



NFDI4Objects

Research Data Infrastructure
for the Material Remains of
Human History

N40 Graph

Round Table Schwerpunktthema

Knowledge Graph: Datenformate, -flüsse und -modelle


Jakob Voß, Josef Heers, Gerald Steilen, Anja Gerber

Aus dem Funding Proposal

“To make all object-related research data findable and accessible via a single access point, [TA5] will create a general **index of all relevant metadata** for a rich metadata discovery service, with an associated search front end. A **graph database and standardised APIs** will enable retrieval and dissemination of metadata” (p. 87)

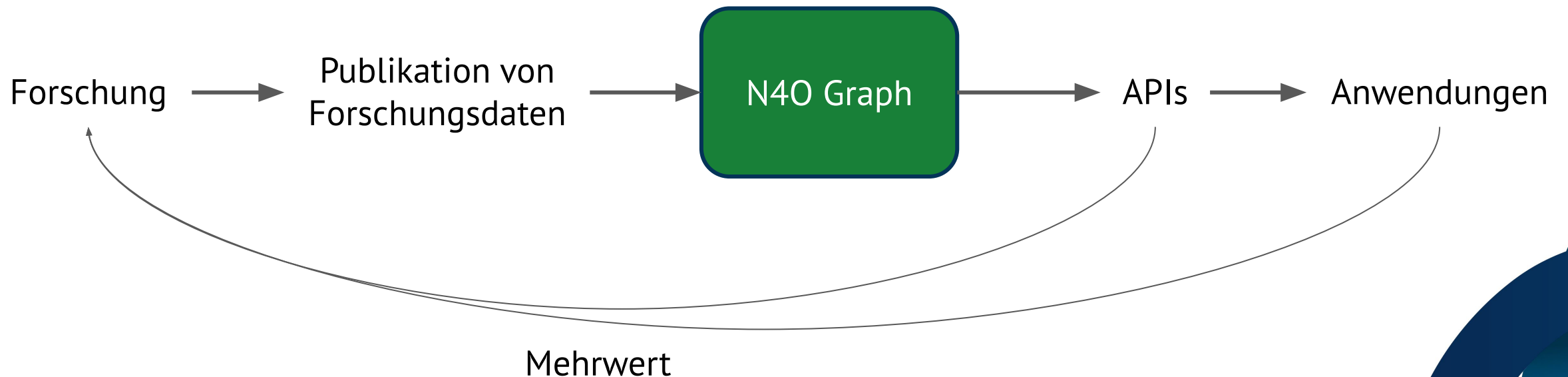
“N40 will create **an integrated knowledge base about objects**, which will be represented in a **knowledge graph** [...]” (p. 95)

Was heißt das?

- **Integration von Daten** die bisher nicht zusammengeführt sind
 - **Knowledge Base** als Ergebnis der Integration in einem Knowledge Graph
 - **Abfragemöglichkeiten** über standardisierte Schnittstellen
- 

Was heißt das?

- **Integration von Daten** die bisher nicht zusammengeführt sind
- **Knowledge Base** als Ergebnis der Integration in einem Knowledge Graph
- **Abfragemöglichkeiten** über standardisierte Schnittstellen



Ziel: Integration aller Datenquellen aus dem Konsortium

Der N40 Graph kann alle Daten aus den verschiedenen Projekten und Arbeitsgruppen aus NFDI4Objects zusammenführen.

- **Verbesserte Interoperabilität:** Der Graph ermöglicht die Verknüpfung von Daten aus verschiedenen Quellen.
- **Effizientere Datensuche:** Durch die semantische Verknüpfung der Daten können Forschende schneller und effizienter die gesuchten Informationen finden.
- **Neue Forschungsmöglichkeiten:** Der Graph ermöglicht die Verknüpfung von Daten aus verschiedenen Fachdisziplinen und die Nutzung von Synergien für neue Forschungsprojekte.

