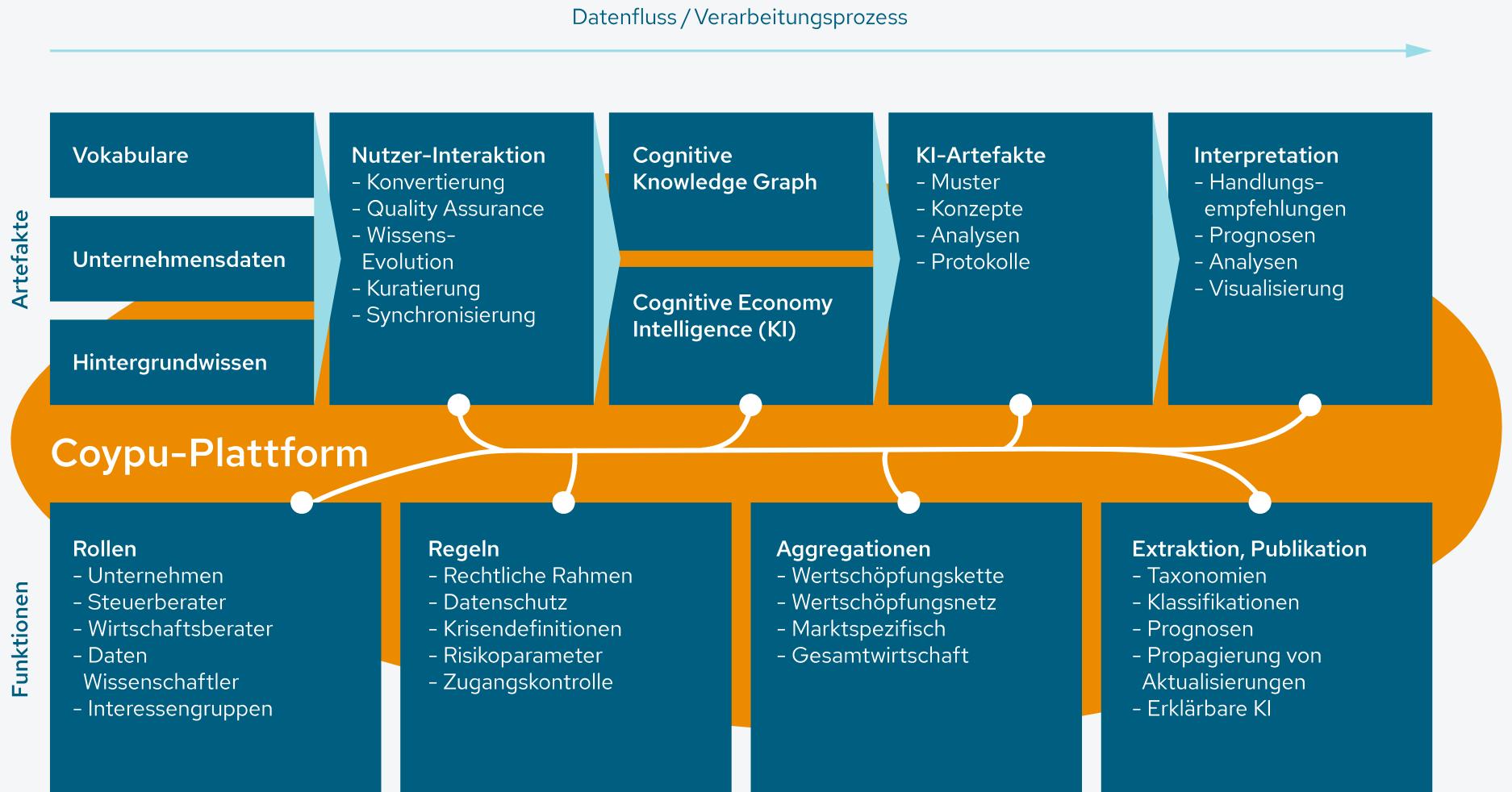


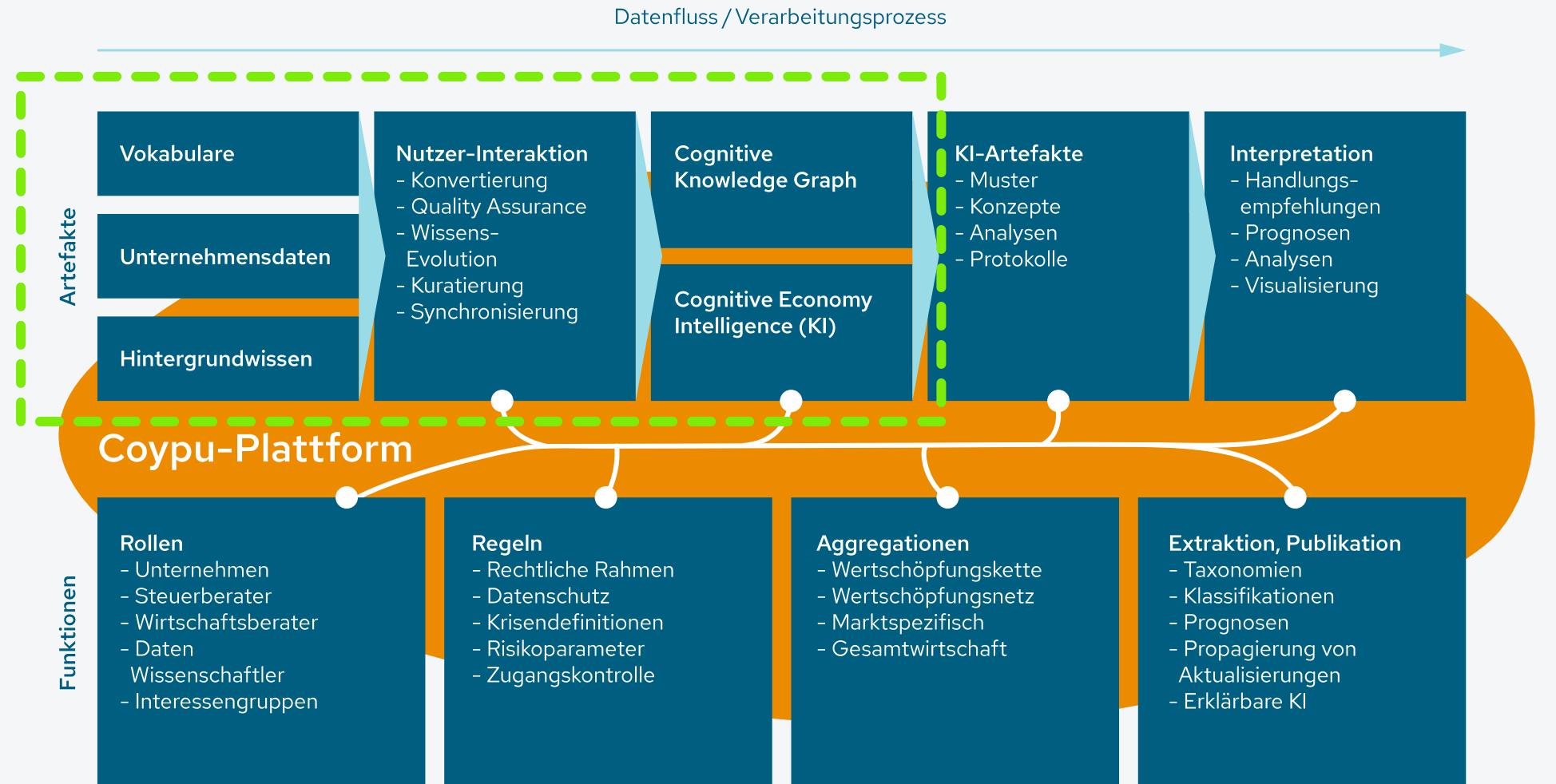
The Coypu Platform Data Integration Workflow

Dr. Natanael Arndt
Senior Linked Data Expert @ eccenca GmbH

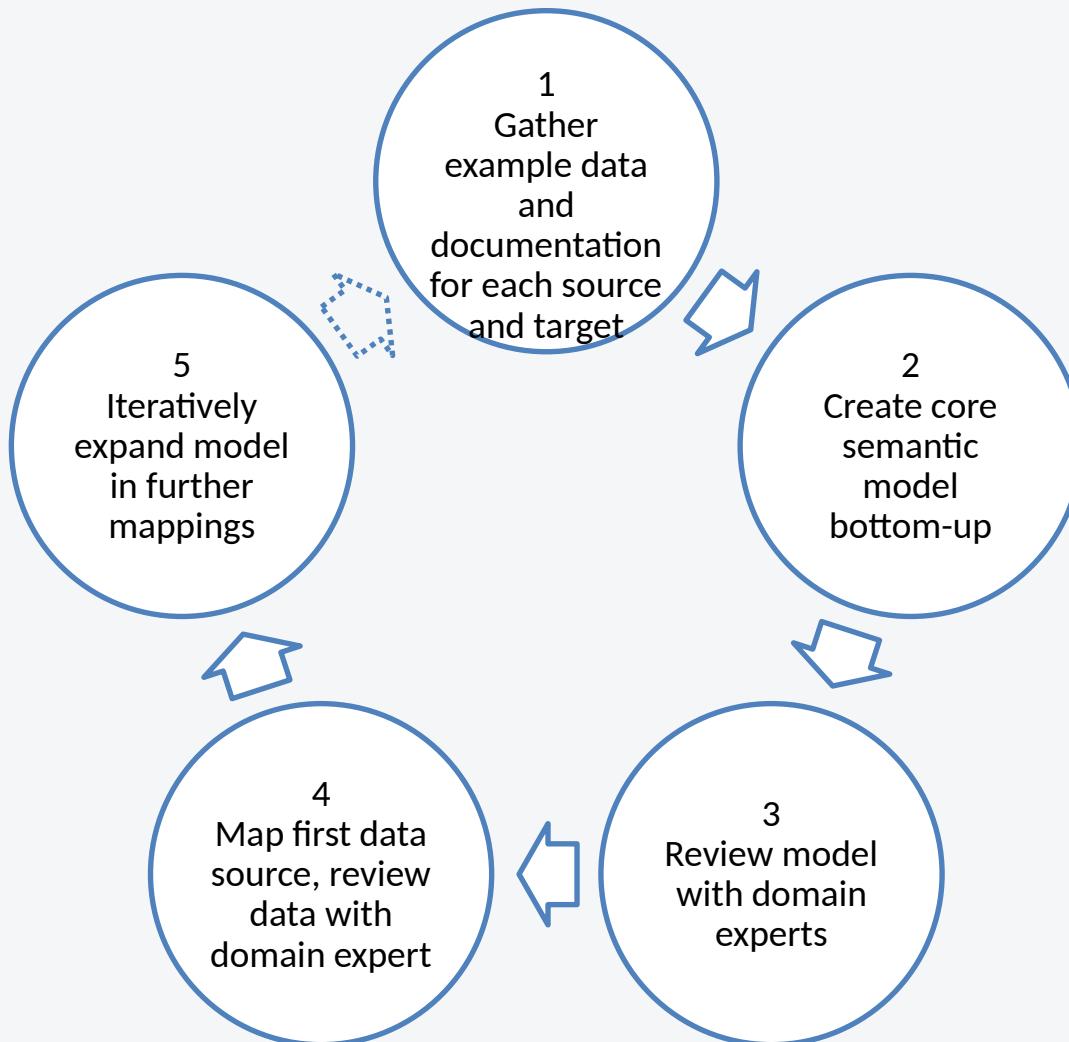
The Coypu-Plattform



The Coypu-Plattform



Enterprise Knowledge Graph Project Flow



1. Analyze data sources and understand domain
2. Find or build a semantic model (ontology/vocabulary)
3. Create the mappings
4. Clean, enrich and link data
5. Explore the results
... Re-iterate if needed ...

