

# Herausforderungen der Datenintegration & semantische Integrationskonzepte

Bernhard Bauer

 **Software Methodologies**  
for Distributed Systems

- Einführung
  - › Motivation und Herausforderungen
  - › Ziele und Ansätze
- Ein Modell für klinische Informationen
- Wissensmodelle
  - › Normalgrößen Spezifikation
  - › Krankheitssymptom Modell
- Automatische Inferenz
- Zusammenfassung und Ausblick



## Daten

### Strukturiert

- Diagnosen
- Behandlungen
- Laborwerte
- Demographie
- Medikation
- Versorger

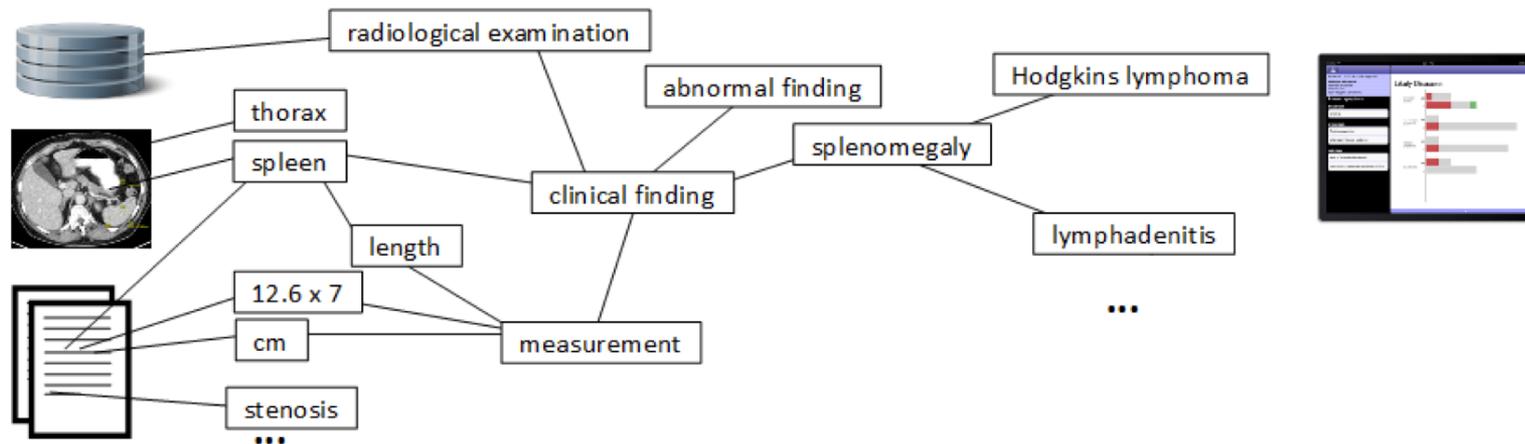
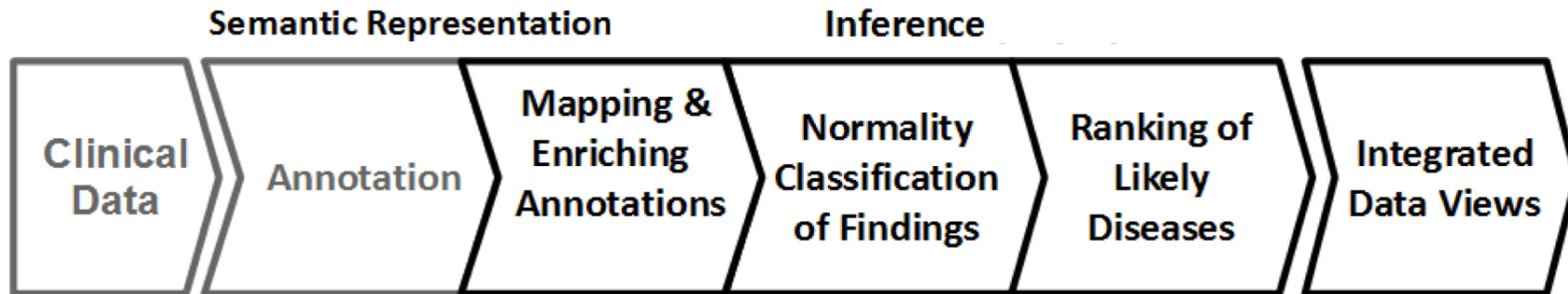
### Unstrukturiert