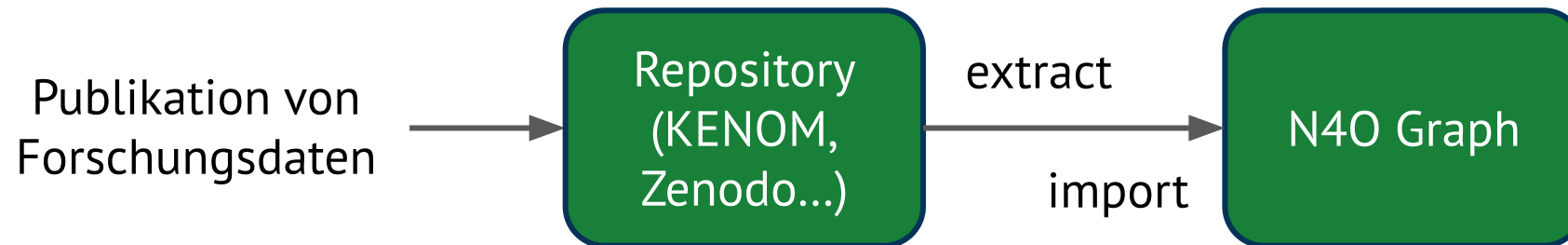
Datenquellen des N40 Graph

Publizierte Forschungsdaten mit Bezug zu materiellen Kulturgütern.

- **Findable:** Liste unter <https://graph.nfdi4objects.net/collection/>
- **Accessible:** Bereits publiziert (der N40 Graph ist kein Repository)
- **Interoperable:** Integriert mit gemeinsamen Ontologien und Vokabularen
- **Reusable:** Freie Lizenzierung

Datenimport

Es wird eine Infrastruktur für den **automatisierten Datenimport** in den N4O Graphen geschaffen. Dies erleichtert die Datenbereitstellung durch Forschungsdatenrepositorien und trägt zur Aktualität des N4O Graphen bei.



Ohne Interoperabilität keine Integration

Einheitliche Datenmodelle und -formate

- CIDOC-CRM (+Erweiterungen) als gemeinsame Ontologie
- LIDO als einheitliches XML-Format
- Abbildung in RDF und als Property-Graph

Einheitliche Verweise auf gleiche Entitäten (Personen, Orte, Materialien...)

- Kontrollierter Vokabulare / Normdaten (nur IDs!)

Herausforderungen der Interoperabilität

Verschiedene Interpretationen und Versionen von Standards, unterschiedliche Modellierung für gleiche Sachverhalte

- Nicht grundsätzlich lösbar
- Konvergenz durch gute Beispiele

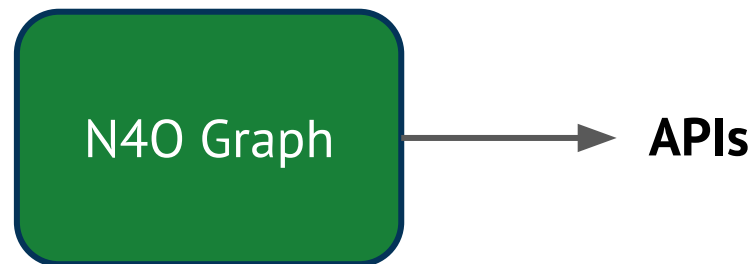
Unterschiede überwinden helfen

- Mapping zwischen verschiedenen Vokabularen mit Webanwendung Cocoda
- Minimaldatensatz in Arbeit (Erweiterung von minimaldatensatz.de)

Nutzung des Knowledge Graphen

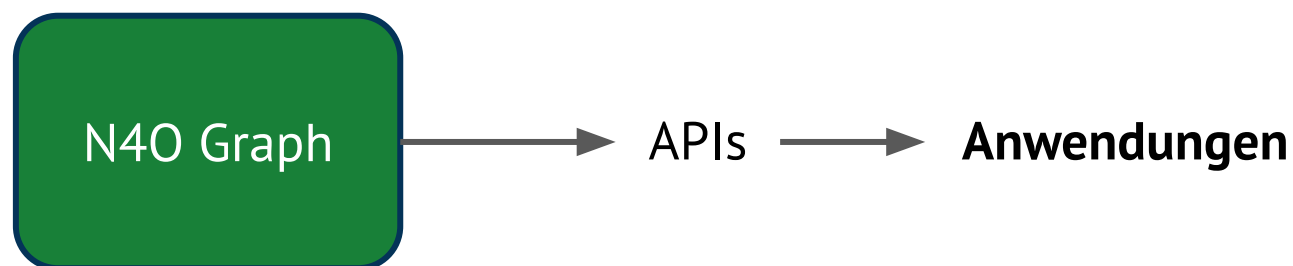
Nutzung über SPARQL- und Cypher-Endpoints

