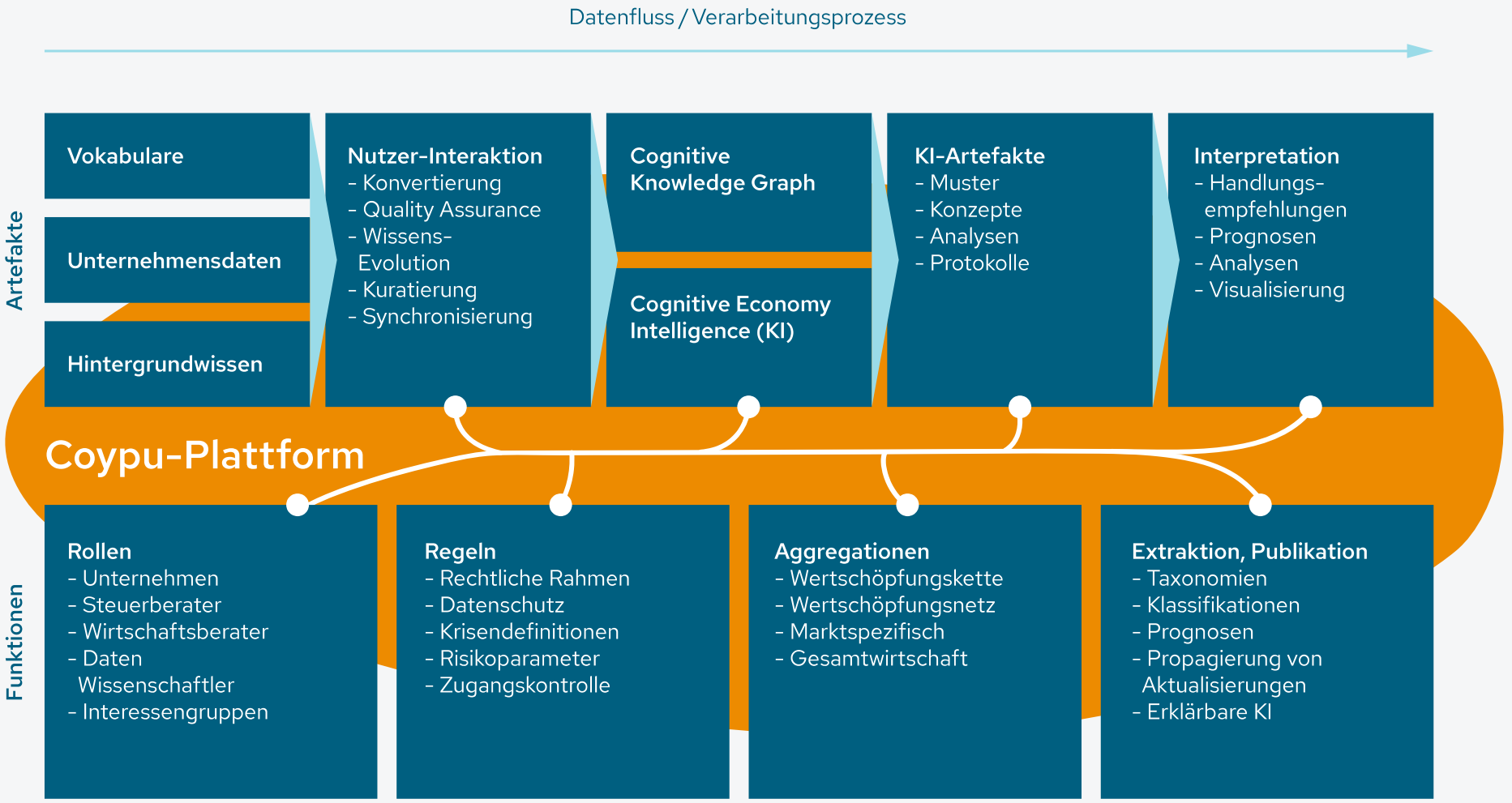
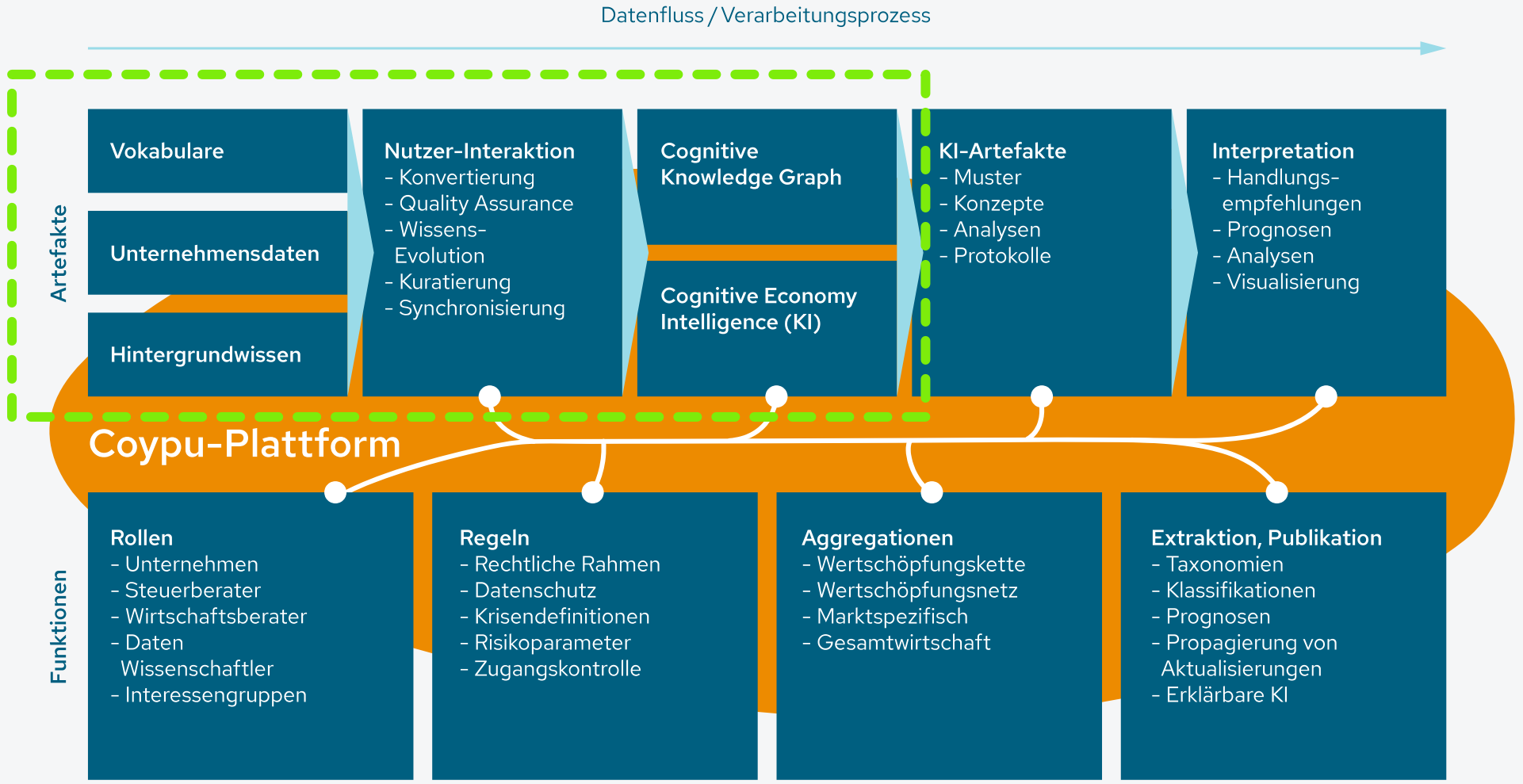
The Coypu Platform
Data Integration Workflow