Data Sources

The screenshot shows the data.europa.eu homepage with a search bar for datasets. Below it, a map of Europe allows filtering by location. A sidebar on the left provides filter settings for operators (AND/OR), data scopes (European Union, International, National), and catalogues (GDI-DE, 352 806). The main content area displays two dataset cards:

- Oficiālā elektroniskā adrese (e-adrese)**: Information about official electronic addresses, last modified on 04.07.2022 at 18:30, created on 01.10.2019 at 11:19, from data.gov.lv.
- COVID-19 apstiprināto gadījumu skaits un 14 dienu kumulatīvā saslimstība pa administratīvajām teritorijām**: Data on COVID-19 cases and cumulative incidence by administrative regions, last updated on 04.07.2022 at 16:22, created on 08.04.2020 at 02:00, from data.gov.lv.

Authoritative Country Dataset from the EU XML Dataset

The screenshot shows the GeoNames website. At the top is a search bar with a dropdown set to "all countries". Below it is a placeholder text: "enter a location name, ex: \"Paris\", \"Mount Everest\", \"New York\"". The page is divided into sections: "Browse the names", "Information", and "Download".

- Browse the names**: Includes links to Countries, Postal codes, Country statistics, Ambassadors and Team, Forum, Blog, Mailing list, and Commercial Support and Consulting.
- Information**: Includes links to About GeoNames, Data Sources, User manual, and Ambassador and Team.
- Download**: Includes links to Info, Free Gazetteer Data, Free Postal Code Data, Premium Data, and Web Services.

Web Services includes links to Overview, Documentation, Client Libraries, and Premium Web Services.

Sponsoring lists various online casinos and sports betting sites.

A note at the bottom states: "This work is licensed under a Creative Commons Attribution 4.0 License".

Open CC-BY licensed global geo information CSV/TSV Files

The Vocabulary

The image shows two screenshots of the CoyPU Ontology interface. The left screenshot displays the 'Properties' tab for the 'country' class, listing properties such as type (Class), subClassOf (location), Label (country@en), Comment (a country as defined in ISO 3166@en), and seeAlso (ontology.Country, schema.org.Country). The right screenshot displays the 'References' tab for the 'country' class, listing predicates like has adaptive capability, has coping capability, has currency, has exposure, has ISO code, has language, has population, has risk level, has susceptibility, and has vulnerability, each associated with a domain and the CoyPU Ontology.

TODO: publish under <https://schema.coypu.org/global>

Transformation of the Datasources

Summary

Label publications.europa.eu Extract Country XML distribution

Description TRANSFORMATION 1 in <https://ns.coypu.org/country>
.. more

Created by unknown user. Last modified 2022/06/07 by white-gecko.

Mapping editor

root

Country https://data.coypu.org/country/{code-3166-1-alpha-3}

Target entity type Country · coy:Country · a country as defined in ISO 3166

URI pattern https://data.coypu.org/country/{code-3166-1-alpha-3}

Examples of target data

Value path	Value	Transformed value
/code-3166-1-alpha-3	AND	https://data.coypu.org/country/AND

Label root

Edit Copy

Mapping rules (19)

Rule	Type	From	To
code-3166-1-alpha- coy:code-3166-1-alpha-2	StringValueType	code-3166-1-alpha-2	
URI of code-3166-1-alpha- owl:sameAs	n/a	n/a	→
code-iana coypu-ref:code-iana	StringValueType	code-iana	
URI of code-iana owl:sameAs	n/a	n/a	→
code-3166-1-num coy:code-3166-1-num	StringValueType	code-3166-1-num	
URI of code-3166-1-num owl:sameAs	n/a	n/a	→
authority-code coy:authority-code	StringValueType	authority-code	
URI of authority-code owl:sameAs	n/a	n/a	→
code-3166-1-alpha-1	StringValueType	code-3166-1-alpha-1	

Transformation of the Datasources

Summary

Label	publications.europa.eu Extract Country XML distribution
Description	TRANSFORMATION 1 in https://ns.coypu.org/country .. more
Created by unknown user. Last modified 2022/06/07 by white-gecko.	

Mapping editor

root

Country
https://data.coypu.org/country/{code-3166-1-alpha-3}

Multiple entities are allowed

Target entity type: Country • coy:Country • a country as defined in ISO 3166

URI pattern: https://data.coypu.org/country/{code-3166-1-alpha-3}

Examples of target data:

Value path	Value	Transformed value
/code-3166-1-alpha-3	AND	https://data.coypu.org/country/AND

Label: root

Mapping rules (19)

Path	Type	Value
code-3166-1-alpha- coy:code-3166-1-alpha-2	StringValueType	code-3166-1-alpha-2
URI of code-3166-1-alpha- owl:sameAs	n/a	n/a
code-iana coypu-ref:code-iana	StringValueType	code-iana
URI of code-iana owl:sameAs	n/a	n/a
code-3166-1-num coy:code-3166-1-num	StringValueType	code-3166-1-num
URI of code-3166-1-num owl:sameAs	n/a	n/a
authority-code coy:authority-code	StringValueType	authority-code
URI of authority-code owl:sameAs	n/a	n/a
code-3166-1-alpha-1	StringValueType	code-3166-1-alpha-1

Mapping editor

Label: URI

Source paths: Country XML distri...

- (custom path)
- @deprecated
- @id
- @IMMC.proposal.date
- @IMMC.approval.date
- @date.creation
- @adm.status
- @celex
- @pub
- code-3166-1-alpha-2

Transformations: Recommended

- Constant
- Default Value
- Lower case
- Tokenize

Mapping editor Transform evaluation Transform execution ✓

```

graph LR
    constant[constant0  
Value: https://data.coypu.org/country/...  
Constant (Transform)] --> path1[path1  
code-3166-1-alpha-3  
Path (Source)]
    path1 --> encode1[encode1  
Encoding: UTF-8  
Encode URL (Transform)]
    encode1 --> buildUri[buildUri  
Glue:  
Missing values as empty strings  
False  
Concatenate (Transform)]
  
```

Explore the Data

The screenshot shows the CoyPu Country and Location dataset in the eccenca platform. The top navigation bar includes a search icon, the URL 'CoyPu Country and Location / Country.list / Germany', a user icon, and a search bar with placeholder 'Go to resource'.

The left sidebar features a navigation tree with icons for Home, User (18), Vocabularies (6), System (1), and a search bar with placeholder 'https://data.coupy.org/country'. Below this is a 'Navigation' section with a search bar and a list of terms: Country (selected), di_Dataset, coy:TerritorialEntityOfSingleCountry, and coy:Country-EU.

The main content area is titled 'Properties' and displays the properties of the 'Germany' resource in Turtle format. The properties listed are:

- rdf:type**: Country (with 'SHOW IN LIST' and 'ADD' buttons)
- owl:sameAs**: geoscheme_Western_Europe, 2_DE, iana_de, num_276, code_DEU (each with 'SHOW IN LIST' and 'ADD' buttons)
- label**: Germany (with 'SHOW IN LIST' and 'ADD' buttons)
- coupu-ref:code-iana**: .de (with 'SHOW IN LIST' and 'ADD' buttons)
- coy:code-3166-1-alpha-2**: DE (with 'SHOW IN LIST' and 'ADD' buttons)
- coy:code-3166-1-num**: 276 (with 'SHOW IN LIST' and 'ADD' buttons)
- coy:authority-code**: DEU (with 'SHOW IN LIST' and 'ADD' buttons)
- coy:code-3166-1-alpha-3**: DEU (with 'SHOW IN LIST' and 'ADD' buttons)
- coy:UNSD-geoscheme**: Western Europe (with 'SHOW IN LIST' and 'ADD' buttons)

A large blue circular button with a '+' sign is located at the bottom right of the main content area.

Explore the Data

The screenshot shows the CoyPU Country and Location dataset in the eccenca RDF Workbench. The interface includes a sidebar with navigation icons, a top bar with a search field and user profile, and three main panels: a left panel for navigation, a middle panel for properties, and a right panel for the Turtle code representation.

Left Panel (Navigation):

- User (18)
- Vocabularies (6)
- System (1)
- Search: https://data.coypu.org/country
- CoyPU Country and Location https://data.coypu.org/country
- Navigation:
 - Country (selected)
 - di_Dataset
 - coy:TerritorialEntityOfSingleCountry
 - coy:Country-EU

Middle Panel (Properties):

Properties: rdf:type, owl:sameAs, label, coy:ref:code-iana, coy:code-3166-1-alpha-2, coy:code-3166-1-num, coy:authority-code, coy:UNSD-geoscheme.

RDFurtle View:

```
1 @prefix coy: <https://schema.coypu.org/global#> .  
2 @prefix coypu-ref: <https://ns.coypu.org/reference/> .  
3 @prefix owl: <http://www.w3.org/2002/07/owl#> .  
4 @prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .  
5 @prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .  
6 @prefix xsd: <http://www.w3.org/2001/XMLSchema#> .  
7  
8 <https://data.coypu.org/country/DEU>  
9   rdf:type          coy:Country ;  
10  rdfs:label        "Germany" ;  
11  owl:sameAs       <https://data.coypu.org/country/code-3166-1-num/276> , <https://data.coypu.org/country/alpha-2> ;  
12  coy:ref:code-iana ".de" ;  
13  coy:UNSD-geoscheme "Western Europe" ;  
14  coy:authority-code "DEU" ;  
15  coy:code-3166-1-alpha-2 "DE" ;
```

Right Panel (Turtle):

Properties: User (18), Vocabularies (6), System (1). Search: https://data.coypu.org/country. CoyPU Country and Location https://data.coypu.org/country.

Navigation:

- Country (selected)
- di_Dataset
- coy:TerritorialEntityOfSingleCountry
- coy:Country-EU

Properties: Properties, Turtle.

