



30. Juni 2017

## Automatically Populating a Preclinical Spinal Cord Injury Knowledge Base to Support Clinical Translation (PSINK)

Matthias Hartung and the PSINK Consortium

## Project and Partners

- Project funded by BMBF for 3+2 years in program iD:Sem
- Uni Bielefeld
  - Philipp Cimiano (PI), Roman Klinger (Co-PI, Uni Stuttgart)
  - Matthias Hartung (Postdoc), Hendrik ter Horst (Ph.D. student)
- HHU Düsseldorf, University Hospital
  - (with Center for Neuroregeneration)
  - Hans Werner Müller (PI), Nicole Brazda (Co-PI, Postdoc)
  - Veronica Estrada, Katinka Fischer
- Coordination Center for Clinical Trials Düsseldorf
  - Christian Ohmann, Wolfgang Kuchinke, Gerd Felder

# Überblick

- 1 Motivation und Ausgangslage
- 2 Ziele des Projekts
- 3 Meilensteine
- 4 Zwischenergebnisse
- 5 Fazit und Ausblick

## Motivation und Ausgangslage

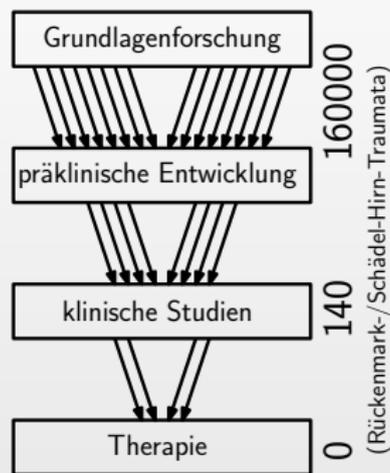
# Motivation

### Ausgangspunkt

- Zu vielfältigen Symptomen und Krankheiten gibt es **umfangreiche präklinische Forschung**, aber nur **unzureichende Therapiemöglichkeiten**

### Beispiele

- Rückenmarksverletzungen, neurodegenerative, Tumor-, Herz-Kreislaufkrankungen



## Motivation und Ausgangslage

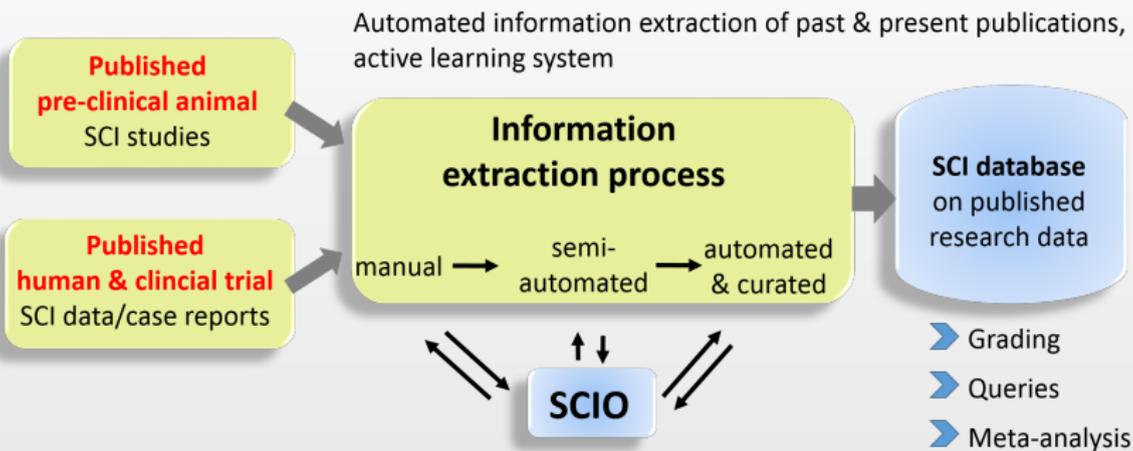
# SCI: Präklinische Studien

### Herausforderungen

- Unstrukturierte Daten: +8000 Publikationen pro Jahr
- Fehlende Reproduktion von Experimenten
- Uneinheitliche Reporting-Standards
- Übertragbarkeit vom Tierexperiment auf den Menschen
- Datenlage für Metastudien

## Ziele des Projekts

# Überblick



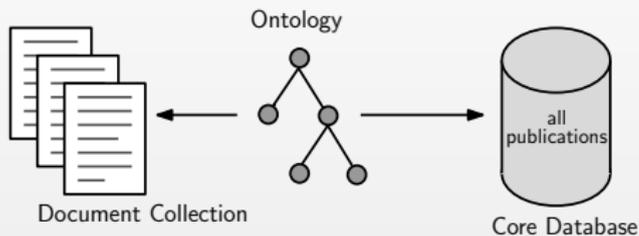
## Ziele des Projekts

# Translation therapeutischer Konzepte

- Entwicklung eines “Grading Schemes” zur Bewertung des Translationspotentials präklinischer Studien
  - Identifikation relevanter Parameter
  - Nutzerspezifisch, manuell konfigurierbar
- Bereitstellung präklinischen Wissens im klinischen Alltag
- Unterstützung von Therapieentscheidungen durch Einbeziehung präklinischer Erkenntnisse

