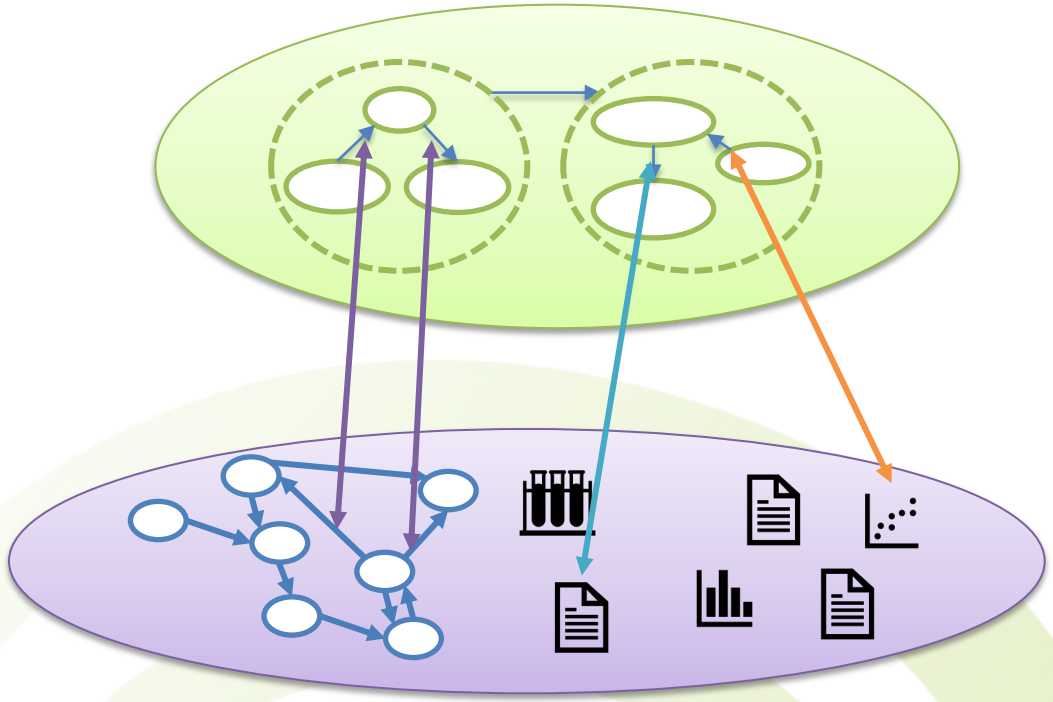
ifis
Institut für Informationssysteme
Technische Universität Braunschweig

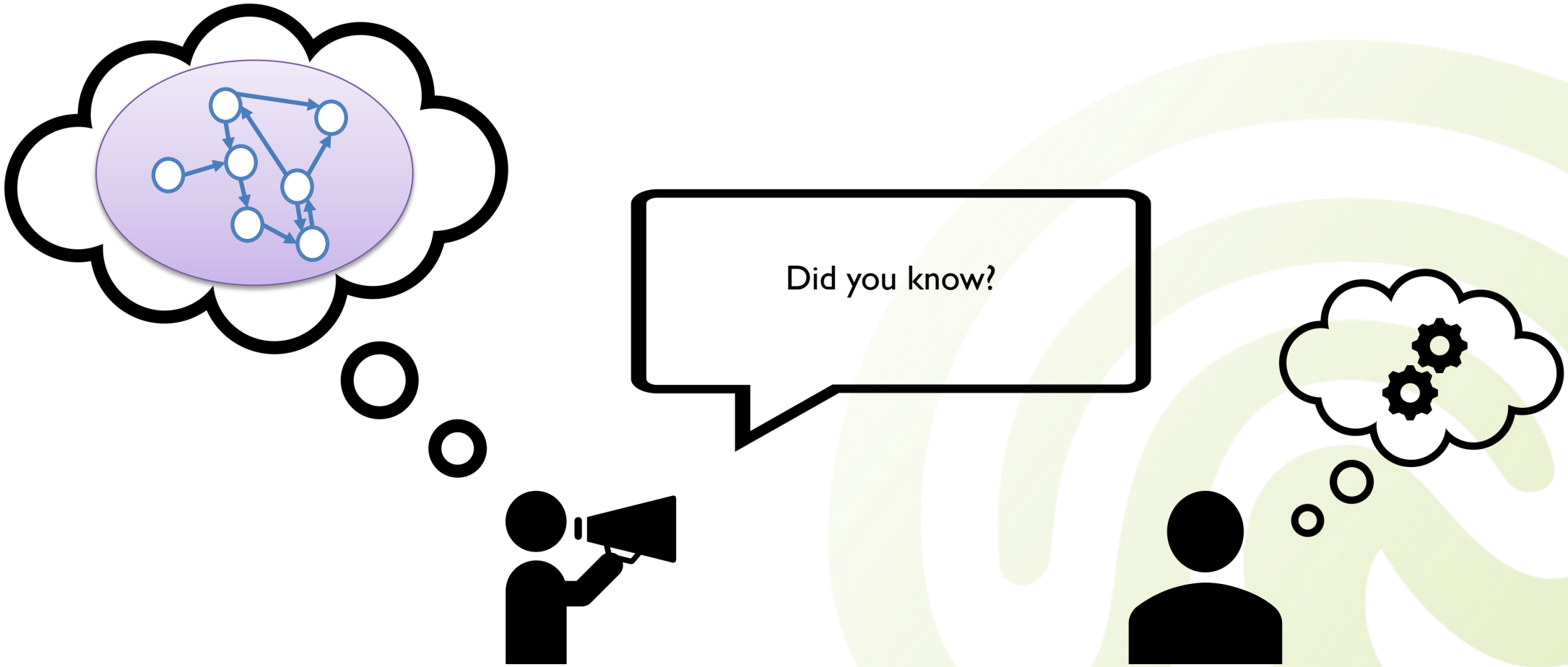


The Art of Binding: Making Narrative Overlays Plausible

Hermann Kroll

Institute for Information Systems
TU Braunschweig, Germany

Motivation

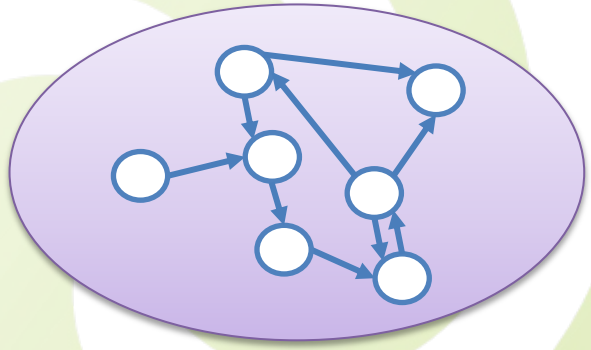


Why Narratives?