Turtle code (partially visible):

```
1 @prefix coy: <https://schema.coypu.org/global#> .  
2 @prefix coypu-ref: <https://ns.coypu.org/reference/> .  
3 @prefix owl: <http://www.w3.org/2002/07/owl#> .  
4 @prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .  
5 @prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .  
6 @prefix xsd: <http://www.w3.org/2001/XMLSchema#> .  
7  
8 <https://data.coypu.org/country/DEU>  
9   rdf:type          coy:Country ;  
10  rdfs:label        "Germany" ;  
11  owl:sameAs       <https://data.coypu.org/country/code-3166-1-num/276> , <https://data.coypu.org/country/alpha-2> ;  
12  coy:ref:code-iana ".de" ;  
13  coy:UNSD-geoscheme "Western Europe" ;  
14  coy:authority-code "DEU" ;  
15  coy:code-3166-1-alpha-2 "DE" ;
```

Linking

Linking editor [Linking editor](#) [Linking evaluation](#) [Linking execution](#) [Reference links](#) [Learning](#) [↗](#)

[↶](#) [↷](#) | [☰](#) [>Show evaluation](#)

[Caches: 4 hours / 4 hours](#) | [▶](#) | [Save](#)

All [☰](#) [Search for operators and patterns](#)

Source path [ⓘ](#)
[Input](#) [Recommended](#)

Target path [ⓘ](#)
[Input](#) [Recommended](#)

Lower case [ⓘ](#)
[Transform](#) [Normalize](#) [Recomm...](#)

Tokenize [ⓘ](#)
[Transform](#) [Tokenization](#) [Recom...](#)

Constant [ⓘ](#)
[Transform](#) [Value](#) [Recomm...](#)

Default Value [ⓘ](#)
[Transform](#) [Value](#) [Recomm...](#)

Levenshtein distance [ⓘ](#)
[Comparison](#) [Characterbased](#) [Re...](#)

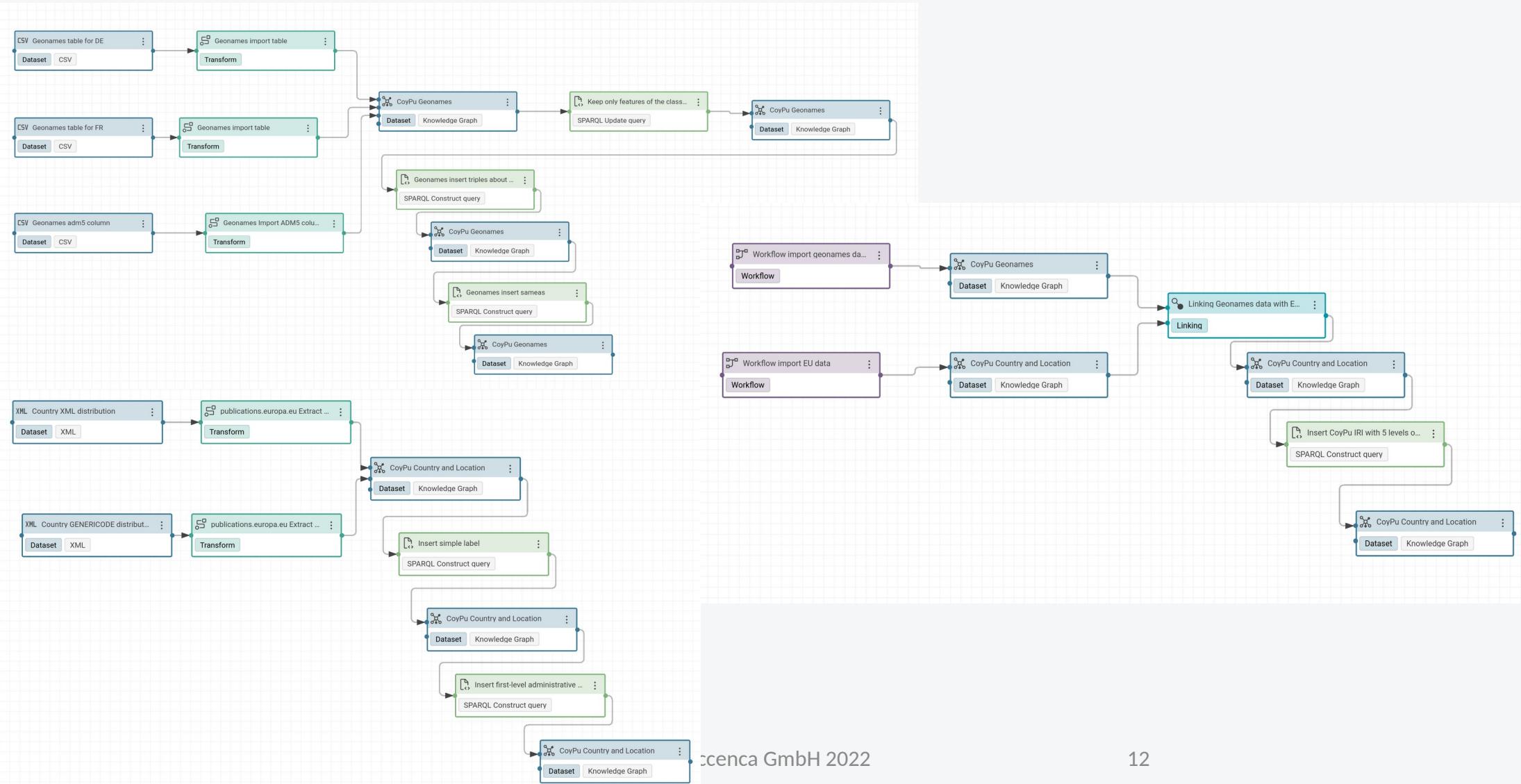
String equality [ⓘ](#)
[Comparison](#) [Equality](#) [Recomm...](#)

Jaccard [ⓘ](#)
[Comparison](#) [Tokenbased](#) [Reco...](#)

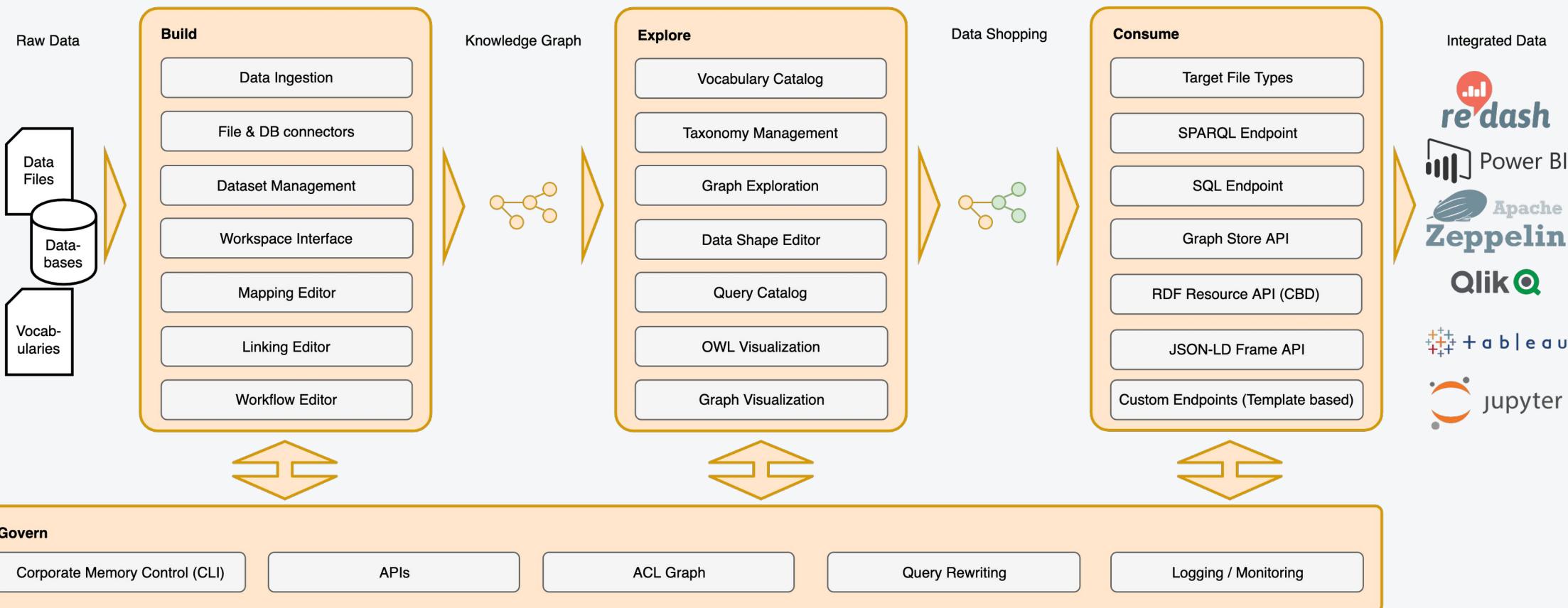
Numeric equality [ⓘ](#)

The diagram illustrates a linking configuration with two parallel paths. The top path starts with a 'Source path' node (Path: 'gn:alternateName') which connects to a 'String equality' node (Threshold: 0.0, Weight: 1). This node then connects to an 'And' node (Weight: 1). The bottom path starts with a 'Source path' node (Path: 'gn:countryCode') which connects to a 'String equality' node (Threshold: 0.0, Weight: 1). This node also connects to the same 'And' node. Additionally, the bottom path includes a 'Regex extract' node (Regex: '(.*)-.*', Extract all: off) which connects to the 'String equality' node. A 'Target path' node (Path: 'label') connects to both 'String equality' nodes. The 'And' node has a summary visualization on its right side.