Dr. Natanael Arndt
Senior Linked Data Expert @ eccenca GmbH

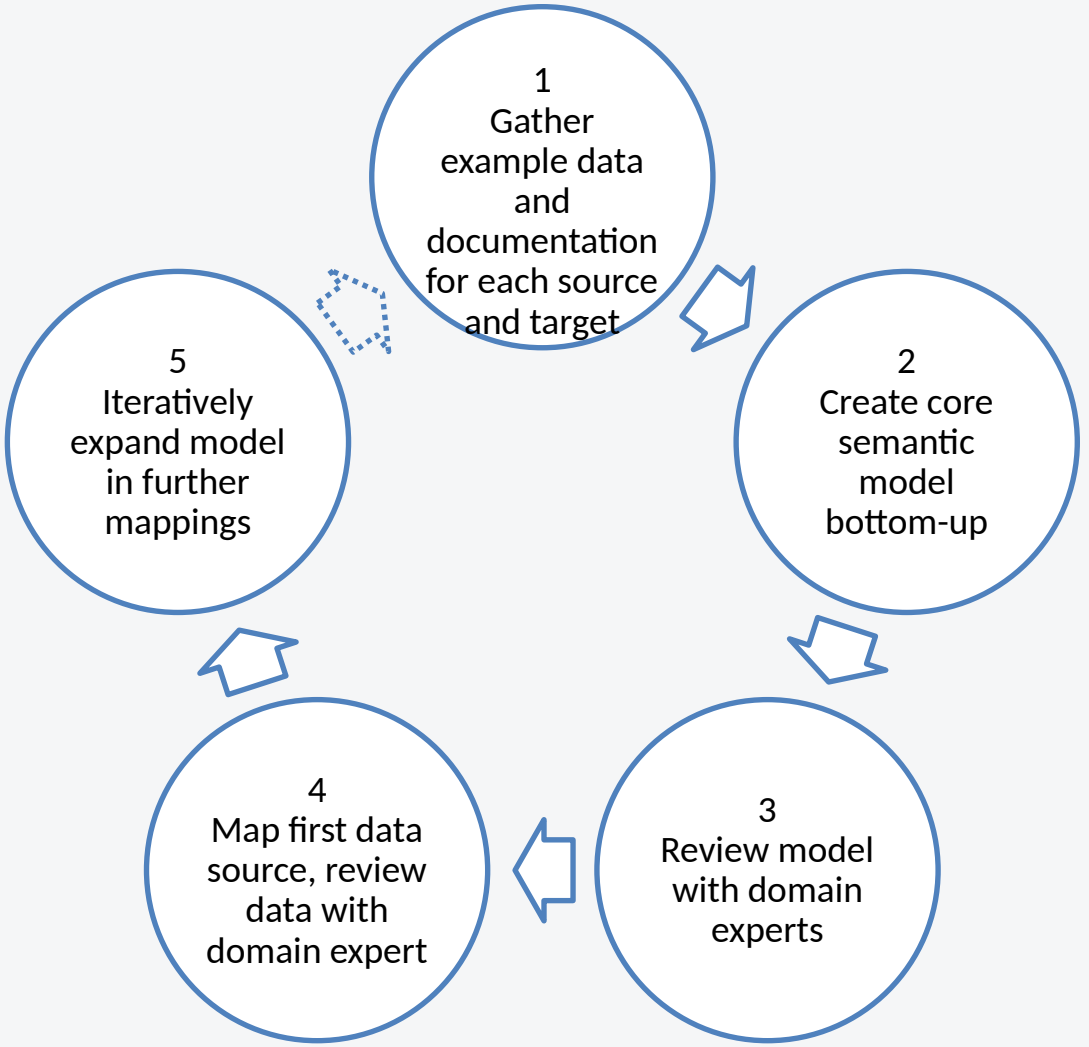
The Coypu-Plattform



The Coypu-Plattform



Enterprise Knowledge Graph Project Flow

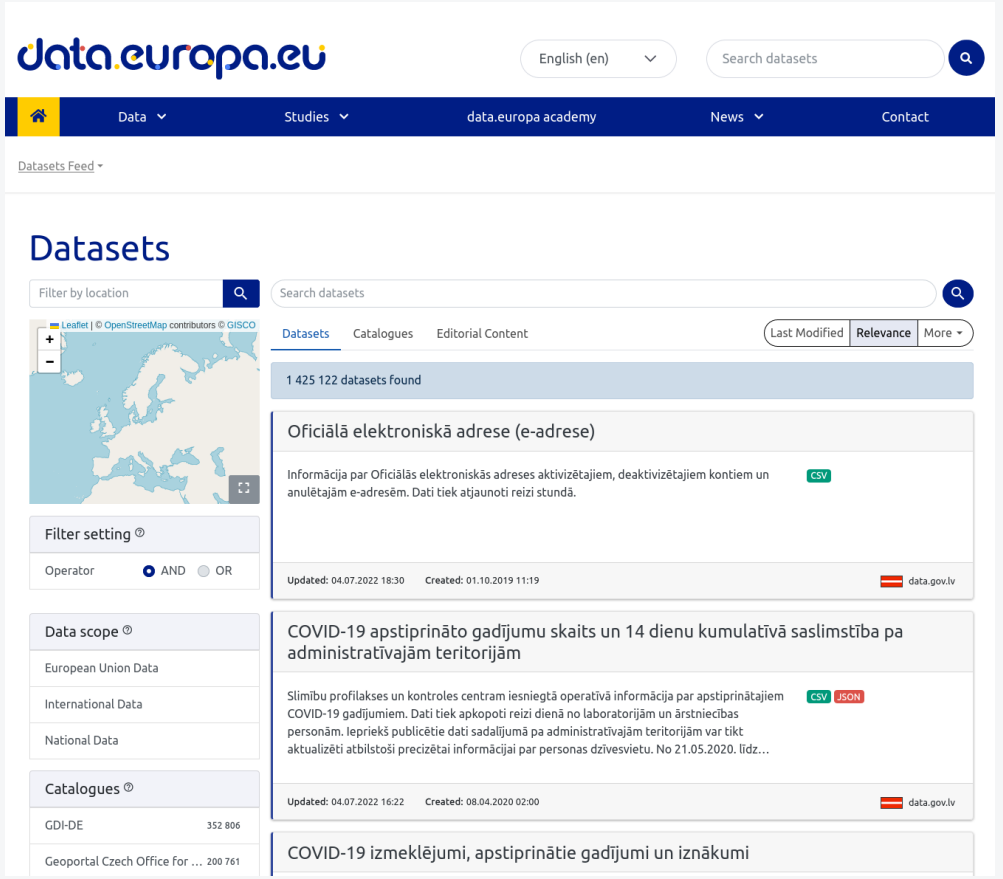


1. Analyze data sources and understand domain
2. Find or build a semantic model (ontology/vocabulary)
3. Create the mappings
4. Clean, enrich and link data
5. Explore the results

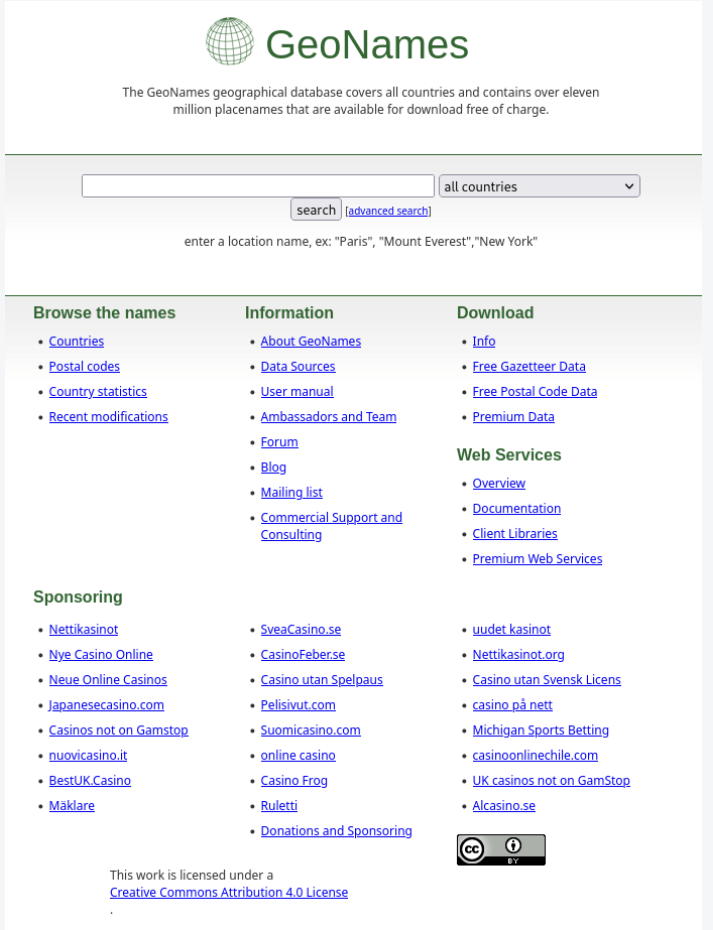
... Re-iterate if needed ...

