



Gefördert durch:



aufgrund eines Beschlusses
des Deutschen Bundestages



Einfacher Zugang zu Umwelt- und Geodaten mittels semantischer Technologien

Dr. Wassilios Kazakos

+49 721 16006-260, wassilios.kazakos@disy.net

Dr. Gergely Lukacs

+49 721 16006-217, gergely.lukacs@disy.net

Projektrahmen

- THESEUS ist ein vom Bundesministerium für Wirtschaft und Technologie (BMWi) initiiertes Forschungsprogramm mit dem Ziel, den Zugang zu Informationen zu vereinfachen, Daten zu neuem Wissen zu vernetzen und die Grundlage für die Entwicklung neuer Dienstleistungen im Internet zu schaffen.
- Das größte deutsche IT-Forschungsprogramm
- HIPPOLYTOS: Sohn von Theseus
 - Also der, der die Ergebnisse von Theseus in die Praxis umsetzt
 - (für den Bereich Umwelt- und Geodaten)
- Projektpartner
 - Einzelantragsteller: disy Informationssysteme GmbH
 - Unterauftragnehmer: Fraunhofer IOSB



Gefördert durch:

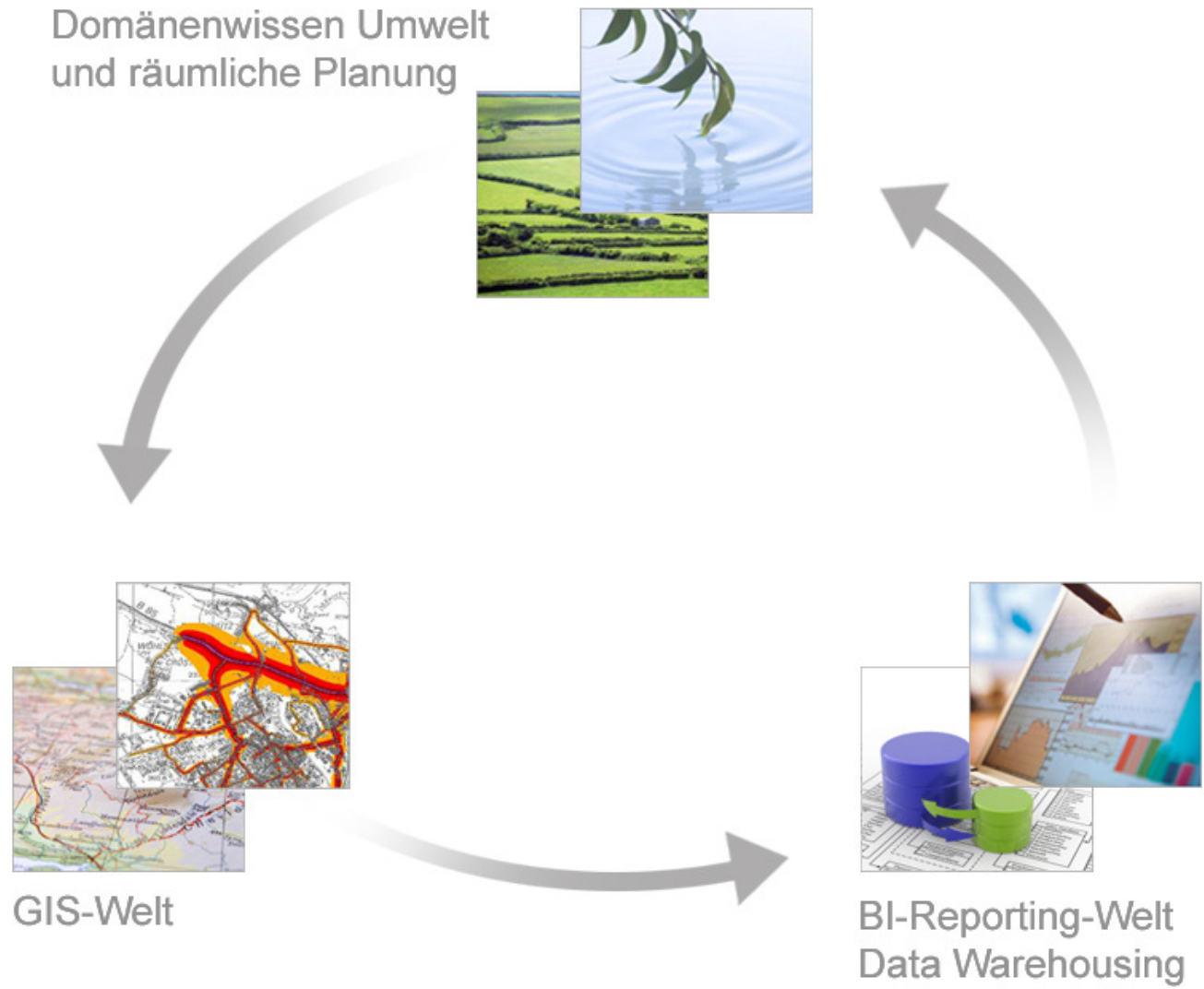


aufgrund eines Beschlusses des Deutschen Bundestages



Vorgeschichte

Zugang zu Umweltdaten und Geodaten



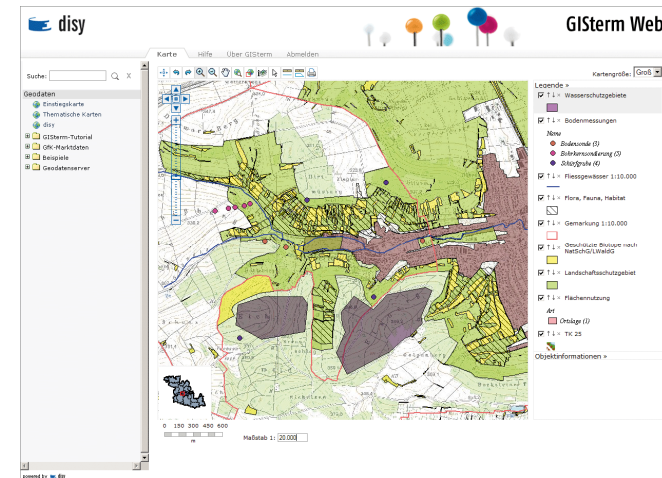
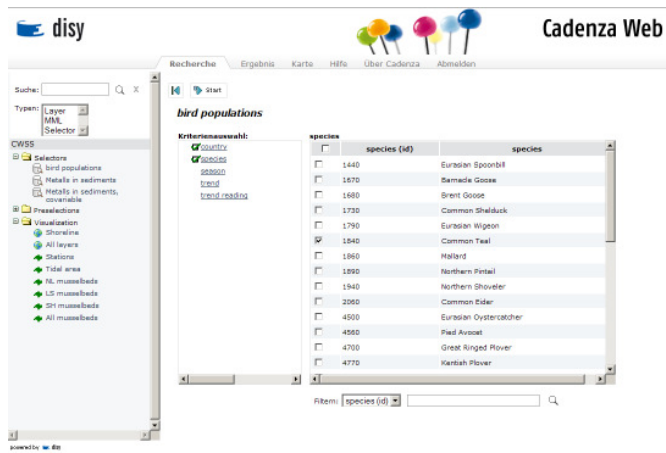
Denken Sie im Folgenden an folgende Fragestellung

- Wie hat sich die Luftverschmutzung in den letzten 10 Jahren in Leipzig entwickelt?

Status Quo bei der Recherche und Analyse von Datenbeständen

Formularbasierte Suche
 Explorative/kontextsensitive Suche
 Beispiel: **disy Cadenza**

Kartenbasierte Suche
 Zuladen von Layern
 Beispiel **disy GISterm**



Volltextsuche
 Meistens über
 Metadaten/Stammdaten
 Beispiel **disy Preludio**

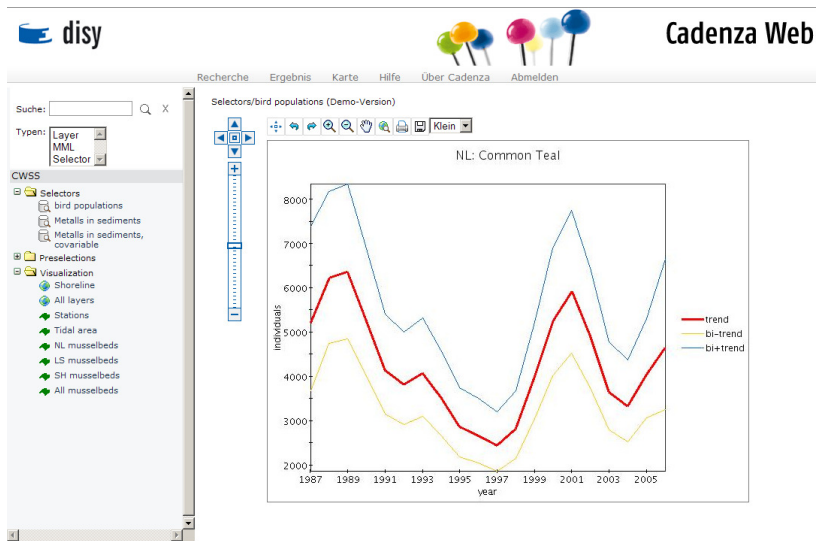
Suche Ansicht Erstellen Bearbeiten Mein Nokis ? Superuser

Geo-Metadaten (15) Kontakte (0) Projekte (0) Plattformen (0) Dienste (4) CSW-Dienste (0) Alle (19) Auswahl (0)

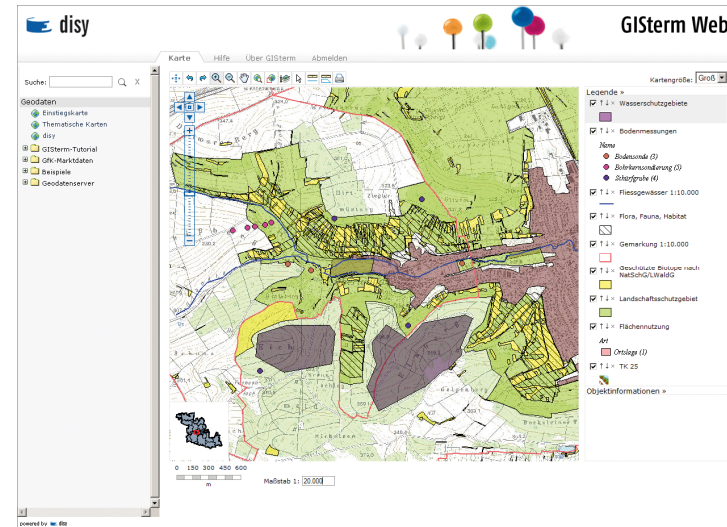
<input type="checkbox"/>			Titel	Letzte Änderung	Eigentümer	Freigabeebene	Gültig (valide)
<input type="checkbox"/>			Barth - Sundische Straße 135	08.06.2008 18:07:29	sa	In Bearbeitung	Ja
<input type="checkbox"/>			Bestandsaufnahme Hochwasserschutz im Küstenabschnitt Ahrenshoop	15.06.2007 14:19:25	User	Bearbeitung abgeschlossen	Ja
<input type="checkbox"/>			Binz Küstenaufnahme	21.09.2007 11:27:58	sa	In Bearbeitung	Ja
<input type="checkbox"/>			Küstenaufnahme im Küstenabschnitt Ahlbeck	18.09.2007 10:46:37	Username	Bearbeitung abgeschlossen	Ja
<input type="checkbox"/>			Lage- und Höhenplan Küstenabschnitt Ahrenshoop	18.09.2007 11:00:30	Username	Intern freigegeben	Ja