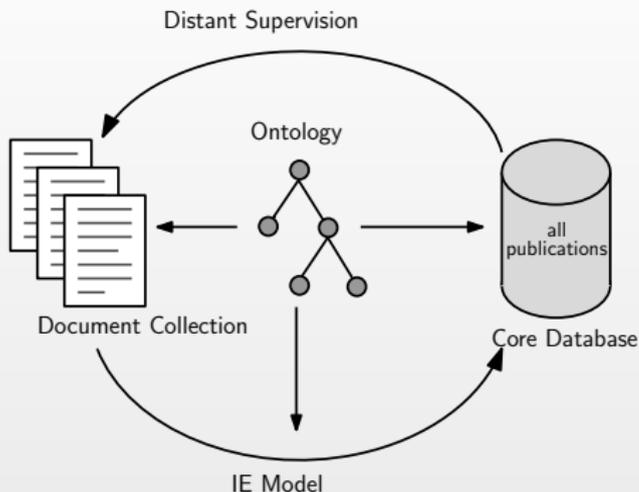
## Ziele des Projekts

# Analyse der bestehenden Literatur



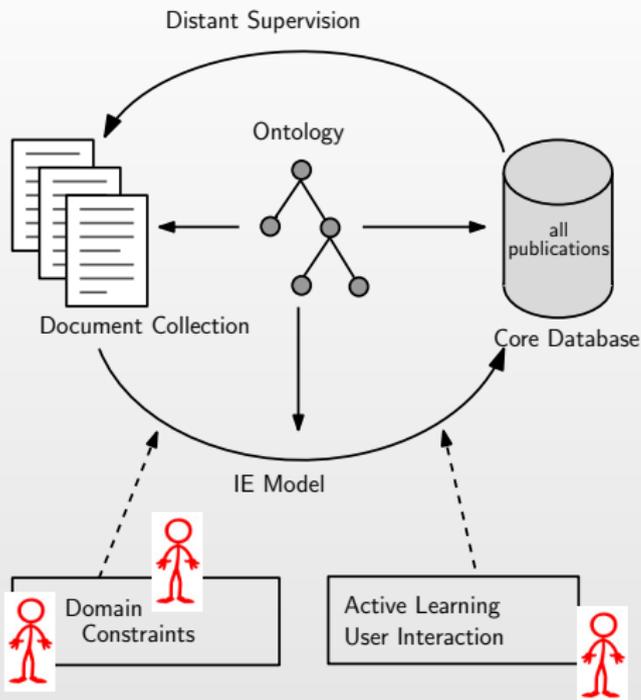
# Ziele des Projekts

## Analyse der bestehenden Literatur



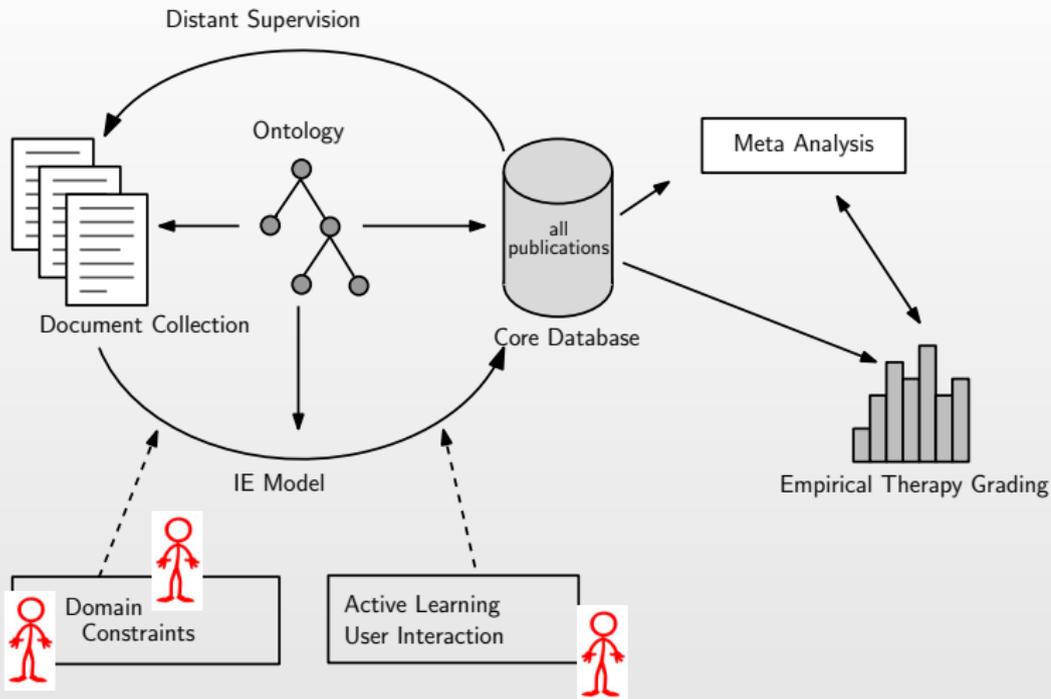
# Ziele des Projekts

## Analyse der bestehenden Literatur



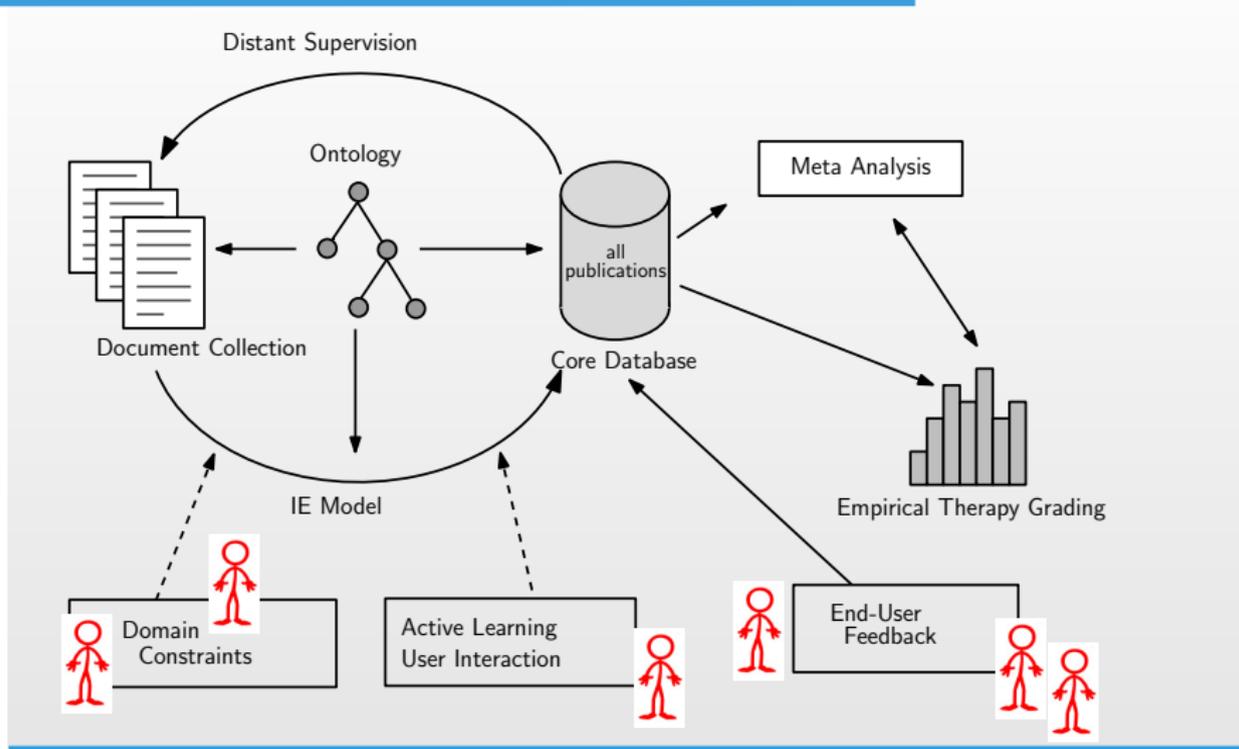
# Ziele des Projekts

## Analyse der bestehenden Literatur



# Ziele des Projekts

## Analyse der bestehenden Literatur



# Meilensteine Übersicht

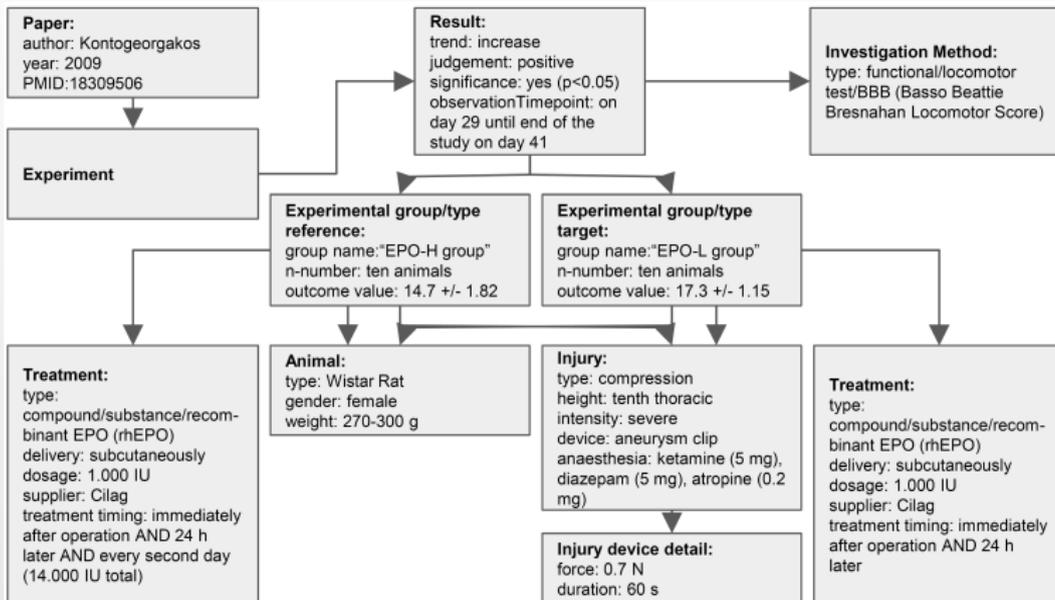
- 1 Motivation und Ausgangslage
- 2 Ziele des Projekts
- 3 Meilensteine**
- 4 Zwischenergebnisse
- 5 Fazit und Ausblick



## Meilensteine Informationsextraktion und interaktiver Lernzyklus

- Entwicklung eines probabilistischen Modells zur automatischen Informationsextraktion aus wissenschaftlichen Publikationen
