

Kolloquium

# Wissensinfrastruktur

**Forschungsdatenmanagement am  
Center of Excellence  
Cognitive Interaction Technology**

**Philipp Cimiano und Dennis Spohr**

# Überblick

## CITEC – Fakten

Cognitive Interaction Technology

Mission und Struktur

## Aktuelle Studie zum Publikations- und Forschungsdatenmanagement

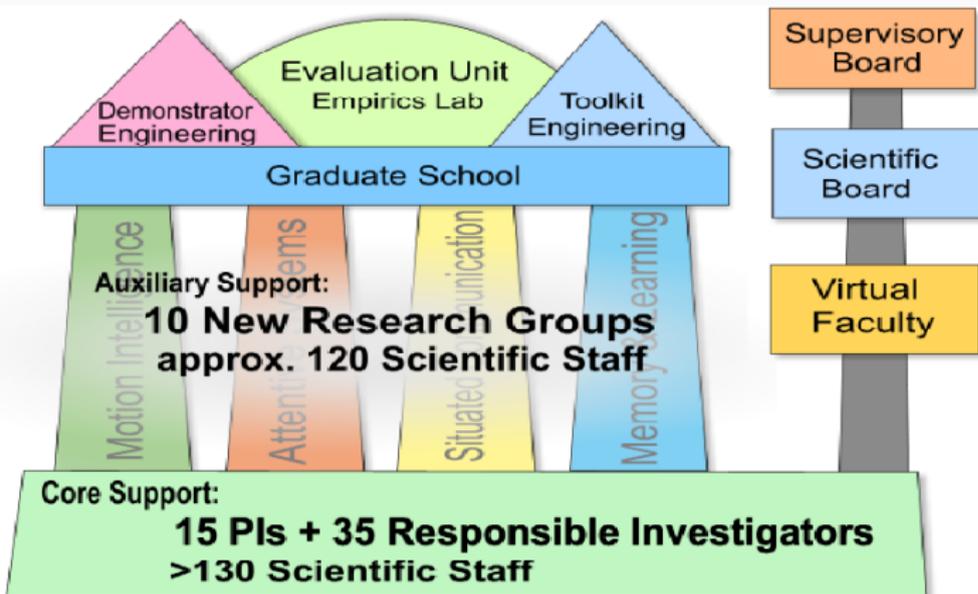
Rahmen und Umfang der Studie

Erste Ergebnisse

## Gegenwärtige und bevorstehende Entwicklungen



# Struktur des CITEC

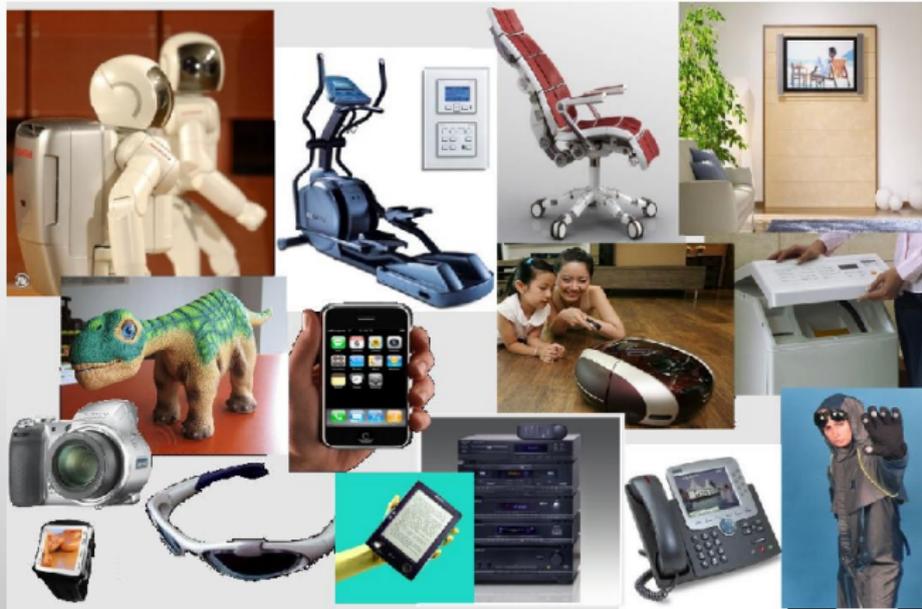


# Cognitive Interaction Technology

## Herausforderungen

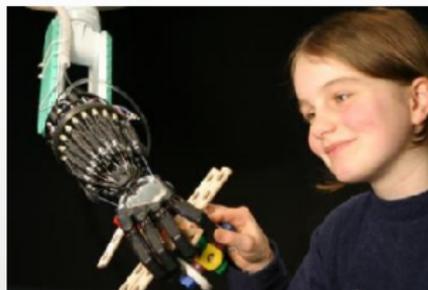
- ▶ **Technologie**
  - ▶ Handhabung technischer Systeme mit der Leichtigkeit und Natürlichkeit menschlicher Kommunikation
- ▶ **Grundlagenforschung**
  - ▶ Neue Einsichten in die Prinzipien und Mechanismen kognitiver Interaktion
- ▶ **Interdisziplinarität**
  - ▶ Verbindung von **geisteswissenschaftlicher** und **naturwissenschaftlicher** Forschung mit **ingenieurwissenschaftlicher** Entwicklung, um Technologie in unseren Alltag zu integrieren