- Befunde/Berichte
- Bilder
- Omics Daten

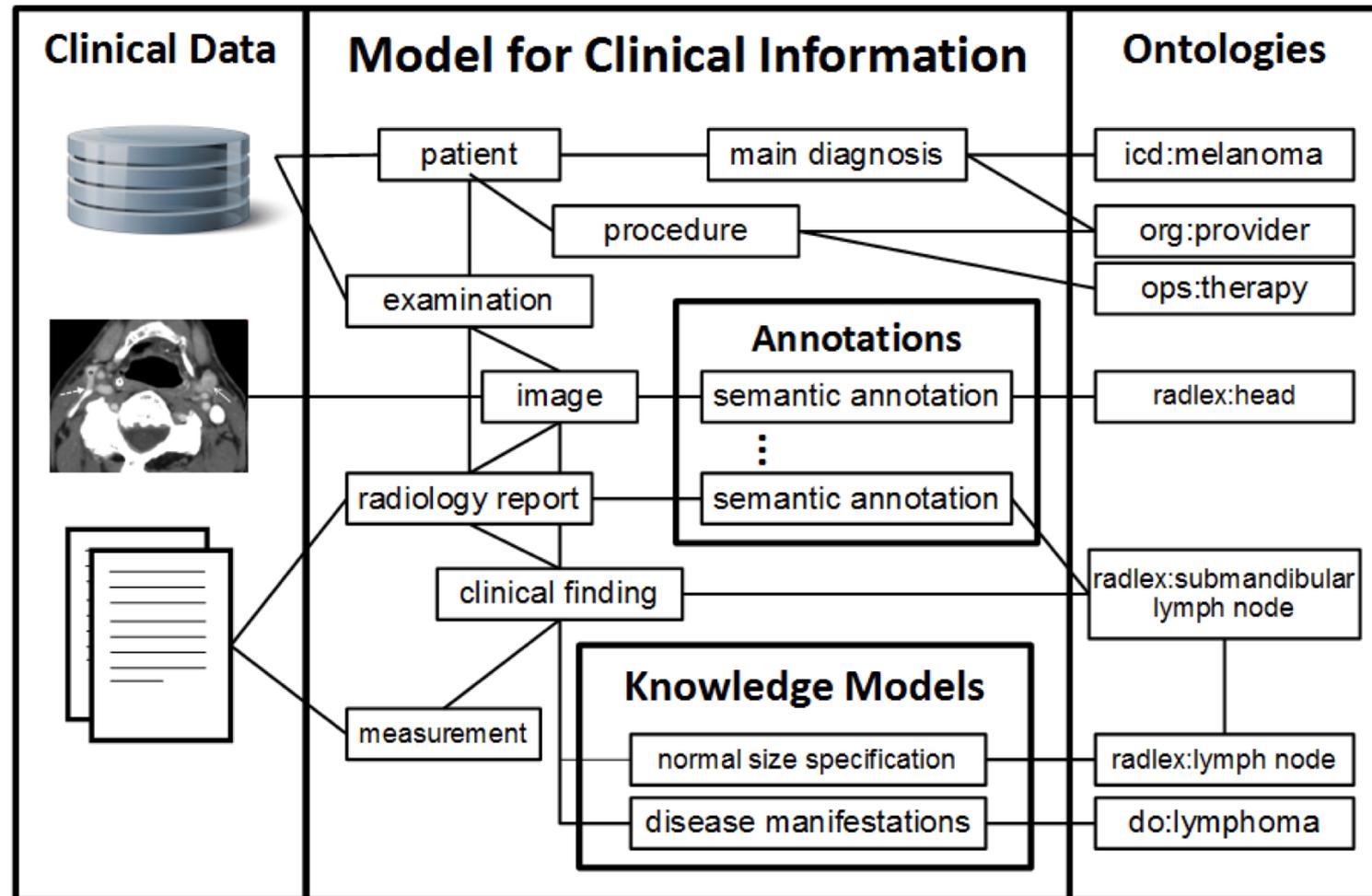
## Herausforderung

- Daten sind vorwiegend unstrukturiert
- Daten sind nicht integriert
- Unterschiedliche Daten müssen im klinischen Entscheidungsprozess verknüpft werden
- Daten müssen bewertet werden um Entscheidungen zu beschleunigen (z.B. normal, auffällig, pathologisch etc.)
- Ärzte nutzen nur einen Bruchteil der verfügbaren Daten