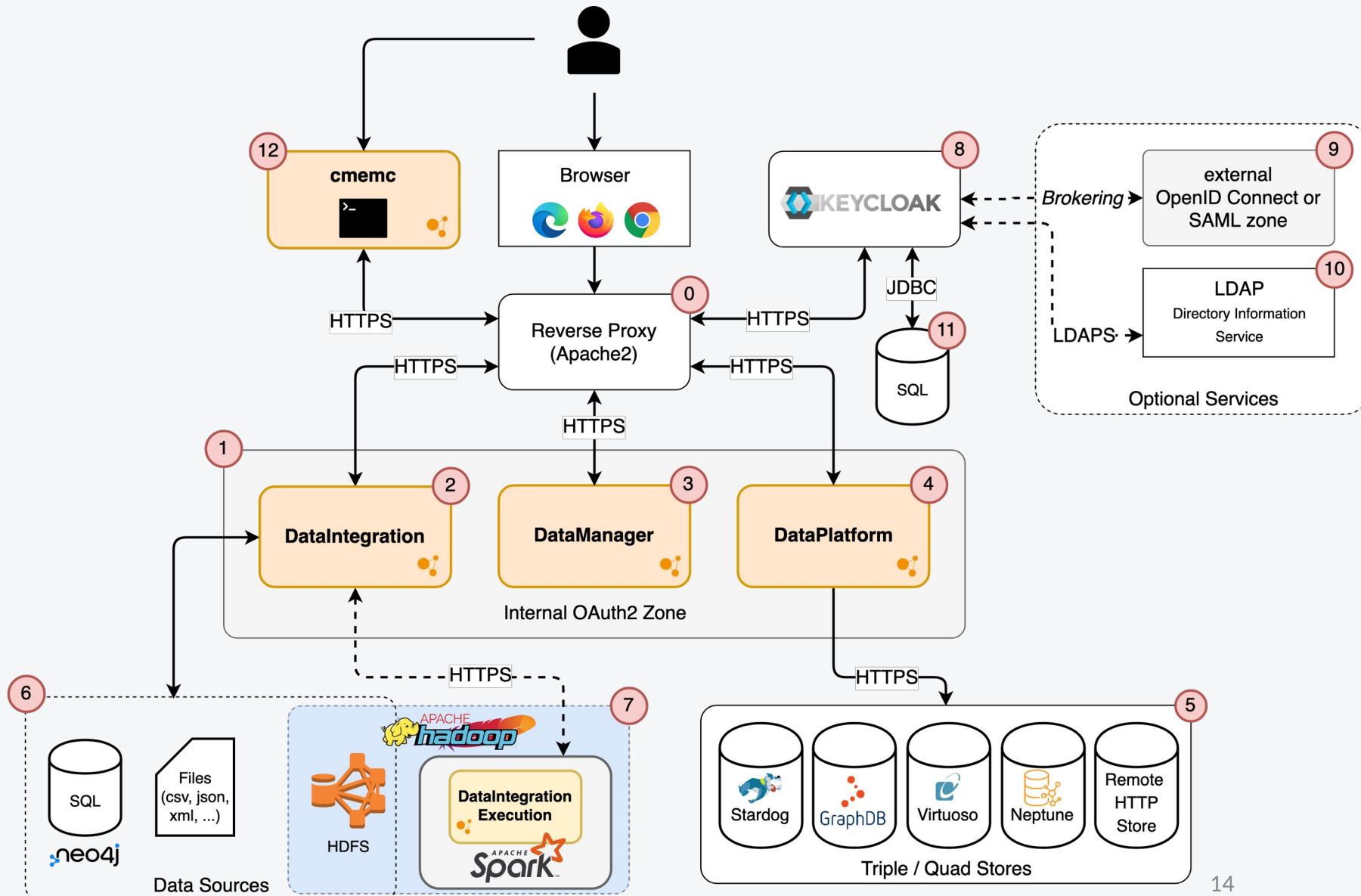
The Mapping Workflow



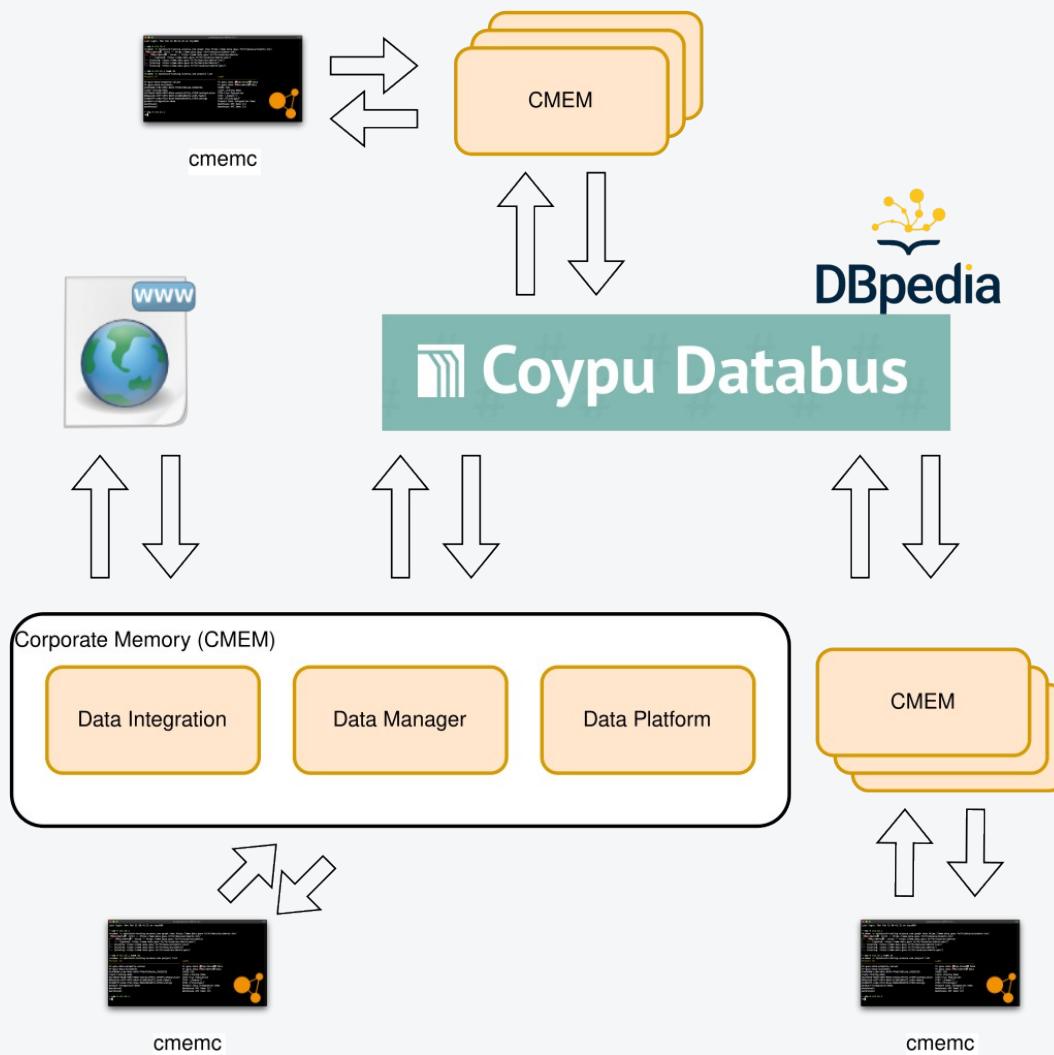
User Journey and Functional Areas



Reference Architecture Corporate Memory



Integration with the Databus

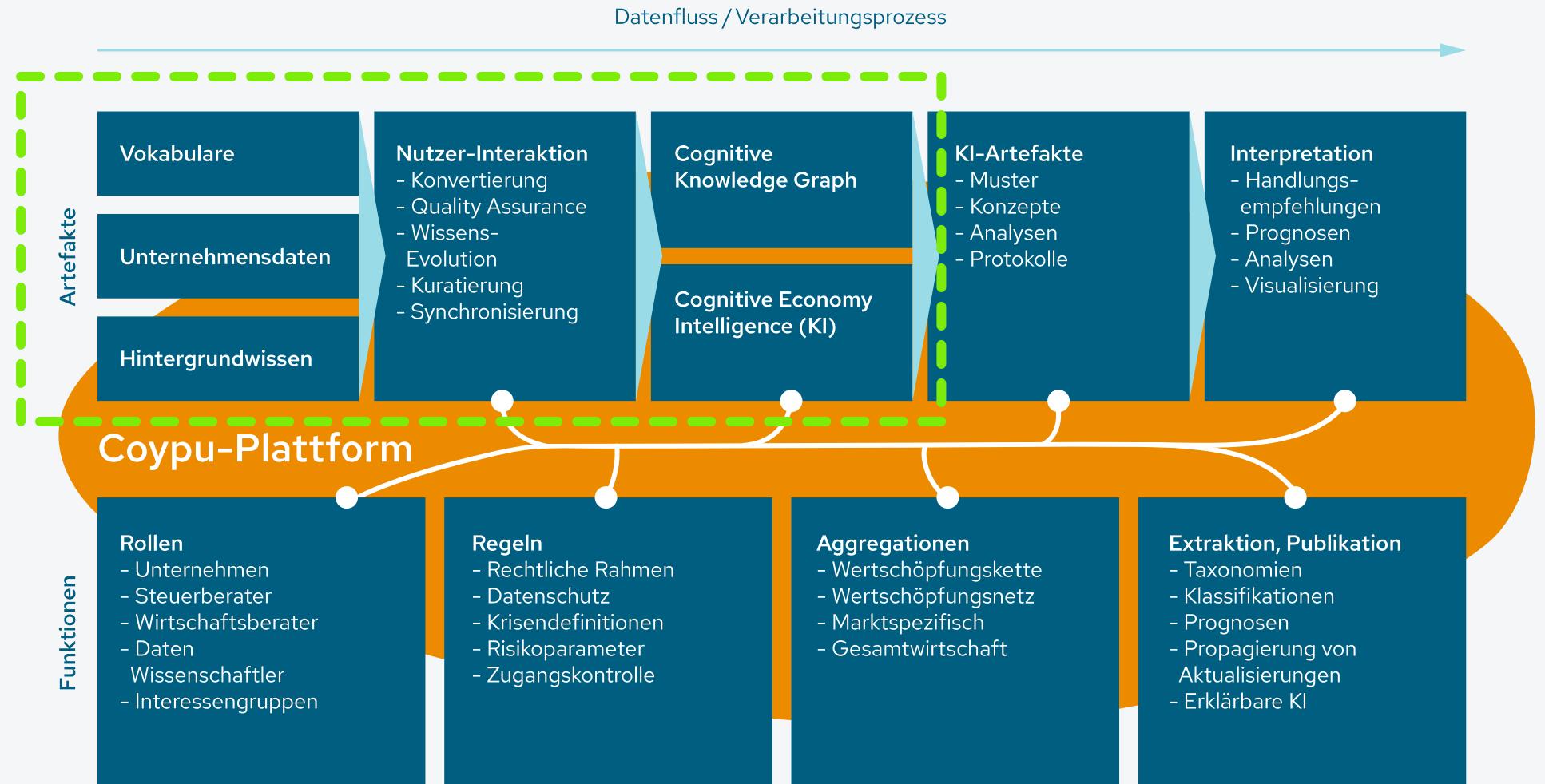


Screenshot of the Coypu Databus interface:

- RECENT ACTIVITY:** A line chart showing 'Uploaded Data (Gbyte)' over time from September 2022 to June 2023. The data shows a significant spike starting in May 2023, reaching approximately 0.0001 Gbyte.
- SEARCH:** A search bar with placeholder text 'Search the Databus...' and icons for file, book, user, and cube.
- NEW ARTIFACTS:**
 - A card for 'Implisense Company Data Dump' by 'nsteinert' (ips) with the sub-dump 'ips-dump'. It notes a 'Test dump of full Implisense company data' from '2022-06-10T13:16:45.257Z'.
- MOST ACTIVE USERS:** A table showing user activity:

User	Uploads	Derived Data
nsteinert	1	107 KB
- DBpedia:** A footer section with the text 'Global and Unified Access to Knowledge Graphs' and '© Copyright 2019 by DBpedia. All Rights Reserved.'