Forscher können den N4O-Graphen über SPARQL- und Cypher-Endpoints abfragen. **SPARQL** und **Cypher** sind standardisierte Abfragesprachen für unterschiedliche Graphdatenbanken (Triple-Store bzw. Property Graph).



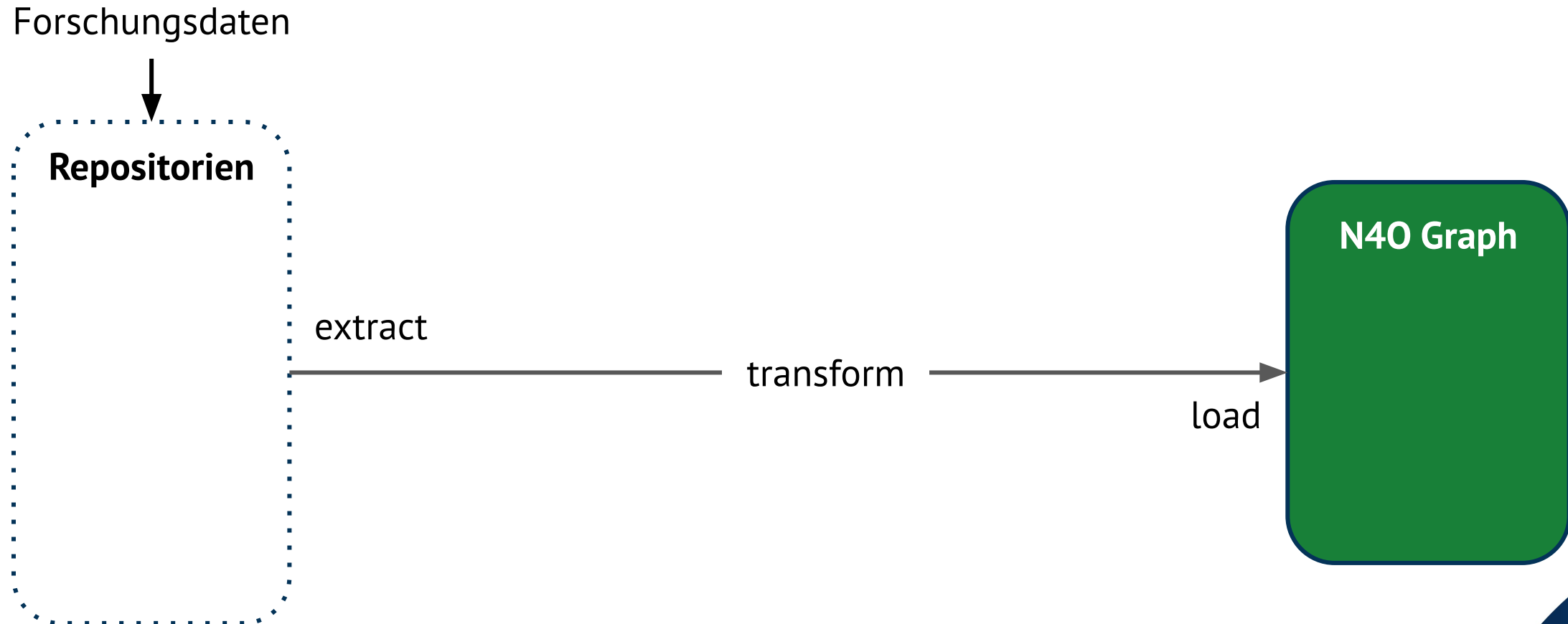
Integration in eigene Rechercheplattformen

Forscher*innen können den N40 Graphen über APIs in ihre eigenen **Rechercheplattformen und Werkzeuge** integrieren. Dies ermöglicht ihnen, die Daten des Konsortiums nahtlos in ihre eigenen Forschungsprozesse einzubinden.