© eccenca GmbH 2022

Data Sources

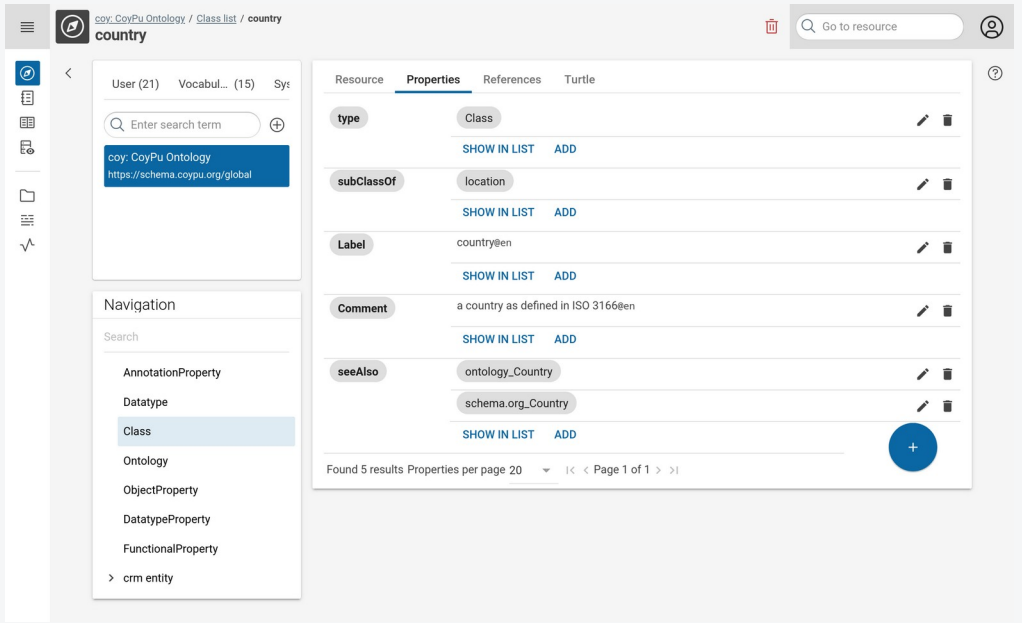


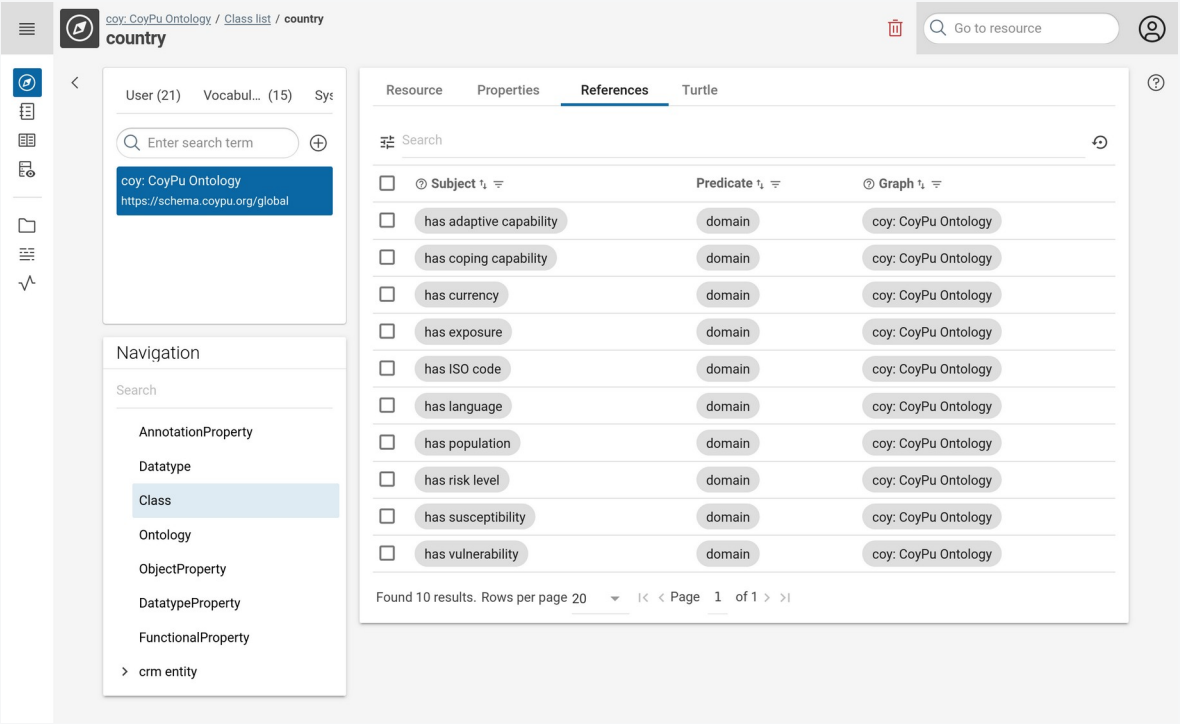
Authoritative Country Dataset from the EU
XML Dataset