Ergebnis einer Suche / Analyse

Trends/Diagramme



Karten/Layer



Tabellen, Listen

Suche Ansicht Erstellen Bearbeiten Mein Nokis ? Superuser

Geo-Metadaten (15) Kontakte (0) Projekte (0) Plattformen (0) Dienste (4) CSW-Dienste (0) Alle (19) Auswahl (0)

<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Titel	Letzte Änderung	Eigentümer	Freigabeebene	Gültig (valide)
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Barth - Sundische Straße 135	08.06.2008 18:07:29	sa	In Bearbeitung	Ja
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Bestandsaufnahme Hochwasserschutz im Küstenabschnitt Ahrenshoop	15.06.2007 14:19:25	User	Bearbeitung abgeschlossen	Ja
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Binz Küstenaufnahme	21.09.2007 11:27:58	sa	In Bearbeitung	Ja
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Küstenaufnahme im Küstenabschnitt Ahlbeck	18.09.2007 10:46:37	Username	Bearbeitung abgeschlossen	Ja
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Lage- und Höhenplan Küstenabschnitt Ahrenshoop	18.09.2007 11:00:30	Username	Intern freigegeben	Ja

Motivation für HIPPOLYTOS

Motivation für HIPPOLYTOS

- Zugänge zu Daten und Informationen sind eher für Fachanwender
- Einführung eines „dritten Informationszugangs“
 - 1. räumlicher Einstieg (GIS)
 - 2. Fachlicher, strukturierter Einstieg
 - **3. intuitiver textbasierter Einstieg und Navigation**

Was bedeutet „intuitiver Informationszugang“

[Web](#) [Bilder](#) [Videos](#) [Maps](#) [News](#) [Shopping](#) [E-Mail](#) [Mehr ▾](#)

[iGoogle](#) | [Sucheinstellungen](#) | [Anmelden](#)



[Erweiterte Suche](#)
[Sprachoptionen](#)

Google-Suche

Auf gut Glück!

[Werben mit Google](#)

[Unternehmensangebote](#)

[Über Google](#)

[Google.com in English](#)

© 2010 - [Datenschutz](#)

Problem: Suchmaschinen arbeiten auf Text-Quellen (HTML, PDF, etc.)
Die fachlichen Schätze liegen jedoch meistens in Datenbanken

Was bedeutet „einfacher Informationszugang“

The screenshot shows the homepage of the HIPPOLYTOS website. At the top left, there are navigation links: [Maps](#), [Dienste](#), [Ontologien](#), and [Neues](#). At the top right, there are links for [Mein Account](#) and [Anmelden](#). The main content area features the HIPPOLYTOS logo, which consists of a stylized green and blue leaf-like shape with a circular pattern of lines behind it. To the right of the logo, the text reads "HIPPOLYTOS" in large blue letters, followed by "Neue Technologien für Umwelt- und Geodienste" in smaller blue text. Below the logo and text is a search bar containing the text "Luftverschmutzung Karlsruhe 10 Jahre". To the right of the search bar are two links: [Durchsuchen](#) and [Räumliche Suche](#). Below the search bar is a button labeled "Suche". Further down, there are links for [über HIPPOLYTOS](#), [Partner](#), [Ontologien](#), and [Einloggen](#). At the bottom left, there is a copyright notice: ©2009 - [Datenschutz](#). At the bottom right, there is a large version of the HIPPOLYTOS logo.

Typische Fragen

- „wie hat sich die Luftverschmutzung in den letzten 10 Jahren in Leipzig verändert“
 - Oder: „Luftverschmutzung“, „verändert“, „letzten 10 Jahre“, „Leipzig“
- Problem 1: Ort „Leipzig“
 - Ansatz: Gazetteer
 - Zuordnung Ort, Geometrie (Punkt, Polygon, Linie), Ober- Unter- und verwandte Begriffe etc.
- Problem 2: Zeit „letzten 10 Jahre“, „verändert“
 - Implizit: Trend-Anfrage
- Problem 3: „Luftverschmutzung“
 - Kennt keine Fachanwendung
 - Eher „Luftqualität“ oder „NOx“, „ozon“ etc.

Worum geht's uns

- Ziel ist es **nicht** eine neue Suchmaschine für Webseiten zu bauen

sondern

- Ziel ist es eine einfache (also textbasierte) Suche
- auf komplexe Datenstrukturen und Zusammenhänge abzubilden

also

- Jemanden der die Daten nicht kennt einen intuitiven Zugang zu verschaffen.

Wie könnte das Ergebnis einer einfachen Suche aussehen

Web Bilder Videos Maps News Shopping E-Mail Mehr ▾

[Webprotokoll](#) | [Sucheinstellungen](#) | [Anmelden](#)



leipzig

Suche

Ungefähr 29.800.000 Ergebnisse (0,10 Sekunden)

[Erweiterte Suche](#)

Alles

Maps

News

Blogs

Bilder

Mehr

Das Web

Seiten auf Deutsch

Seiten aus
Deutschland

Alle

Neueste
Letzte 2 Tage

Standardansicht

Websites mit Bildern

Mehr Optionen

[75 Hotels in Leipzig](#)

Anzeige

www.booking.com/Hotels-Leipzig

Schnell und sicher online buchen. Sparen Sie jetzt bis zu 50%!

[Leipzig](#) maps.google.de



[leipzig.de](#) - Das offizielle Portal der Stadt Leipzig

Das offizielle Internet-Portal für **Leipziger** Bürger, Touristen und Unternehmen. Ein Service der Stadt **Leipzig**.

www.leipzig.de/ - Ähnlich

[Leipzig Touristservice eV](#)

Unterkunft, Gastronomie und Szene ist hier zu finden.

www.leipzig.de/de/tourist/ - Ähnlich

[Weitere Ergebnisse anzeigen von www.leipzig.de](#)

[Leipzig](#) – Wikipedia

Leipzig ist eine kreisfreie Stadt im Nordwesten des Freistaates Sachsen und mit über 515.000 Einwohnern die bevölkerungsreichste Stadt in den fünf neuen ...

[Geographie](#) - [Geschichte](#) - [Politik](#) - [Wirtschaft und Infrastruktur](#)

de.wikipedia.org/.../Leipzig - vor 16 Stunden gefunden - [Im Cache](#) - Ähnlich

[News zu leipzig](#)

Anzeigen

[Hotels in Leipzig ab 20 €](#)

pro Person, 81 Hotels verglichen. Hier Ihr Hotel in **Leipzig** buchen!

www.Hotelreservierung.de/Leipzig

[Leipzig ab 18.-€](#)

Über 85 Hotels in **Leipzig** direkt beim **Leipzig** Spezialisten buchen!

www.Hotels-Leipzig.Reisen.de

[90% Rabatt in Leipzig](#)

Alles außer teuer - die besten Rabatte in **Leipzig** - Hol sie dir.

www.citydeal.de/Leipzig

[Schalten Sie hier Ihre Anzeige »](#)

Wie könnte das Ergebnis einer einfachen Suche aussehen


Texteingabe

Erste Trends, Karten, Bilder die dazu passen

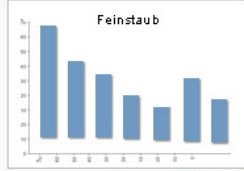
Weitere mögliche „sinnvolle“ Kombinationen aus den Daten



The screenshot shows the Hippolytos search interface. At the top, there are navigation links for 'Maps', 'Dienste', 'Ontologien', and 'Neues', along with 'Mein Account' and 'Anmelden'. The search bar contains the text 'Luftverschmutzung Karlsruhe 10 Jahre' and a 'Suche' button. Below the search bar, three main results are displayed: a bar chart titled 'Feinstaub', a map of Karlsruhe, and a photograph of an air quality measurement station. Each result has a 'browse' link and icons for a home button and a window icon. Below these are several text-based search results, each starting with a small Hippolytos logo icon. The results include 'Trend Ozone', 'Trend Feinstaub', 'Karte - Aktuelle Ozonwerte im Landkreis Karlsruhe', 'Plaketten: Stadt Karlsruhe' (with a PDF icon), and 'Ozonbelastung - Stadtwiki Karlsruhe'. At the bottom, there are 'Verwandte Suchvorgänge' and a pagination control.

Maps Dienste Ontologien Neues Mein Account Anmelden


 **HIPPOLYTOS**



Feinstaub






[browse](#)  





Karlsruhe



[browse](#)  



[browse](#)  

-  [Trend Ozone](#)
Entwicklung der Ozonwerte nach Regionen
[www.disy.net/hippolytos/.../region_trend_ozone...178623](#)
-  [Trend Feinstaub](#)
Naturschutz Baden-Württemberg. **Feinstaub** und Stickstoffoxide: Emissionen, Immissionen und **Trends** in Baden-Württemberg. R Fachtagung in **Karlsruhe** ...
[www.disy.net/hippolytos/.../trend_feinstaub...45866954](#)
-  [Karte - Aktuelle Ozonwerte im Landkreis Karlsruhe](#)
Besondere Informationen. Hinweise. Sie sind hier: Startseite **Aktuelle** Immissionsdaten > Ozon. **Aktuelle** 1-Stundenmittelwerte Ozon. 15.02.2010 06:00 ...
[www.disy.net/hippolytos/.../karten/aktuelle-ozonwerte_lk-karlsruhe/...588756-ay](#)
-  [Plaketten: Stadt Karlsruhe](#)
Dateiformat: PDF/Adobe Acrobat -
Regelung für **Plaketten** zur Verbesserung der Luftqualität in der Stadt Karlsruhe
[www.rp.baden-wuerttemberg.de/servlet/PB/show/.../rpt-ref54-1-flyer.pdf](#)
- [Ozonbelastung - Stadtwiki Karlsruhe](#)
Stadtwiki Karlsruhe. In Karlsruhe gibt es drei Ozonmessstationen. Die Messwerte ...
[ka.stadtwiki.net/Ozonbelastung](#)

