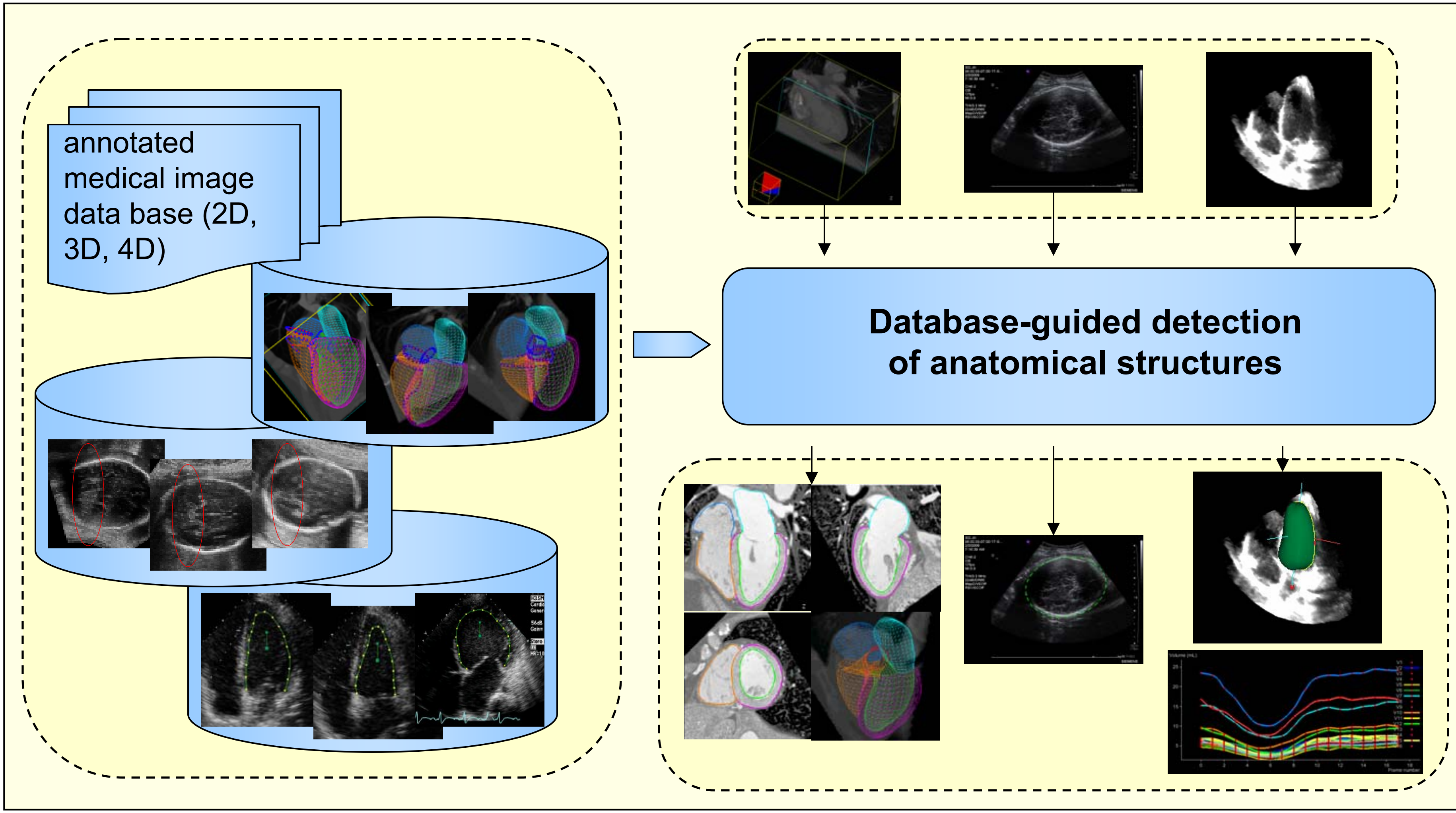
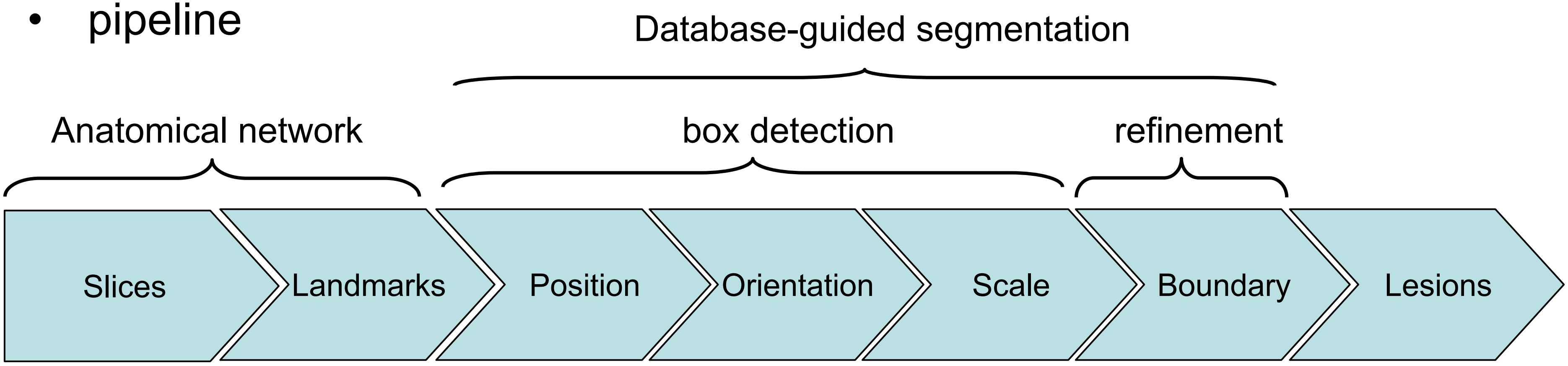