ABER: Eine große Anzahl an medizinischen Ontologien



1. Erstellung eines semantischen Modells für klinische Informationen (MCI)
2. Abbildung der Daten nach MCI und Anreicherung semantischer Annotationen
3. Anreicherung klinischer Daten durch Benutzung von formalisiertem medizinischem Wissen



### Basic Formal Ontology (BFO2)

Vollständig Domänen unabhängig

### Ontology for General Medical Science (OGMS)

Grundlegende klinische Konzepte, z.B. Diagnose, Befund, pathologische anatomische Struktur etc.

### Information Artifact Ontology (IAO)

Z.B. Dokument, Bericht, Grafik, Messung etc.

### Units Ontology (UO)

Auszugsweise wiederverwendet: Z.B. Einheiten für Länge, Gewicht, Zeit etc.

### Phenotypic Quality (PATO)

Auszugsweise wiederverwendet: Z.B. Größe, Länge, Gewicht, Farbe etc.

### Ontology for Biomedical Investigations (OBI)

Auszugsweise wiederverwendet: Z.B. Studie, Gewebeprobe, Plan, Voraussage, Färbung, ...

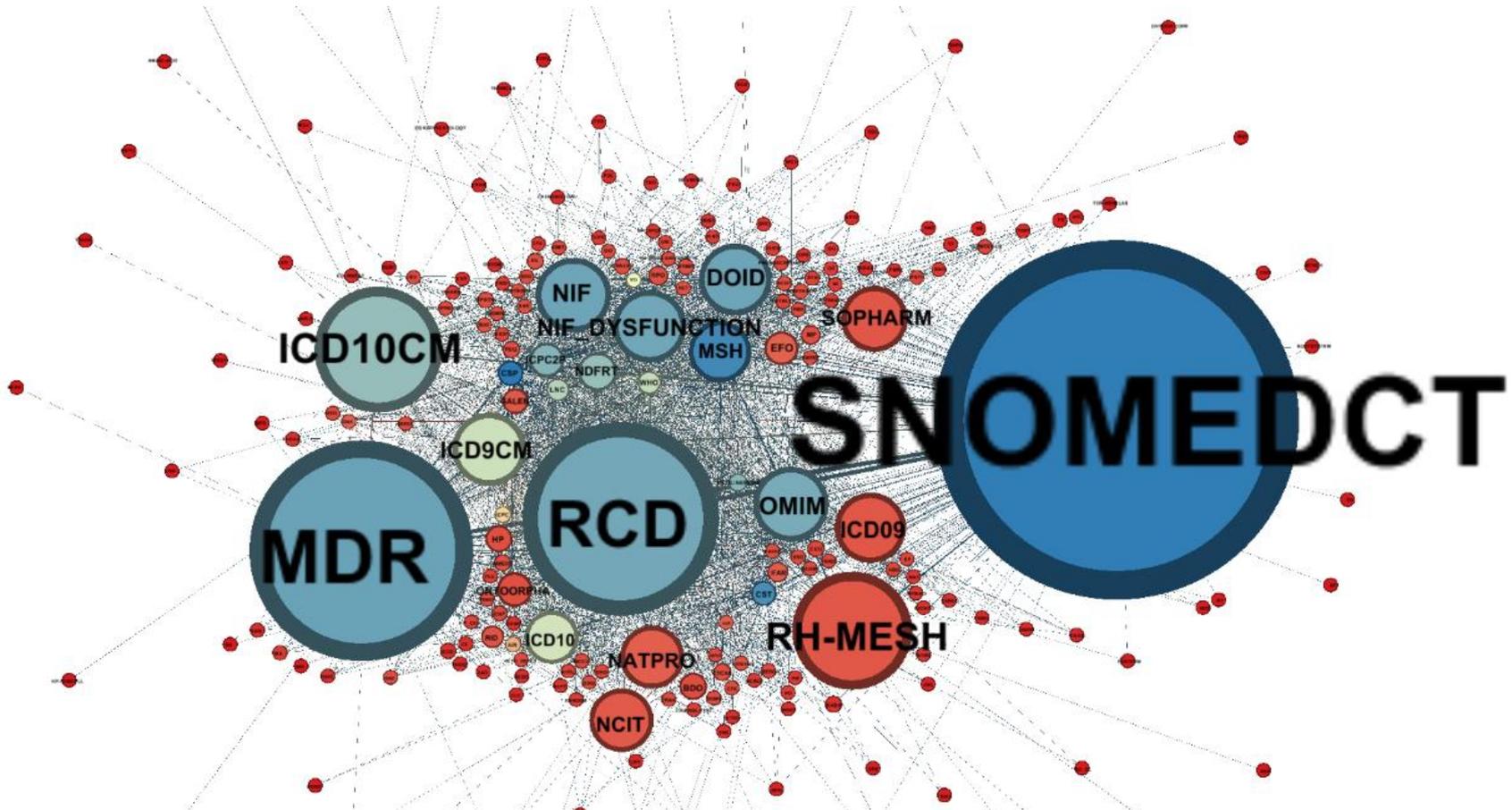
### Ontology of Medically Related Social Entities (OMRSE)