- Workflow zur automatisierten Population einer Datenbank
- Feedback-Algorithmen zur effizienten Kuratierung der extrahierten Wissensstrukturen durch Experten

# Meilensteine Population der Datenbank



## Zwischenergebnisse Übersicht

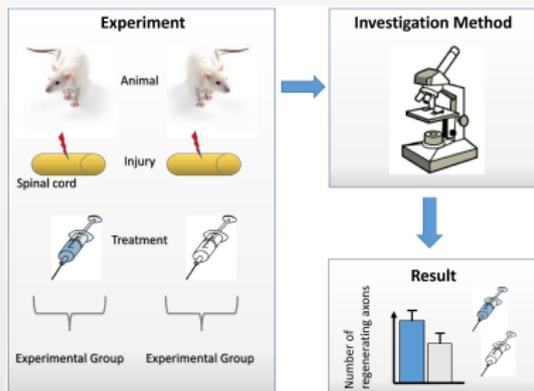
- 1 Motivation und Ausgangslage
- 2 Ziele des Projekts
- 3 Meilensteine
- 4 Zwischenergebnisse**
- 5 Fazit und Ausblick

## Zwischenergebnisse

# Übersicht

- Spinal Cord Injury Ontology (SCIO)
- Annotation wissenschaftlicher Publikationen
- Informationsextraktion mit probabilistischen graphischen Modellen
- Datenexploration

## Zwischenergebnisse Spinal Cord Injury Ontology (SCIO)



- Repräsentation des Wissens, das zur Modellierung präklinischer Studien erforderlich ist
- Methodik: Bottom-Up aus Fachpublikationen
- ca. 500 Entitäten, 80 Properties
- Integration verschiedener Referenzontologien: Time-Ontology, QUDT, SNOMED