- “A narrative, story or tale is any account of a series of related events or experiences, whether nonfictional (memoir, biography, news report, […]) or fictional (fairy tale, […]).”

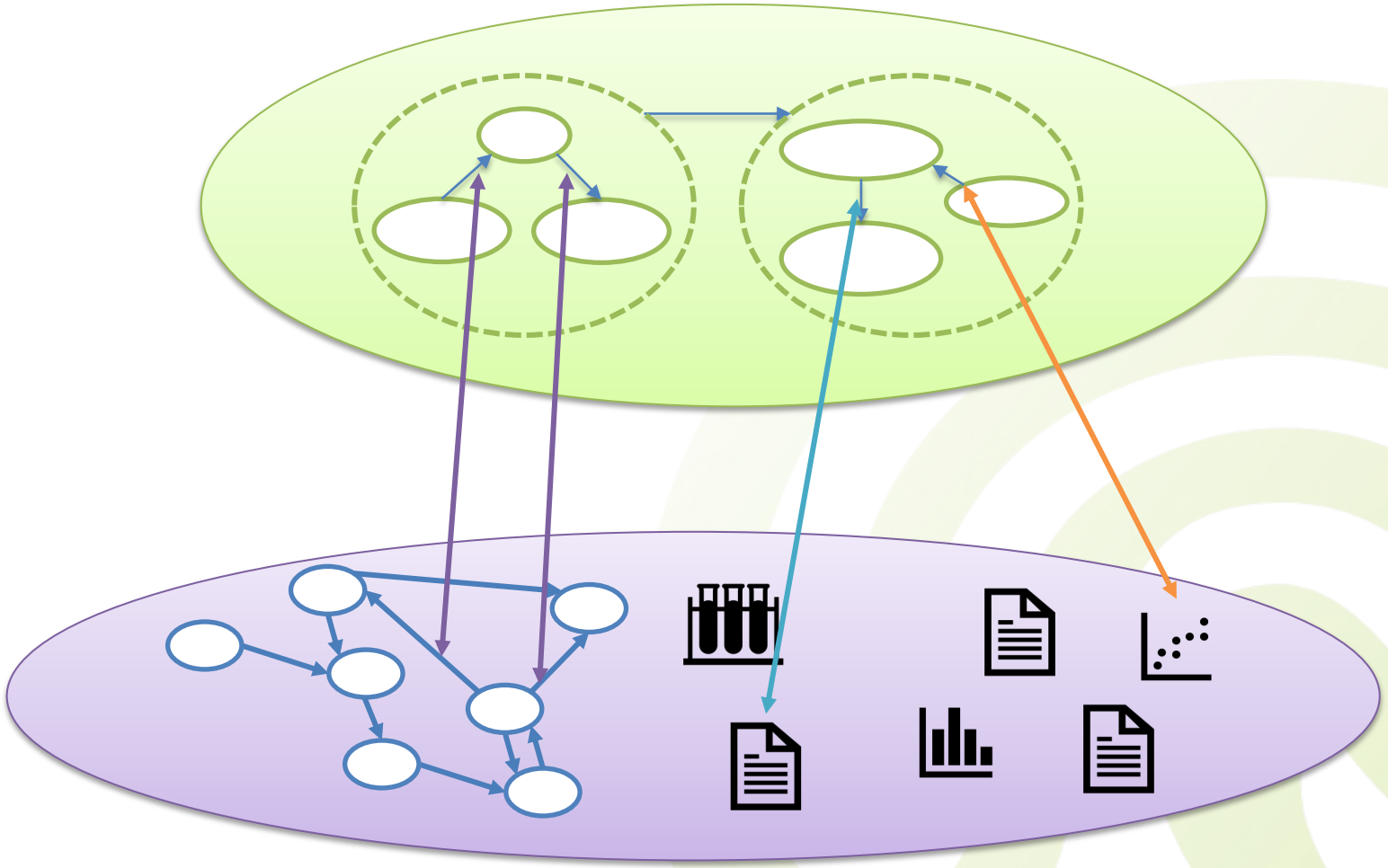
- “A presentation of real-world events that connects them in a storylike way”

- “A spoken or written account of connected events; a story.”