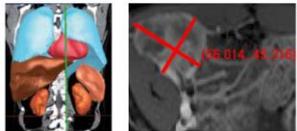
Auszugsweise wiederverwendet: Z.B. Patienten Rolle, Versorgungsstelle, Geschlecht etc.

	Total	MCI
Classes	538	99
Object Properties	88	21
Datatype Properties	27	16
Individuals	102	5

# Disease Graph mit ca. 250,000 Disease Classes

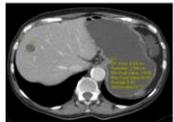
Integrated available disease-symptom information from ontologies and linked data.





## Größe

Länge: 1D, 2D, 3D  
Fläche, Volumen  
Index (z.B. Milz-Index = Breite x Tiefe x Länge)



## Dichte, gemessen in Hounsfield Units (Hu)

Vor allem in CT Bildern  
Minimal, maximal und durchschnitts Dichte Werte für ausgewählte Bereiche



## Winkel

Z.B. Knochenanordnungen oder Brüche

## Blutfluss

Z.B. PET: Blutfluss im Gehirn

...

1) Source: <http://www.recist.org/recist-in-practice/19.html>

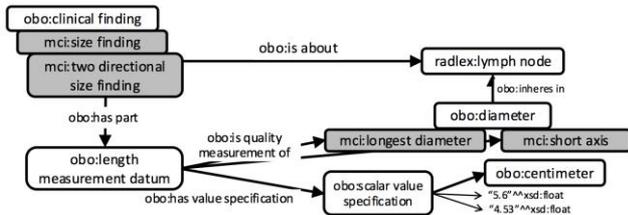


## RECIST

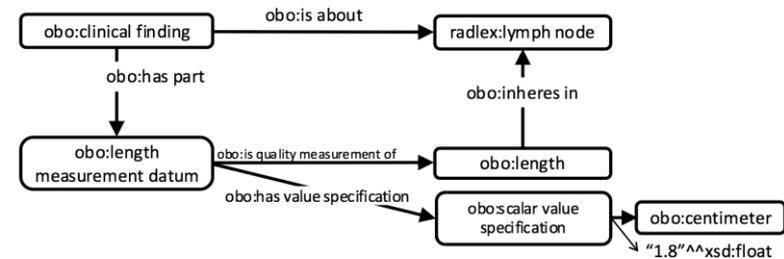
(Response Evaluation Criteria In Solid Tumors)

"First you need to identify the longest [in plane] diameter of a lymph node or nodal mass (here 56.0mm) and then choose the longest perpendicular diameter to that as the short axis (here 45.3mm)."

Source: <http://www.recist.org/recist-in-practice/19.html>



## Beispiel: Vergrößerter Lymphknoten mit 1,8 cm Durchmesser.



Die Abbildung zeigt das Schema eines Messungs-Befund einer Länge. Übersichtshalber wurde auf eine Unterscheidung von Klassen und Instanzen verzichtet.