# Cognitive Interaction Technology Anwendungen



# Forschungsfelder (1/2)

- ▶ **A: Bewegungsintelligenz**
  - ▶ Kopplung von Wahrnehmung und Handlung
  
- ▶ **B: Systeme mit Aufmerksamkeit**
  - ▶ Lernen der Fähigkeit, Aufmerksamkeit mit Menschen zu teilen und unwichtige Details zu ignorieren



## Forschungsfelder (2/2)

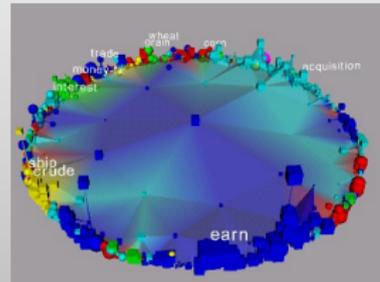
### ► C: Situierete Kommunikation

- Koordination von Sprache, Wahrnehmung und Aktionen für möglichst natürliche Kommunikation zwischen technischen Systemen und Menschen



### ► D: Gedächtnis und Lernen

- Erforschung kognitiver Architekturen und Mechanismen für das Lernen, Speichern und Abrufen von Wissen



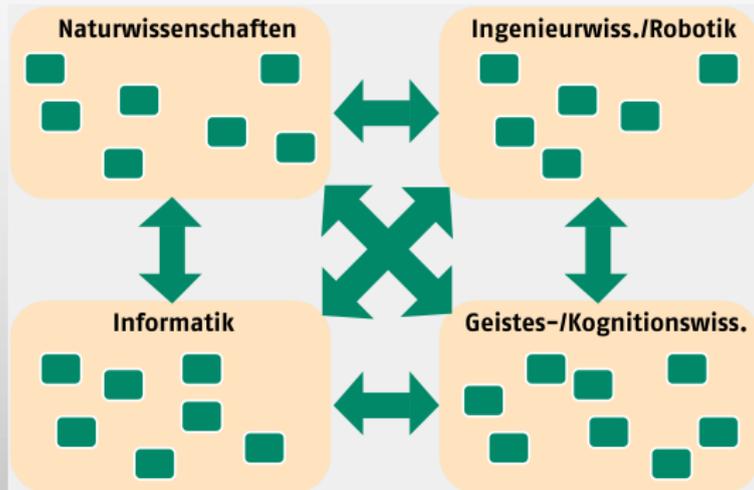
# Mission des CITEC

“Wir glauben, dass das duale Ziel der Kombination von Grundlagenforschung mit der technologischen Anwendung, welches eine signifikante Erweiterung der Kognition an sich zum Ziel hat, nur durch **intensive interdisziplinäre Kooperationen** erreicht werden kann.”

<http://www.cit-ec.de/de/CITECMission>

# Struktur des CITEC

**32 Arbeitsgruppen aus unterschiedlichen Forschungsdisziplinen,**  
strikte Einordnung in traditionelle Disziplinen oftmals schwer



# Anforderungen durch Interdisziplinarität

- ▶ **Diverse Forschungsumgebungen und -instrumente**
  - ▶ Labs mit Videokameras, Audiogeräten, Eyetrackern, hochsensiblen Messinstrumenten, ...
- ▶ **Diverse Forschungsprodukte**
  - ▶ Literatur, Methodologien
  - ▶ Experimentelle Primärdaten und angereicherte Sekundärdaten
  - ▶ Formate (Textdaten, Binärformate)
  - ▶ Software (Algorithmen, Modelle, ...)
- ▶ **Gelebte Heterogenität**
  - ▶ Alteingesessene Institute und Praktiken neben neu gegründeten Arbeitsgruppen

# Überblick

## CITEC – Fakten

Cognitive Interaction Technology

Mission und Struktur

## Aktuelle Studie zum Publikations- und Forschungsdatenmanagement

Rahmen und Umfang der Studie

Erste Ergebnisse

Gegenwärtige und bevorstehende Entwicklungen



# OpenAIRE

## EU-Projekt "Open Access Infrastructure for Research in Europe"

### ► Ziele

- Unterstützung von Wissenschaftlern beim Ablegen ihrer FP7-Publikationen (European Helpdesk)
- Einrichtung einer e-Infrastruktur zum Umgang mit wissenschaftlichen Artefakten (Archivierung, Suchfunktionen etc.)