The Coypu-Plattform





eccenca GmbH
Hainstraße 8
D-04109 Leipzig
Germany

+49 341 2650 8028
info@eccenca.com
<https://eccenca.com>



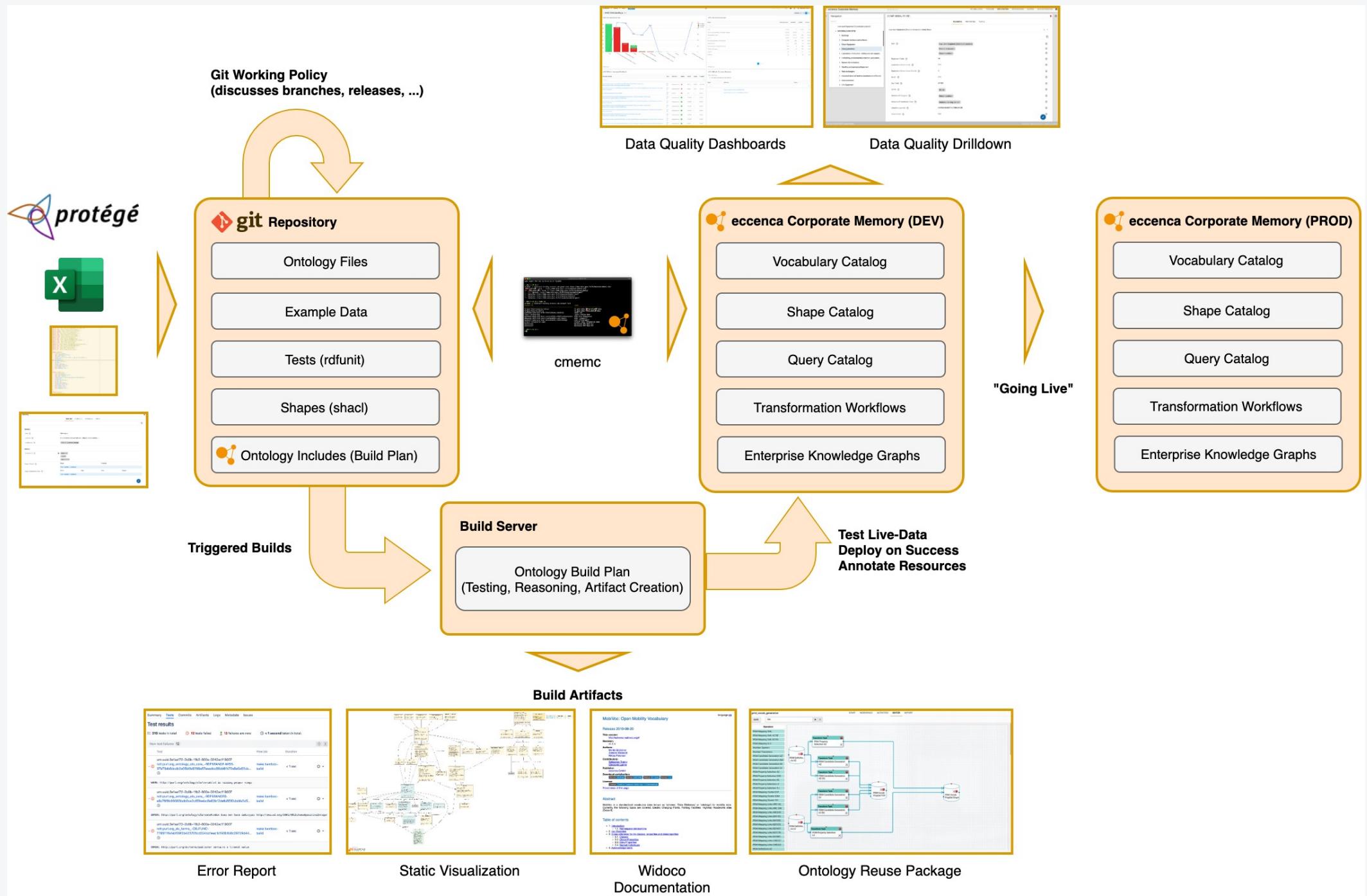
Hans-Chr. Brockmann
Geschäftsführer

Hainstraße 8
D-04109 Leipzig
Germany

+49 511 3365 2810
+49 173 3698 610

brockmann@eccenca.com
<https://eccenca.com>

BACKUP



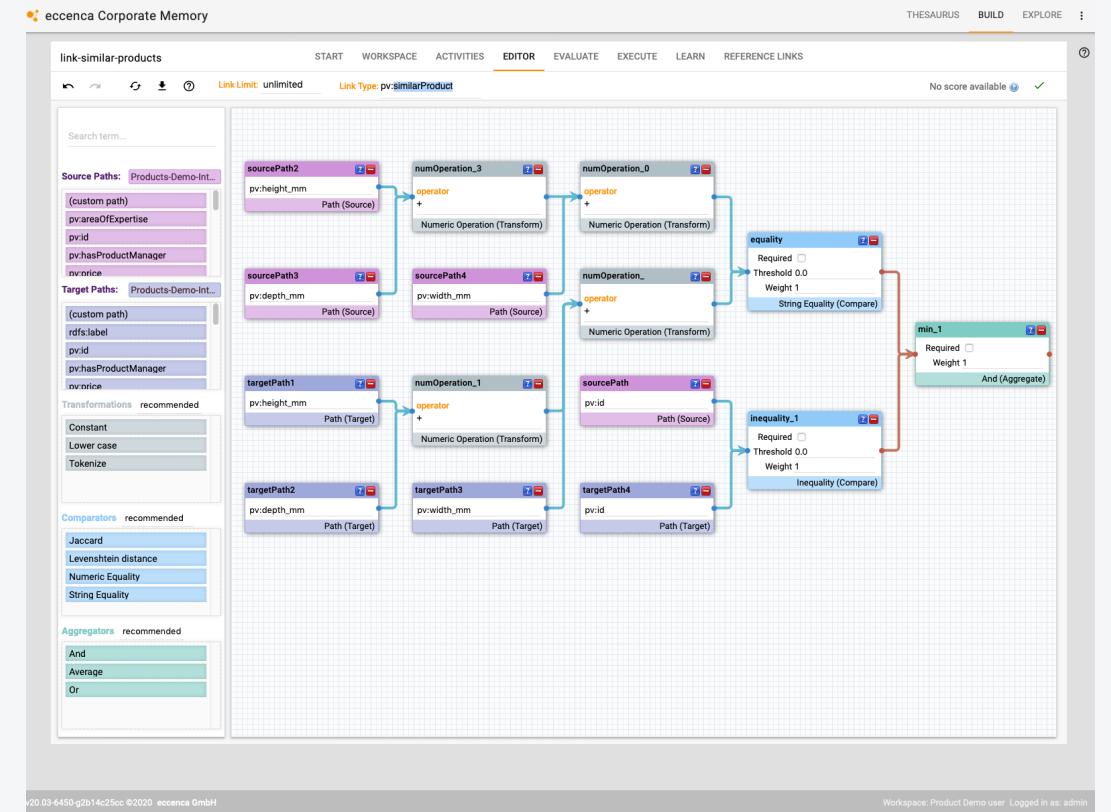
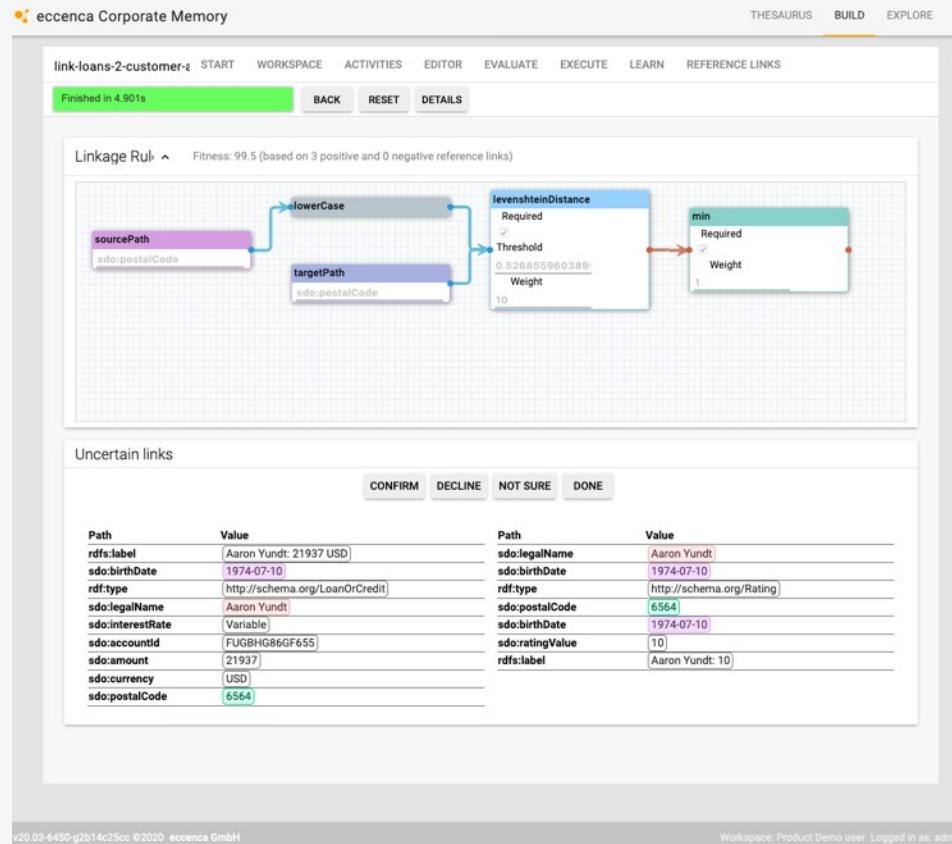
Data defined UI Configuration

- Framework for interactive data editors that uses W3C SHACL
- Supporting trees of graphs to partition data
- Application UI can be flexibly configured, e.g. including navigation
- Multiple such application configurations can be provided, to support multiple perspectives on the same knowledge graph
- Custom functionality by event driven query hooks

<https://documentation.eccenca.com/latest/explore/building-a-customized-user-interface>

The top screenshot shows the 'eccenca Corporate Memory' interface. On the left, a navigation tree includes categories like Agent, Employee, Manager, Product, Hardware, Service, Department, Price, Dataset, and Product Category. On the right, a detailed view of a 'Meter Transducer' resource is shown, with fields for ID (Z249-1364492), Name (Meter Transducer), Price (amount 4.12, currency EUR), and category (Transducer). The bottom screenshot shows the 'Documentation' page for 'Building a customized User Interface'. It includes a sidebar with links like Back, Explore, Dataset Catalog, Vocabulary Catalog, Taxonomy Management, Graph Exploration, Query Module, and Building a customized User Interface. The main content area provides an introduction to using shapes for creating a customized Linked Data user interface, defining forms using SHACL rules, and creating new resources.

Linking as general rules mechanism

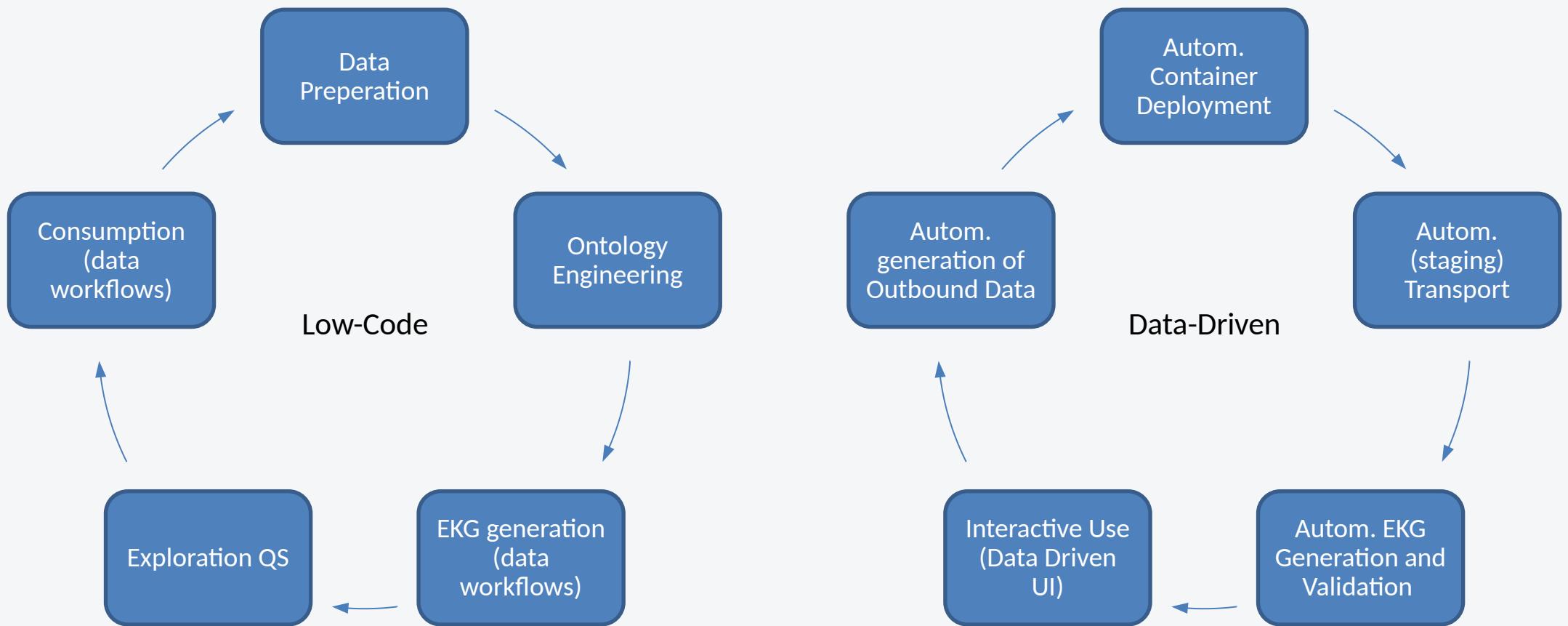


DevOps based Ontology Engineering Pipeline

- We apply code and testing principles to RDF datasets, esp. ontologies Engineering
- We use git to manage revisions of our ontologies and bamboo as the DevOps tool
- Any Ontology tool / IDE possible that consumes/produces RDF
- Automatic test generation and quality checks using builds with each commit
- cmemc is used to automate the workflow between ontology engineering, corporate memory development and instance provisioning
-