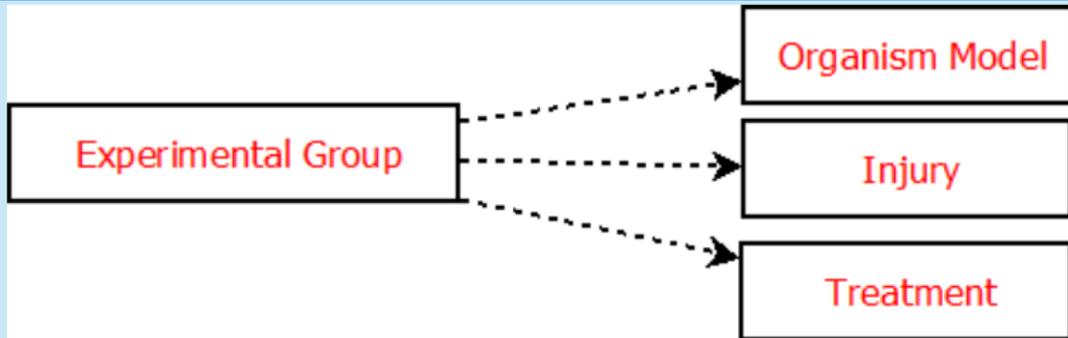
## Zwischenergebnisse Spinal Cord Injury Ontology (SCIO)

### Übersicht der Hauptklassen



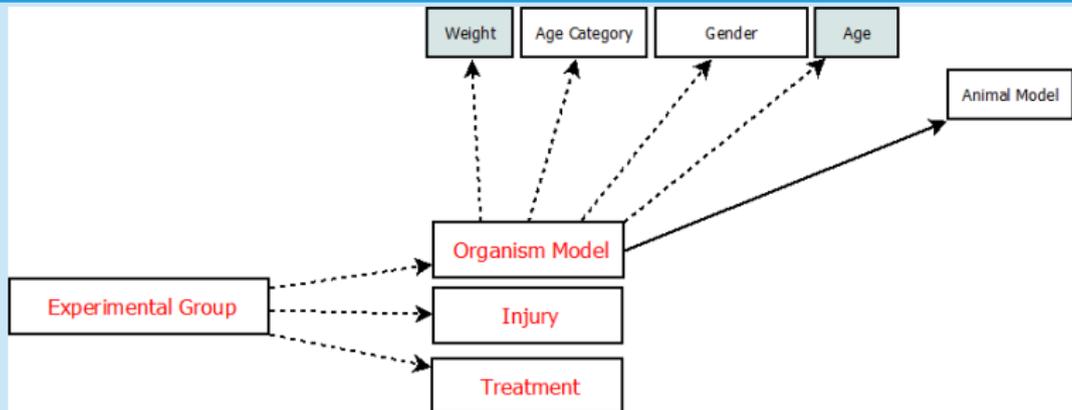
## Zwischenergebnisse Spinal Cord Injury Ontology (SCIO)

### Modellierung der Experimentellen Gruppen



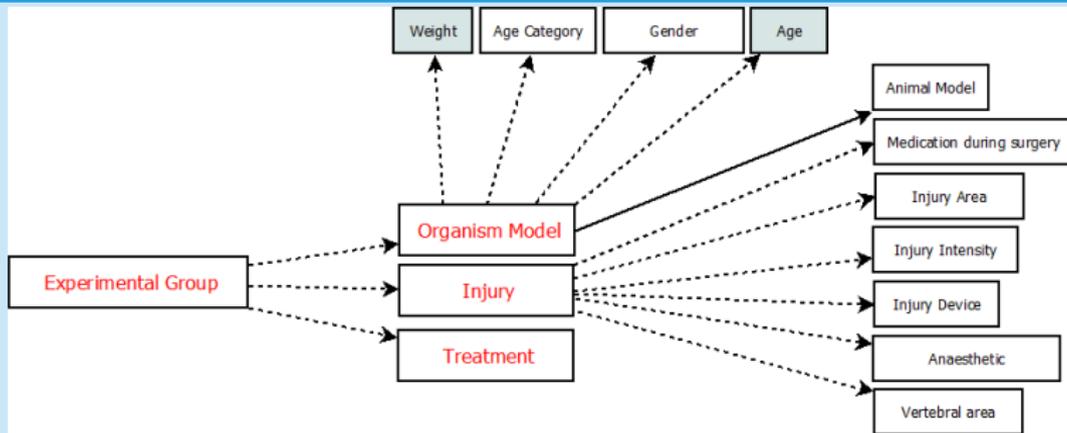
## Zwischenergebnisse Spinal Cord Injury Ontology (SCIO)

### Modellierung der Experimentellen Gruppen



## Zwischenergebnisse Spinal Cord Injury Ontology (SCIO)

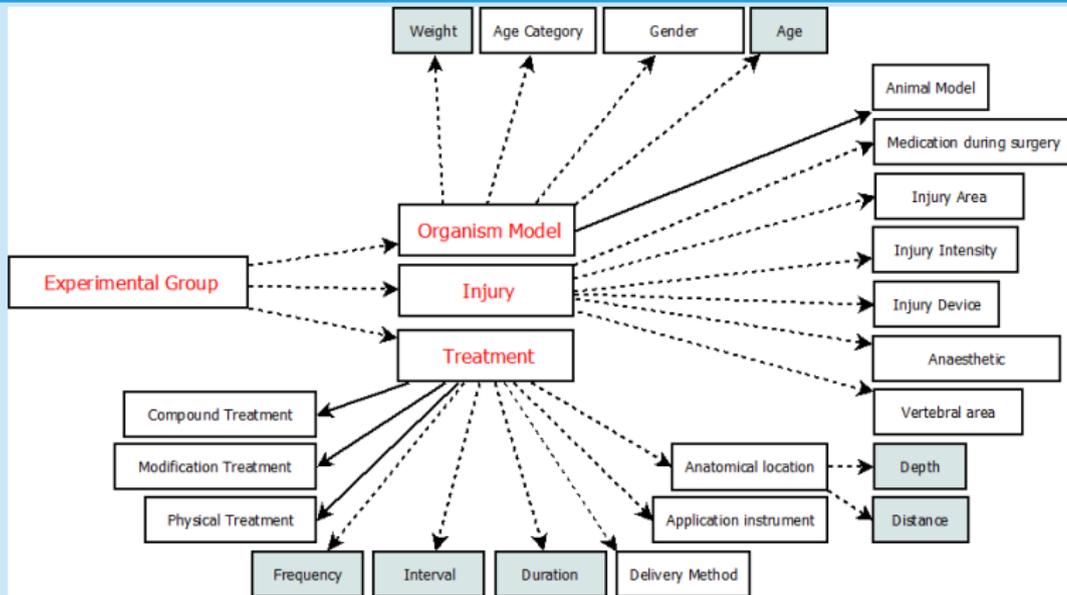
### Modellierung der Experimentellen Gruppen