Verwandte Suchvorgänge: [Luftwerte Karlsruhe](#) [Trend Ozone](#) [Karlsruhe](#) [Luftqualitätsparameter](#)

[|<](#) [<<](#) [Zurück](#) [5](#) [6](#) [7](#) [8](#) [9](#) [10](#) [11](#) [12](#) [13](#) [14](#) [15](#) [10](#) [11](#) [12](#) [13](#) [14](#) [15](#) [10](#) [11](#) [12](#) [13](#) [Vorwärts](#) [>>](#) [|>](#)

Wie könnte das Ergebnis einer einfachen Suche aussehen

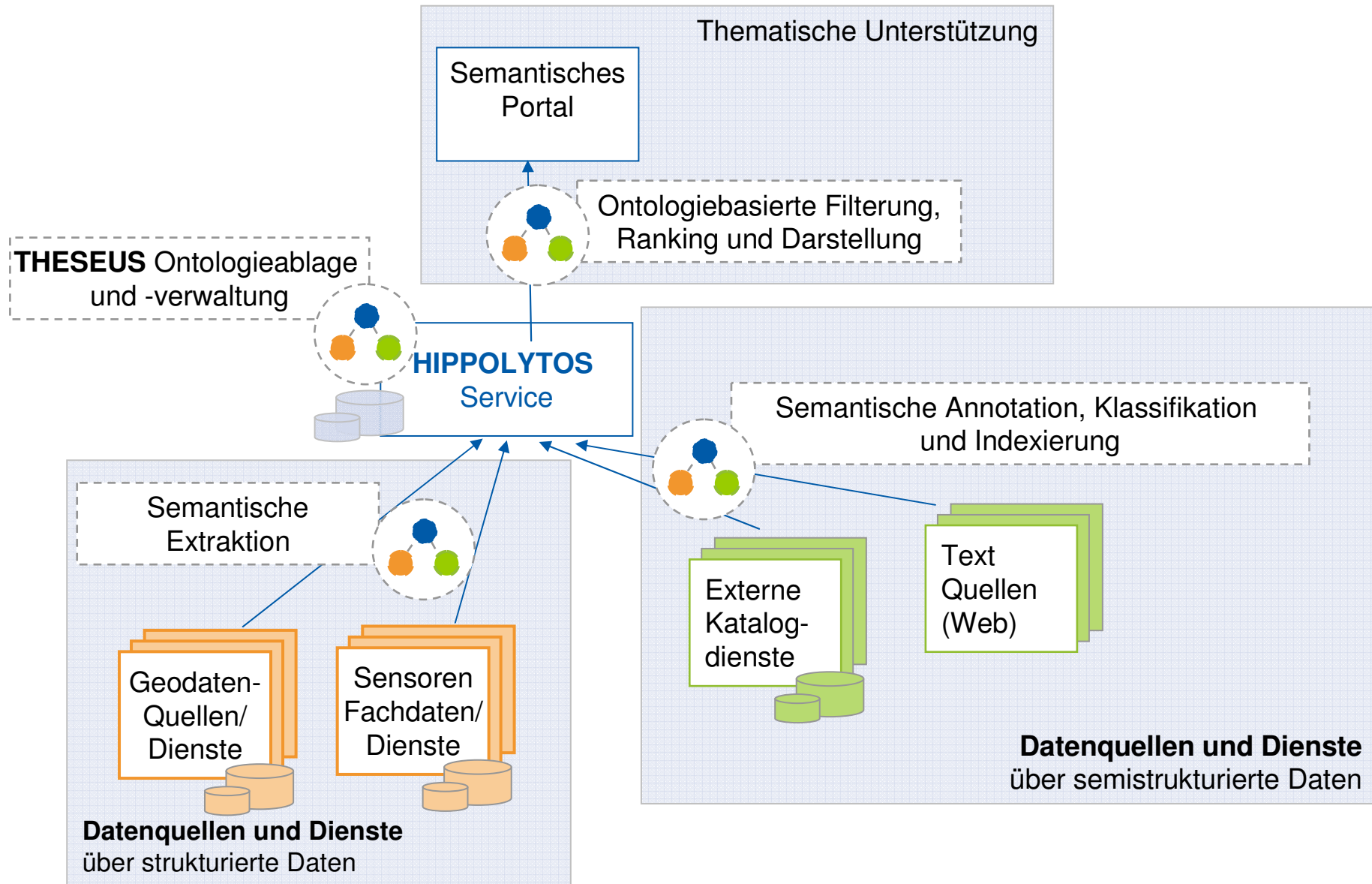
The screenshot displays the HIPPOLYTOS web application interface. At the top, there are navigation links for 'Maps', 'Dienste', and 'Ontologien Neues', along with 'Mein Account' and 'Anmelden'. The search bar contains the query 'Luftverschmutzung Karlsruhe 10 Jahre'. Below the search bar, there are sections for 'Daten auswählen:' and 'Datendarstellung:'. The 'Daten auswählen:' section includes 'Verwandte Begriffe' (Luftqualität, Luftverschmutzung, Lufttemperatur, Windrichtung, Windgeschwindigkeit), 'Unterbegriffe' (Benzol, Kohlenmonoxid, Ozone, Schwefeldioxyd, Stickstoffdioxyd), 'Datum' (a bar chart showing ozone levels from 1990 to 2002), and 'Karte' (a map of Germany). The 'Datendarstellung:' section shows a map of the Karlsruhe region with various stations marked. A table lists stations with their numbers and names, and a detailed view for station 4441 (Karlsruhe-Mitte) shows ozone levels for the year 2010 as 33 µg/m³. A line chart shows the ozone levels for station 4441 from 2001 to 2010.