## Typen von Größen-Beschreibungen von anatomischen Entitäten:

<b>Intervall</b>	<ul style="list-style-type: none"> <li>Anterior-posterior Durchmesser der Leber normalerweise 10-13 cm</li> <li>Dicke der Gallenblasenwand normalerweise 0,1 - 0,3 cm</li> </ul>
<b>Normalwert mit Abweichung</b>	<ul style="list-style-type: none"> <li>Truncus pulmonalis: 1,4 cm +/- 0,4 cm</li> </ul>
<b>Obere Grenze</b>	<ul style="list-style-type: none"> <li>Normaler Lymphknoten &lt; 1 cm</li> </ul>
<b>Untere Grenze</b>	<ul style="list-style-type: none"> <li>Normaler Aorta Durchmesser &gt; 4 cm an der Wurzel</li> <li>Lymphknoten vergrößert ab &gt; 1 cm</li> </ul>

**Grundlegende Form:** Anatomische Entität, Eigenschaft, Wert-Spezifikation  
**Wissensmodell:** 50 Größen-Spezifikationen für 38 verschiedene anatomische Entitäten; einige davon alters- oder geschlechts-abhängig

Drei Beispiel-Sätze:

- Leber mit kranio-kaudalem Durchmesser von **15,5 cm**.
- Größenprogrediente, unscharf abgrenzbare Hypodensität links temporal nach kranial bis nach parietobasal reichend (IMA 7-22; aktuell etwa **8 x 7 x 6 cm** - Voruntersuchung etwa **4,5 x 3,5 cm**) mit einzelnen, neuaufgetretenen, stippchenförmigen Hyperdensitäten (IMA 11-14).
- Etwas kaudal hiervon im Unterlappen am Lappenspalt zentral ein **1.3 cm** (VU **1.3 cm**) großer Rundherd mit weiterhin deutlich vermehrtem FDG-Uptake (SUV max. 3.9; VU 5.7; IMA 182) im Oberlappen lappenspaltnah ein **1.0 cm** (VU **1.0 cm**) großer Rundherd mit vermehrtem FDG-Uptake (SUV max. **0.8**; VU **1.5**; IMA 199) sowie auf gleicher Höhe im Unterlappen dorsal paravertebral zwei Rundherde mit Ausläufern von **1.5 cm** (VU **1.3 cm**) und lateral hiervon zwei verschmolzene Lymphknoten von zusammen **1.7 cm** Durchmesser (VU **1.5 cm** + Satellit von **0.9 cm**) mit deutlich vermehrtem FDG-Uptake (SUV max. **4.0**; VU **3.2** bzw. SUV max. **6.6**; VU **4.8**; IMA 207) und im costophrenischen Winkel dorsal ein **0.9 cm** (VU **0.5 cm**) großer Rundherd mit vermehrtem FDG-Uptake (SUV max. **1.7**; VU **1.7**; IMA 234).

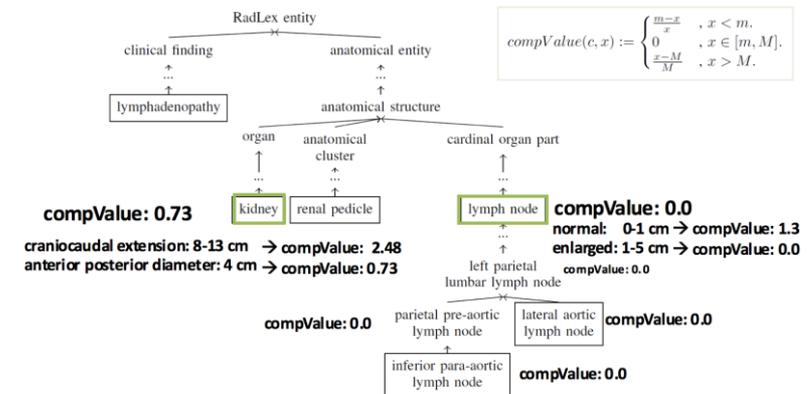
**Befund:** "Vergrößerter Lymphknoten rechts paraaortal unterhalb des Nierenstiels jetzt 23 mm."