A Narrative Model

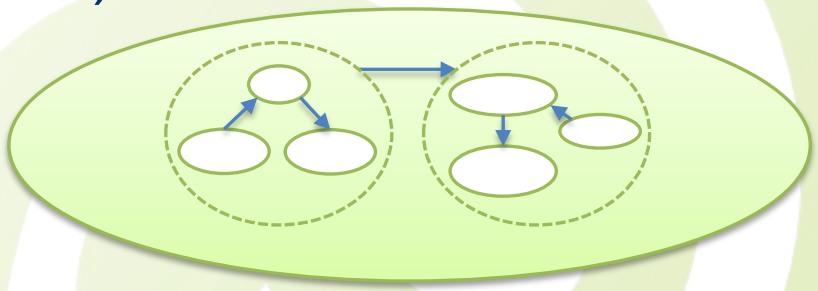
Narratives as Logical Overlays



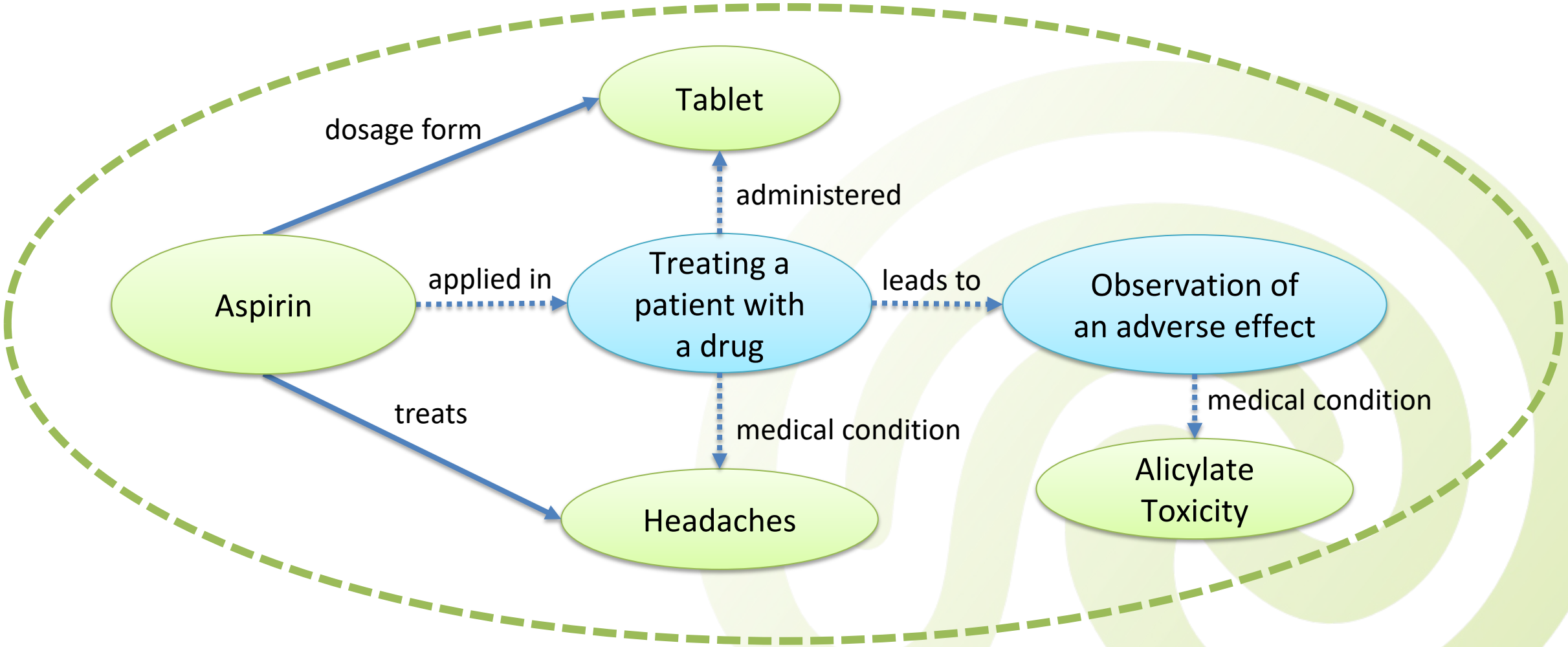
A Narrative Model

- "*A scientific narrative is a story, or more precisely, an argumentation, in which knowledge is shared.*"

- Components of a Narrative:
  - Factual knowledge (properties, types, natural laws)
  - Events (labeled states or changes of states)
  - Narrative relations (basic structure)
  - Narratives (narratives are defined inductively)



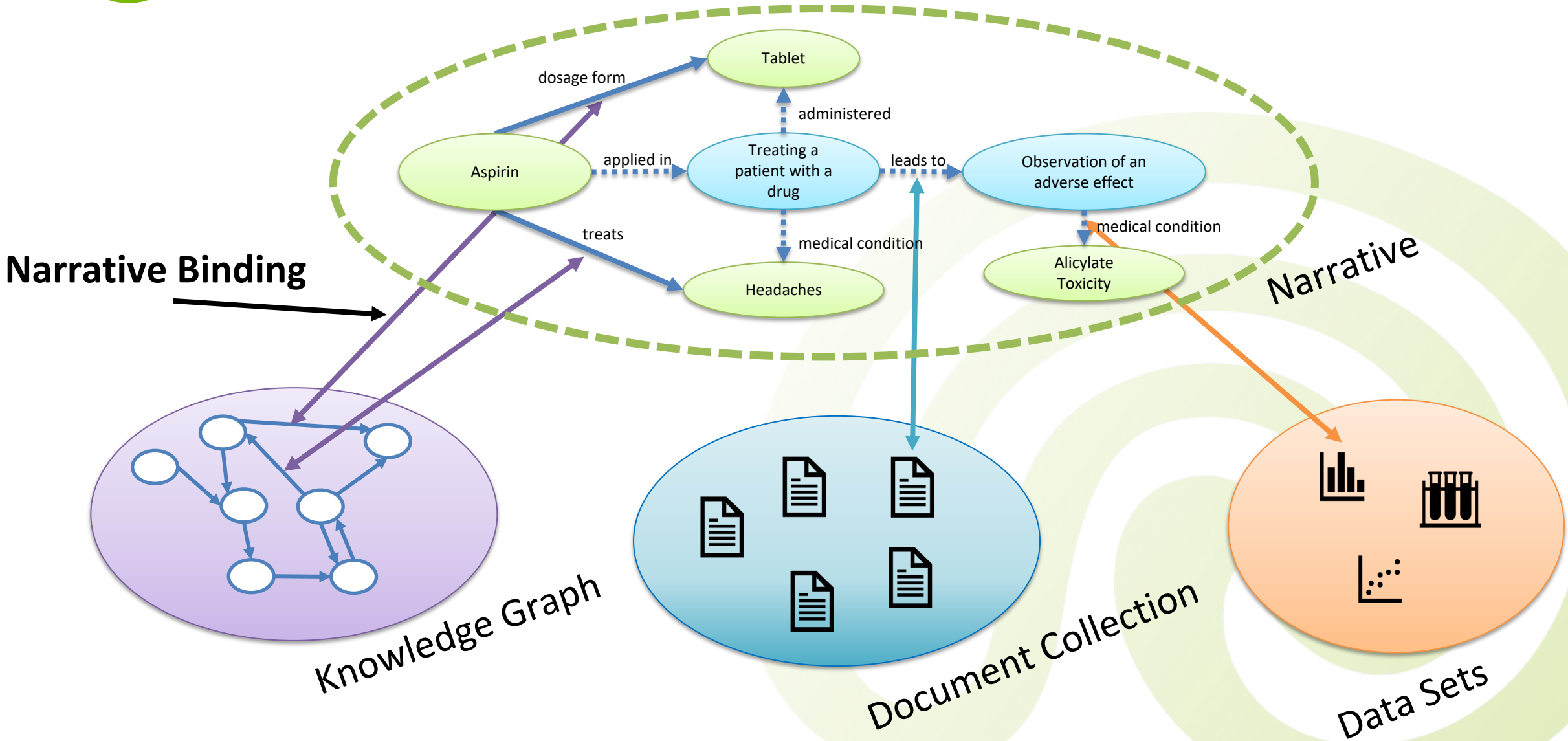
An Example Narrative



Narrative Bindings

ER2020: "Modeling Narrative Structures in Logical Overlays on Top of Knowledge Repositories"
Text2Story2021: "Demonstrating Narrative Bindings: Linking Discourses to Knowledge Repositories"