Drucken,
Bericht erstellen
versenden

Langsam hinführen
zu komplexeren
Visualisierungs- und
Analysemöglichkeiten

Lösungsweg

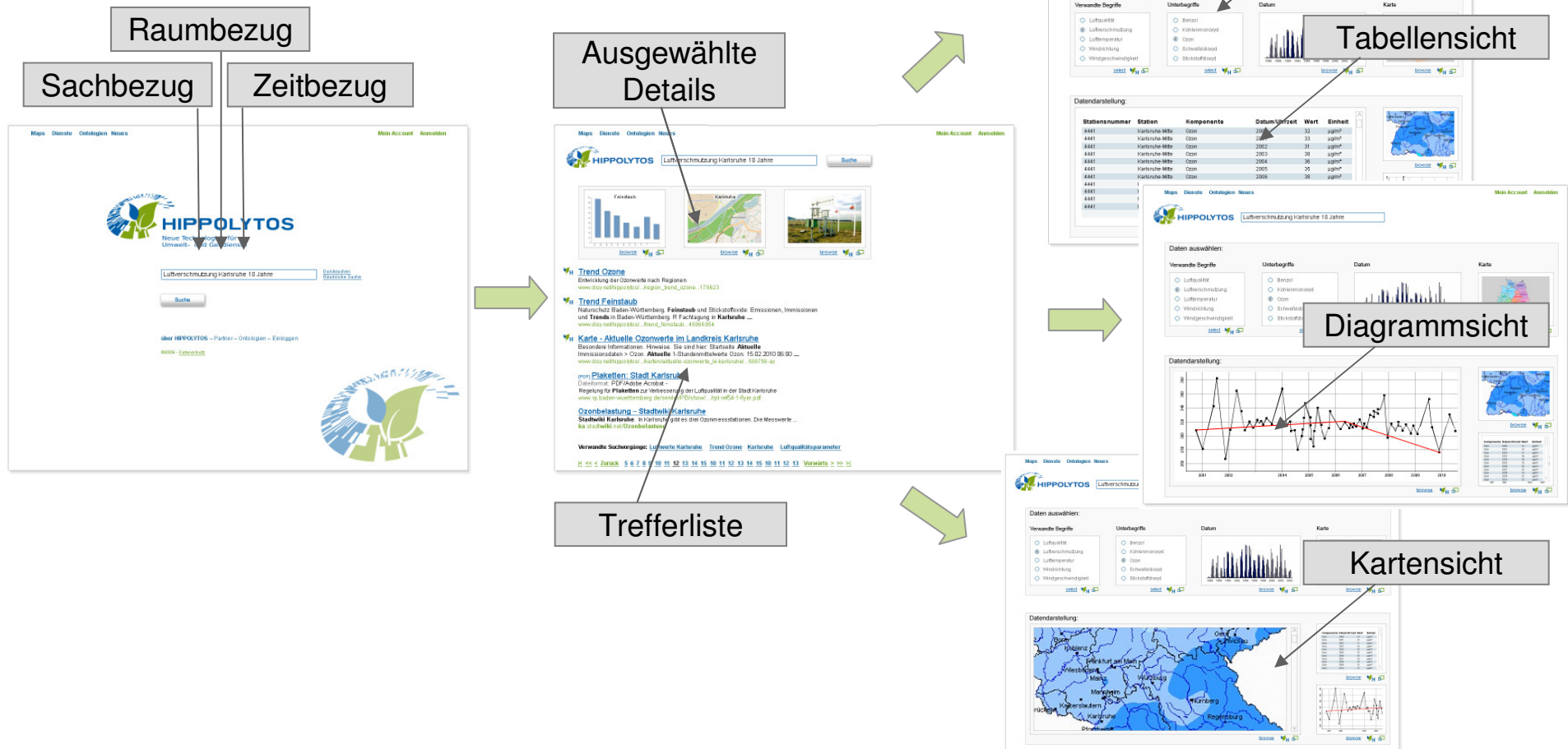
Grundidee



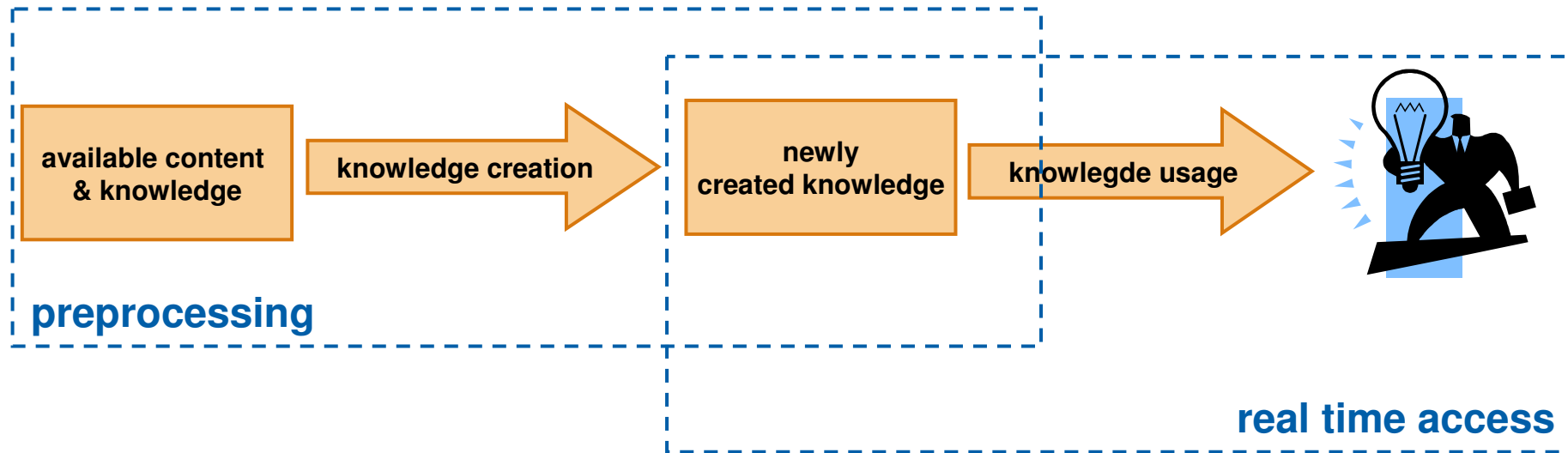
Aufbau

Intelligente, integrierte Datenrecherche

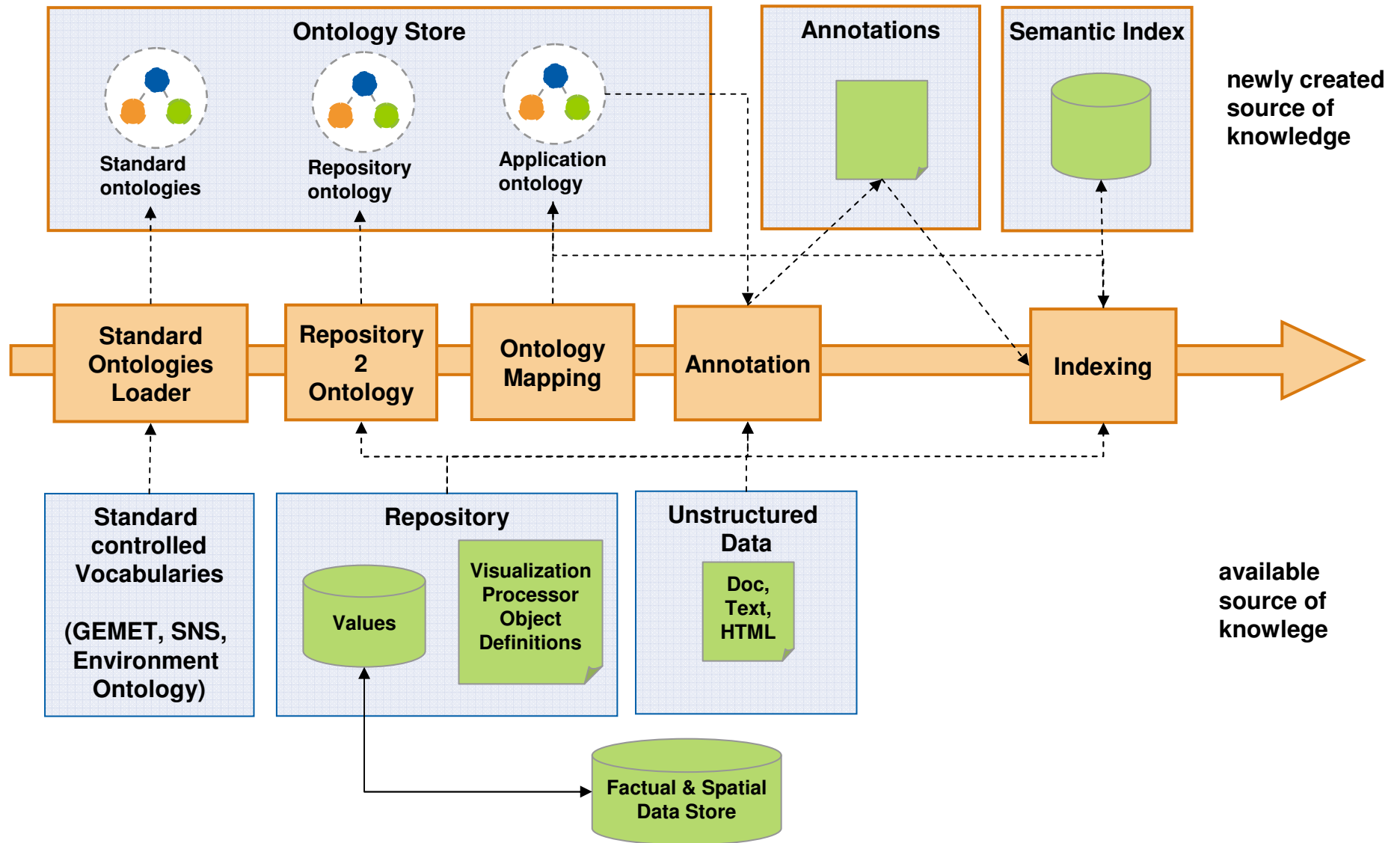
- „Hat sich die Luftqualität in den letzten 10 Jahren in Karlsruhe verschlechtert?“



Gesamtprozess



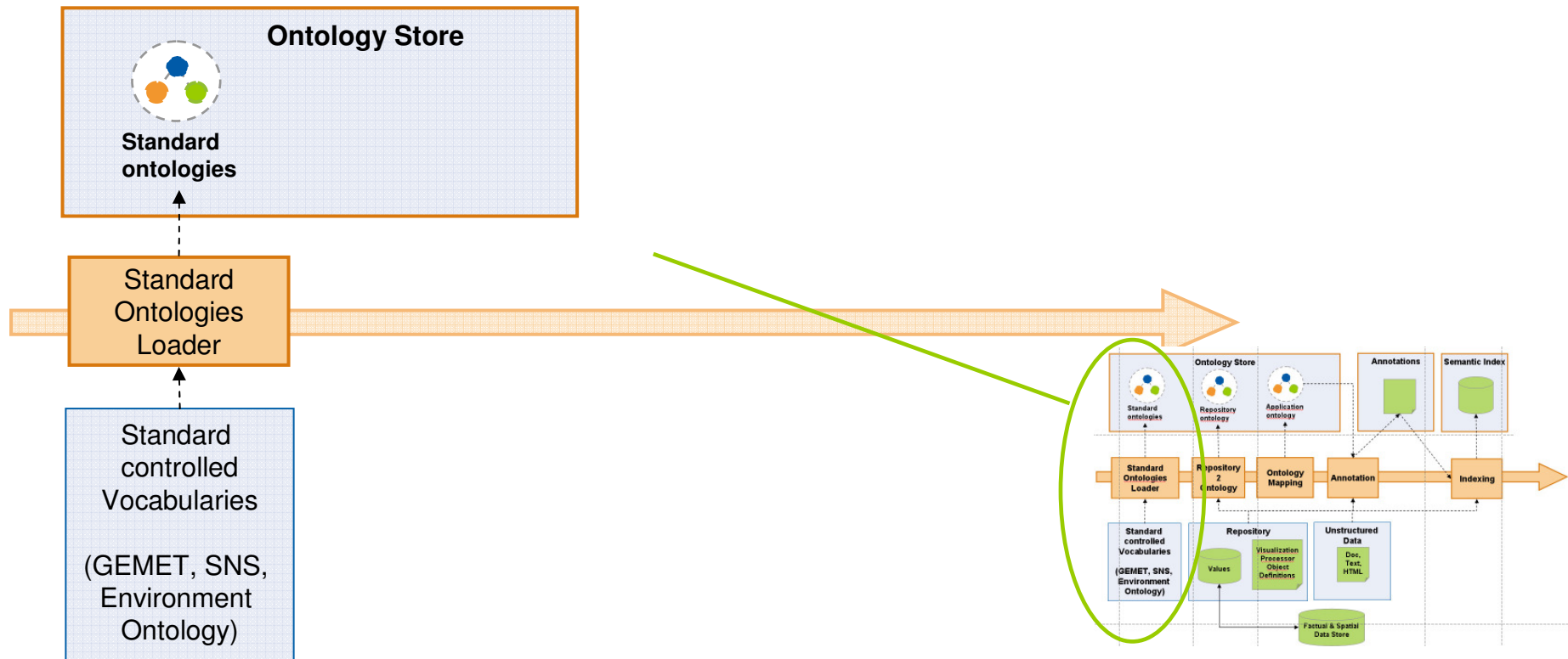
Preprocessing/Knowledge Creation



Standard Ontology Loader Modul

Liest Standardontologien in das Ontology-Store ein.

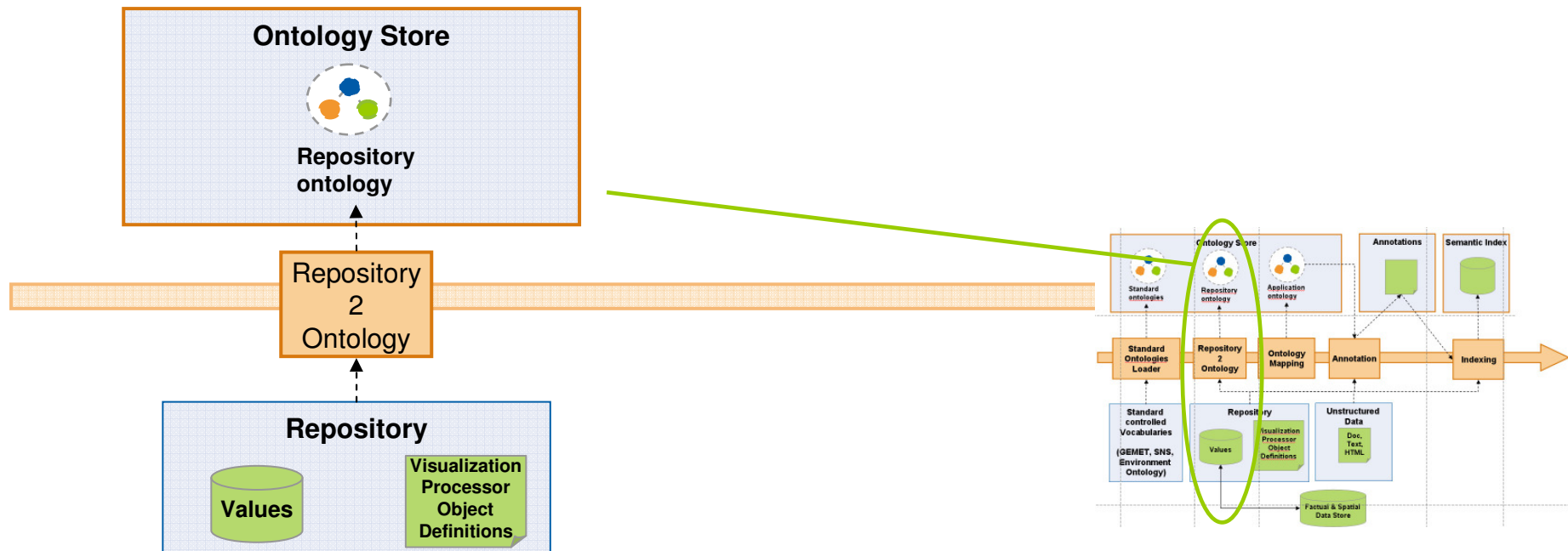
- Eingabe: Standardontologien
- Ausgabe: Ontology Store