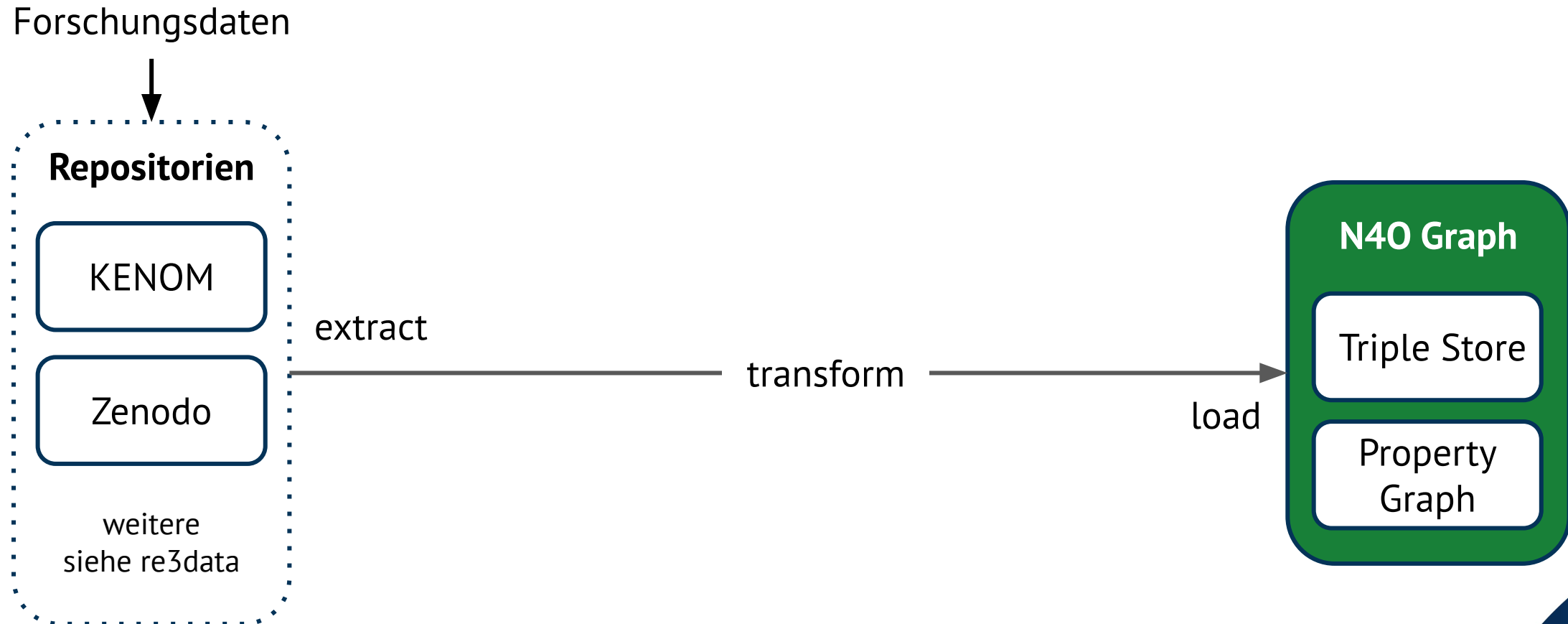


Technische Umsetzung

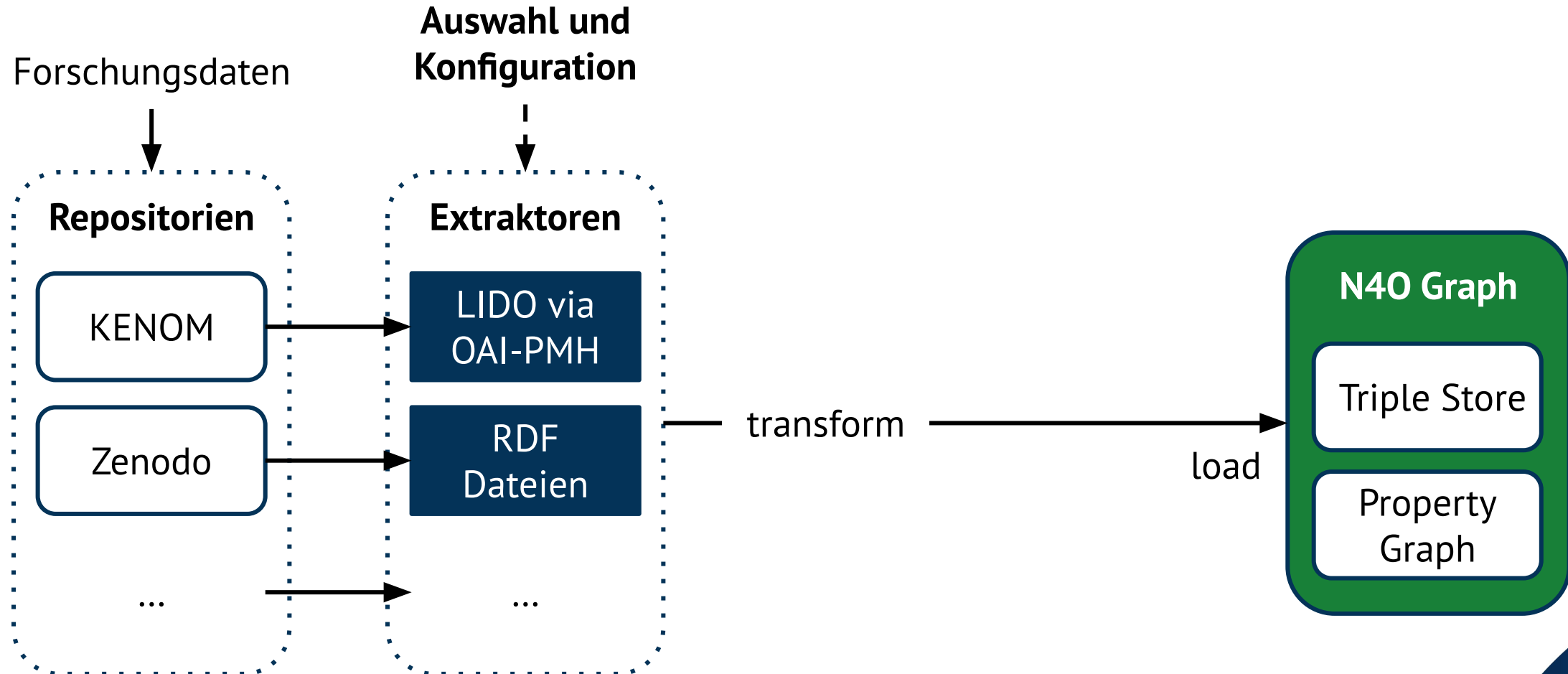
Systemarchitektur und Datenflüsse



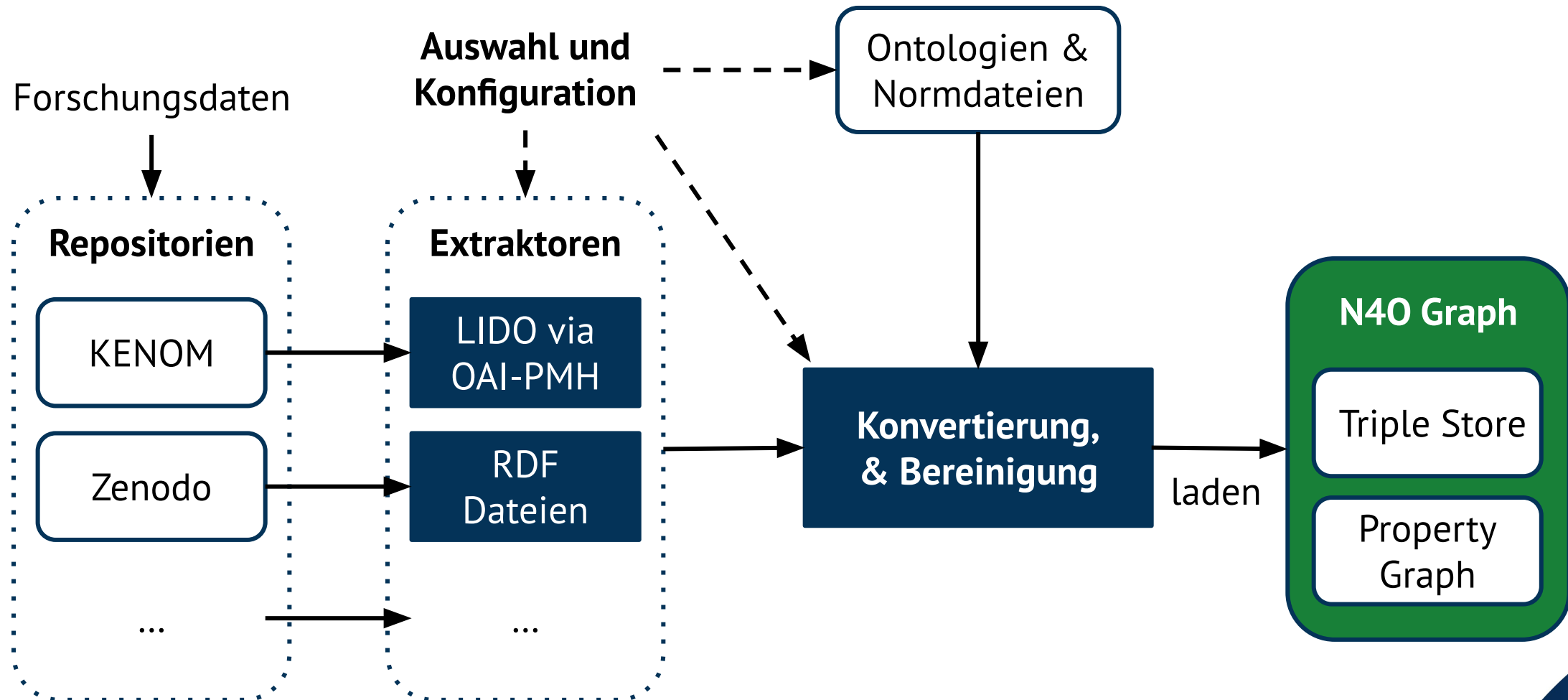
Systemarchitektur und Datenflüsse



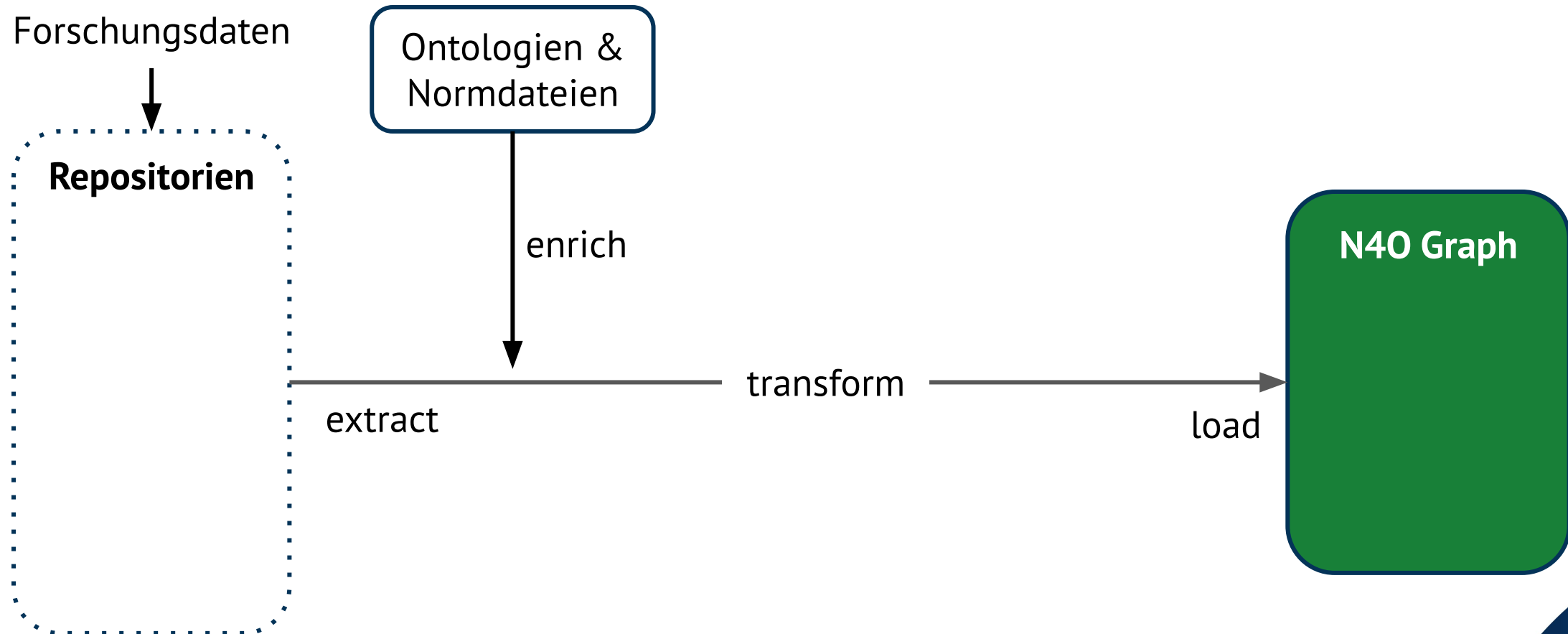