Narrative Bindings



Technical Challenges

- Each knowledge repository **type** requires its own methods:
  - Extraction/Retrieval for textual sources
  - Data set matching & explorations for data sets

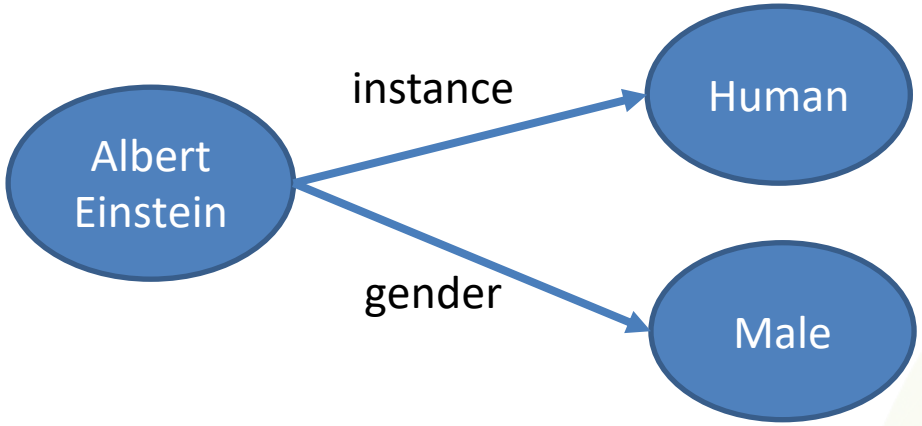
- But even if we solve these technical challenges, can we then *simply* make narratives plausible?

Validity of Information

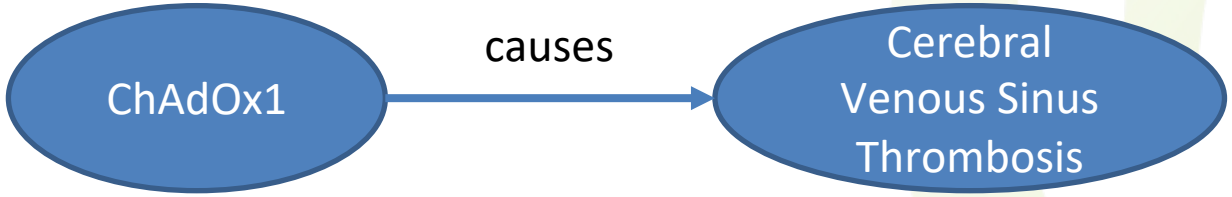
TPDL2020: "Context-Compatible Information Fusion for Scientific Knowledge Graphs"
JCDL2022: "What a Publication Tells You - Benefits of Narrative Information Access in Digital Libraries"

Validity of Information

- Some facts are **valid in general**:



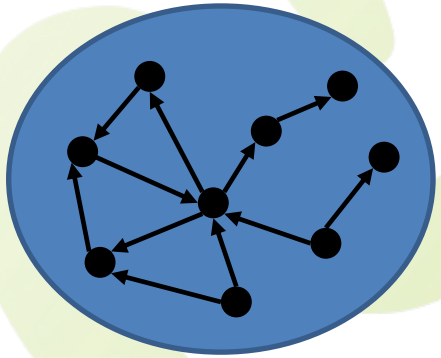
- But some are **not valid in general**:



DBpedia

WIKIDATA

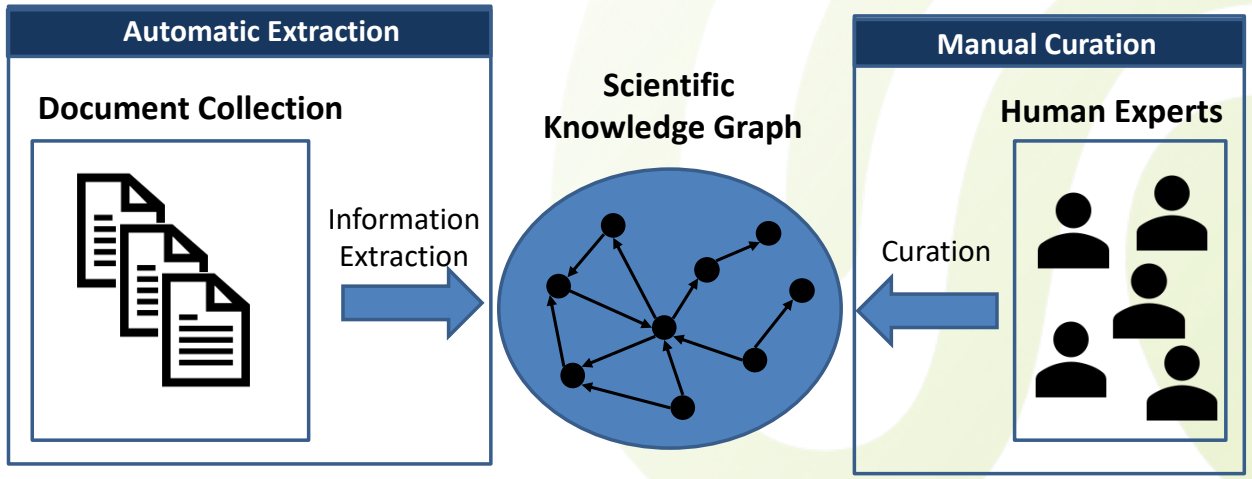
Scientific Knowledge Graph



Are we Lost in Extraction?