Repository2Ontology Modul

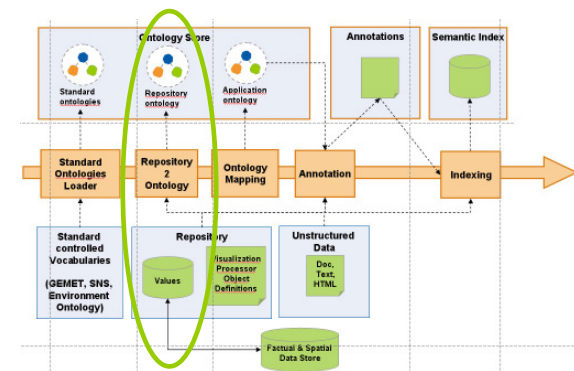
Erstellt projektspezifische Ontologie aus dem Repository + Datenbanken

- Eingabe: Repository
(Objekte, Prozessoren, Visualizations, DB-Schlüssellisten)
- Ausgabe: Repository-Ontologie
- Relevante Basistechnologie
 - MNEMOSYNE -- DB-Persistenz für OWL Ontologien



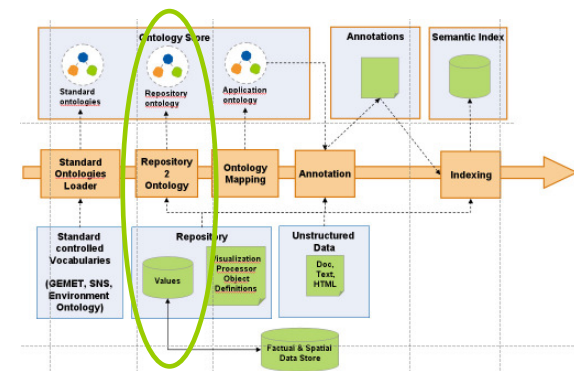
Repository2Ontology Modul - 2

- Format der Zielontologie
 - OWL-Lite
- Repository-Elemente für Ontologie
 - Begriffe
 - Navigator-Zwischenknoten
 - Selektornamen
 - (Kriteriennamen)
 - Objektnamen
 - Attributnamen
 - Beziehungen
 - Assoziationen
 - Vererbung
 - Hierarchie
 - Navigator-Hierarchie



Repository2Ontology Modul - 3

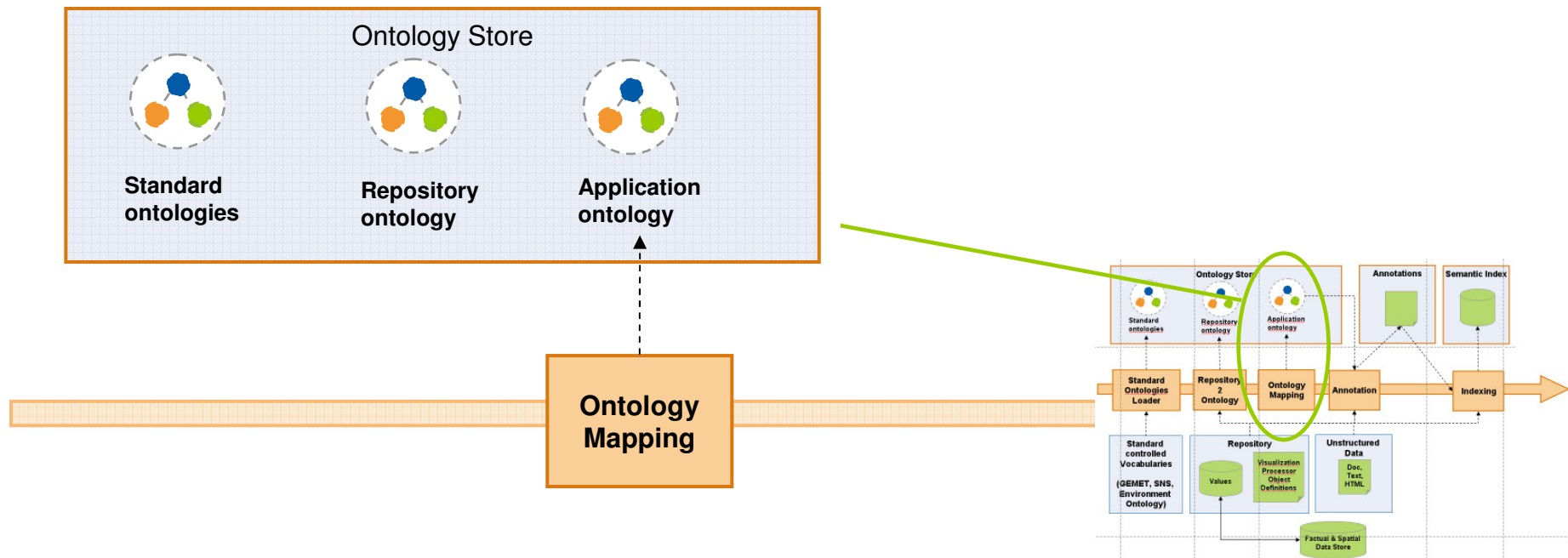
- Notwendige Erweiterungen
 - Repository-Objekte,
 - Hierarchie
 - Vererbung
 - Schlüsselliste
 - Technische Metadaten
 - Visualisations - sinnvolle Datengrundlage



Ontology Mapping Modul

Erstellt Mapping zwischen dem Repository-Ontology und Standardontologien

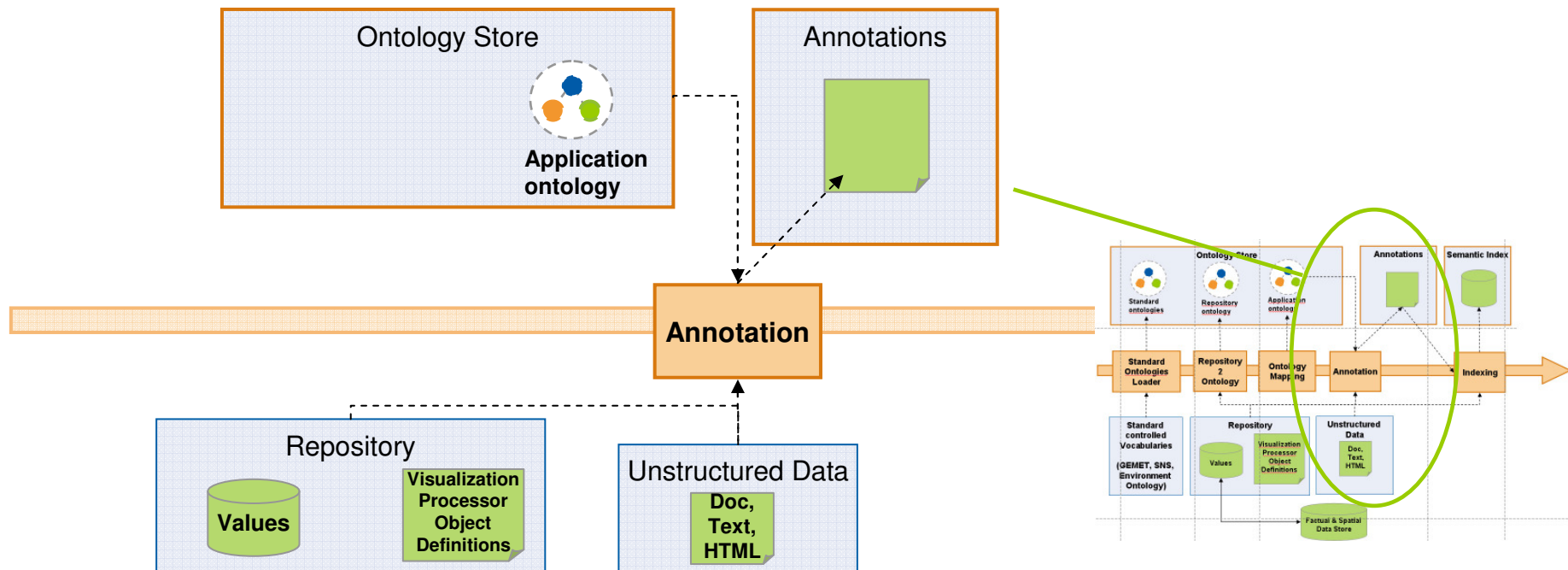
- Eingabe: Repository-Ontology, Standardontologien
- Ausgabe: integrierte Application-Ontology
- Relevante Basistechnologie:
 - HARMONIA: Ontology Mapping