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

The Vocabulary

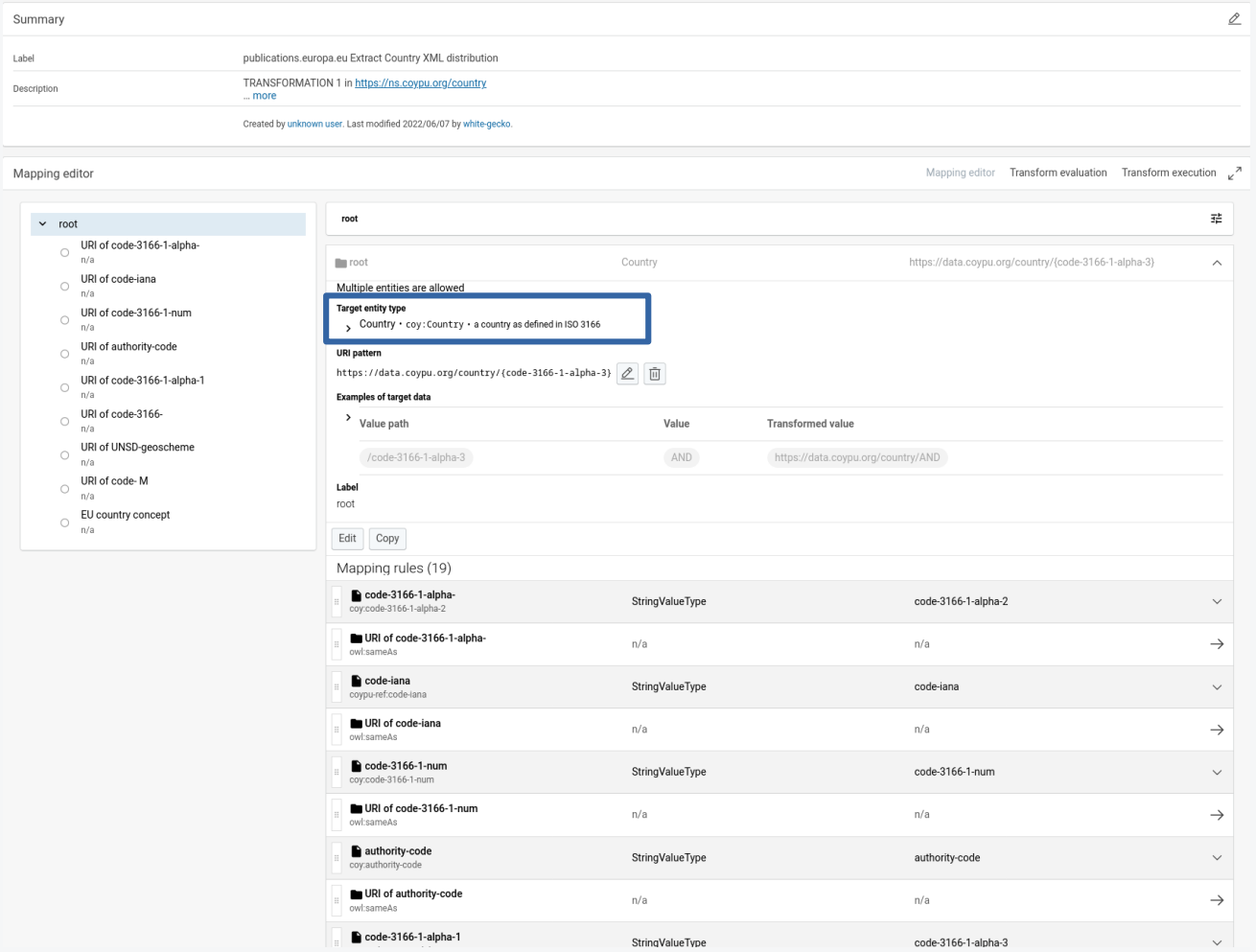




TODO: publish under https://schema.coypu.org/global

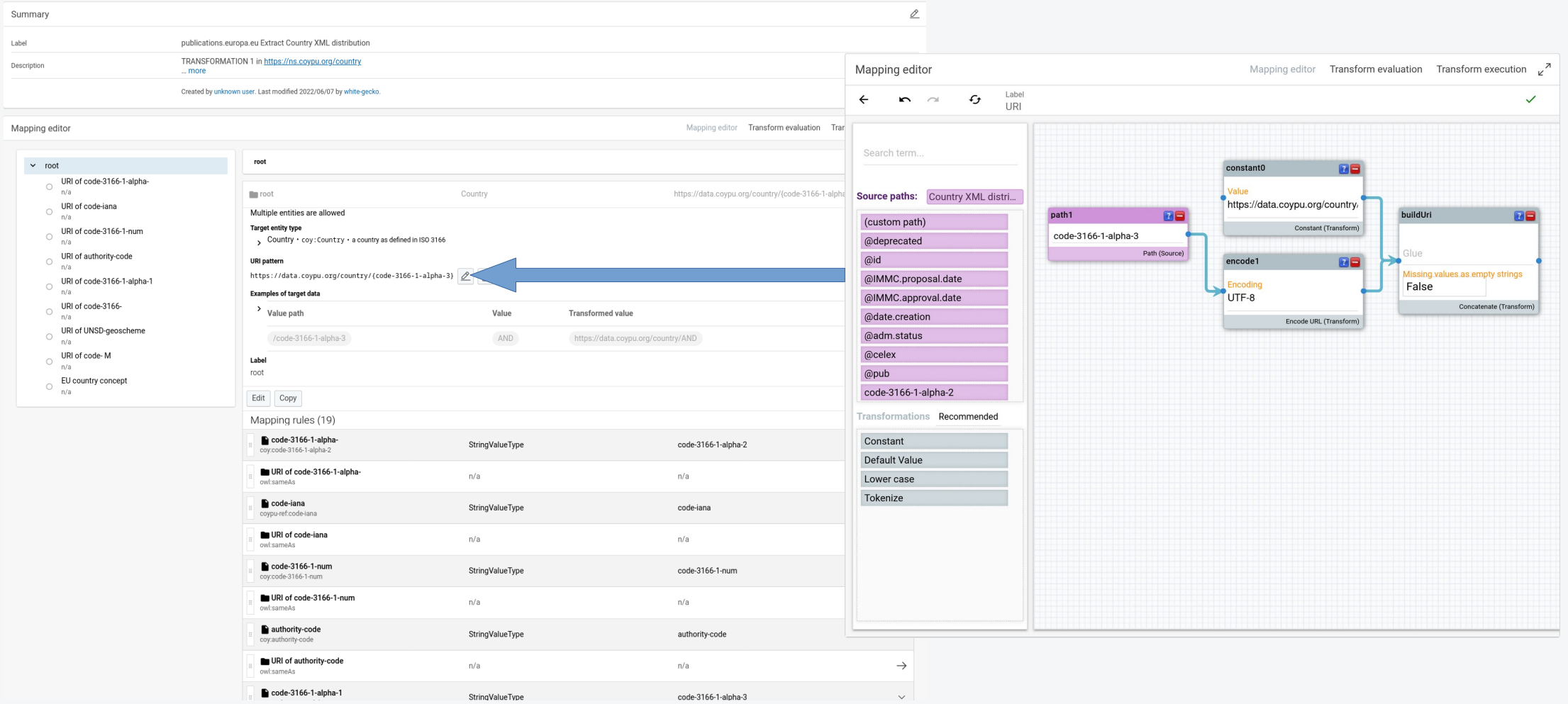
© eccenca GmbH 2022

Transformation of the Datasources



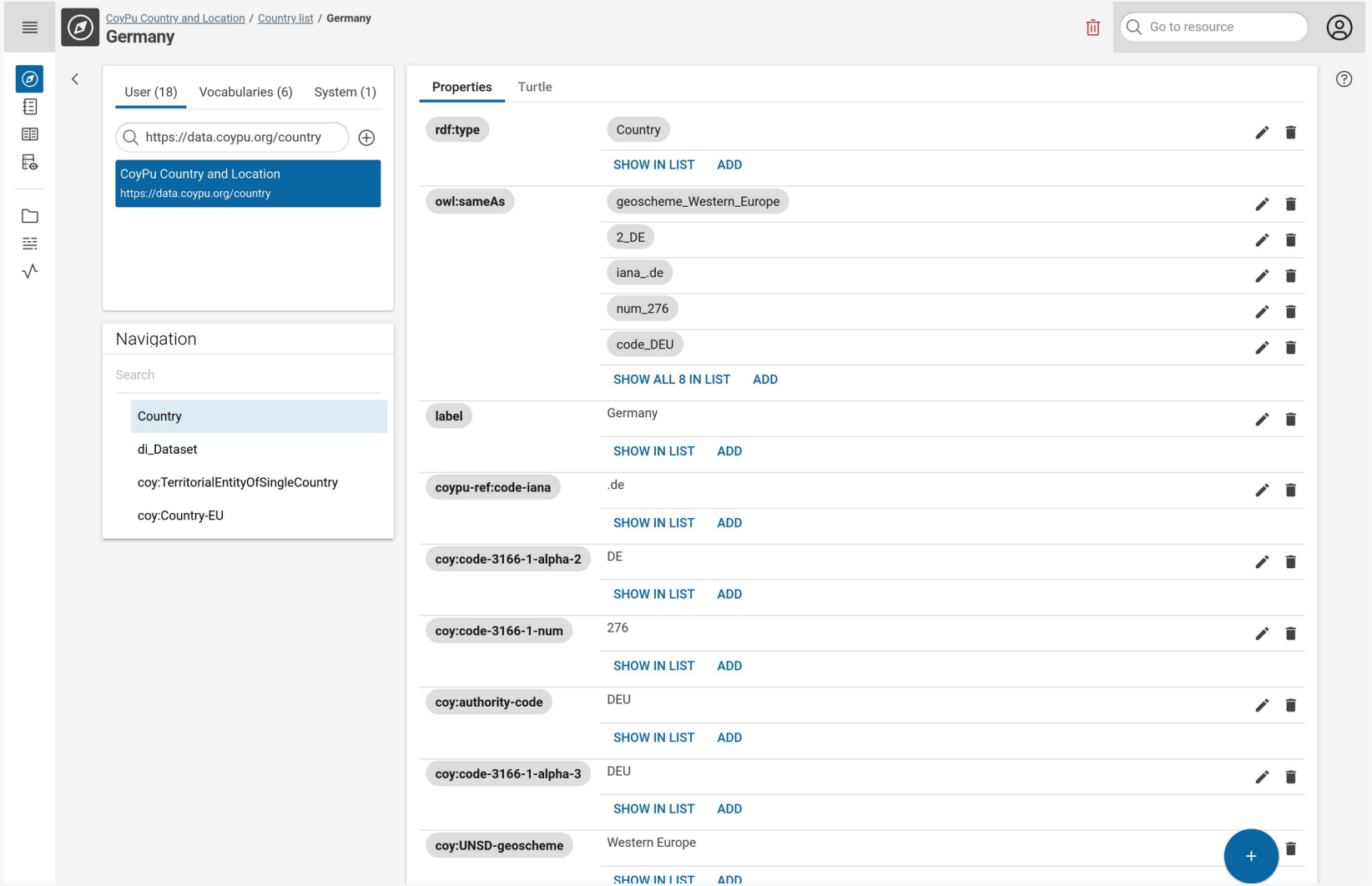
© eccenca GmbH 2022

Transformation of the Datasources



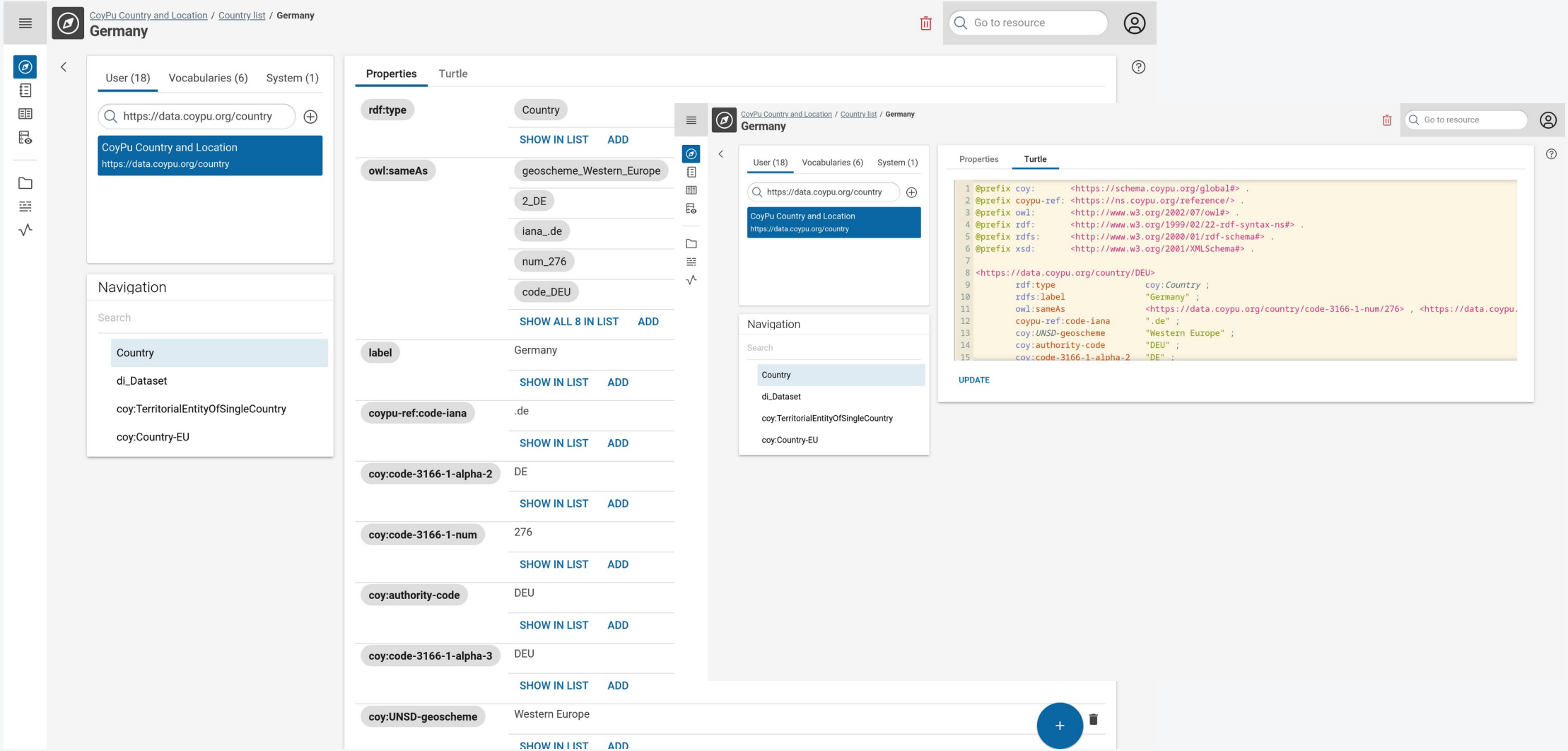
© eccenca GmbH 2022

Explore the Data



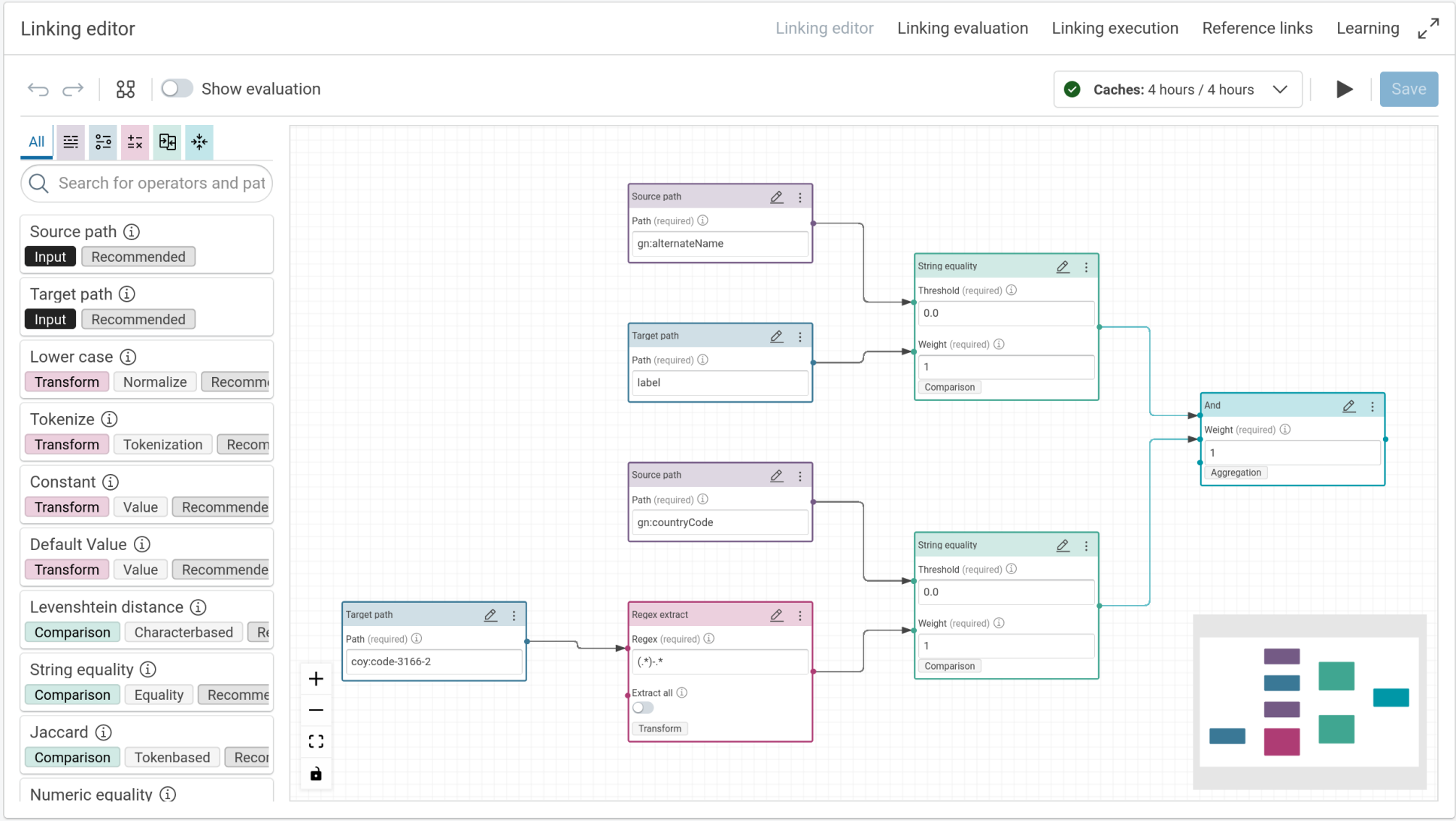
© eccenca GmbH 2022