Image Parsing System

- objective: given a CT scan (e.g. whole body), we want to automatically abstract as much information as possible, about:
  - **body region**
  - **landmarks**
  - **organs and tissues**
  - **lesions**

- pipeline

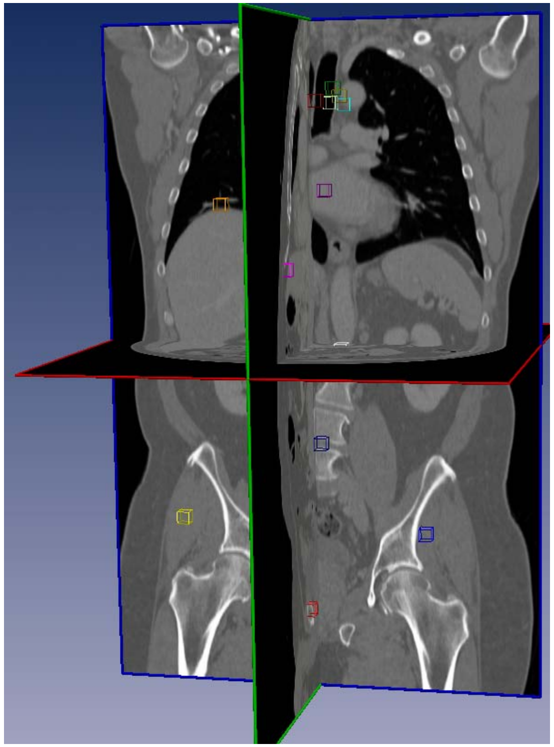


[S. Seifert, A. Barbu, K. Zhou, D. Liu, J. Feulner, M. Huber, M. Suehling, A. Cavallaro, and D. Comaniciu. Hierarchical Parsing and Semantic Navigation of Full Body CT Data, SPIE Medical Imaging, 2009]

Body Region and Landmarks

- Technologies
  - Marginal Space Learning (MSL)
  - 2D/3D Haar-like features
  - Steerable features
  - Discriminative Anatomical Network (DAN)

Performance
Body Region Detection
- 99,7% accuracy
- 380ms computation time



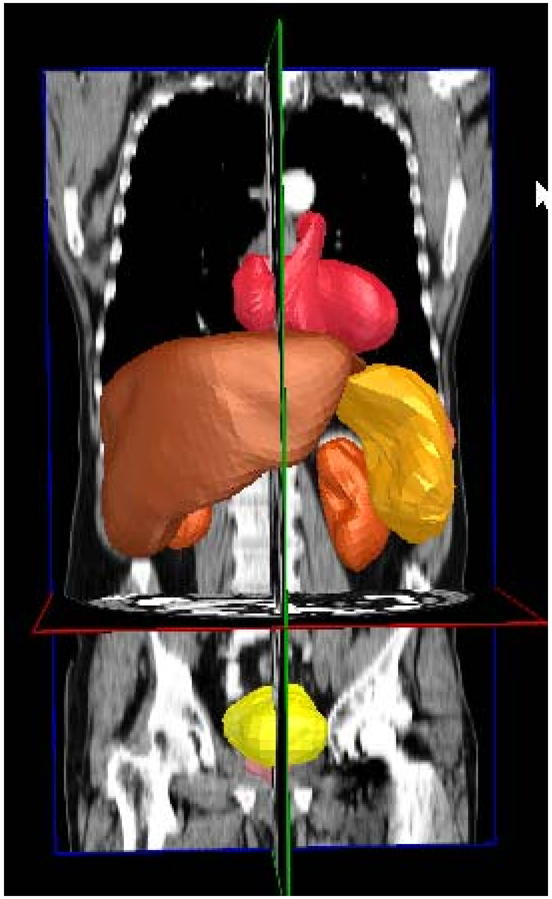
Performance
Landmark Detection
- 19 body landmarks
- <4mm average error
- 14.7sec

Organs

- Technologies
  - Constrained MSL
  - Hierarchical Active Shape Models
  - Patch-based Deformable Models
  - Trainable Boundary Detector
- Performance
  - 7 organ, total <1min

3-fold	min	median	max	std	mean
Kidney Right	0.61	1.00	14.60	1.30	1.31
Kidney Left	0.76	1.14	7.91	0.79	1.35
Spleen	1.15	2.10	16.63	2.22	2.80
Bladder	0.71	2.34	10.41	1.92	2.96

1-fold	min	median	max	std	mean
Lung Left	0.86	1.25	2.23	0.24	1.30
Lung Right	1.13	1.56	3.25	0.29	1.58
Prostate	1.13	1.89	3.34	0.61	2.16



Speed: 53 sec for 9 organs (detection + refinement)

THESEUS MEDICO

Thank you for your attention!