- Focus often lies on extraction accuracy …
  - **But** what about … Negations? Beliefs? Assumptions? Assertions? Complex Statements? **What about context?**

Must read: Fabian Suchanek - "The Need to Move beyond Triples"



Constraining Contexts

- *"We report a case of a 62-year-old man who developed cerebral venous sinus thrombosis with subarachnoid hemorrhage and concomitant thrombocytopenia, which occurred 13 days after ChAdOx1 nCov-19 injection."*
  - *(patient, vaccinated by, ChAdOx1 nCov-19)*
  - *(patient, suffered from, cerebral venous sinus thrombosis)*

Corresponding Contexts

- *"Secondary analyses found increased risk of CVST after ChAdOx1 nCoV-19 vaccination (4.01, 2.08 to 7.71 at 8-14 days), after BNT162b2 mRNA vaccination (3.58, 1.39 to 9.27 at 15-21 days), and after a positive SARS-CoV-2 test."*
  - *(ChAdOx1 nCov-19, observed condition, CVST)*
  - *(BNT162 Vaccine, observed condition, CVST)*
  - *(CVST, risk after vaccination, 4.01)*
  - *(CVST, risk after vaccination, 3.58)*

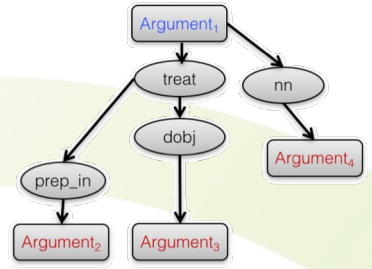
Is it a domain-specific problem?

- Wikidata: *(Barack Obama, born in, Kenya)*
  - Qualifier: „Stated in a Conspiracy Theory”

- DBpedia:
  - (Barack Obama, was, Senator of Illinois)
  - (Barack Obama, predecessor, Peter G. Fitzgerald)
  - (Barack Obama, was, U.S. President)
  - (Barack Obama, predecessor, George W. Bush)

Context Models

- N-Ary Relations:
  - *vaccinated_patients_suffer: (patient, suffered from, cerebral venous sinus thrombosis, **ChAdOx1 nCov-19**)*



Read:"HighLife: Higher-arity Fact Harvesting"

- Explicit Context and Provenance Models:
  - *Context: (old man, vaccinated by, ChAdOx1 nCov-19)*
    For: (patient, suffered from, cerebral venous sinus thrombosis)

PROV O

Reification

Named Graphs

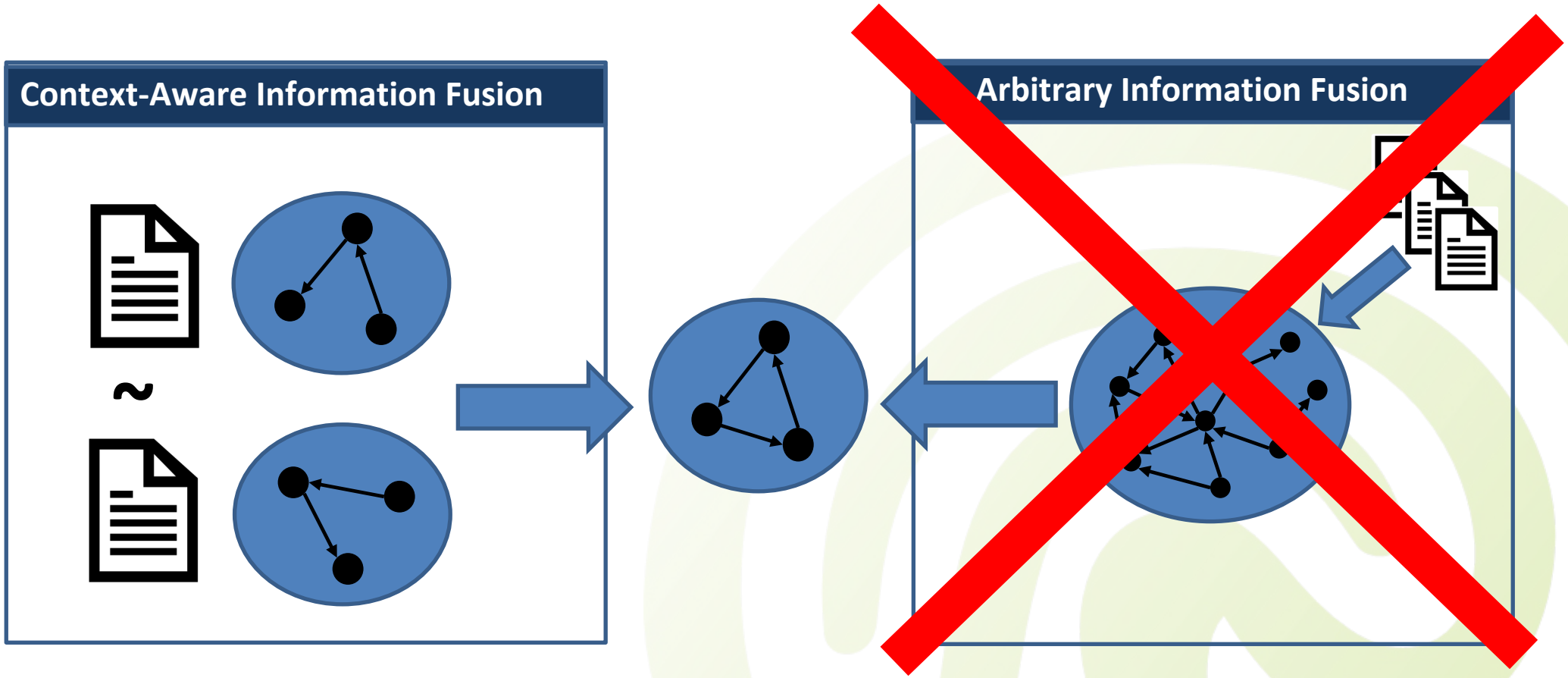
Qualifier

- Implicit Contexts Models:
  - (patient, suffered from, cerebral venous sinus thrombosis)
    Stated in source X which implicitly describes the context