Annotation Modul

Erstellt Annotationen für das Repository und für die Dokumente

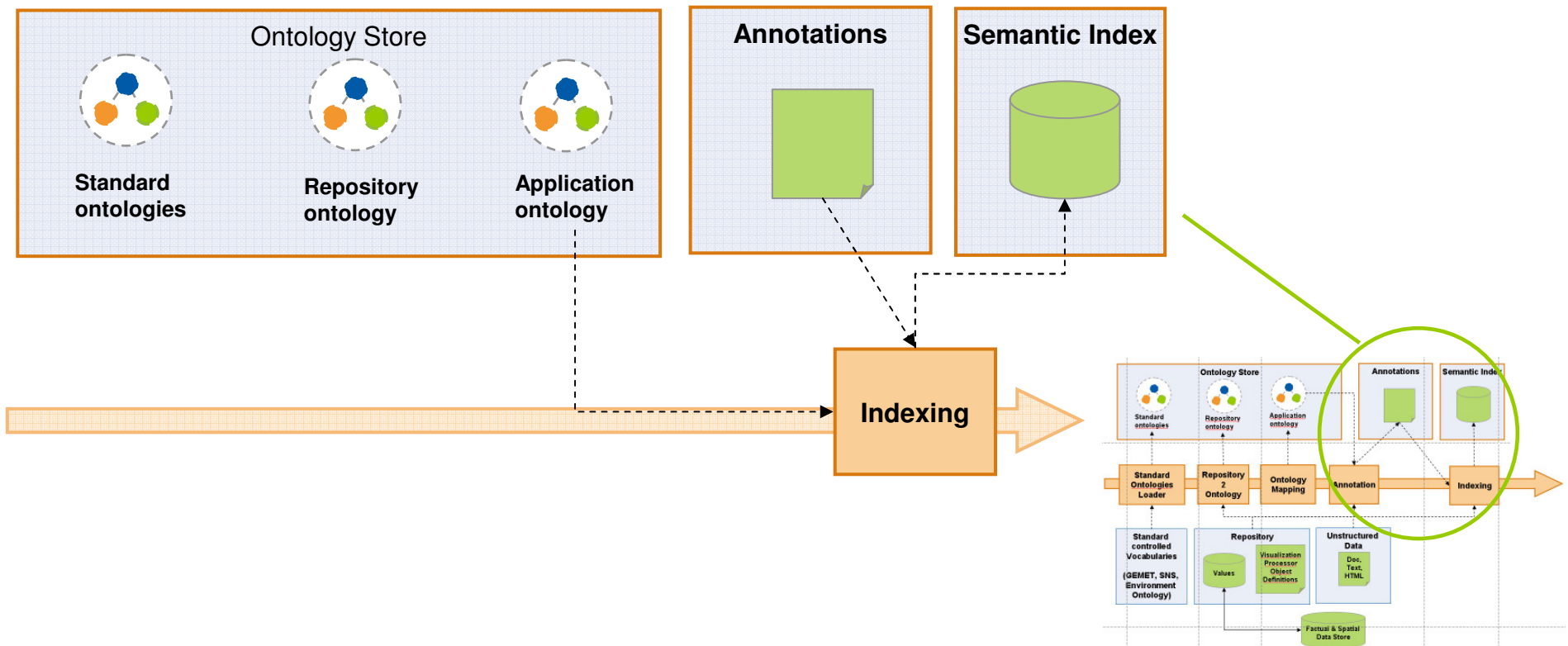
- Eingabe: Application-Ontologie, Repository, Dokumente
- Ausgabe: Annotationen vom Repository und Dokumente mit der Application-Ontologie
- Relevante Basistechnologie
 - ARETE – Annotierung von Textdokumenten



Semantic Indexing Modul

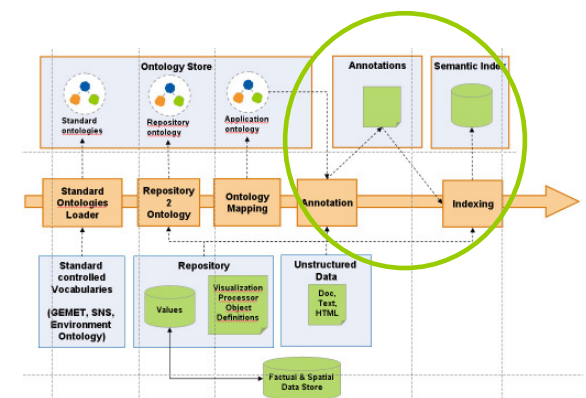
Erstellt semantisches Index für die Application-Ontologie und die Annotations

- Eingabe: Application-Ontologie, Repository, Annotationen
- Ausgabe: Semantisches Index

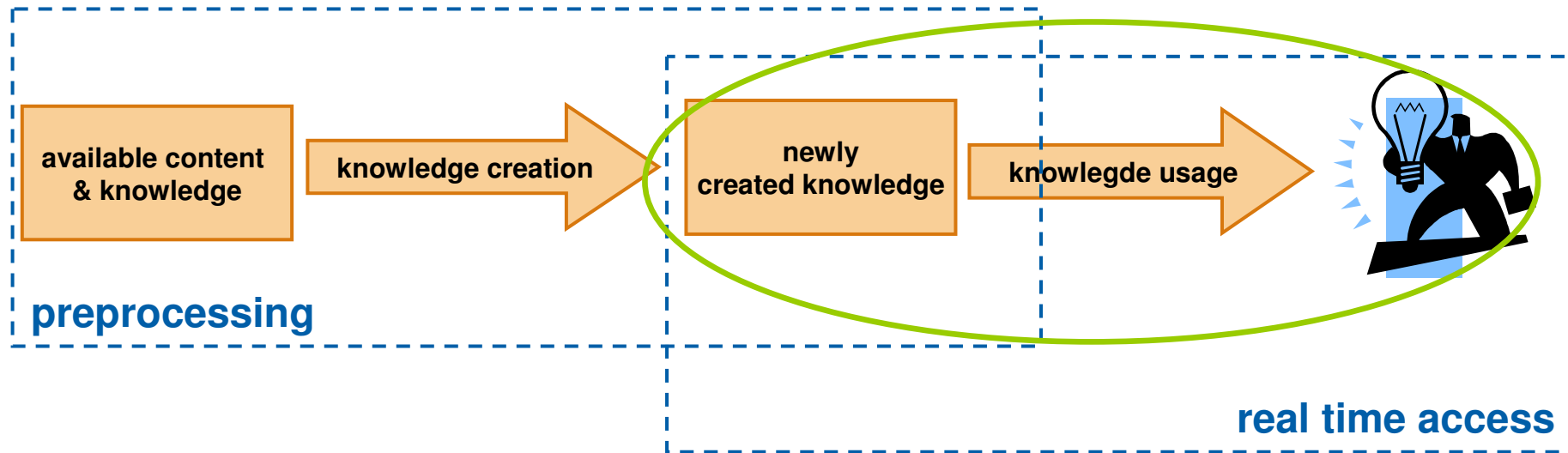


Semantic Indexing Modul - 2

- Technologie, Werkzeug für Index
 - e.g. Lucene
- Synonyme im Index direkt berücksichtigt
- Informationen für Ranking berücksichtigt



Gesamtprozess

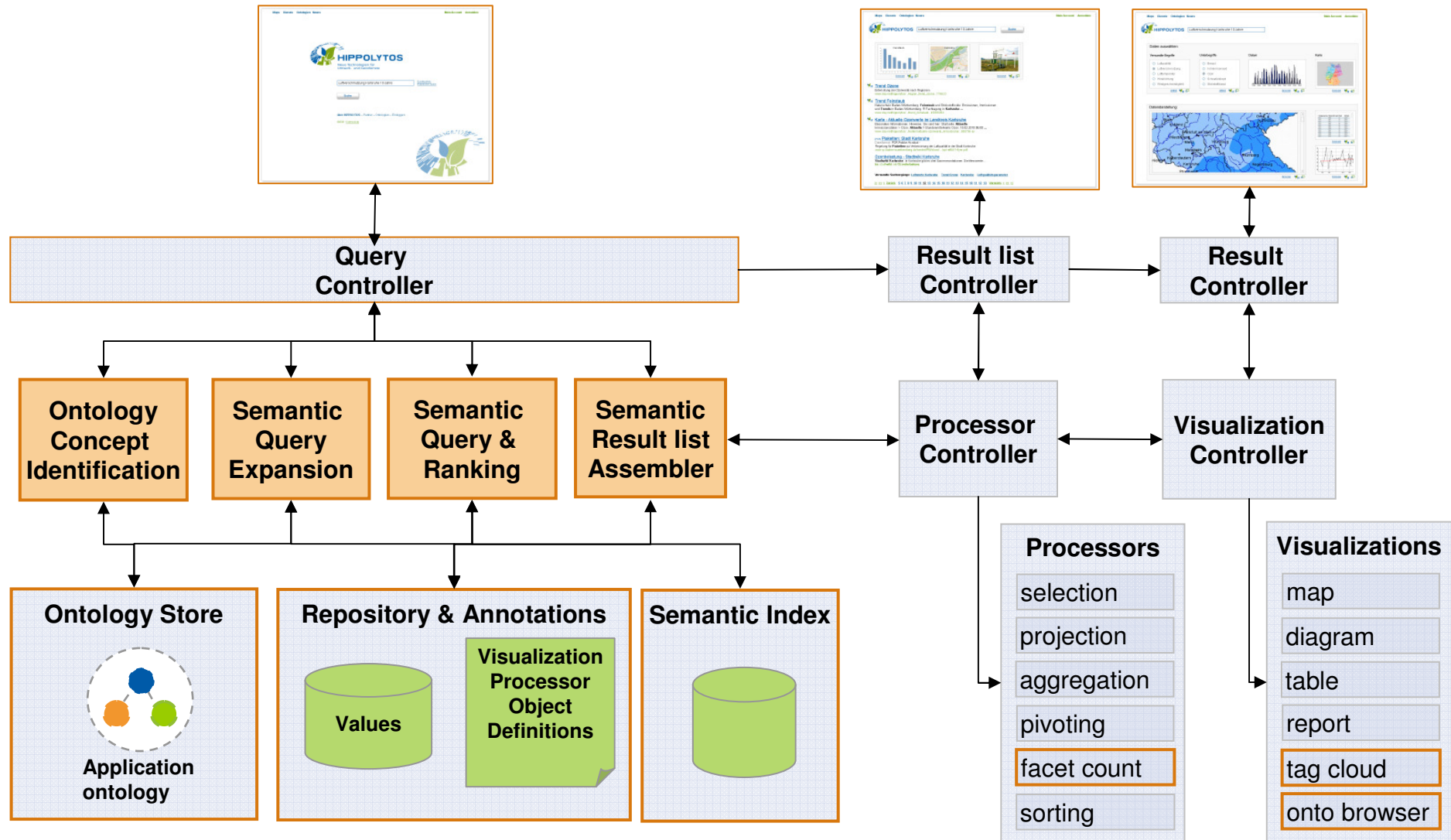


Intelligente, integrierte Datenrecherche

- „Hat sich die Luftqualität in den letzten 10 Jahren in Leipzig verändert?“

Stationsnummer	Station	Komponente	Datum/Uhrzeit	Wert	Einheit
4441	Karlsruhe-Mitte	Ozon	2001	33	µg/m³
4441	Karlsruhe-Mitte	Ozon	2001	33	µg/m³
4441	Karlsruhe-Mitte	Ozon	2002	31	µg/m³
4441	Karlsruhe-Mitte	Ozon	2002	30	µg/m³
4441	Karlsruhe-Mitte	Ozon	2004	36	µg/m³
4441	Karlsruhe-Mitte	Ozon	2005	35	µg/m³
4441	Karlsruhe-Mitte	Ozon	2008	38	µg/m³