# Zwischenergebnisse

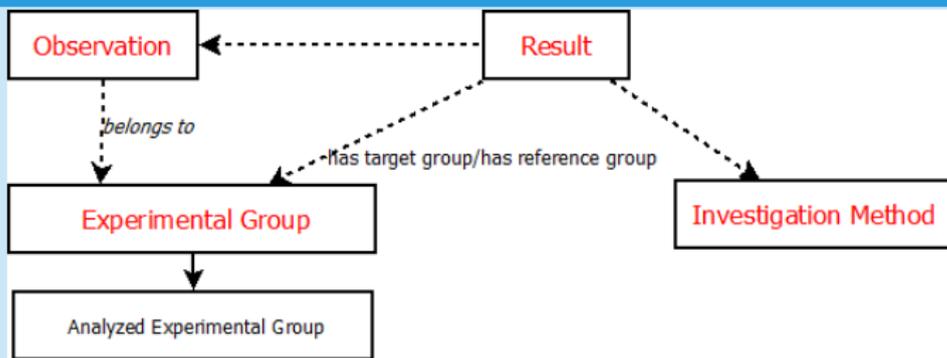
## Spinal Cord Injury Ontology (SCIO)

### Modellierung der Experimentellen Gruppen



## Zwischenergebnisse Spinal Cord Injury Ontology (SCIO)

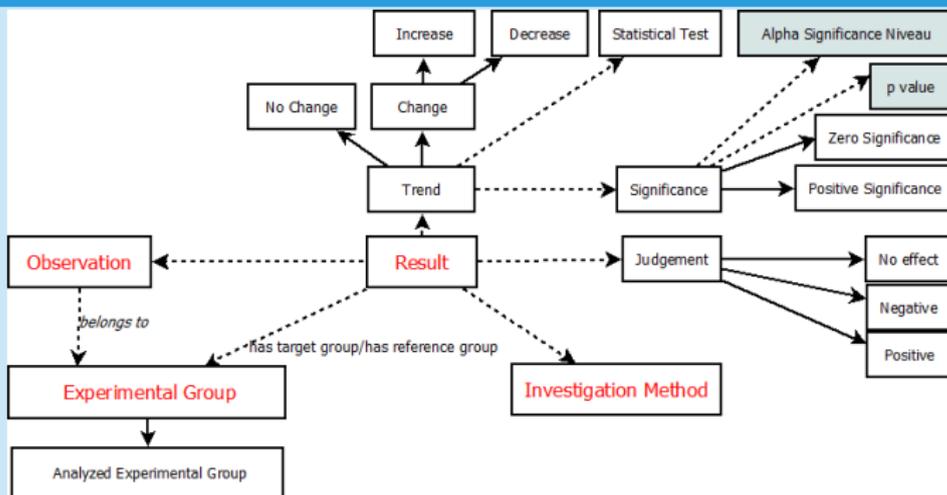
### Modellierung der Results



# Zwischenergebnisse

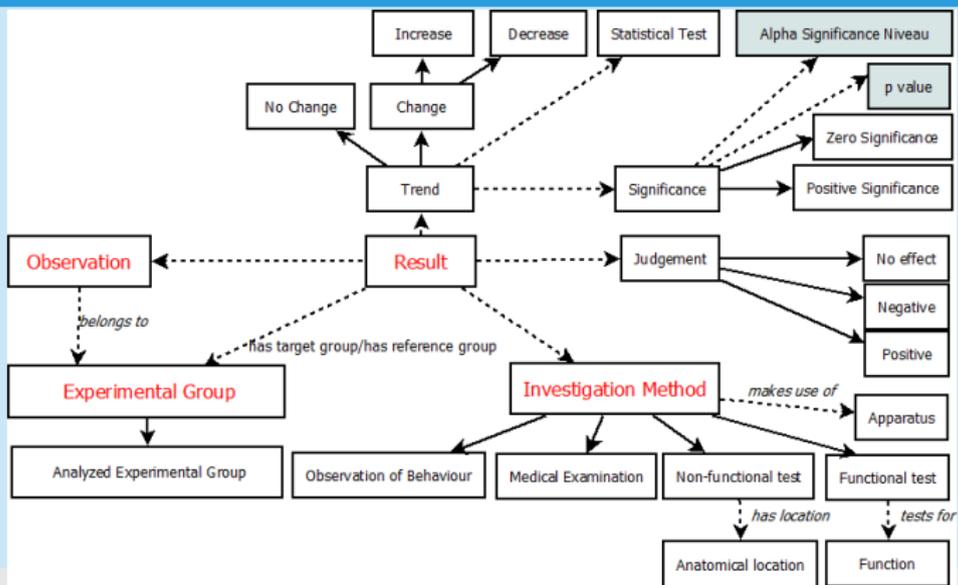
## Spinal Cord Injury Ontology (SCIO)