# OpenAIRE

## EU-Projekt "Open Access Infrastructure for Research in Europe"

### ▶ Ziele

- ▶ Unterstützung von Wissenschaftlern beim Ablegen ihrer FP7-Publikationen (European Helpdesk)
- ▶ Einrichtung einer e-Infrastruktur zum Umgang mit wissenschaftlichen Artefakten (Archivierung, Suchfunktionen etc.)

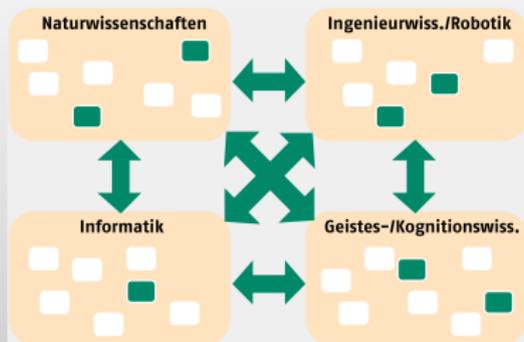
### ▶ WP7: Subject-specific requirements and data challenges

- ▶ Erforschung bestehender Praktiken, Workflows etc. zum Umgang mit Forschungsdaten
- ▶ Definition disziplinspezifischer Anforderungen an eine OpenAccess-Infrastruktur
- ▶ Parallele Studie an 7 europäischen Forschungseinrichtungen

# Methoden (1/2)

## Vorstudie zum Methodeneinsatz am CITEC

- ▶ Durchführung: 08–12/2010 in Kooperation mit der Universitätsbibliothek (Wolfram Horstmann, Christian Meyer zu Verl)
- ▶ Auswahl: 7 von 32 Gruppen aus verschiedenen Bereichen

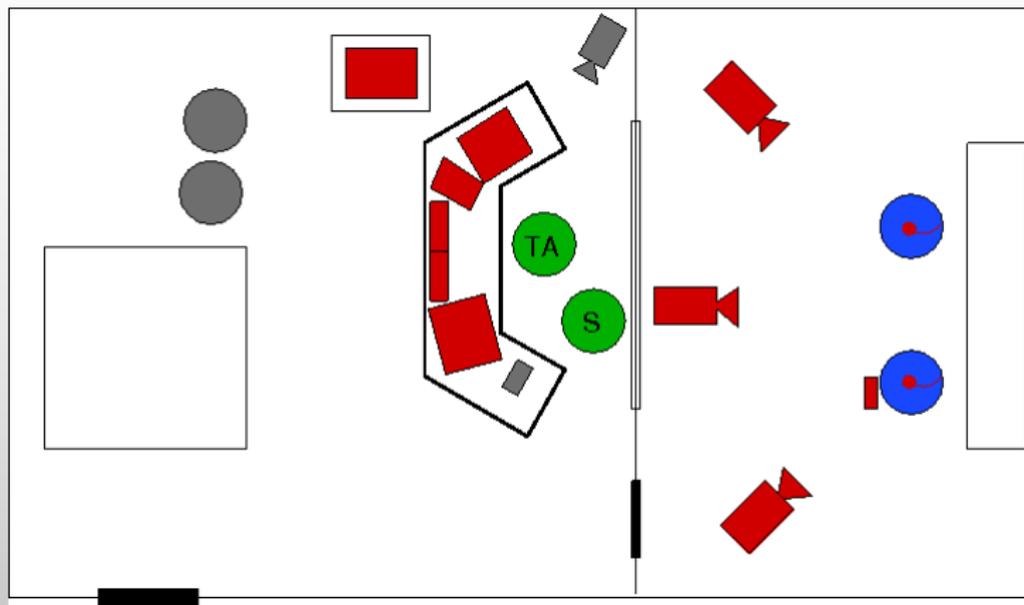


- ▶ **Seit Anfang Januar 2011:**  
Ausweitung der Studie auf alle Forschungsgruppen

# Methoden (2/2)

- ▶ **Vorgespräch**
- ▶ **Beobachtung** von Forschungsinstrumenten und Workflow
- ▶ **Strukturierte Interviews** mit Detailfragen zum Publikations- und Datenmanagement
- ▶ **Fragebogen** auf Basis der strukturierten Interviews
- ▶ **Quantitative Analyse des Publikationsverhaltens**