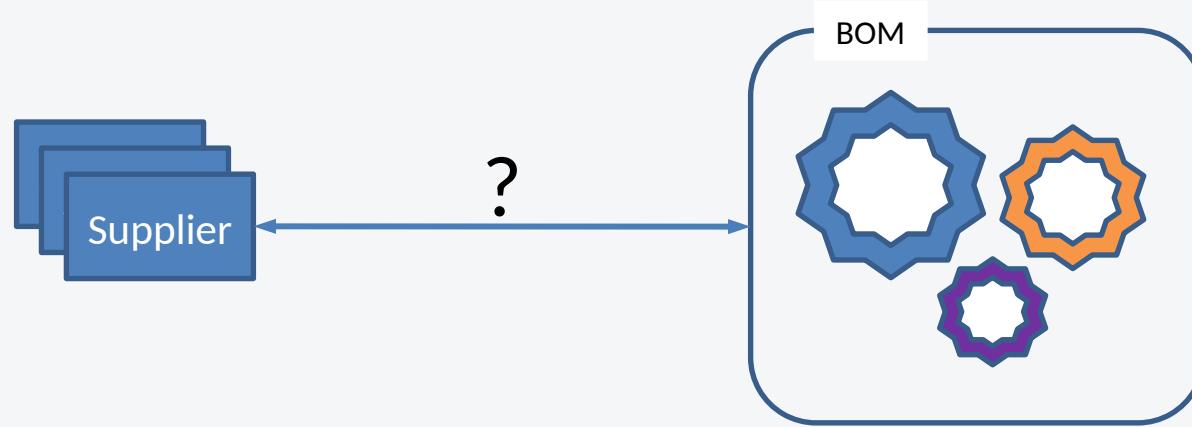
<https://bamboo.eccenca.com/browse/SCHEMA-ECCDSM-99/test>

eccenca Build- and Run-Cycles



SAMPLE USE CASES

Apply Machine Learning to perform “Form Fit Functional” Material Linking



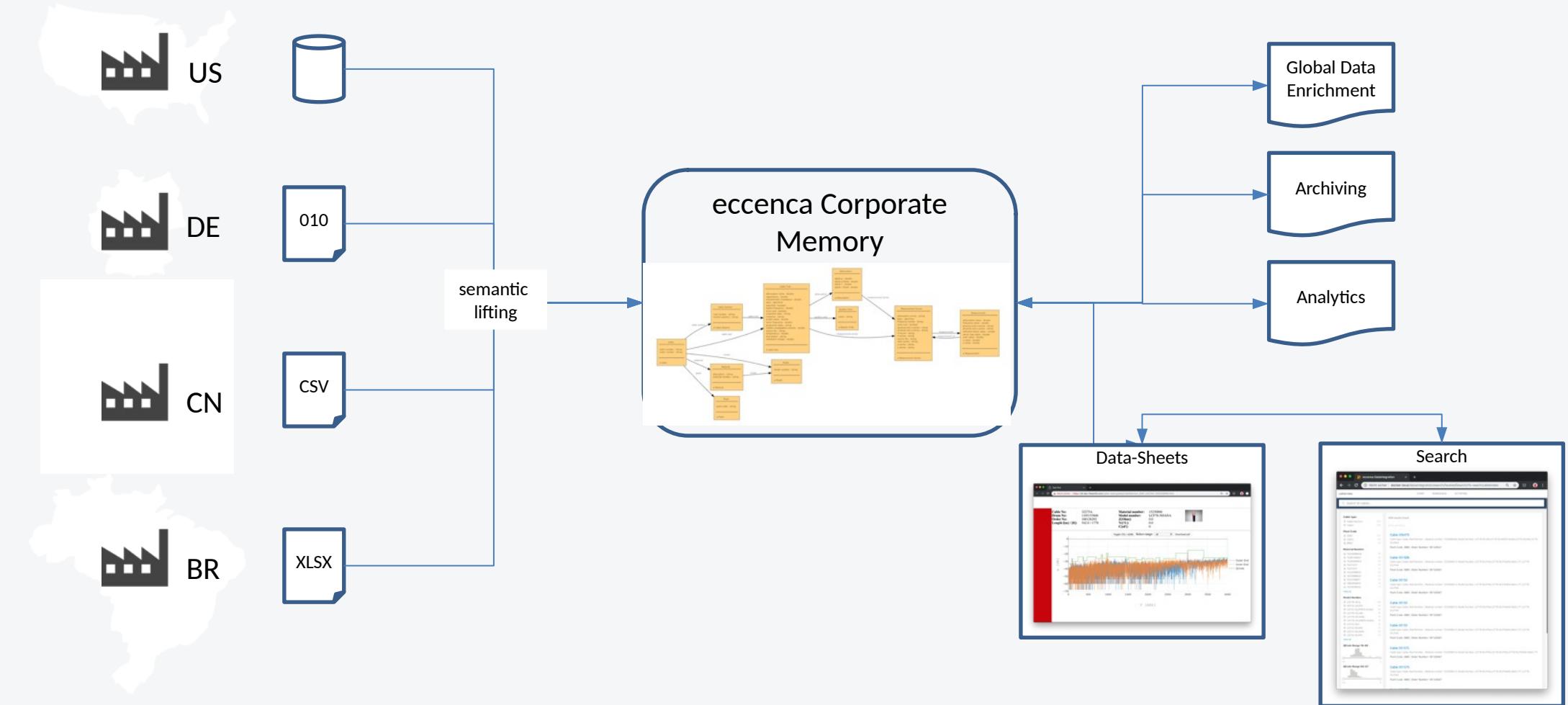
Today based on part number matching and man-made part lists:

MPN	Supplier
293D105X9016A2##E3	VISHAY
B45196E3105K10	KEMET
T491A105K016AS	KEMET
T491A105K016AT	KEMET
TAJA105K016RNJ	AVX

Machine learning based on granular material properties like:

Capacitor Type	TANTALUM CAPACITOR
Capacitance	1.0 µF
Dielectric Material	TANTALUM (DRY/SOLID)
Mounting Feature	SURFACE MOUNT
Neg. Tolerance	10.0 %
Op. Temp.-Min	-55.0 Cel
Op. Temp.-Max	85.0 Cel
Package Shape	RECTANGULAR PACKAGE
Package Style	SMT
...	

Cross Site Semantic Data Harmonization



Project Benefits

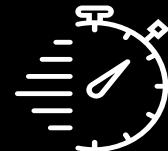
- 12% Inventory reduction contribution within 3 months
- 200% Project ROI Data-as-Service creates automation and productivity gains
- 70% of effort analyzing data
Instead of finding, aligning, integrating and cleaning data

BENEFIT Cases



Global S&OP
Capacity Planning

- Less global inventory
- Better factory balancing



Customer Service Teams

- Improvement lead-times
- Automation



Customer Experience

- Self-service, on-demand production data

About eccenca GmbH

brox IT-Solutions GmbH

Gegründet: 1998
Fokus: IT-Consulting
IPR: Initiierung/Leitung eclipse.org/SMILA
eccenca Enterprise Search
Key Accounts: Volkswagen, Audi, Skoda, MAN, Telekom,
Daimler, Bosch, Siemens, Continental



eccenca GmbH

Gründung: 2013
Fokus: Produkte/Lösungen
Team Size: 35
IPR: Linked Data, M2N Synchronization,
Linking, Authentication/Data Security
Kunden: Volkswagen, Bosch, Nokia, Infineon,
Ericsson, Telekom, Daimler
F&E Projekte: LUCID, ELDS, GeoKnow, Diachron
Initiativen: MOBIVOC, OSFP



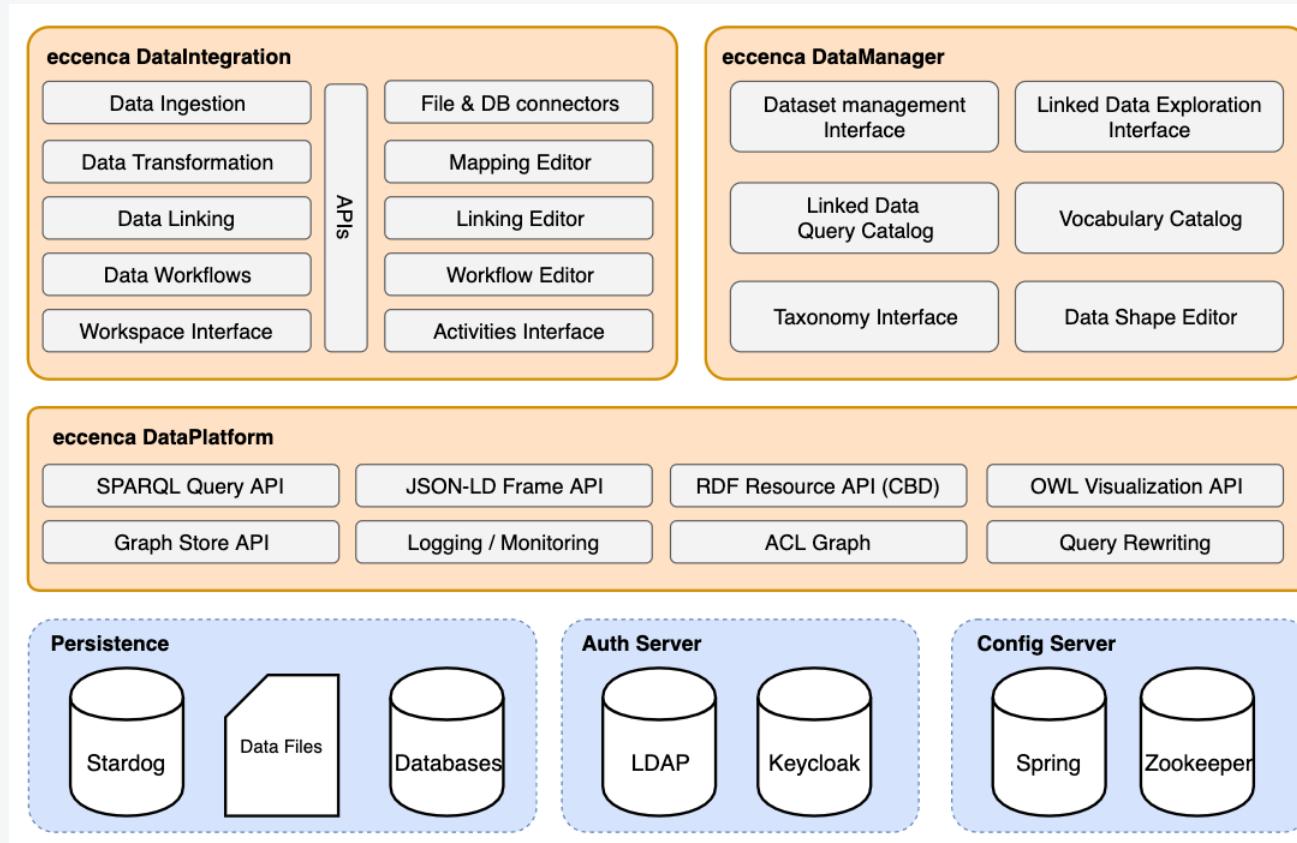
AKSW – Universität Leipzig & Fraunhofer IAIS