### Modellierung der Results

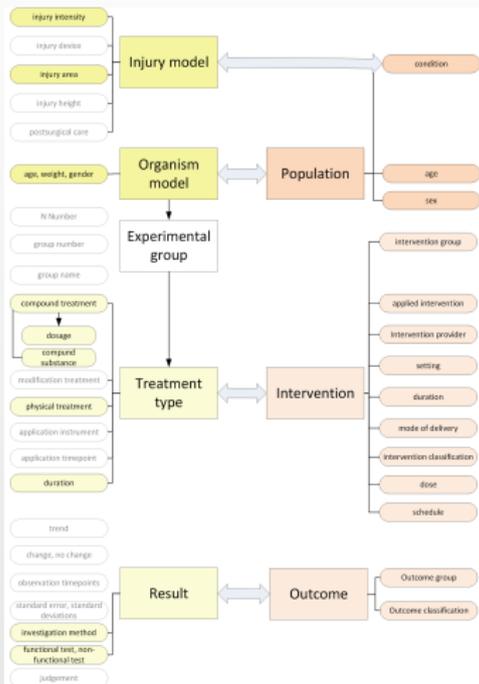


# Zwischenergebnisse Spinal Cord Injury Ontology (SCIO)

## Modellierung der Results



# Zwischenergebnisse Spinal Cord Injury Ontology (SCIO)



- Erweiterung des Einsatzbereiches von SCIO durch Mapping auf biomedizinische Ontologien
- Interoperabilität mit klinischen Studien (PICO-Ontologie)
- Mapping SCIO ↔ PICO auf Top Level
- Hauptunterschiede:
  - InjuryModel (induzierte Verletzung) vs. ClinicalCondition (Mischung von Verletzungen)
  - Result vs. Outcome (z.B. Analyse von Subgruppen)

## Zwischenergebnisse

# Übersicht

- Spinal Cord Injury Ontology (SCIO)
- Annotation wissenschaftlicher Publikationen
- Informationsextraktion mit probabilistischen graphischen Modellen
- Datenexploration

## Zwischenergebnisse Annotation präklinischer Studien



- Ableitung eines Annotations-schemas aus SCIO
- Annotationsrichtlinie
- 6 AnnotatorInnen bearbeiten ein Corpus aus 300 Publikationen
- Monitoring durch Abgleich gegen Gold-Standard und Inter Annotator Agreement

## Zwischenergebnisse

# Annotation präklinischer Studien

## Annotation mit Verankerung im Text (WebAnno)

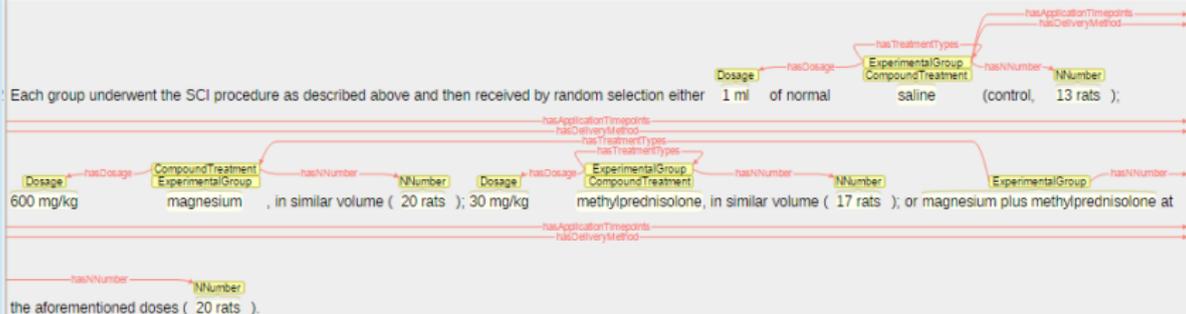
- 10 Here, we compared the effects of KS- or CS-digestion plus **Treatment** rehabilitation **RehabilitativeTraining** on **Recovery** recovery from spinal cord **Injury** injury.
- 11 Keratanase II or chondroitinase ABC was locally administered at the lesion **Enzyme** after spinal cord injury at C3/4 **ChondroitinaseABC** **DeliveryMethod** **LesionSite** **Timepoint** **Injury** **CS**.
- 12 Task-specific rehabilitation training, i.e., a single pellet reaching task using a Whishaw apparatus, was done for **RehabilitativeTraining** 3 **Duration** weeks before injury, and then again at 1-6 weeks after injury. **SinglePelletReachingTest** **Apparatus** **Timepoint** **Duration** **Timepoint**.
- 13 The combination of KS-digestion and rehabilitation yielded a **DefinedExperimentalGroup** better rate of **Increase** pellet removal **Judgement** than either **SinglePelletReachingTest** KS-digestion alone **DefinedExperimentalGroup** or **Treatment** rehabilitation alone **DefinedExperimentalGroup**, although these differences were not statistically significant. **ZeroSignificance** **RehabilitativeTraining**.
- 14 The combination of CS-digestion and rehabilitation showed similar results. **DefinedExperimentalGroup**
- 15 Strikingly, both KS-digestion/rehabilitation **DefinedExperimentalGroup** and CS-digestion/rehabilitation **DefinedExperimentalGroup** showed **PositiveSignificance** significant **Increase** increases in **AxonalSproutingTest** neurite growth in vivo as estimated by 5-hydroxytryptamine and GAP43 staining.

- Entitäten mit Verankerung im Text ✓
- überlappende und diskontinuierliche Entitäten ✓

## Zwischenergebnisse

# Annotation präklinischer Studien

### Annotation mit Verankerung im Text (WebAnno)



- Entitäten mit Verankerung im Text ✓
- überlappende und diskontinuierliche Entitäten ✓
- Relationen sehr unübersichtlich ✗

# Zwischenergebnisse

## Annotation präklinischer Studien

### Annotation als "Slot Filling" mit Referenzauflösung

49 (immediately after hypothermia) and late (after 24 h) stage following hypothermia treatment.

21 Materials and Method Experimental Procedure 5.

22 Bilgiç Department of Emergency Medicine, School of Medicine, Gulhane Military Medical Academia, Ankara, Turkey  
Animals and Study Groups A.

Korkmaz Department of Physiology, School of Medicine, Gulhane Military Medical Academia, Ankara, Turkey

23 (NNumber) (Adult) (Male) (SpragueDawleyRat) (Weight)  
Thirty-five adult, male, **Sprague Dawley rats** weighing 200-300 g were provided by the GAIA Research Center (Ankara, Turkey).

24 (Nutrition) (Hydration) (Hydration)  
They were allowed free access to food and water before and after surgery.

25 Rats were divided into five groups each comprising of (NNumber) animals as follows:  
Group 1a: Clip applied-non-treated group: In this group, the clip-applied segments were harvested 1 h after the injury.