**RadLex** (Radiologisches Lexikon): *Terminologie* für die Radiologie mit hauptsächlich taxonomischen, aber auch ontologischen Strukturen; 34.446 Klassen, 58.065 Terme (Englisch und Deutsch)

**Annotationen:**

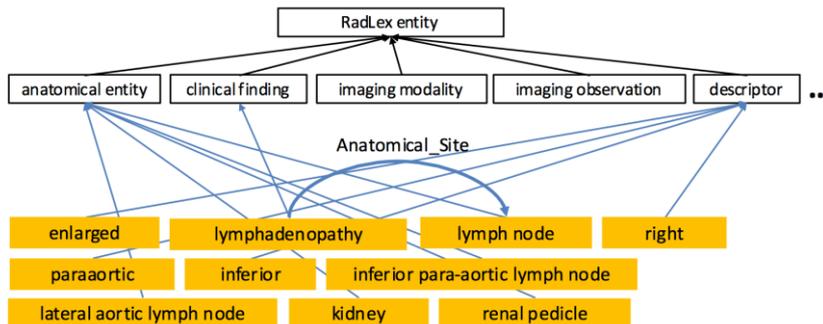
- radlex:enlarged
- radlex:lymphadenopathy
- radlex:lymph node
- radlex:right
- radlex:paraaortic
- radlex:inferior
- radlex:inferior para-aortic lymph node
- radlex:lateral aortic lymph node
- radlex:kidney
- radlex:renal pedicle
- 2.3
- uo:centimeter

Für jeden Knoten aus dem Spannbaum werden Normalwerte aus dem Wissensmodell hinzugefügt.



**Befund:** "Vergrößerter Lymphknoten rechts paraaortal unterhalb des Nierenstiels jetzt 23 mm."

- Filtern und Erweitern der Menge der Annotationen
- Nutzen des Wissens aus RadLex



## Daten:

### Radiologie-Berichte

382 Lymphoma Patienten

- 2700 Berichte (CT, MRI, US, ...)
- Bild-Annotationen

Div. Internistische Patienten

- 6007 Berichte (CT)

Div. Chirurgie Patienten

- 7713 Berichte (CT)

Div. Neuroradiologische Pat.

- 5949 Berichte (CT)

10 Mamma-Patienten

- ~50 Berichte (CT)

Div. Englische Berichte

- ~3000 Berichte

### Vollständige Daten für 6 Melanoma Patienten

- Demographie
- Diagnosen (ICD-10 Kodes)
- Laborwerte (LOINC Kodes)
- Behandlungen (OPS Kodes)
- Medikamente (ATC Kodes)
- Radiologie Berichte
- Pathologie Berichte
- Alle Daten mit Zeitstempel

### Strukturierte Daten für ca. 4000 Nephrologie Pat.

- Zeitintervall 15-20 Jahre

## Methode:

Evaluierung durch einen Radiologen

## Daten:

### Radiologie-Berichte

382 Lymphoma Patienten

- 2700 Berichte (CT, MRI, US, ...)
- Bild-Annotationen

Div. Internistische Patienten

### Vollständige Daten für 6 Melanoma Patienten

- Demographie
- Diagnosen

### Strukturierte Daten für ca. 4000 Nephrologen

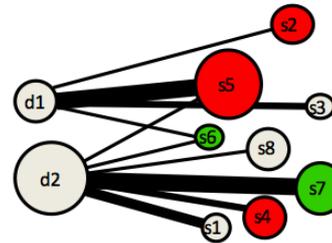
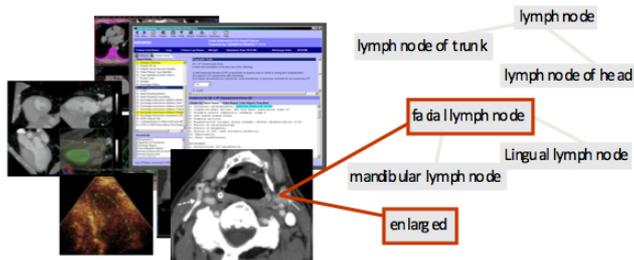
Evaluationsergebnisse für jeweils 500 zufällig ausgewählte Sätze je Datensatz.

Datensatz	Accuracy
Radiologie-Berichte von Lymphoma Patienten (CT, MRT, US...)	80,8 %
CT Radiologie-Berichte von div. internistischen Patienten	80,6 %
Englische Berichte	74,4 %

N

Evaluierung durch einen Radiologen

## Zur Unterstützung in der Differential-Diagnose.



### Unstrukturierte klinische Data

### Annotationen

- Nutzung medizinischer Ontologien RadLex, FMA, SNOMED
- Semi-automatische Erstellung mittels Bild-Annotationen und Informationsextraktion

### Krankheits-Symptom Modell

- Krankheits-Symptom Relationen
- Unterscheidung von Leitsymptomen
- Heuristik zur Wichtigkeit von Symptomen
- Symptom-Untersuchung Relationen