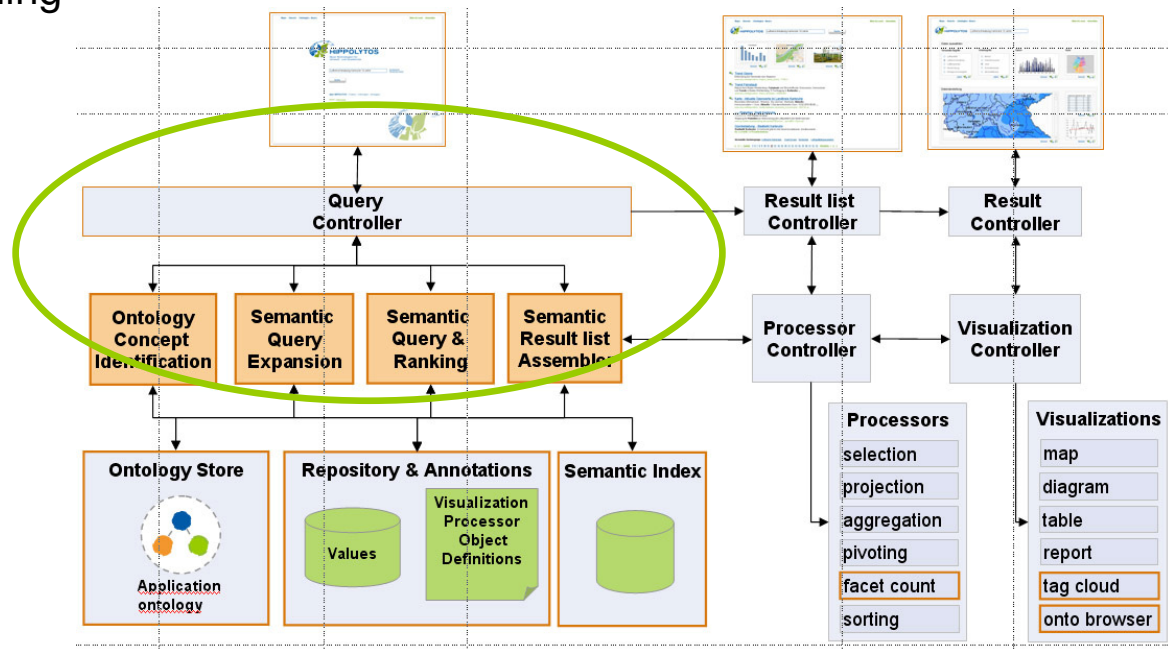
Real time access / Knowledge usage



Query Controller

Erstellt aus einer Suchanfrage die Ergebnisliste

- Eingabe: Benutzeroberfläche, Ontology-Store, Semantic Index
- Ausgabe: Ergebnisliste
- Relevante Basistechnologien
 - HERAKLES: Reasoning-Broker
 - AQA: approximate reasoning



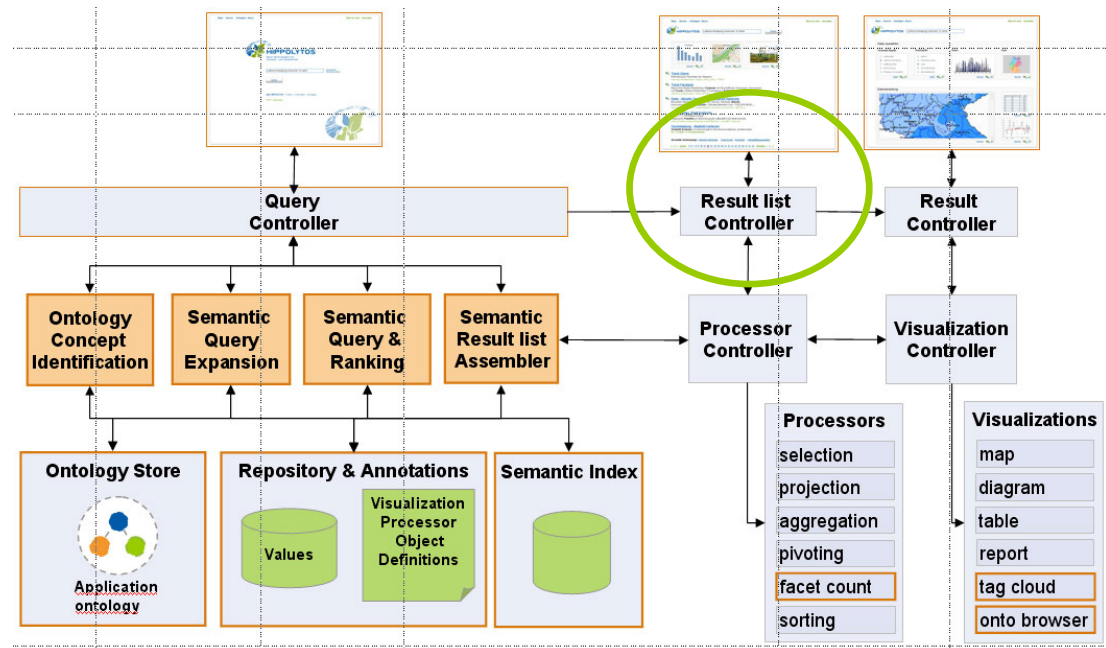
Query Controller - 2

- Ontology Concept Identification Submodul
 - Parsen der Query
- Semantic Query Expansion Submodul
 - Automatisierung riskant
 - Gefahr nach oben: zu allgemein
 - Gefahr nach unten: Performanz, zu breit
 - Benutzerinteraktion
 - Browsing der Ontologie
 - (facet count Informationen anbieten)
 - Benutzerauswahl
- Semantic Query and Ranking Submodul
 - Ausführung der Suchanfrage mit Hilfe des Semantic Index
- Semantic Result List Assembler Submodul
 - Erstellung der Ergebnisliste mit möglichen Prozessoren und Kriterienbelegungen

Result List Controller Modul

Erstellt aus einer Ergebnisliste die Ergebnisdarstellung (Liste + ausgewählte Details)

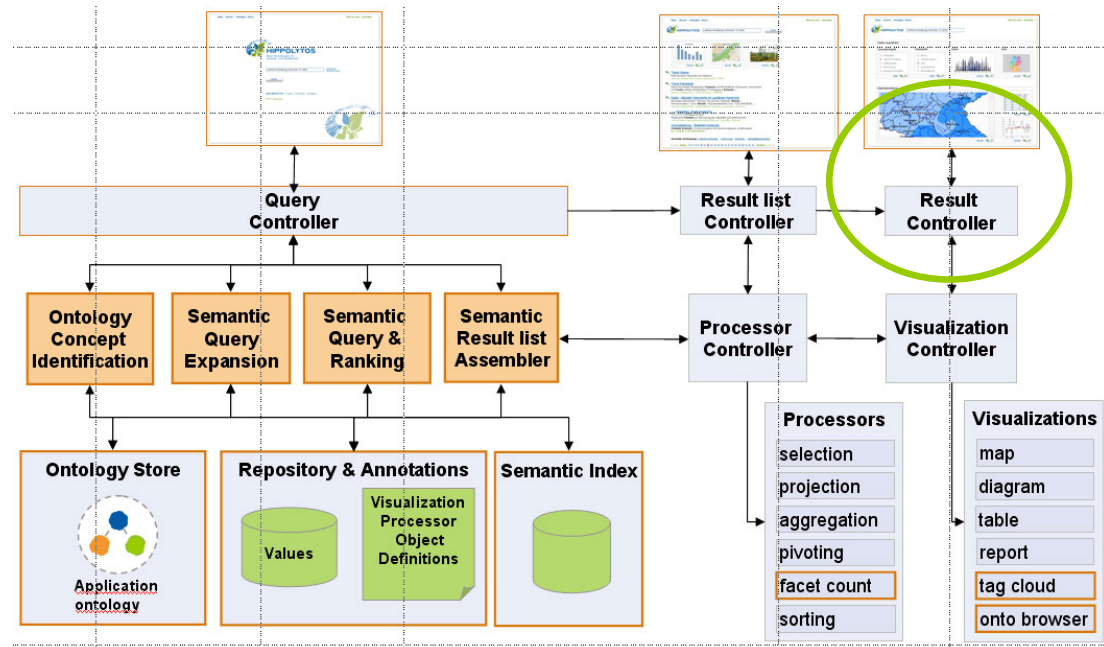
- Eingabe: Result List, Processor Controller, Visualization Controller
- Ausgabe: Ergebnisliste, ausgewählte Details



Result Controller Modul

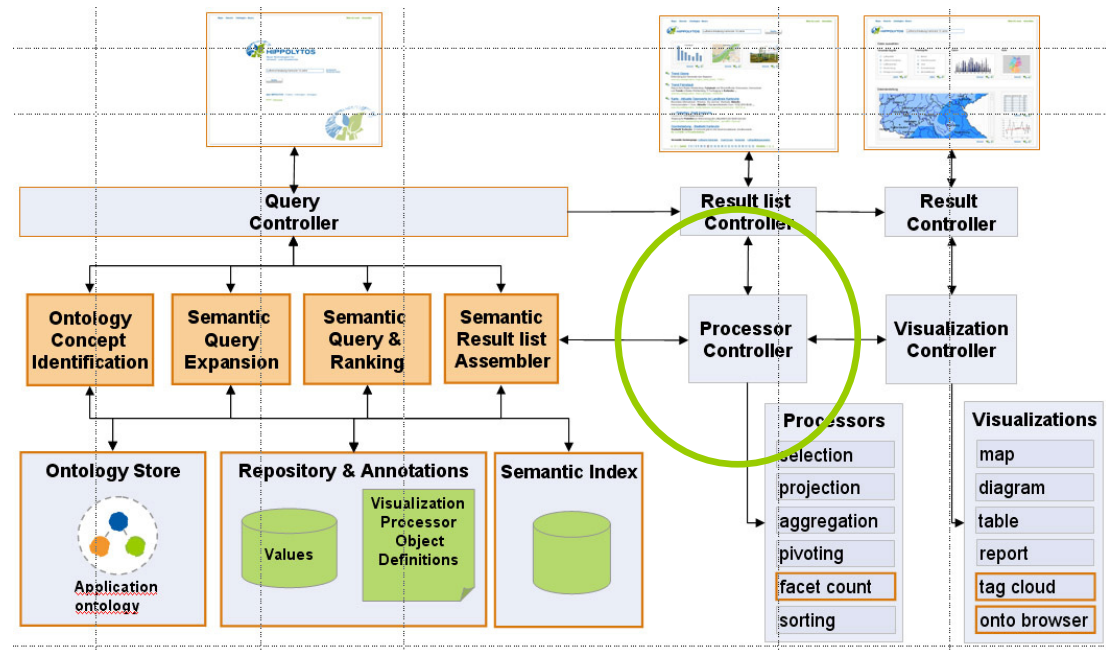
Erstellt das Detailergebnis für einen Treffer aus der Liste

