Talk 3:
The Whole is Greater Than the Sum of Its Parts

Analyzing Aristotle Commentaries in Collaboration between Philology and Data Science

Michael Krewet[1], Danah Tonne[2], Germaine Götzelmann[2],
Philipp Hegel[3] and Sibylle Söring[1]

[1] Freie Universität Berlin, Germany
[2] Karlsruhe Institute of Technology, Germany
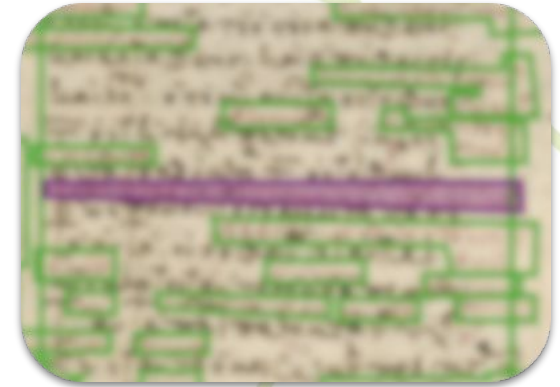[3] Technical University of Darmstadt, Germany
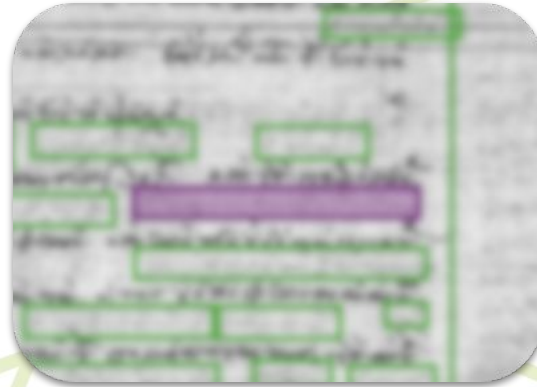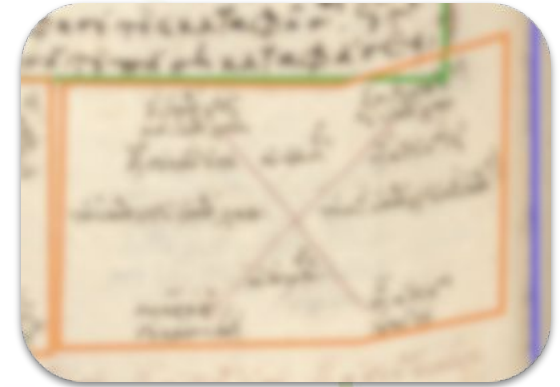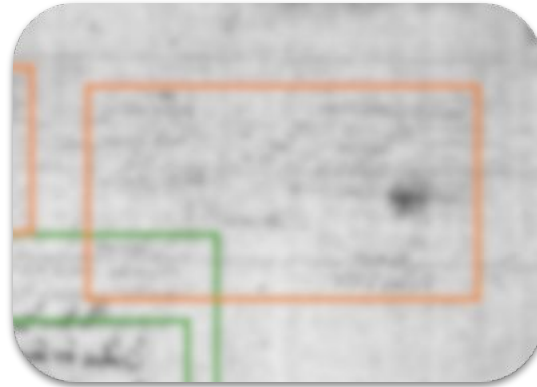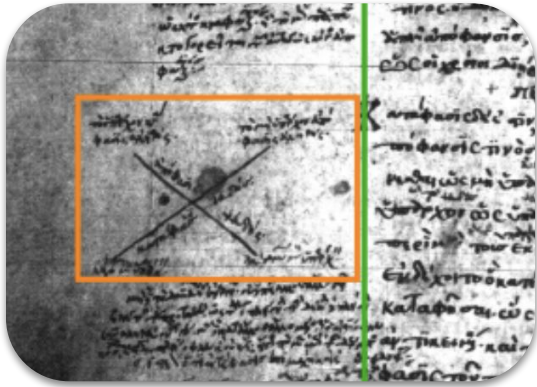
Aristotle's 'De Interpretatione'

- Copies left: ~150
- Dated: 9th – 16th century
- Origin: mostly unknown
- Transmission: highly contaminated

Example of Knowledge Transfer

Paris, BNF, Par. Gr. 1971,
f. 34v and 33r (detail)

Vatikan, BAV, Urb. Gr. 56,
f. 88r and 82r (detail)

Wien, ÖNB, Suppl. Gr. 67,
f. 117v and 113r (detail)

Find relations
and transfers

Search & Analyze
matching words, copying errors,
faulty diagrams, ...