Context Compatibility

- **Explicit** context models:
  - \+ High quality
  - \+ Explainability (Also for Compatibility)
  - \+ May build upon existing methods
  - – Require manual curation
  - – Require rules to determine compatibility
- **Implicit** context models:
  - – Low/moderate quality
  - – Not easy to explain (Compatibility may be based on textual measures)
  - \+ Cheap & Easy (reference to a source)
  - \+ Compatibility based on sim. measures

Context-Compatible Information Fusion



Narrative Query Processing

Narrative Query Processing

- “Given a narrative query and a set of knowledge bases,
  the query processing has to:

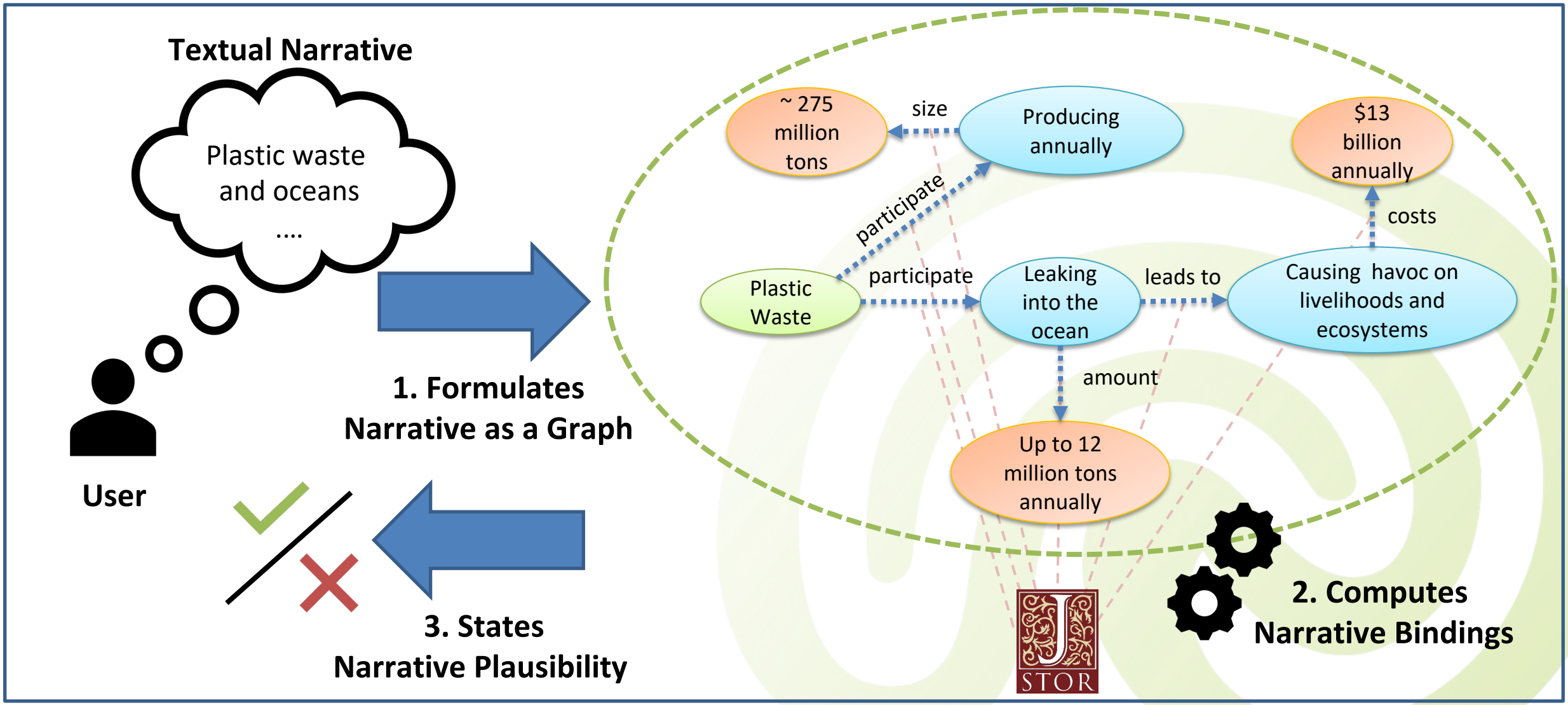
  a) **bind each** individual **query statement** against underlying data of the knowledge base(s) and

  b) check the **context-compatibility** of the computed bindings.

  The result of the query process is thus a set of valid bindings, individually binding all query statements and being context-compatible.”

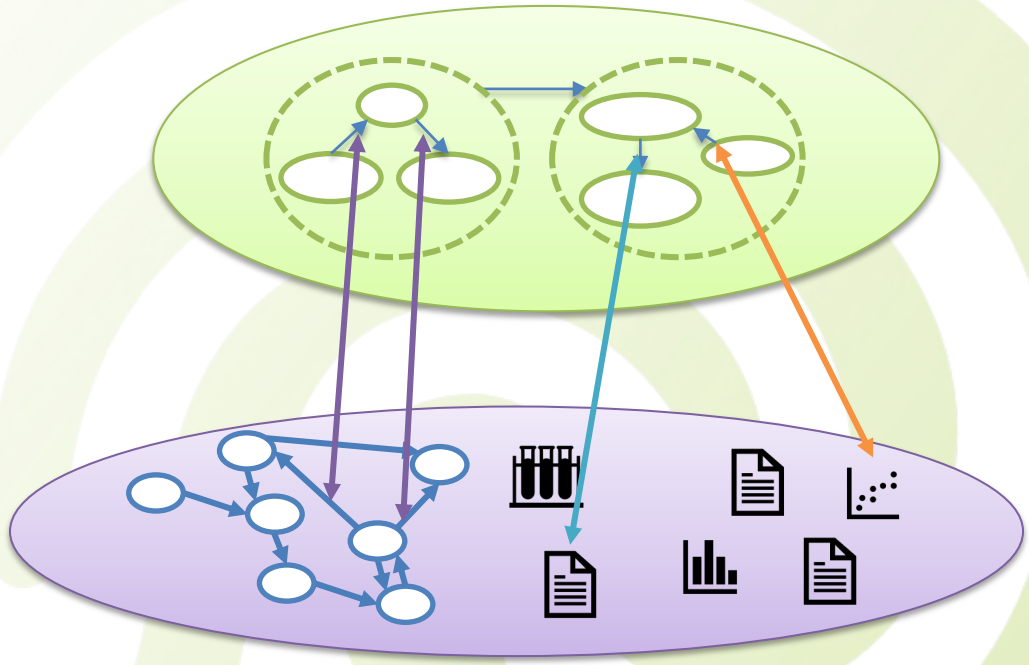
Narrative Plausibility

On Narrative Plausibility



Dimensions of Narrative Plausibility

- **Narrative Structure**
  - Transformation ambiguity (from text to graph)
  - What happens if one representation can be validated and another not?
- **Validation Approach**
  - Which repositories are considered?
  - Which methods are applied?
  - What do these methods guarantee?