Führende Linked Data Forschungseinheit in
Europa. Initiator des nationalen Industrial Data Space

Initiator: DBpedia, Linked GeoData etc.
Betreiber: Datenportal der EU-Kommission
Team-Leitung: Prof. Dr. Sören Auer



Technical Architecture



eccenca DataManager – Features

- Management of Knowledge Bases (Named Graphs, Linked Data access optional)
- Tree, list and resource views
- Versioning (triple based)
- User management and access control
- Query Catalog
- Dataset Schema Browser
- Inline authoring
- Detailed edit view
- Add new resources and properties
- Search
- Facet based filtering
- Complex navigation hierarchies (spatial, class based, organization structure based)
-

eccenca DataIntegration – Features

- UI allows to view and edit linkage rules
- Linkage rules are shown as a tree
- Editing using drag & drop
- DataIntegration provides a high level data manipulation and linking engine
 - Execution of linkage and integration rules on arbitrary datasets provided by eccenca DataPlatform
 - Manual creation of rules with an integrated editor
 - Automatic learning of rules based on training data (positive / negative lists)
-

eccenca DataPlatform – Features

- DataPlatform is a semantic middleware which provides a unified access to structured data
 - W3C standards such as RDF / Linked Data and SPARQL
 - Authorization based on an RDF Access Control Model
 - Authentication via OAuth2 protocol
 - Integration with external User Management Systems (e.g. LDAP, Active Directory)
 - Versioning Support (tracking of changes on triples and graphs)
 - Integration of non-RDF sources via mapping and query rewriting technologies (RDB2RDF component)

Semantic Data Management

Data in Context

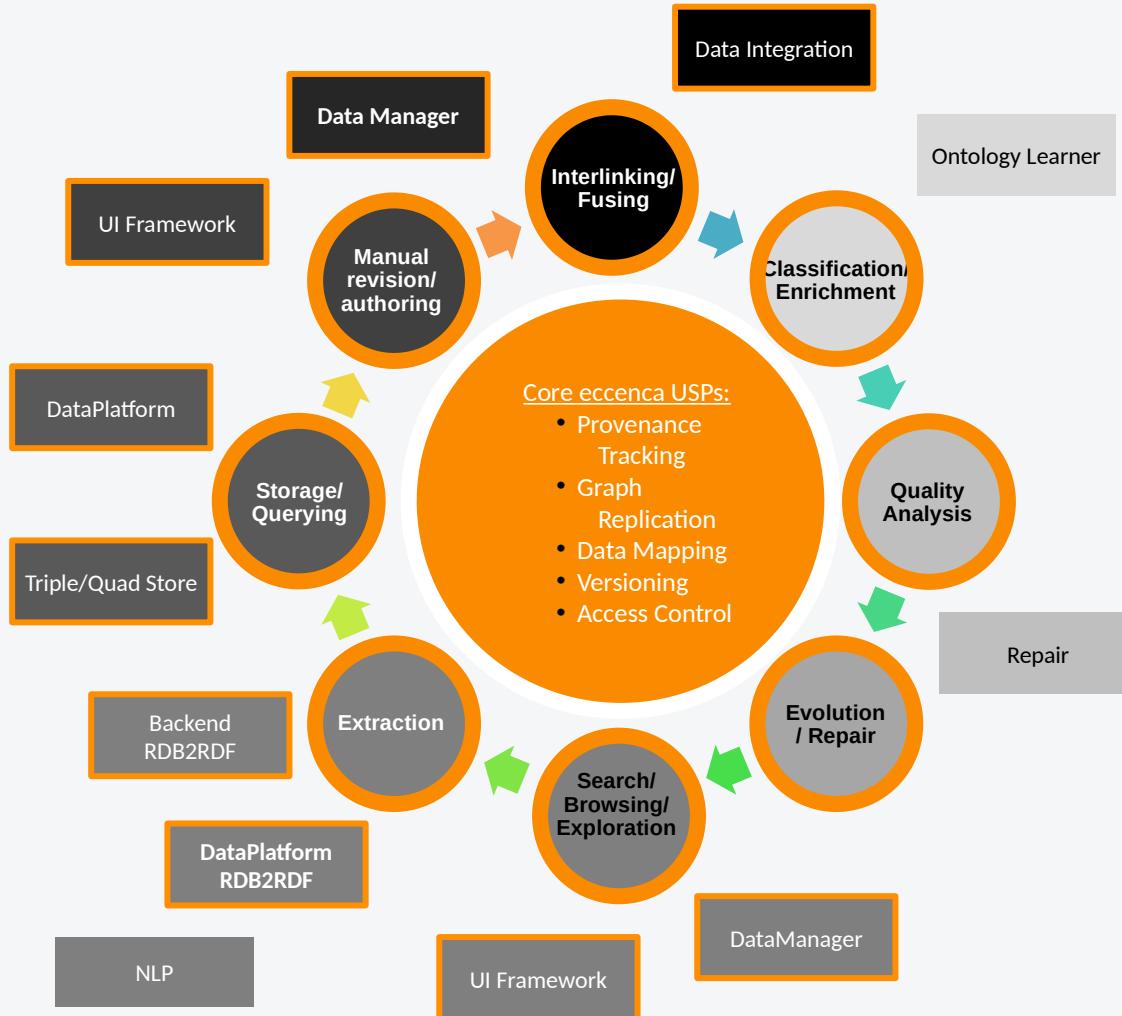
What

- § Create knowledge graphs by connecting datasets and metadata to logical models
- § Physical data models unchanged!
- §
- Ø Explore metadata & structures
- Ø Query & Access data via models
- Ø Integrate data on model level
- Ø Share data on model level

How

- § Leverage linked data principles
 - § Schema as data (RDF)
 - § Global identifiers (URIs)
 - § Linked data graphs (W3C)
- §
- § Catalog your data assets: datasets, vocabularies (models), ...
- § Publish-subscribe for sharing
- § Machine learning for integration

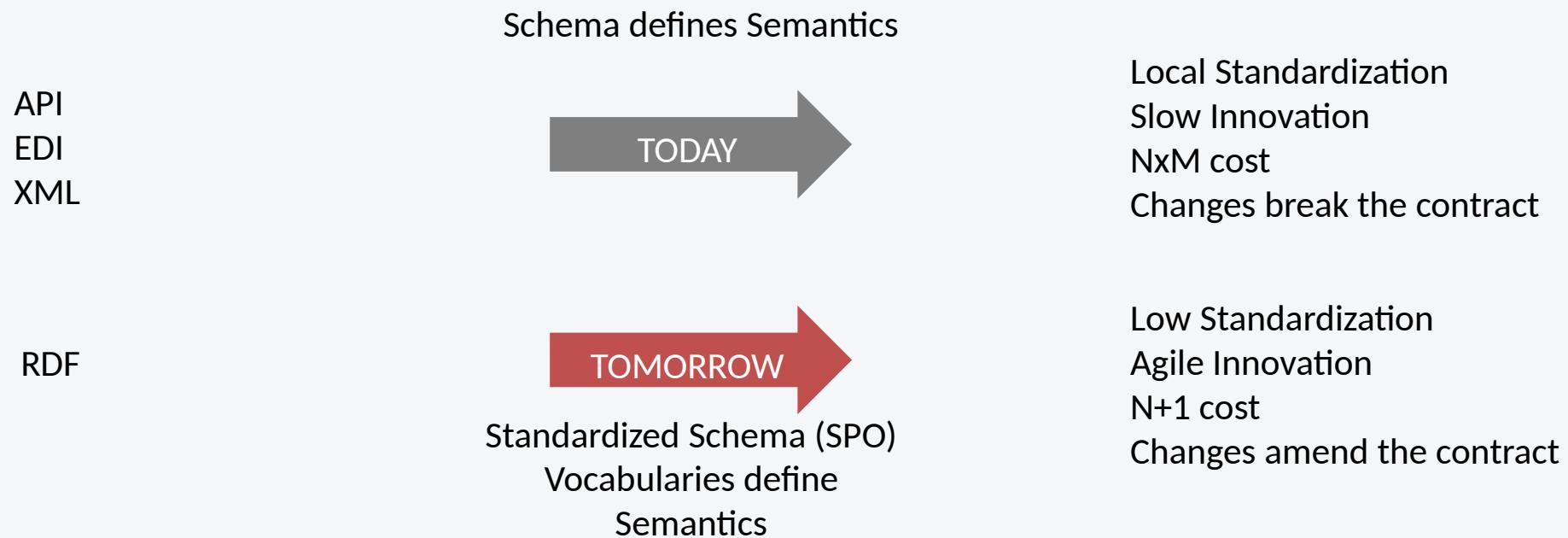
Linked Data Life Cycle



- Extraction / Mapping
- Storage / Querying
- Manual Revision / Authoring
- Linking / Fusion
- Classification / Enrichment
- Quality / Evolution
- Search / Browse / Explore