# Beispiel eines Versuchsaufbaus



# Forschungsdatenmanagement (1/4)

## Beispiel eines Data lifecycles

**Datenerhebung**  
Aufnahme

**Archivierung**  
Mitschnitte  
auf DVD und DAT

**Anreicherung**  
Annotation,  
z.B. ELAN, PRAAT

**Archivierung**  
SVN-Repository  
der Abteilung

**Weiterverarbeitung**  
Stat. Analyse mit R,  
SPSS, eig. Software

**Archivierung**  
Server für  
XML-Korpora

# Forschungsdatenmanagement (2/4)

## Projektarbeit

- ▶ **CITEC OpenSource Server:**  
Repository für am CITEC entwickelte OpenSource-Software
- ▶ **CITEC Subversion Server:**  
System zur Versionsverwaltung von Projektdaten
- ▶ **Redmine:**  
Tool für kollaborative Verwaltung von Projekten
- ▶ **Social network:**  
Austausch zwischen Wissenschaftlern, Wissensmanagement

# Forschungsdatenmanagement (3/4)

## Literatur

### Speicherung von Literatur

Server  
(vereinzelt PUB)

Workstation

### Metadaten- management

Software:  
Reference  
Manager,  
oder Jabref/  
Bibtex

### Erzeugung von Literatur

Software:  
MS Word,  
oder LaTeX

# Forschungsdatenmanagement (4/4)

## Datenaustausch

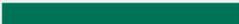
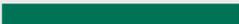
### Datenaustausch denkbar?

	Engerer Kollegenkreis	Anderen Projekten	Öffentlichkeit
Software			
Primärdaten			
Sekundärdaten			

# Forschungsdatenmanagement (4/4)

## Datenaustausch

### Datenaustausch denkbar?

	Engerer Kollegenkreis	Anderen Projekten	Öffentlichkeit
Software			
Primärdaten			
Sekundärdaten			

- Für einige Gruppen ist es denkbar, zumindest Software und Sekundärdaten der Öffentlichkeit zugänglich zu machen

# Forschungsdatenmanagement (4/4)

## Datenaustausch

### Datenaustausch de facto.

	Engerer Kollegenkreis	Anderen Projekten	Öffentlichkeit
Software			
Primärdaten			
Sekundärdaten			

- ▶ Für einige Gruppen ist es denkbar, zumindest Software und Sekundärdaten der Öffentlichkeit zugänglich zu machen
- ▶ De facto tun dies (bisher) nur sehr wenige

# Erste Analyse

- ▶ **Fehlende Infrastruktur?**
  - ▶ Teilweise **Mangel an zentralen Repositories genannt**
  - ▶ Für Software vorhanden, bisher vereinzelt genutzt
  - ▶ Für Publikationen im Aufbau
  - ▶ Für andere Forschungsdaten nicht vorhanden, aber **Erweiterung des OpenSource Servers in Planung**

# Erste Analyse

- ▶ **Fehlende Infrastruktur?**
  - ▶ Teilweise **Mangel an zentralen Repositories genannt**
  - ▶ Für Software vorhanden, bisher vereinzelt genutzt
  - ▶ Für Publikationen im Aufbau
  - ▶ Für andere Forschungsdaten nicht vorhanden, aber **Erweiterung des OpenSource Servers in Planung**
- ▶ **Fehlende Ressourcen?**
  - ▶ Vereinzelte Gruppen haben Personal, das sich explizit mit Forschungsdatenmanagement befasst

# Erste Analyse

- ▶ **Fehlende Infrastruktur?**
  - ▶ Teilweise **Mangel an zentralen Repositories genannt**
  - ▶ Für Software vorhanden, bisher vereinzelt genutzt
  - ▶ Für Publikationen im Aufbau
  - ▶ Für andere Forschungsdaten nicht vorhanden, aber **Erweiterung des OpenSource Servers in Planung**
- ▶ **Fehlende Ressourcen?**
  - ▶ Vereinzelte Gruppen haben Personal, das sich explizit mit Forschungsdatenmanagement befasst
- ▶ **Fehlende Anreize?**

# Überlegungen

- ▶ Zentralisierte Infrastruktur- und Ressourcenfragen nicht unmittelbar zu lösen (insbes. Finanzierung)
- ▶ **Dezentral beginnen:**  
Wissenschaftlern die Möglichkeit bieten, ihre Daten zu archivieren und zu veröffentlichen
- ▶ **Verlinkung** vs. Inseln des Wissens (cf. **Linked Data**)
- ▶ Linked Data auch bei **digitalen Bibliotheken** zentrales Thema

# Überlegungen

- ▶ Zentralisierte Infrastruktur- und Ressourcenfragen nicht unmittelbar zu lösen (insbes. Finanzierung)
- ▶ **Dezentral beginnen:**  
Wissenschaftlern die Möglichkeit bieten, ihre Daten zu archivieren und zu veröffentlichen
- ▶ **Verlinkung** vs. Inseln des Wissens (cf. **Linked Data**)
- ▶ Linked Data auch bei **digitalen Bibliotheken** zentrales Thema