- Eingabe: Treffer vom Result List Controller
- Ausgabe: Detailergebnis



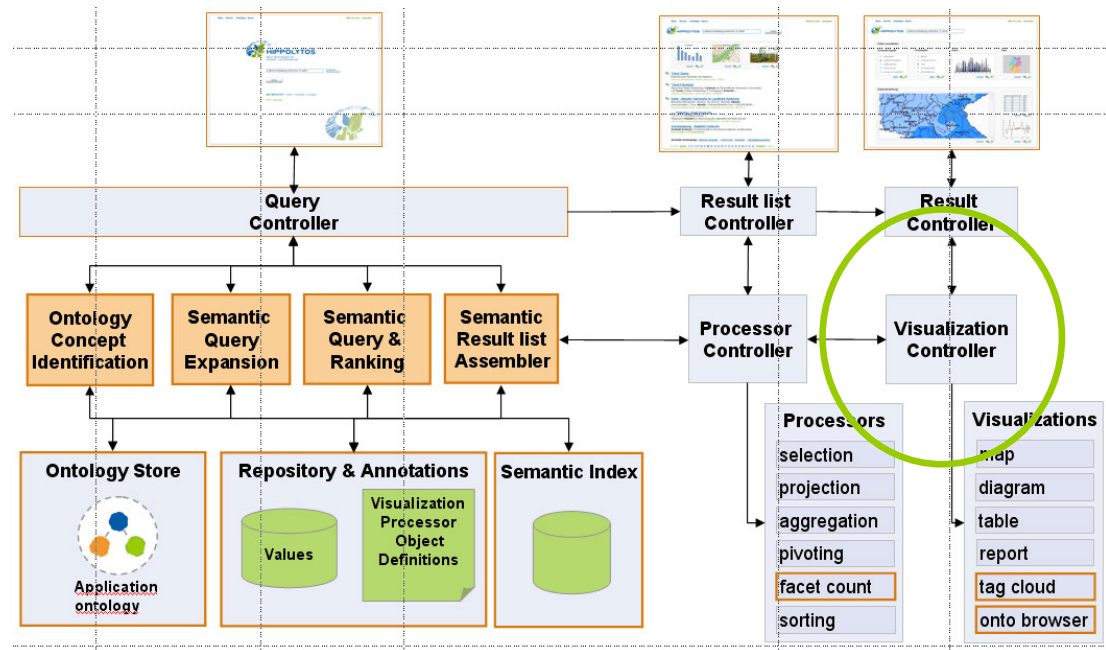
Visualization Controller Modul, Visualization Modul

Kontrollieren der Ausführung eines Prozessors, Prozessordefinition



Processor Controller Modul, Processor Modul

Kontrollieren der Ausführung einer Visualisierung, Definition einer Visualisierung



Umsetzung und Fortschritt

HIPPOLYTOS – Projektstand

- Aktuell: Konzeption und erste technische Experimente / Prototypen
- Konkrete Szenarien spezifiziert
 - Unter Nutzung von THESEUS Technologien
 - auf der Basis von disy Produkten
 - Im Bereich Umwelt- und Geodaten
- Vorauswahl von THESEUS Technologien
- Architektur erstellt und wesentliche Module identifiziert

Nächste Schritte

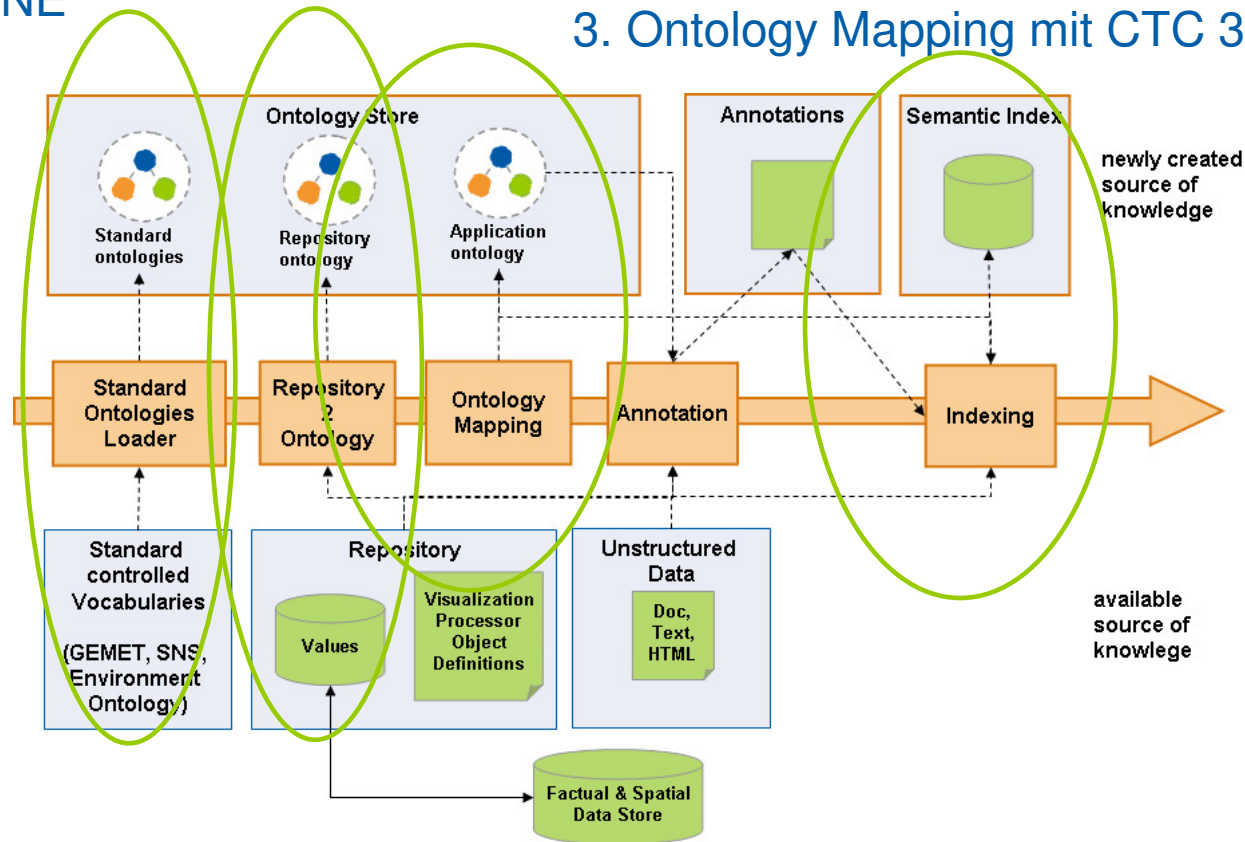
Nächste Schritte (1)

1. Laden von Standards in MNEMOSYNE

2. Laden des Repositories in MNEMOSYNE

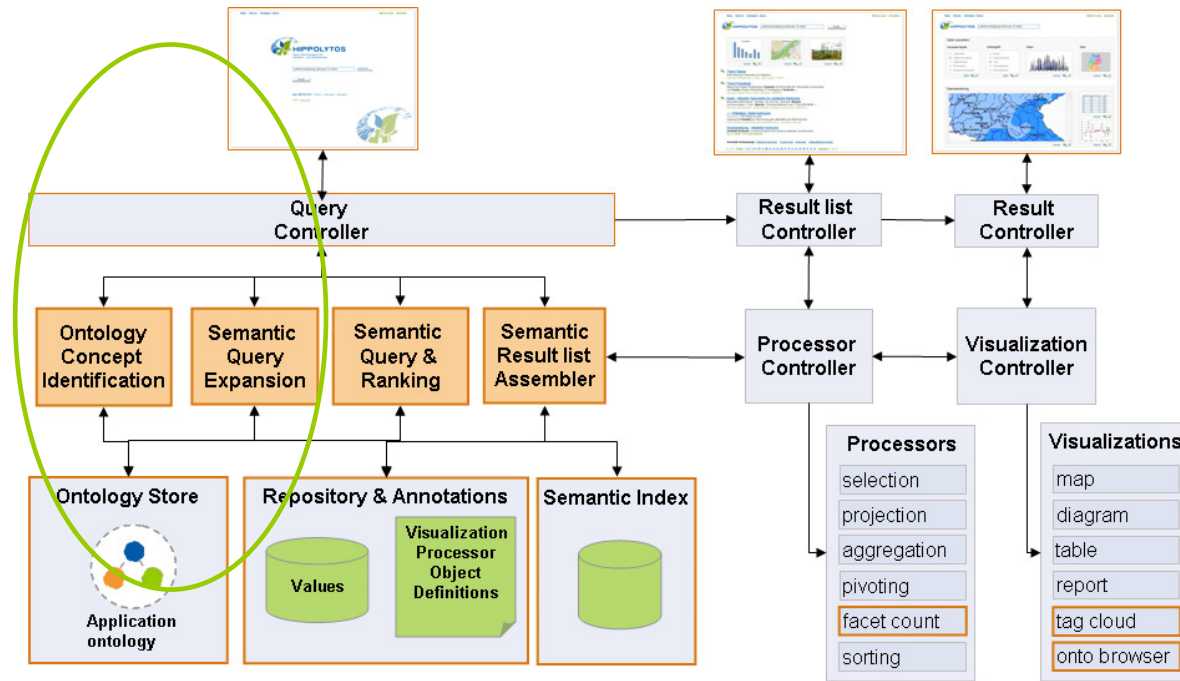
3. Ontology Mapping mit CTC 3

4. Erster Semantic Index mit echten Daten

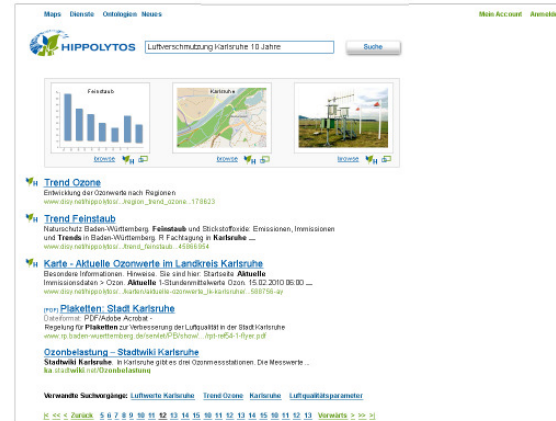


Nächste Schritte (2)

5. Erste Umsetzung Suche und Expansion



Zusammenfassung



- Forschung und Entwicklung für einen 3. Zugang zu Umwelt- und Geodaten
- Über einen einfachen „textbasierter“ Zugang zu komplexen Informationen und Sachverhalten hinführen
- Erster Pilot im September geplant
- Abschluss des Projekts Juli 2011
- Integration von erfolgreichen Konzepten und Ansätzen in disy Cadenza
- Aktuelles zum Projekt: <http://www.disy.net/hippolytos>

Vielen Dank für Ihre Aufmerksamkeit



disy Informationssysteme GmbH
Erbprinzenstr. 4–12
76133 Karlsruhe

Tel.: +49 721 1 6006-260
Fax: +49 721 1 6006-05

E-Mail: wassilios.kazakos@disy.net

www.disy.net



Dr. Wassilios Kazakos



Dr. Gergely Lukacs