Dimensions of Narrative Plausibility

- **Types of Evidence**
  - How can we handle **indirect evidence**?
  - Are we searching for **counter-evidence**?
  - What about the **absence of evidence**?

- **Confidence of Bindings**
  - **Trustworthiness** of sources (fact-checked vs. website)
  - **Quality** of bindings (e.g., score of a retrieval method)

FACHINFORMATIONSDIENST
PHARMAZIE
TU Braunschweig

www.narrative.pubpharm.de

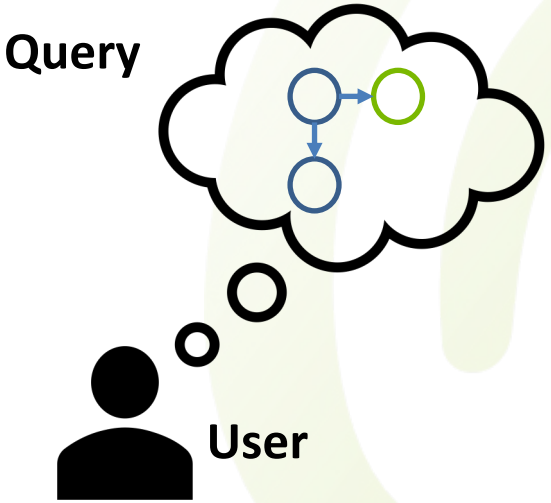
Demo:
Narratives in Pharmacy

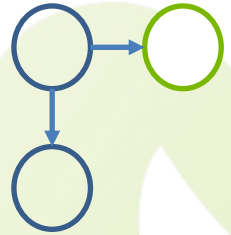
ICADL2021: “Narrative Query Graphs for Entity-Interaction-Aware Document Retrieval”

Narrative Query Graphs

- Limitations of keyword-based retrieval:
  - **Challenging** to specify **interactions** between keywords
  - **Do not feature placeholders/variables**

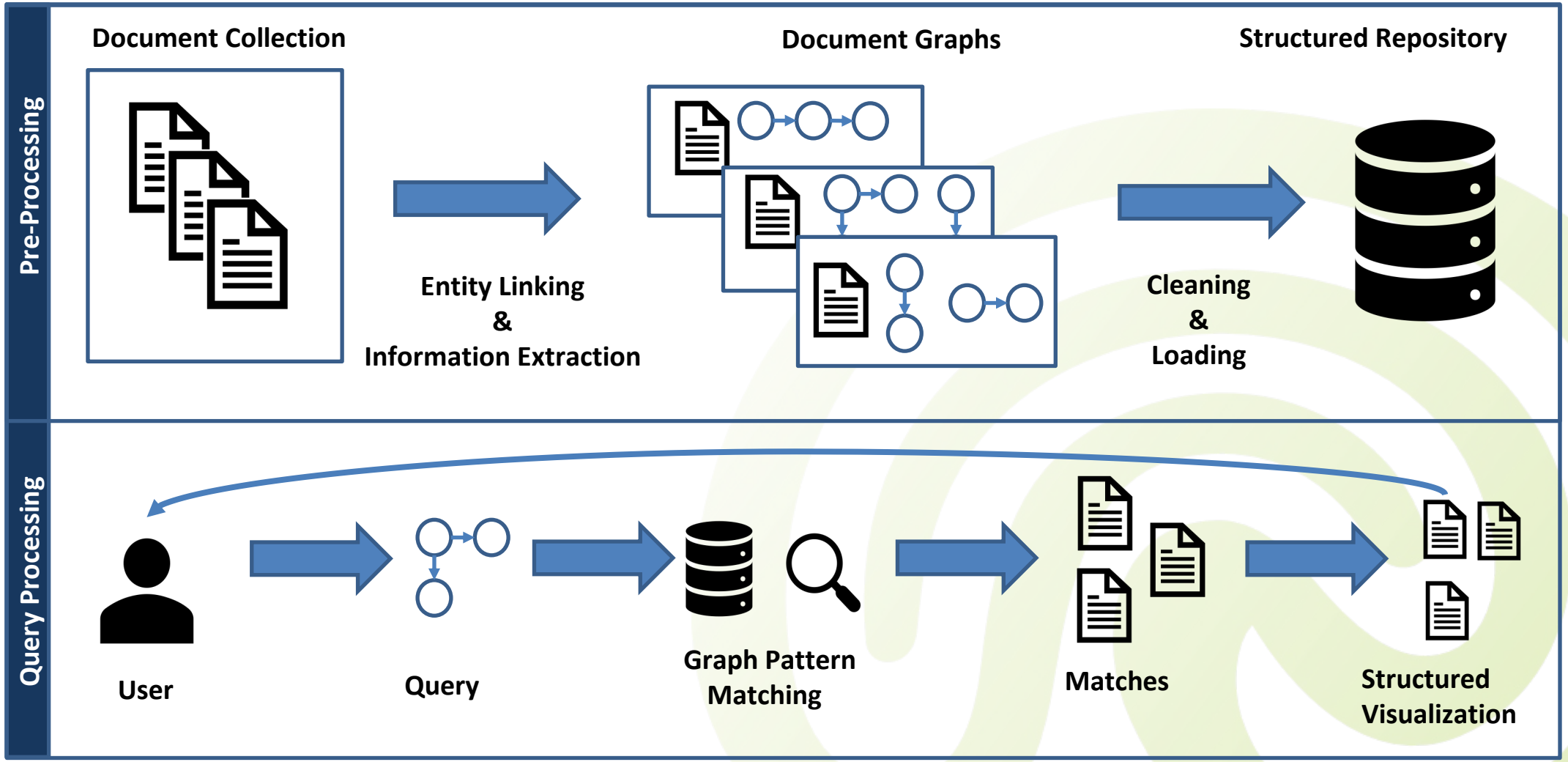
Keyword-based Retrieval





Narrative Query Graphs

A Discovery System



www.narrative.pubpharm.de

Detour

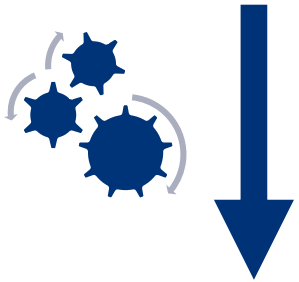
Nearly-Unsupervised Information Extraction Workflows