Changing the Data-Collaboration Paradigm

... by turning **STRINGS** into **THINGS**



Other Use Cases

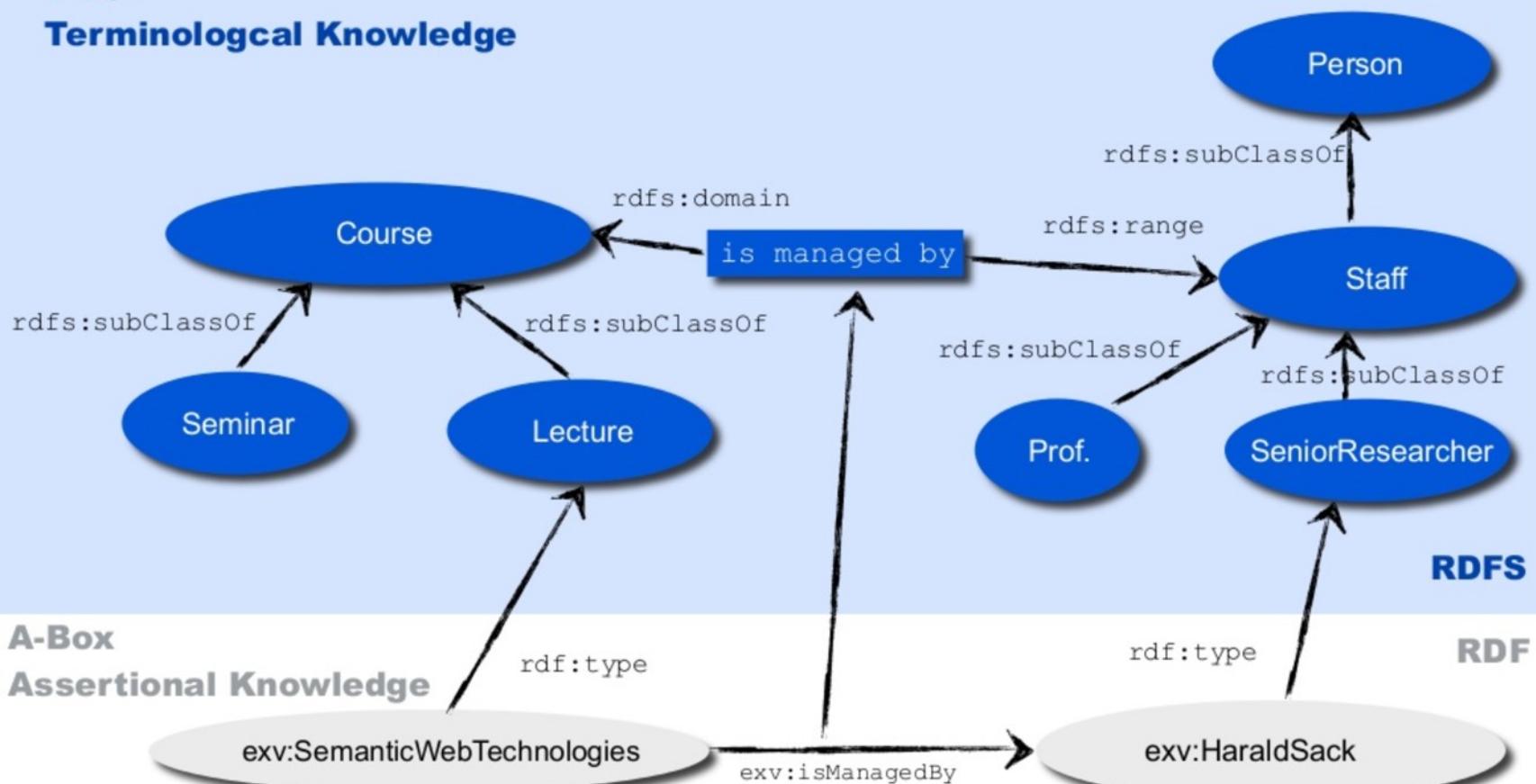
- Conceptual / semantic data model based data preparation for BI and analytics
- Yield management / lead time based dynamic pricing
- Data integration and central data hub for the Software Monetarization Platform project
- Semantic Enterprise Information Model
 - MDM, simplification, data preparation, analytics, etc.

RDF +RDFS

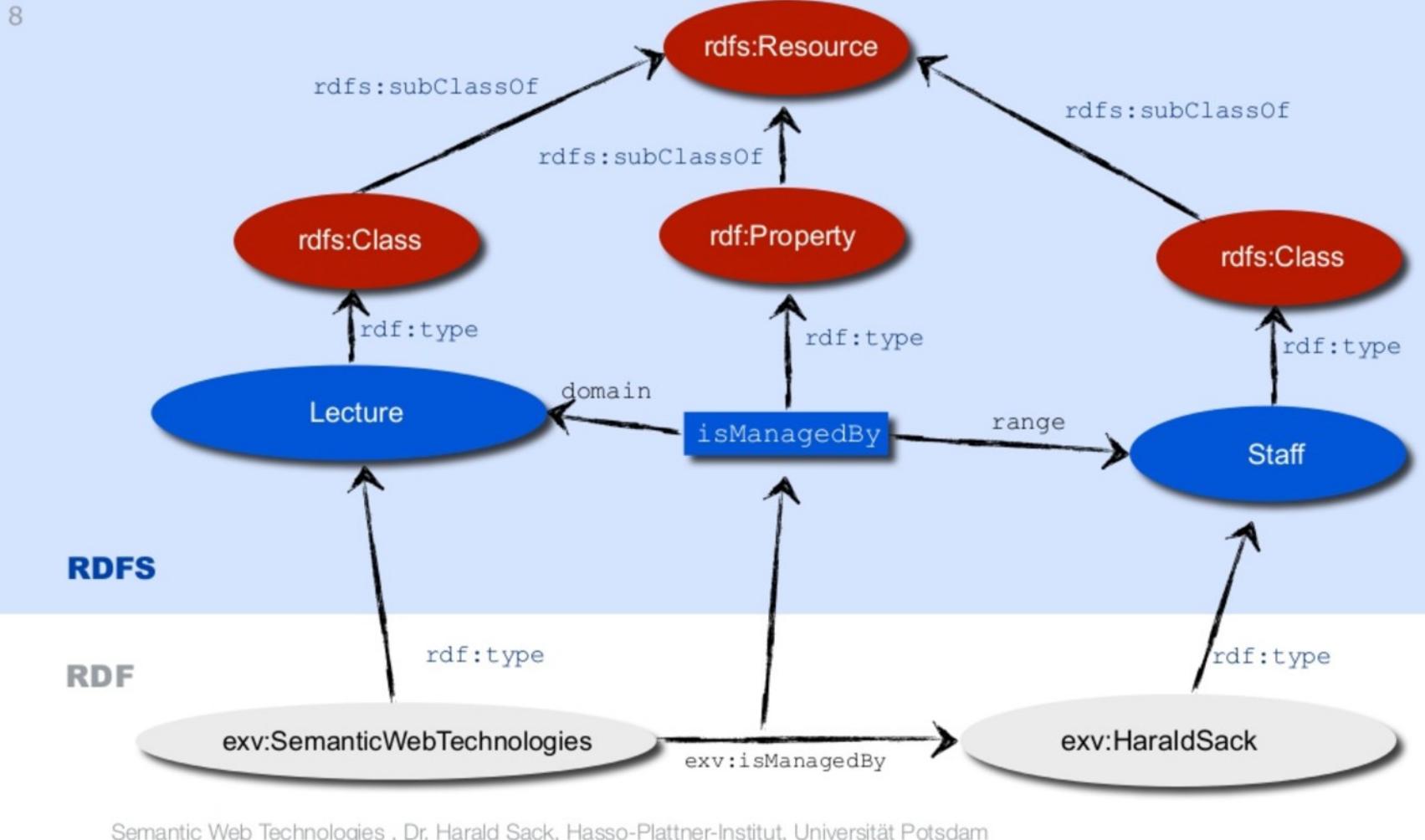
13

T-Box

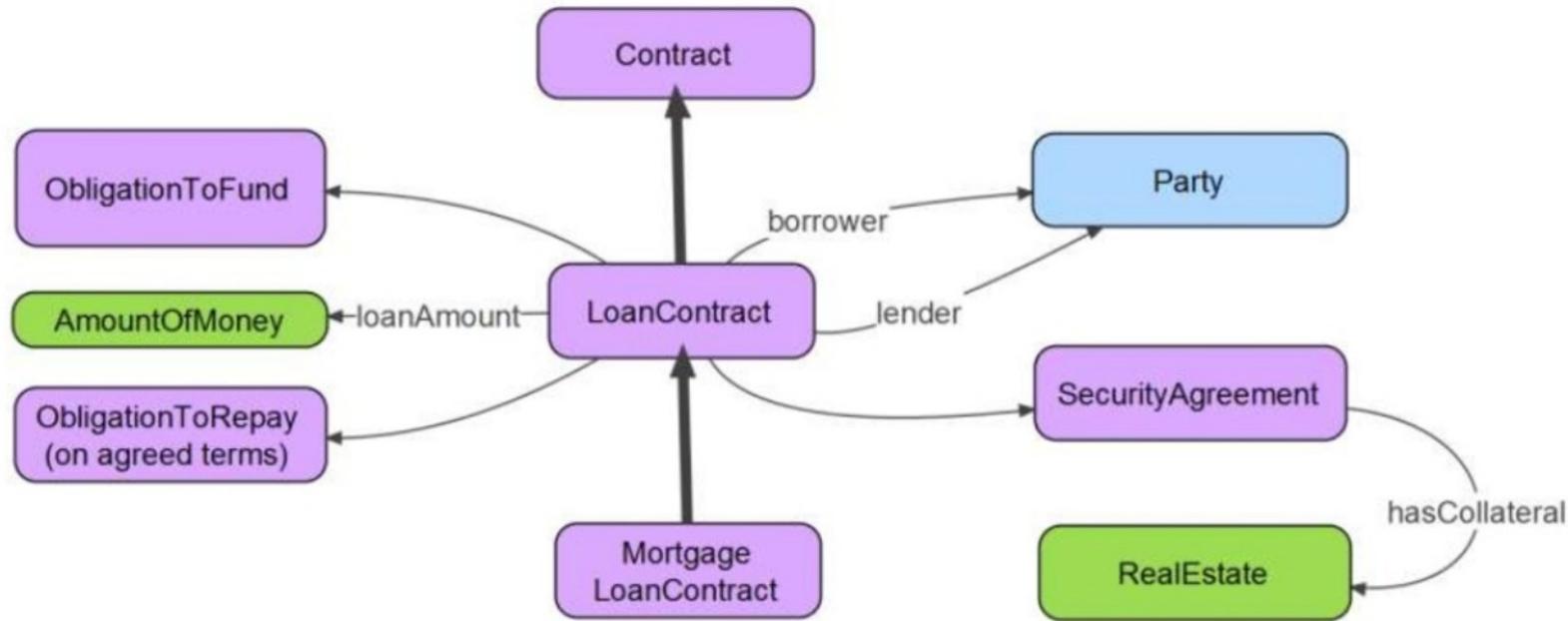
Terminological Knowledge



Classes, Properties, Instances



OWL: Inferencing classification.



Mortgage: A **LoanContract** that has a **SecurityAgreement** where the collateral is **RealEstate**.
Can infer into this class.

SHACL: Checking Graph Patterns

Constraints on values with another shape

Constraint	Description
node*	All values of a given property must have a given shape Recursion is not allowed in current SHACL

```
:User a sh:NodeShape, rdfs:Class ;  
  sh:property [  
    sh:path schema:worksFor ;  
    sh:node :Company ;  
  ] .  
  
:Company a sh:Shape ;  
  sh:property [  
    sh:path      schema:name ;  
    sh:datatype xsd:string ;  
  ] .
```

```
:alice a :User;  
  schema:worksFor :OurCompany .  
  
:bob   a :User;  
  schema:worksFor :Another .  😞  
  
:OurCompany  
  schema:name "OurCompany" .  
  
:Another  
  schema:name 23 .
```