Systemarchitektur und Datenflüsse



Systemarchitektur und Datenflüsse



Systemarchitektur und Datenflüsse



Technische Komponenten

N4O-Graph

- Triple-Store: **Fuseki**
- Property Graph: **Neo4J**

Normdateien und Ontologien

- **BARTOC, Cocoda** und deren
Programmbibliotheken

Bereitstellung von APIs: **n4o-graph-apis**

Extraktoren

- *in Arbeit* (bisher per Hand)

Konvertierung, Bereinigung und
Datenimport: **n4o-import**

- **X3ML** (LIDO-Konvertierung)
- **rdffilter** (RDF-Bereinigung)
- **pgraph** (Property-Graph-Daten)

Alle Komponenten Free/Libre Open Source Software (FLOSS)
Quellcode-Repositories unter <https://github.com/nfdi4objects/>

Zusammenfassung & Ausblick

Voraussetzungen

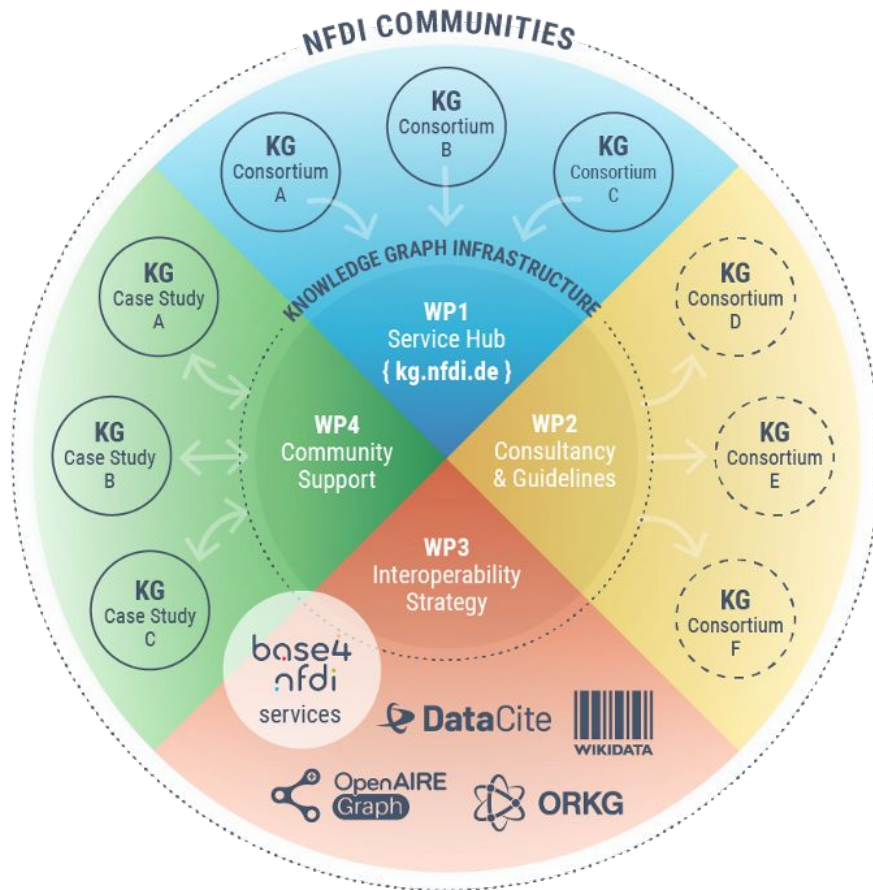
Anforderungen an Datenlieferungen sind in Entwicklung