Explore the Data

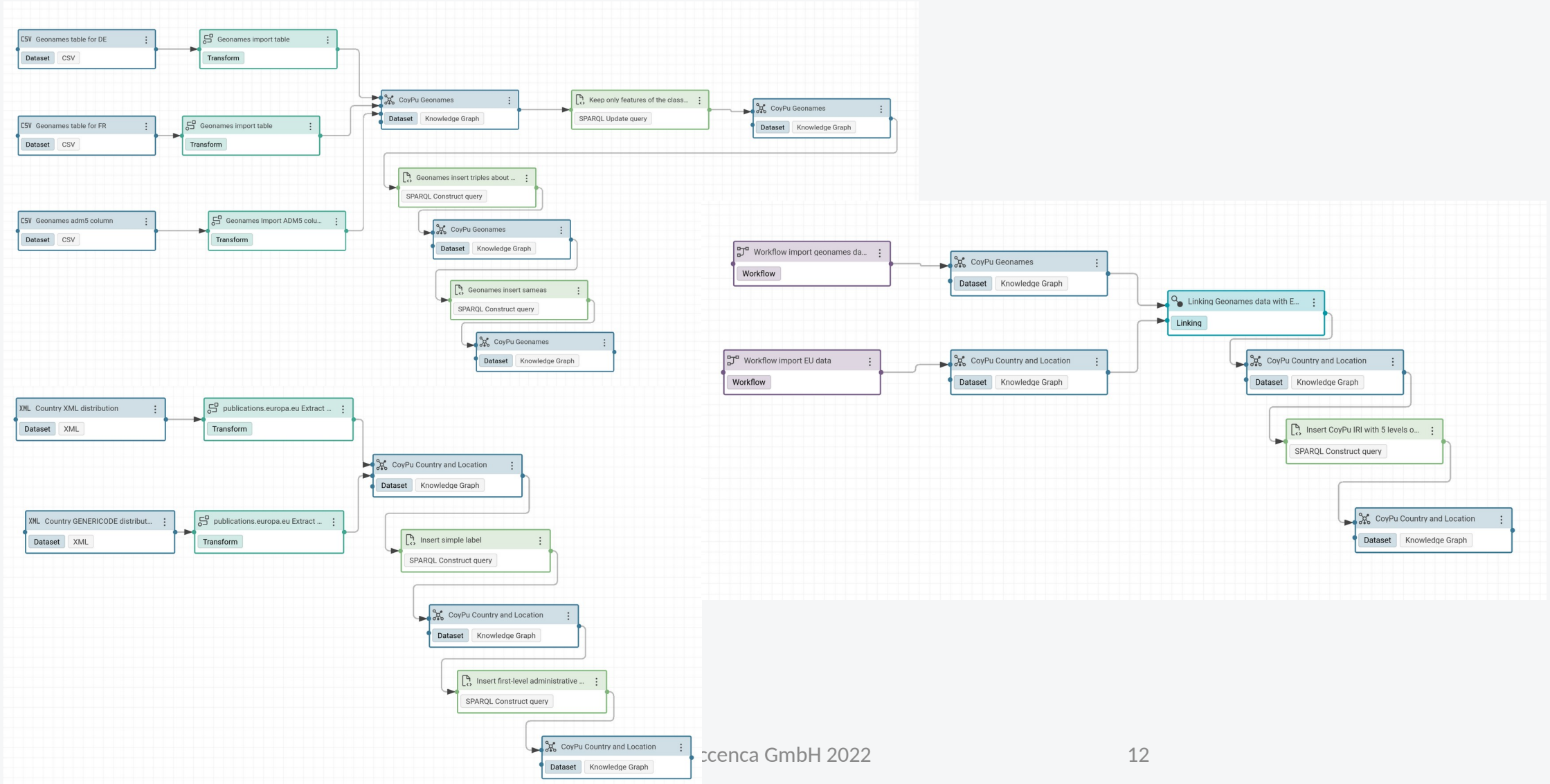


© eccenca GmbH 2022

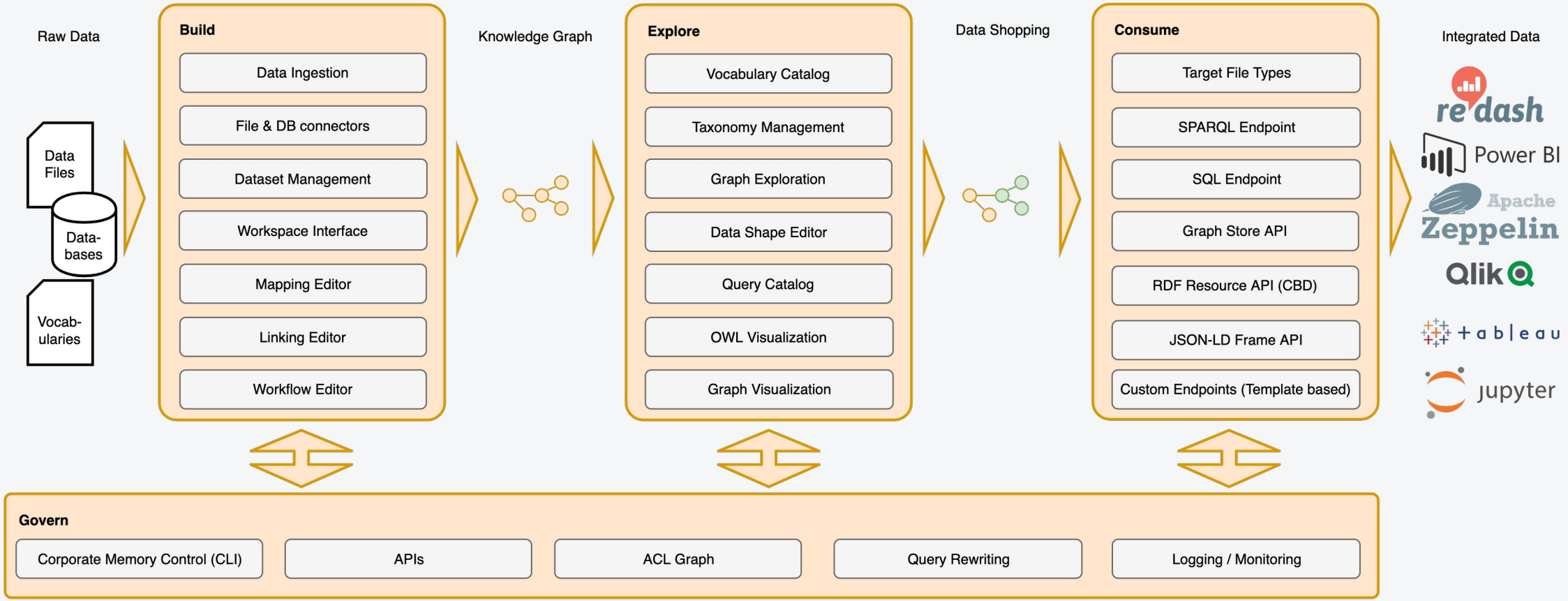
Linking


Linking editor
Linking editor
Linking evaluation
Linking execution
Reference links
Learning
Show evaluation
Caches: 4 hours / 4 hours
Save
All
Search for operators and pat
Source path
Input
Recommended
Target path
Input
Recommended
Lower case
Transform
Normalize
Tokenize
Transform
Tokenization
Constant
Transform
Value
Default Value
Transform
Value
Levenshtein distance
Comparison
Characterbased
String equality
Comparison
Equality
Jaccard
Comparison
Tokenbased
Numeric equality
Source path
Path (required)
gn:alternateName
Target path
Path (required)
label
String equality
Threshold (required)
0.0
Weight (required)
1
Comparison
And
Weight (required)
1
Aggregation
Source path
Path (required)
gn:countryCode
String equality
Threshold (required)
0.0
Weight (required)
1
Comparison
Target path
Path (required)
coy:code-3166-2
Regex extract
Regex (required)
(.*)-.*
Extract all
Transform