26 (DefinedExperimentalGroup) (Clip) (Timepoint)  
Group 1b: Clip applied-treated group: In this group the injured segments were treated with hypothermia for (Duration) 1 h.

27 (Timepoint) (HypothermicTreatment)  
The injured segments were harvested at the end of the hypothermic treatment.

28 (DefinedExperimentalGroup) (Timepoint)  
Group 2a: Clip applied, non-treated group: In this group the injured segments were harvested 24 h after injury.

29 (DefinedExperimentalGroup) (SpinalCord) (HypothermicTreatment)  
Group 2b: Clip applied-treated group: In this group the injured segments were treated with hypothermia for (Timepoint) 1 h after clip application.

30 (Timepoint) (HypothermicTreatment)  
The injured segments were harvested 24 h following the initial hypothermia treatment.

31 (DefinedExperimentalGroup) (Laminectomy)  
Group 3: Sham-operated group: In this group, laminectomy was performed as in the other groups but no clip was applied.

32 In the study, Group 1a is designed as the control Group of 1b and Group 2a is the control Group of 2b.

33 Spinal cord segments of 5-6 mm were resected for examination.

34 The spinal cord with clip applied was centered in the resected segments.

(IntramuscularDelivery) (Ketamine&Anesthesia)

#### Results

#### Animal models

Adult Male Sprague Dawley Rats +

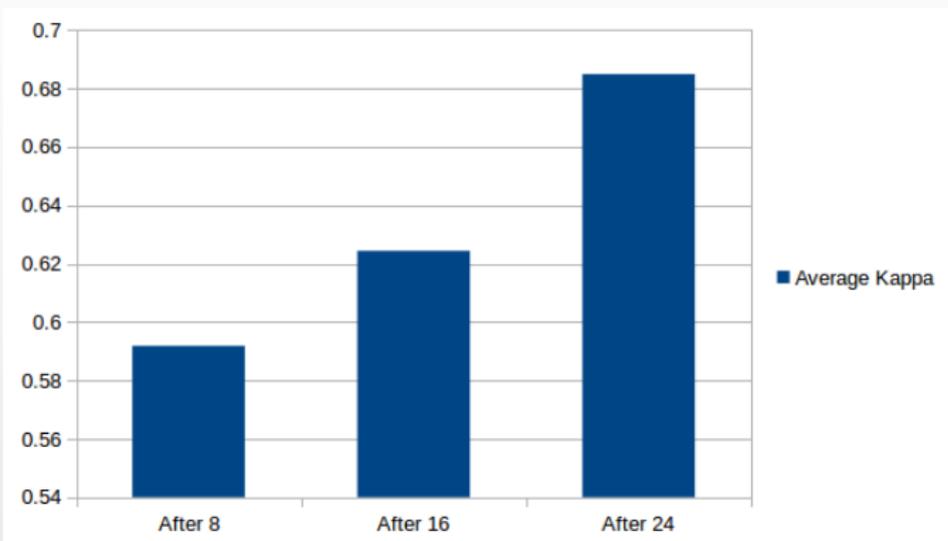
Type	OrganismModel	
Gender	Male (male)	*
Age		*
AgeCategory	Adult (adult)	*
AnimalSpecies	SpragueDawleyRat (Sprague Dawley rats)	*
Weight	Weight (200-300 g)	*

#### Injuries

#### Treatments

#### Experimental groups

## Zwischenergebnisse Inter Annotator Agreement (IAA)



- Fleiss' Kappa auf übereinstimmend annotierten Spans
- Lernkurve nach 24 paarweise annotierten Publikationen

## Zwischenergebnisse

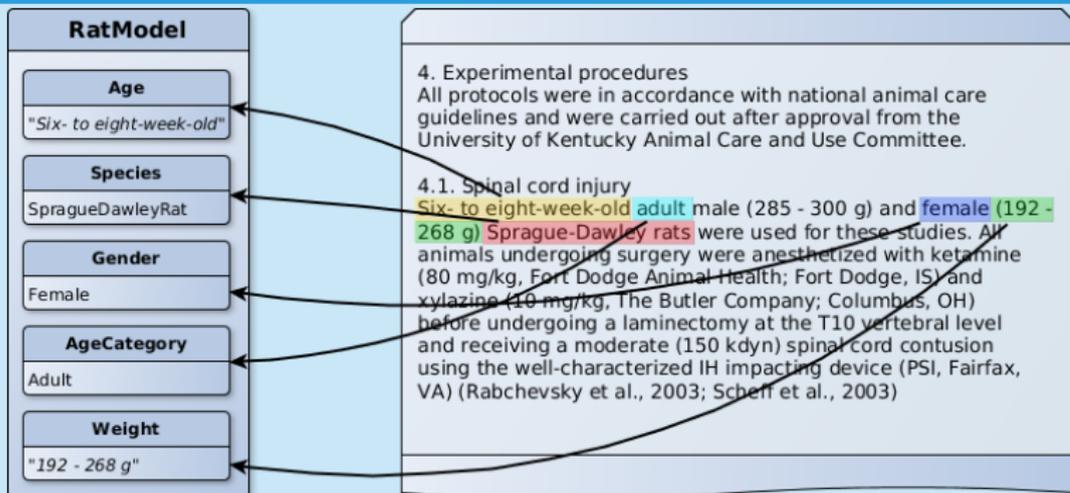
# Übersicht

- Spinal Cord Injury Ontology (SCIO)
- Annotation wissenschaftlicher Publikationen
- Informationsextraktion mit probabilistischen graphischen Modellen
- Datenexploration