**Was wird hierfür benötigt?**

**Was existiert bereits?**

# Gegenwärtige und bevorstehende Entwicklungen

## Modellierung und Verlinkung von Publikationen und anderen Forschungsdaten mittels Semantic-Web-Formalismen

- ▶ WebIDs/URIs für alle Entitäten
- ▶ Digital Object Identifiers
- ▶ Open Researcher & Contributor ID
- ▶ Ontologien zur Beschreibung von
  - ▶ Forschungseinrichtungen und -projekten z.B. AKT
  - ▶ Publikationen und Relationen, z.B. FaBIO, BIBO
  - ▶ Forschungsdaten, ansatzweise in CiTO

# Ontologien zur Beschreibung von Forschungsprojekten

date: Bearbeiten Ansicht Chronik Zusammenfassung Hilfe

http://www.rkbexplorer.com/explorer/#display=project-{http://wiki.rkbexplorer.com/id/resist}

Meistbesuchte Se... Getting Started Latest Headlines The Initiative | ORCID TaxiiOary Google Ngram Viewer

RKBExplorer » Projects » ReSIST Resilience for Survivability in IST

Search... Search about system requirements help data sources contact acknowledgements

**Related projects**

**Details**

**URI:** Resilience for survivability in IST

**Other Names:** ResIST Resilience for Survivability in IST

**Project Overview:** ResIST is a NoE that addresses the strategic objective "Towards a global dependability and security framework" of the Work Programme, and responds to the stated "need for resilience, self-healing, dynamic content and viable environments". It will integrate leading researchers active in the multidisciplinary domains of Dependability, Security, and Human Factors, in order that Europe will have a well-focused coherent set of research activities aimed at ensuring that future "ubiquitous computing systems", the increase systems of ever-evolving networks of computers and mobile devices which are needed to support and provide Ambient Intelligence (AMI), have the necessary resilience and survivability.

**People**

- Jean-Claude Leduc
- Abdelhak Khell
- Rafal Pfitzman
- Enghin Asari
- Benj Barad
- Carl Gerstel
- Carlo Donato
- Chengkun Liu
- Christian Cechin
- Christian Lanzetta

**Organizations**

- The University of Newcastle upon Tyne
- LAAS-CNRS
- CITY UNIVERSITY LONDON
- Laboratoire d'Analyse et d'Architectures des Systèmes
- The University of Pau
- Institut Eurcom
- Technische Universität Darmstadt
- Sepepa University of Rome/Cnr
- Université de Sherbrooke

**Publications**

- Conceptual Model and Architecture
- Malicious and Accidental-Fault Tolerance in Internet Applications, conceptual model and architecture
- Conceptual model and architecture of HAFTA
- The ResIST Resilience Knowledge Base
- A Deniable Fault-Tolerant Architecture for Real-Time Dependable Systems
- ASBERT: Automated proof based system and software engineering for real-time applications
- Final Version of OISD Conceptual Model (CGM1)
- Malicious and Accidental-Fault Tolerance in Internet Applications, reference model and use cases

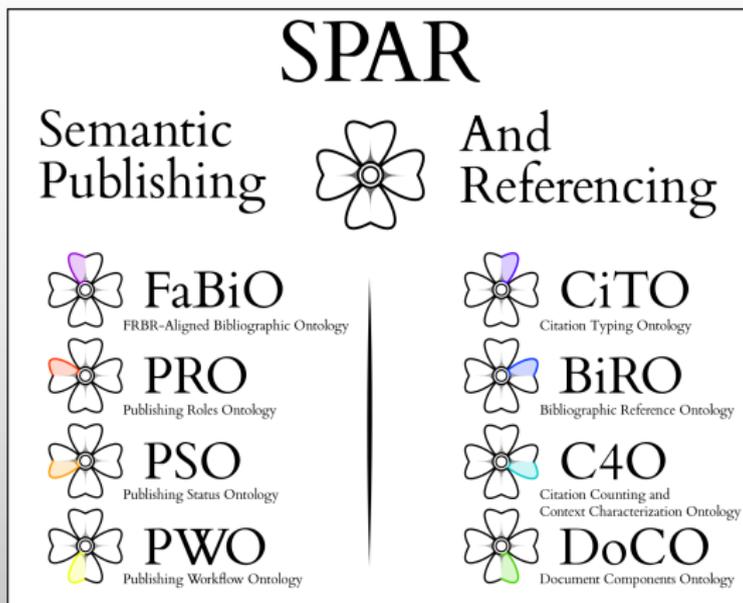
**Courses & Materials**

- Testing, Verification and Validation
- Fault Tolerance via Diversity Against Design Errors, Design Errors and Resilience Assessment
- Software Reliability, Basic Concepts and Assessment Methods
- System and Network Security
- Computer-assisted Modelling and Analysis
- Dependability of computer systems
- Dependable systems and architectures (Architecture Masters' thesis)
- Experimental Assessment of Computer System Dependability