Big Data Methods

Manual enrichment
transcribing, translating,
semantic tagging

Semantic Querying
RESTful Services, SPARQL

Scholarly
manuscript description

**Analysis Tools
Data Visualization**

35,000 annotations*
Web Annotation Data Model
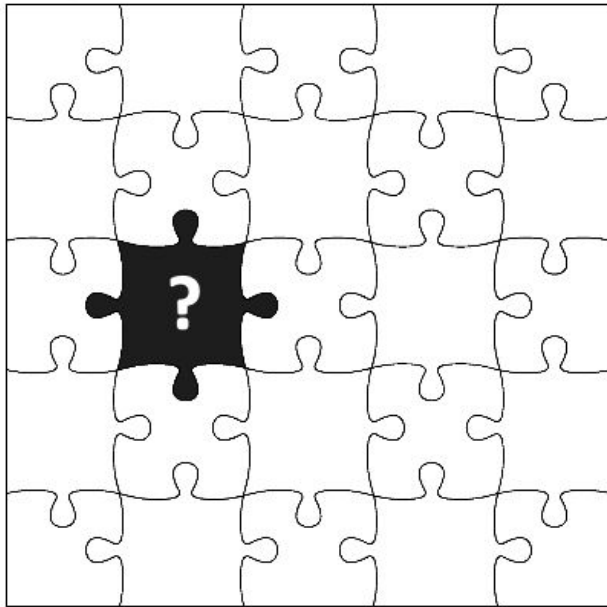
Digitization
of codices

100+ metadata files
TEI xml

**Research Data
Infrastructure**

~4200 images
jpeg, tiff

* ~10,000 manual annotations, ~25,000 annotations by algorithms for layout analysis
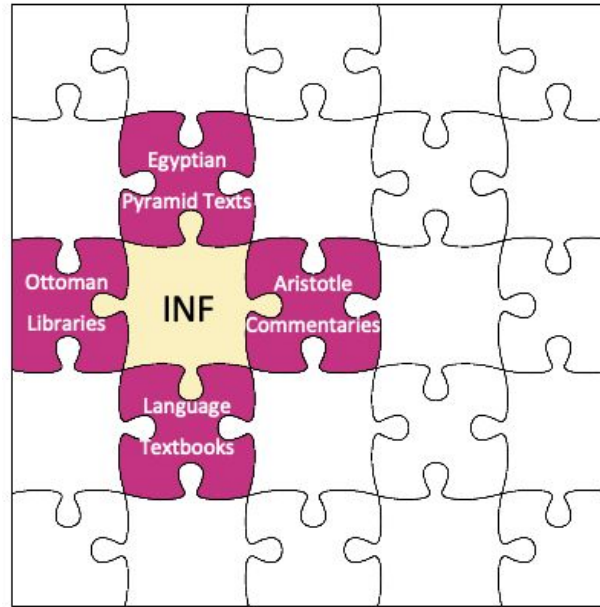
Collaboration in a Collaborative Research Center
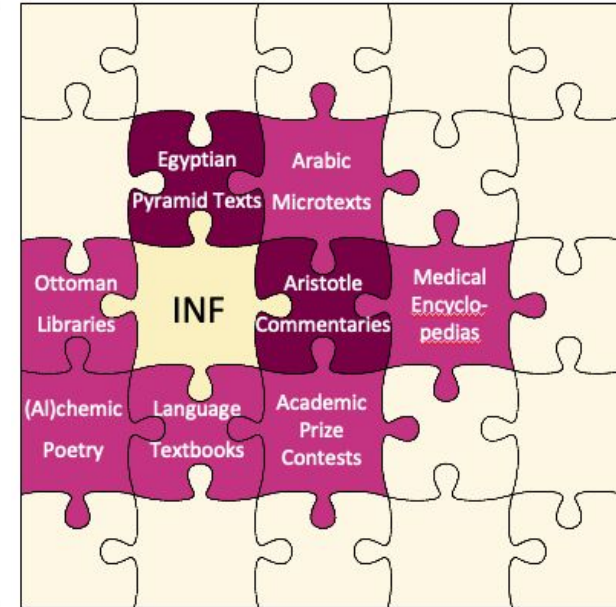## (CRC 980: Episteme in Motion)

Phase I (2012-2016):
Heterogenous projects
from '**small** disciplines'
Well-established collabs

DFG recommendation:
Information Infrastructure
project (INF)

Phase II (2016-2020):
INF project
implemented with
4 *pilot* projects
(Egyptology, Classics,
Historical Linguistics,
Greek Studies)

Phase III (2020-2024):
Extending collaborations
Deepening collaborations
Operation and **dissemination**
for all CRC projects

The Tedious Bits of Joint Research

The quotes below are a work of fiction. Any resemblance to actual persons, living or dead, or actual events is purely coincidental.

"We don't have any data at all."

Start small. Prototype. Iterate.

"We would have a lot of data if we would have more people to collect it."

Match resources on both sides.

"We do it with Excel / we do it on paper - is that alright?"

Some things can be parsed. Some things have to be changed!

"Michael is doing all that great stuff we will never be able to do"

The initial effort can be scary. Lead by example.