JCDL2021: "A Toolbox for the Nearly-Unsupervised Construction of Digital Library Knowledge Graphs"
DISCO2021: "Open Information Extraction in Digital Libraries: Current Challenges and Open Research Questions"
JCDL2022: "A Library Perspective on Nearly-Unsupervised Information Extraction Workflows in Digital Libraries"

Nearly-Unsupervised Extraction

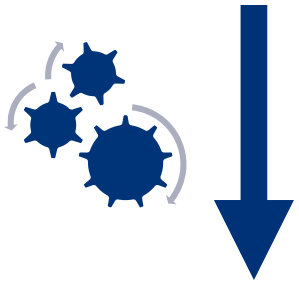
Detour

- *"The MUHAI project cares about research on narratives."*



Open Information Extraction

- *(The **MUHAI** project; cares about; research on **narratives**)*

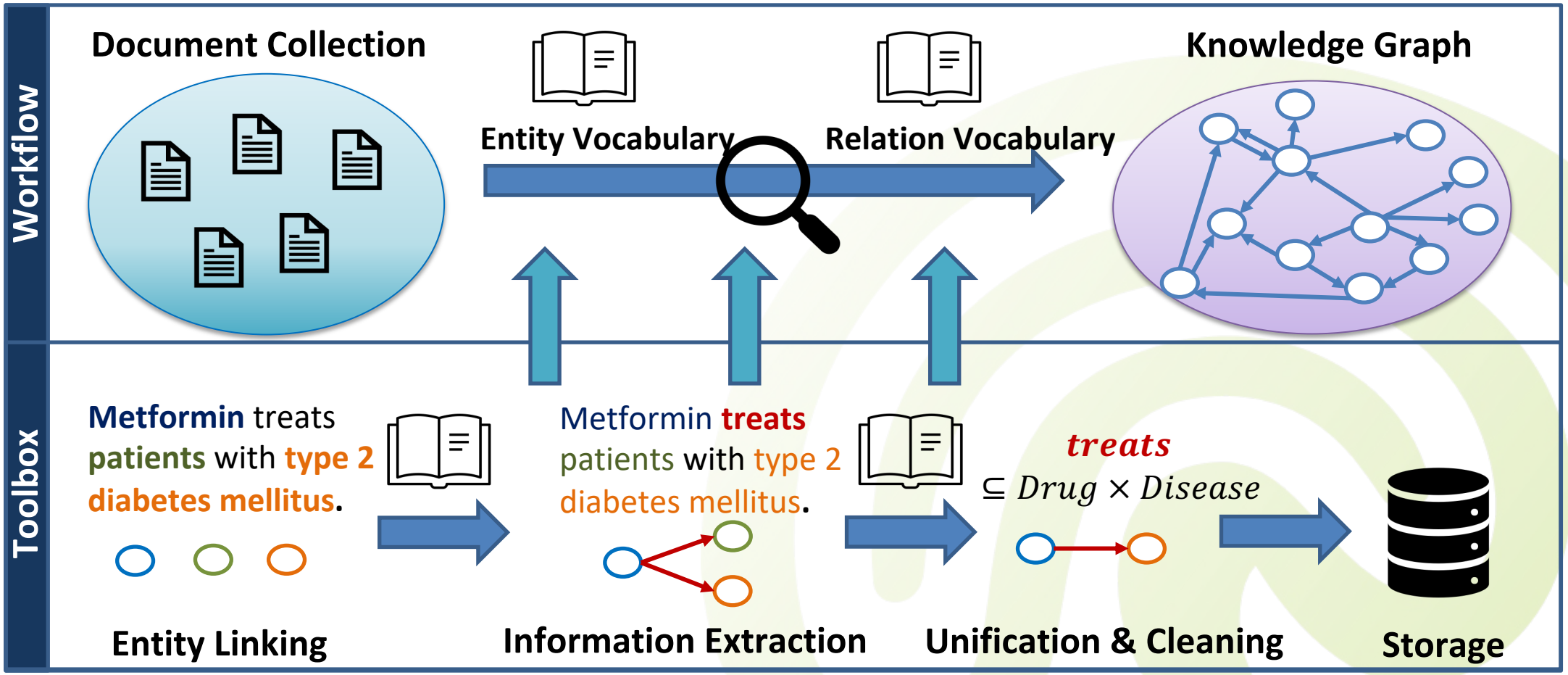


Filtering

- *(**MUHAI**; cares about; **narratives**)*

An Extraction Toolbox

Detour



https://github.com/hermannkroll/kgextractiontoolbox

Conclusion

Detour

- Nearly-Unsupervised workflows **are worth studying** because they
  - **Bypass training data** in the extraction phase completely
  - **But require** extensive filtering in practice

FACHINFORMATIONSDIENST PHARMAZIE
TU Braunschweig

www.narrative.pubpharm.de

Conclusion

Conclusion

- Narratives allow us to connect *things* in a story
  - But is that story plausible – in the sense of evidence?
  - How can we get a ***suitable*** notion of plausibility?

- Where is the **sweet spot** between formalism & application?
  - Textual narratives vs. formal models?
  - Reasoning on structured vs. real-world data?

Thank You!

FACHINFORMATIONSDIENST
PHARMAZIE
TU Braunschweig

Technische Universität Braunschweig

If you have any questions, contact me via:

kroll@ifis.cs.tu-bs.de

@HermannKroll