Print

<http://www.rkbexplorer.com/explorer/#display=project-{http://wiki.rkbexplorer.com/id/resist}>

# Ontologien zur Beschreibung von Publikationen



<http://opencitations.wordpress.com/2010/10/14/introducing-the-semantic-publishing-and-referencing-spar-ontologies/>

# Bevorstehende Entwicklungen innerhalb des CITEC

- ▶ **Ontologische Modellierung des CITEC** unter Verwendung bestehender Ontologien
- ▶ (Lokale) **semantische Anreicherung von PUB-Publikationen**
  - ▶ Verknüpfungen zwischen Publikationen, Autoren, Forschungsdaten, Einrichtungen, ...
- ▶ **Entwicklung von Algorithmen** zur Identifikation von Relationen zwischen Publikationen und Datensätzen

# Semantische Anreicherung als Anreiz

## Verbesserte Datenanalyse durch vielfältige Abfragemöglichkeiten

- ▶ Welche Arbeiten (z.B. Publikationen oder andere Daten) **verwenden bestimmte Datensätze**?
- ▶ Welche Datensätze wurden **aus anderen Datensätzen abgeleitet**, oder bauen auf diesen auf?
- ▶ Welche Arbeiten **widerlegen/bestätigen meine Hypothese**?

# Semantische Anreicherung als Anreiz

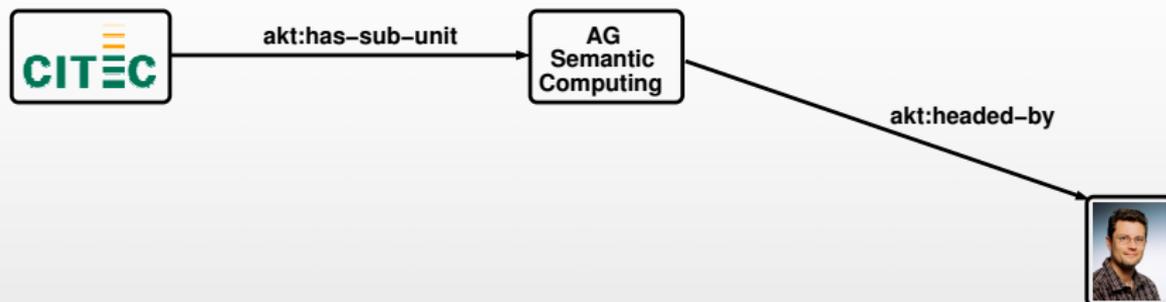
## Verbesserte Datenanalyse durch vielfältige Abfragemöglichkeiten

- ▶ Welche Arbeiten (z.B. Publikationen oder andere Daten) **verwenden bestimmte Datensätze**?
- ▶ Welche Datensätze wurden **aus anderen Datensätzen abgeleitet**, oder bauen auf diesen auf?
- ▶ Welche Arbeiten **widerlegen/bestätigen meine Hypothese**?