## Zwischenergebnisse

# Informationsextraktion

### SCIO-Teilausschnitt: OrganismModel



## Zwischenergebnisse Informationsextraktion

### Probabilistische Graphische Modelle

- Faktorgraph  $\mathcal{G}$ :  
beobachtete Variablen  $\vec{x}_i$ , latente Variablen  $\vec{y}$ , Faktoren  $\psi_i$
- Faktoren berechnen einen Score aus einem Merkmalsvektor  $f_i(\vec{x}_i, \vec{y}_i)$  und einem Vektor gelernter Parameter  $\theta_i$ :

$$\psi_i = e^{f_i(\vec{x}_i, \vec{y}_i) \cdot \theta_i}$$

- Inferenz:

$$p(\vec{y} | \vec{x}; \theta) = \frac{1}{Z(\vec{x})} \prod_{\psi_i \in \mathcal{G}} \psi_i = \frac{1}{Z(\vec{x})} \prod_{\psi_i \in \mathcal{G}} e^{f_i(\vec{x}_i, \vec{y}_i) \cdot \theta_i}$$

## Zwischenergebnisse

# Informationsextraktion

### SCIO-Teilausschnitt: OrganismModel

RatModel
<b>Age</b> "Six- to eight-week-old"
<b>Species</b> SpragueDawleyRat
<b>Gender</b> Female
<b>AgeCategory</b> Adult
<b>Weight</b> "192 - 268 g"

- regelbasierte Entitätserkennung zur Generierung von Kandidaten
- maschinell gelerntes Faktorgraph-Modell zum Slot Filling
- Evaluierung (25 Publikationen, 10-fold CV):
  - Accuracy: 0.78
  - Baseline: 0.58
- Ziel: sukzessive Erweiterung auf weitere Teilbäume der SCIO

## Zwischenergebnisse

# Übersicht

- Spinal Cord Injury Ontology (SCIO)
- Annotation wissenschaftlicher Publikationen
- Informationsextraktion mit probabilistischen graphischen Modellen
- Datenexploration

## Zwischenergebnisse Datenexploration

- Interaktion des Endnutzers mit extrahiertem Wissen
- Inhalt relevanter Studie(n) “auf einen Blick”
- Exploration der Wissensbestände zur Beantwortung von Kompetenzfragen

## Zwischenergebnisse Kompetenzfragen

- In wie vielen Fällen ergab die Therapie mit Erythropoietin eine Verbesserung im Tierexperiment bzw. keinen Unterschied im Vergleich zur Kontrollgruppe?
- In welchen Läsionsmodellen waren keine positiven Effekte von Erythropoietin zu sehen?
- Bei welchem sensorischen Test war Erythropoietin wirkungslos im Tiermodell?

## Zwischenergebnisse Datenexploration

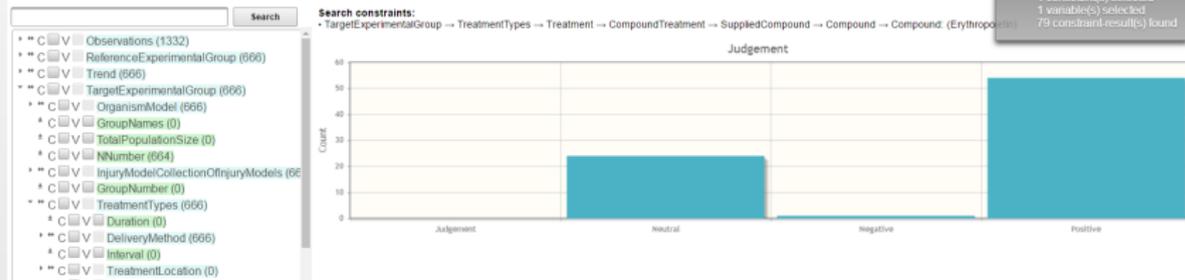
### SCIO Explorer

- Web-Applikation
- RDF-Datenschema, aus der Ontologie abgeleitet
- Einschränkung der Suchergebnisse mittels semantischer Filter über Variablen aus der Ontologie
- Anwendung auf Kompetenzfragen
- Usability-Studie; gleichzeitig Rückschlüsse auf User Requirements

## Zwischenergebnisse Datenexploration

In wie vielen Fällen ergab die Therapie mit Erythropoietin eine Verbesserung im Tierexperiment bzw. keinen Unterschied im Vergleich zur Kontrollgruppe?

SCIO-Explorer



## Fazit und Ausblick

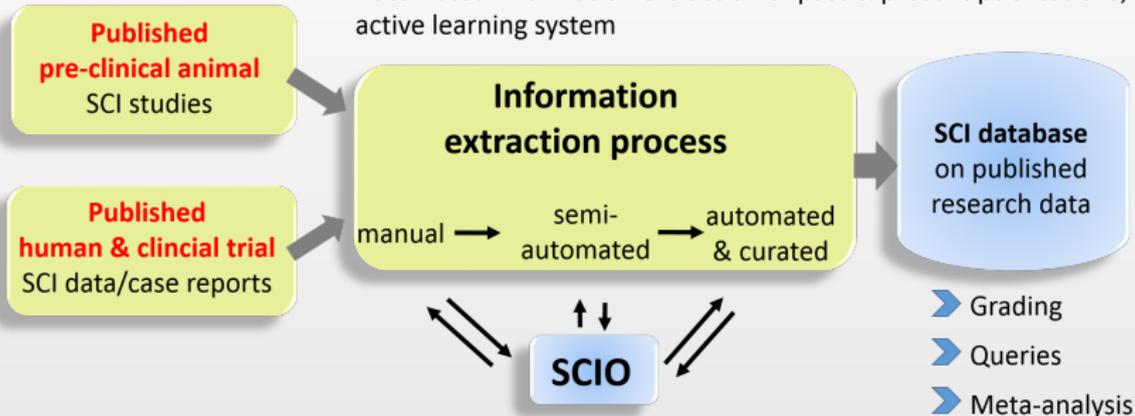
# Übersicht

- 1 Motivation und Ausgangslage
- 2 Ziele des Projekts
- 3 Meilensteine
- 4 Zwischenergebnisse
- 5 **Fazit und Ausblick**

## Fazit und Ausblick

# Nächste Schritte

Automated information extraction of past & present publications,  
active learning system



## Fazit und Ausblick Vielen Dank...

...für Ihre Aufmerksamkeit!



<http://www.psink.de>