### Ranking Algorithmus

- Abbildung der Symptome des Patienten zu den typischen Symptomen von Krankheiten
- Ranking abzuklärender Symptome

### Publikationen:

1. **Interpreting Patient Data using Medical Background Knowledge** Heiner Oberkampff, Sonja Zillner, Bernhard Bauer, Matthias Hammon. ICBO proceedings. Graz, Austria, 2012
2. **Towards a Ranking of Likely Diseases in Terms of Precision and Recall** Heiner Oberkampff, Sonja Zillner, Bernhard Bauer, Matthias Hammon. ECAI International Workshop on Artificial Intelligence and NetMedicine. Montpellier, France, 2012

- Erstellung des Datenmodells für klinische Informationen
- Entwicklung eines wissensbasierten Ansatzes zur Extrahierung von strukturierten Messungs-Informationen von Freitext Radiologie Befunden
- Automatische wissensbasierte Anreicherung von klinischen Daten um verschiedene Zugänge und Sichten zu ermöglichen (krankheits-zentrisch, Verlaufsdarstellung, Selektion von Auffälligkeiten etc.)

“Enterprise architecture is the process of translating business vision and strategy into effective enterprise change by creating, communicating and improving key principles and models that describe the enterprise’s future state and enable its evolution.”

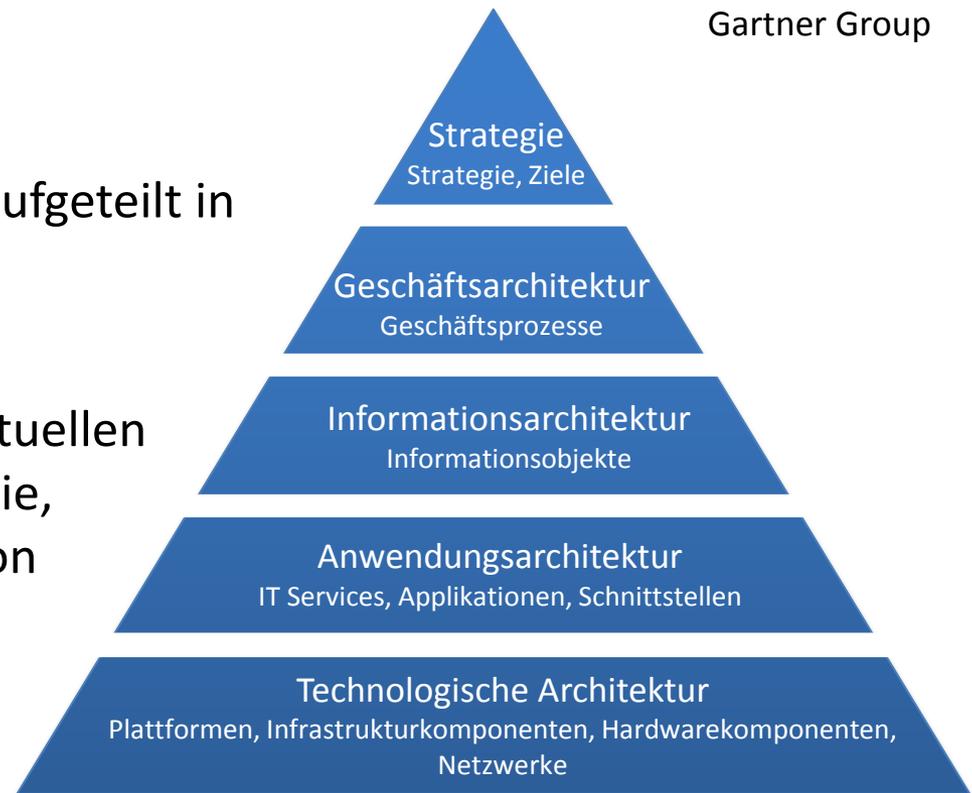
Gartner Group

## Statische Sicht:

Unternehmensarchitekturmodell, aufgeteilt in verschiedene Ebenen

## Dynamische Sicht:

Unterstützung der Evolution des aktuellen Modells ausgehend von der Strategie, durch Definition und Kommunikation zukünftiger Prinzipien und Modelle



Danke für Ihre Aufmerksamkeit!

 **Software Methodologies**  
for Distributed Systems