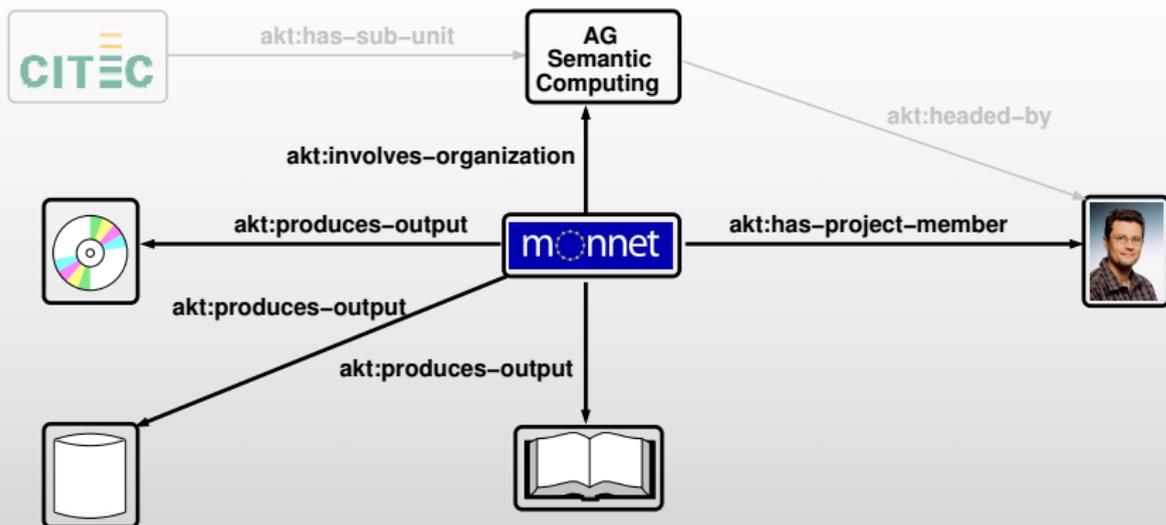
## Verbesserte Informationsextraktion

- ▶ Vereinfachte **Extraktion von Datensätzen und Publikationen** eines Projekts (z.B. für Fortsetzungsanträge von Projekten)
- ▶ Verbessertes Messen z.B. des **Impacts von Projekten**, bzw. der **wissenschaftlichen Präsenz** einer Forschungseinrichtung