- Einhaltung von Standards
- FAIRe Forschungsdaten aus Repositorien
- Datenmodelle CIDOC-CRM und LIDO
- Verwendung kontrollierter Vokabulare

Was brauchen wir?

- **Freie Lizenzierung** zur Nutzung und Weitergabe der Forschungsdaten
- **Mapping auf CRM** von Datenmodellen & Ontologien
- **Beispielabfragen** und dazu passende **Beispieldatensätze** mit URL
- **Relevante Terminologien** melden mit BARTOC-Eintrag.
Aktuelle Liste siehe <https://nfdi4objects.github.io/n4o-terminologies/>

Gesamtinfrastruktur der NFDI



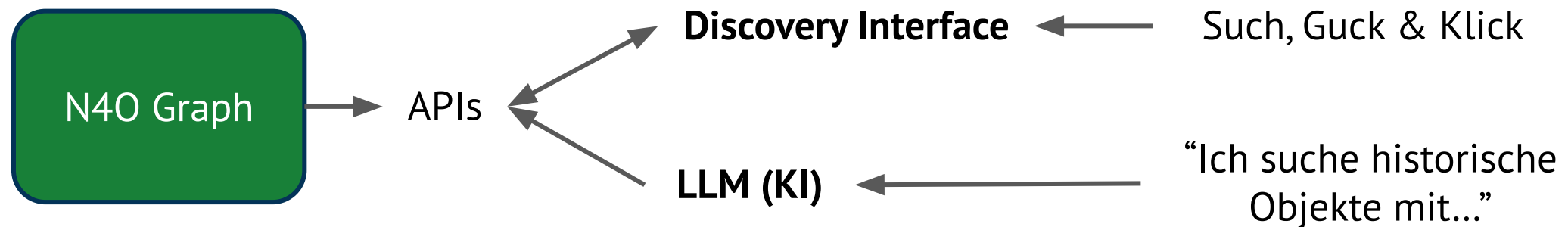
Legend: ○ Existing KGs ◌ KGs in the making

Der N40 Graph wird in die globale **NFDI-Infrastruktur** integriert. Dies ermöglicht die Verknüpfung von Daten aus verschiedenen Disziplinen und die Nutzung von Synergien zwischen den verschiedenen NFDI-Konsortien.

Discovery Interface und sprachliche Anfragen

Interface mit Suchmaske und ansprechender Präsentation von Ergebnissen (Karten, Bilder, Listen...) sowie natürlichsprachige Anfrage.

Erfordert saubere Integration von Daten, daher erst später.



Dokumentation

Handbuch zum N4O Graph

<https://nfdi4objects.github.io/n4o-graph/>

Lebendes Dokument

Gerne Feedback!



NFDI4Objects

Research Data Infrastructure
for the Material Remains of
Human History

Vielen Dank für Ihre Aufmerksamkeit!

Kontakt

nfdi4objects.net



<https://creativecommons.org/licenses/by-sa/4.0/>

Hinweis: Das gesprochene Wort ist ausschlaggebend für diese Präsentation.