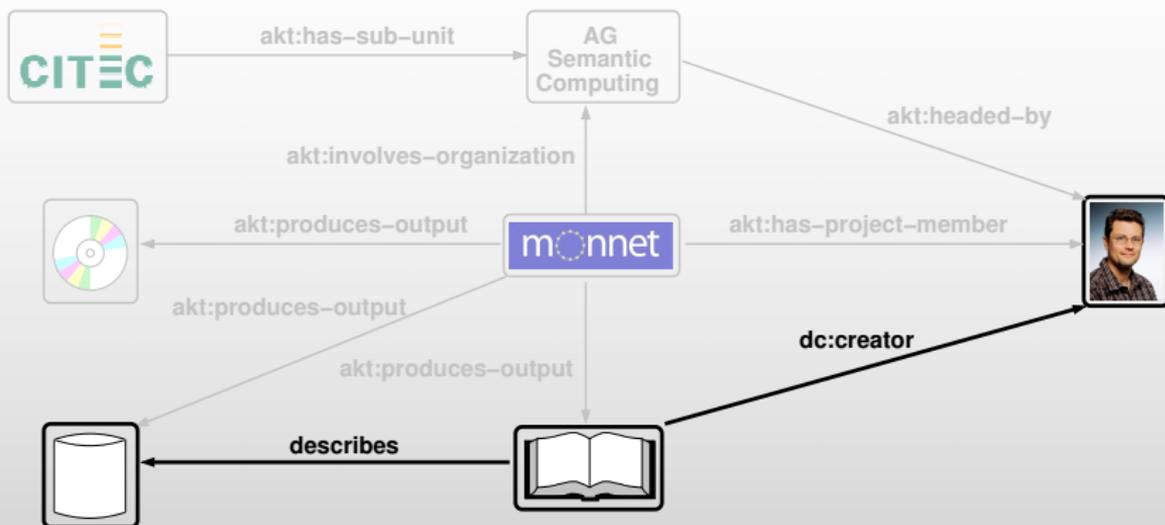
# Mögliches Szenario



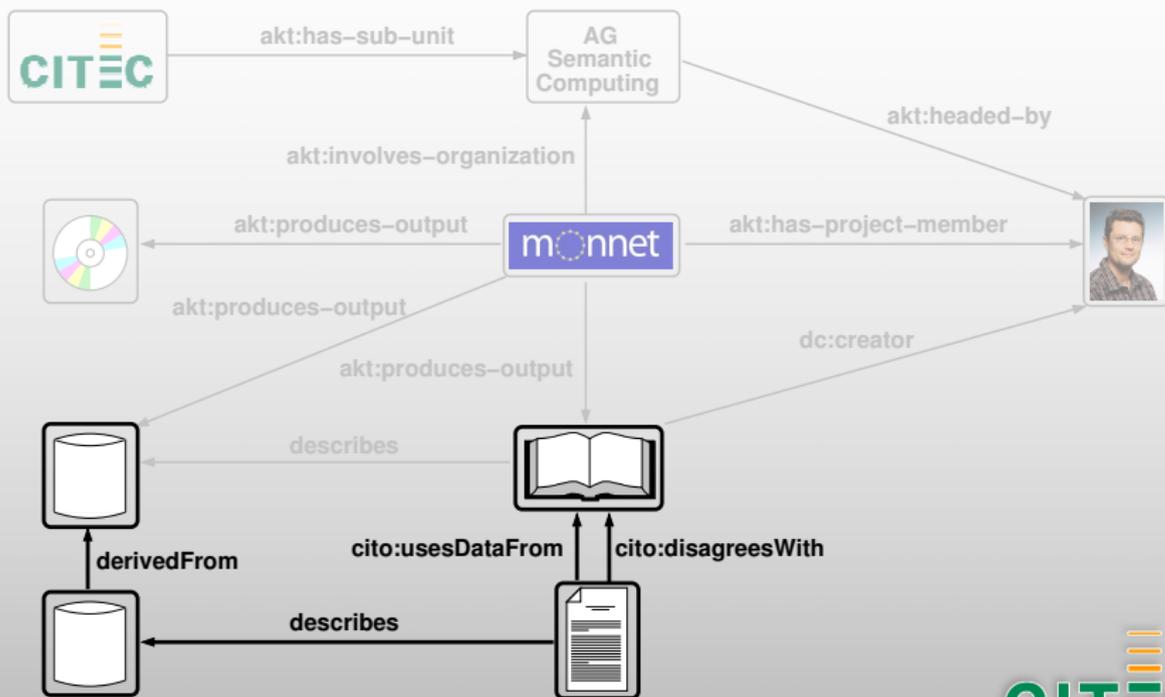
# Mögliches Szenario



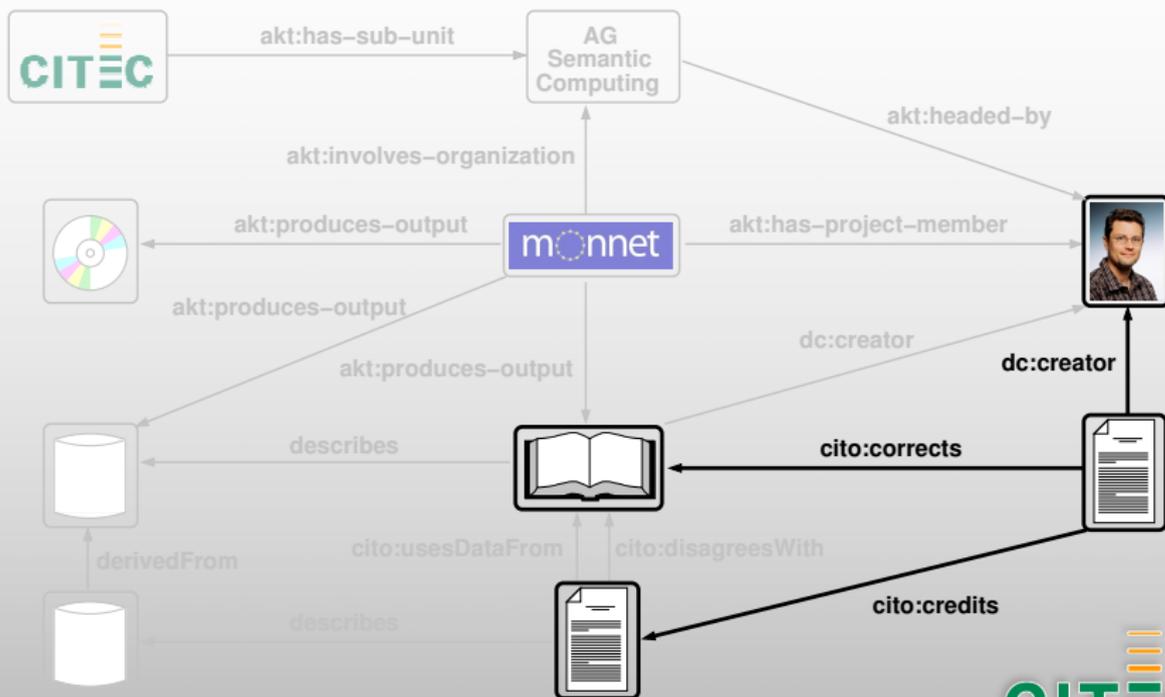
# Mögliches Szenario



# Mögliches Szenario



# Mögliches Szenario



# Zusammenfassung (1/2)

## Status Quo

- ▶ CITEC OpenSource-Server als zentrales Software-Repositoryum
- ▶ Erste Entwicklungen zur Interaktion mit universitätsweitem PUB-System
- ▶ (Noch) kein CITEC-weites Forschungsdatenmanagement, aber kürzlich gegründete Taskforce
- ▶ OpenAIRE-Studie

# Zusammenfassung (2/2)

## Ausblick

- ▶ Entwicklung eines Konzeptes für das Management von Forschungsdaten (Whitepaper)
- ▶ Erarbeitung eines Schemas zur Repräsentation von Metadaten (Provenienz, Lizenz, etc.)
- ▶ Semantische Anreicherung und Verlinkung von Daten (Projekte, Publikationen, Datensätze, etc.)
- ▶ Abstimmung mit dem Bielefeld Data Informium
- ▶ Zugriff vom Arbeitsplatz auf eine **Black-Box**

**Vielen Dank für Ihre Aufmerksamkeit!**

# Links

**AKT reference ontology**

<http://www.aktors.org/ontology>

**CITEC OpenSource Server**

<https://opensource.cit-ec.de>

**Digital Object Identifiers**

<http://www.doi.org>

**OpenAIRE**

<http://www.openaire.eu>

**Open Researcher & Contributor ID**

<http://www.orcid.org>