The Mapping Workflow

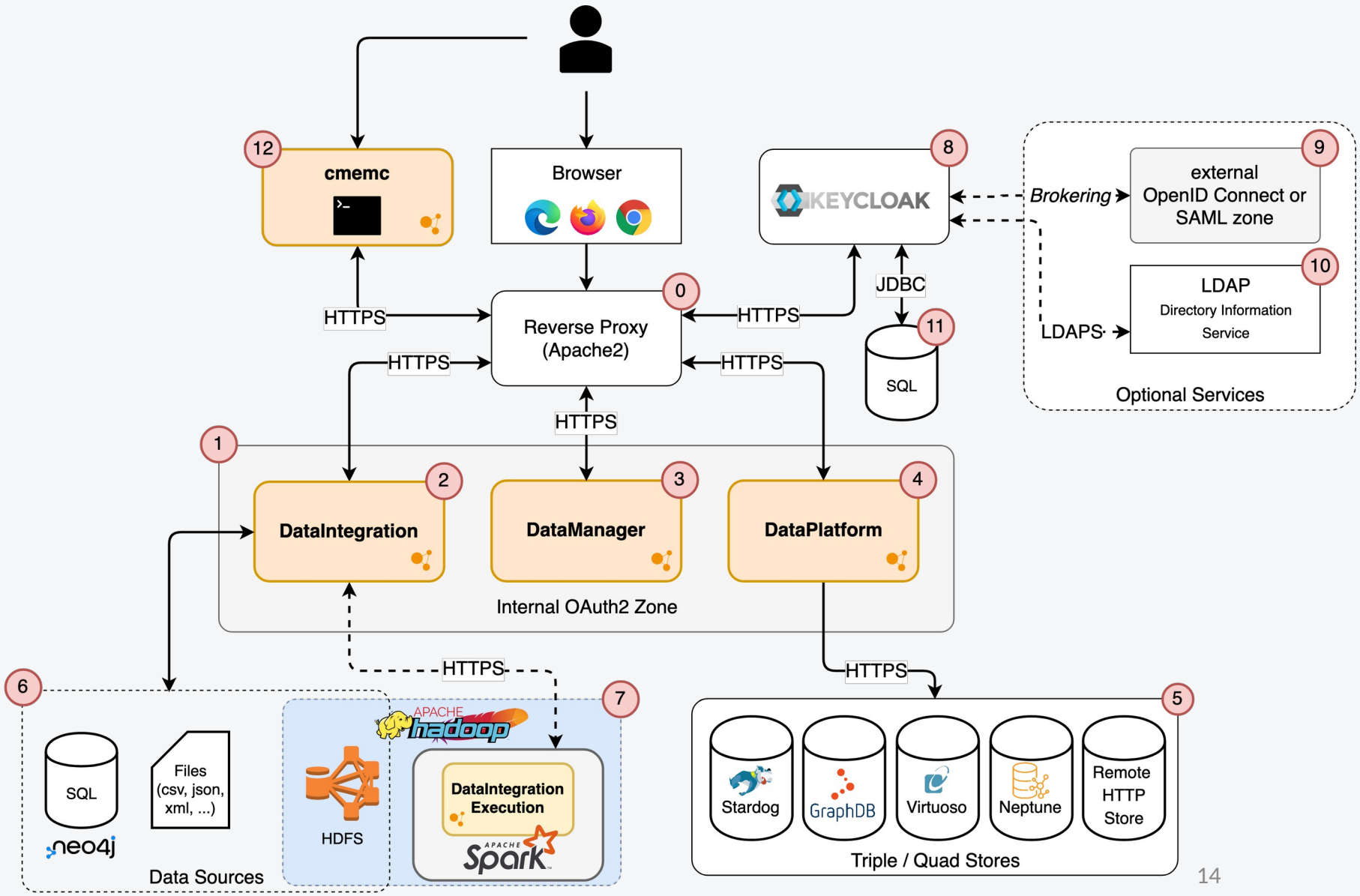


User Journey and Functional Areas

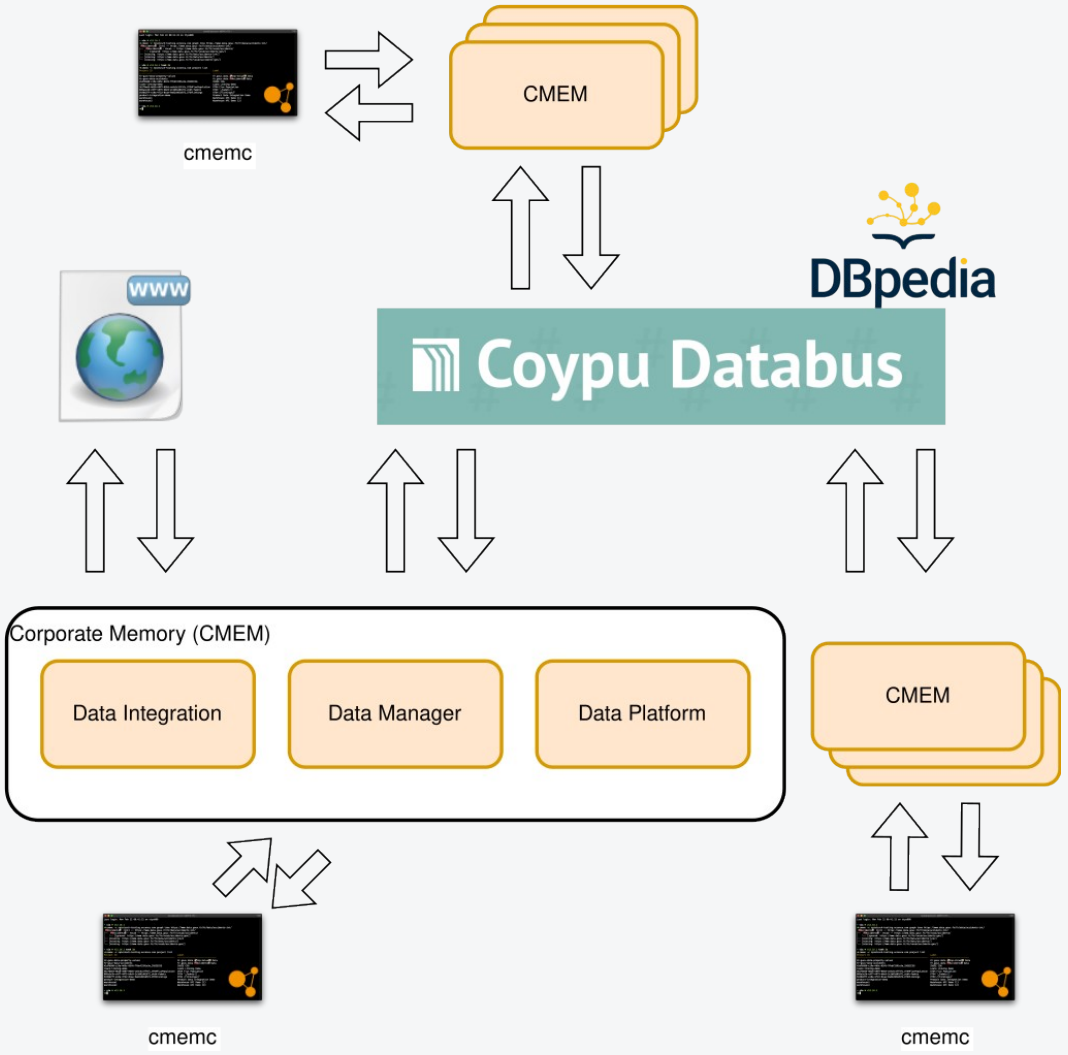


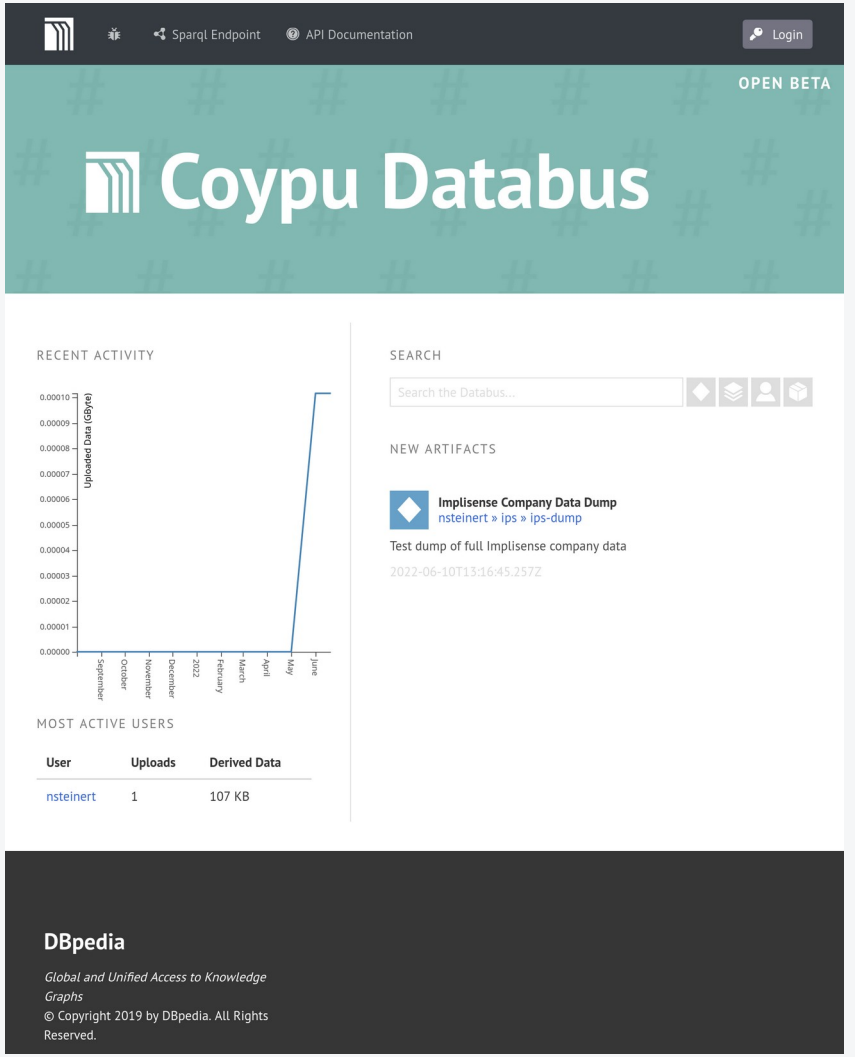
© eccenca GmbH 2022

Reference Architecture Corporate Memory

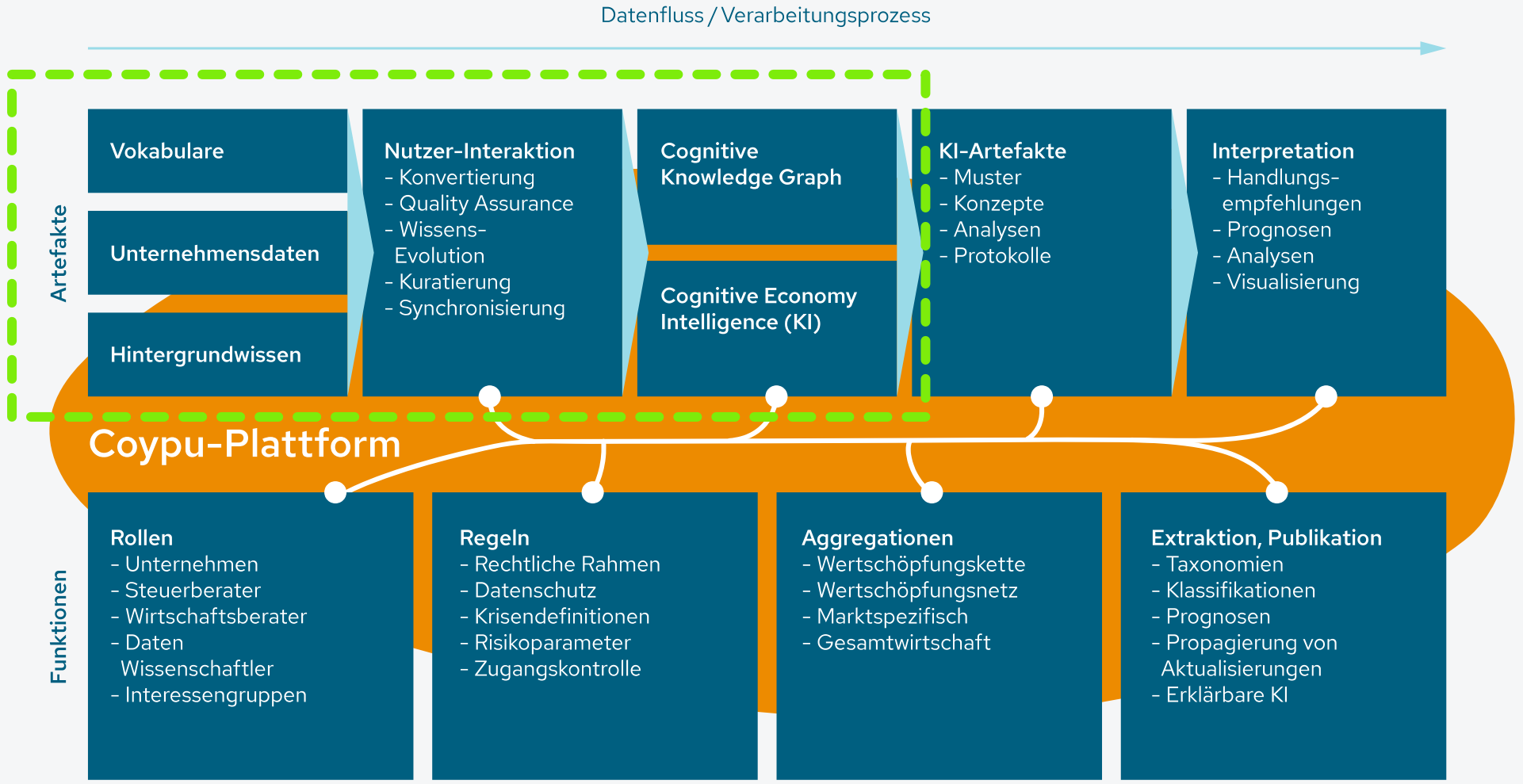


Integration with the Databus





The Coypu-Plattform



eccenca GmbH
Hainstraße 8
D-04109 Leipzig
Germany

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

eccenca
mastering complexity

© eccenca GmbH 2021

Hans-Chr. Brockmann
Geschäftsführer

Hainstraße 8
D-04109 Leipzig
Germany

+49 511 3365 2810
+49 173 3698 610

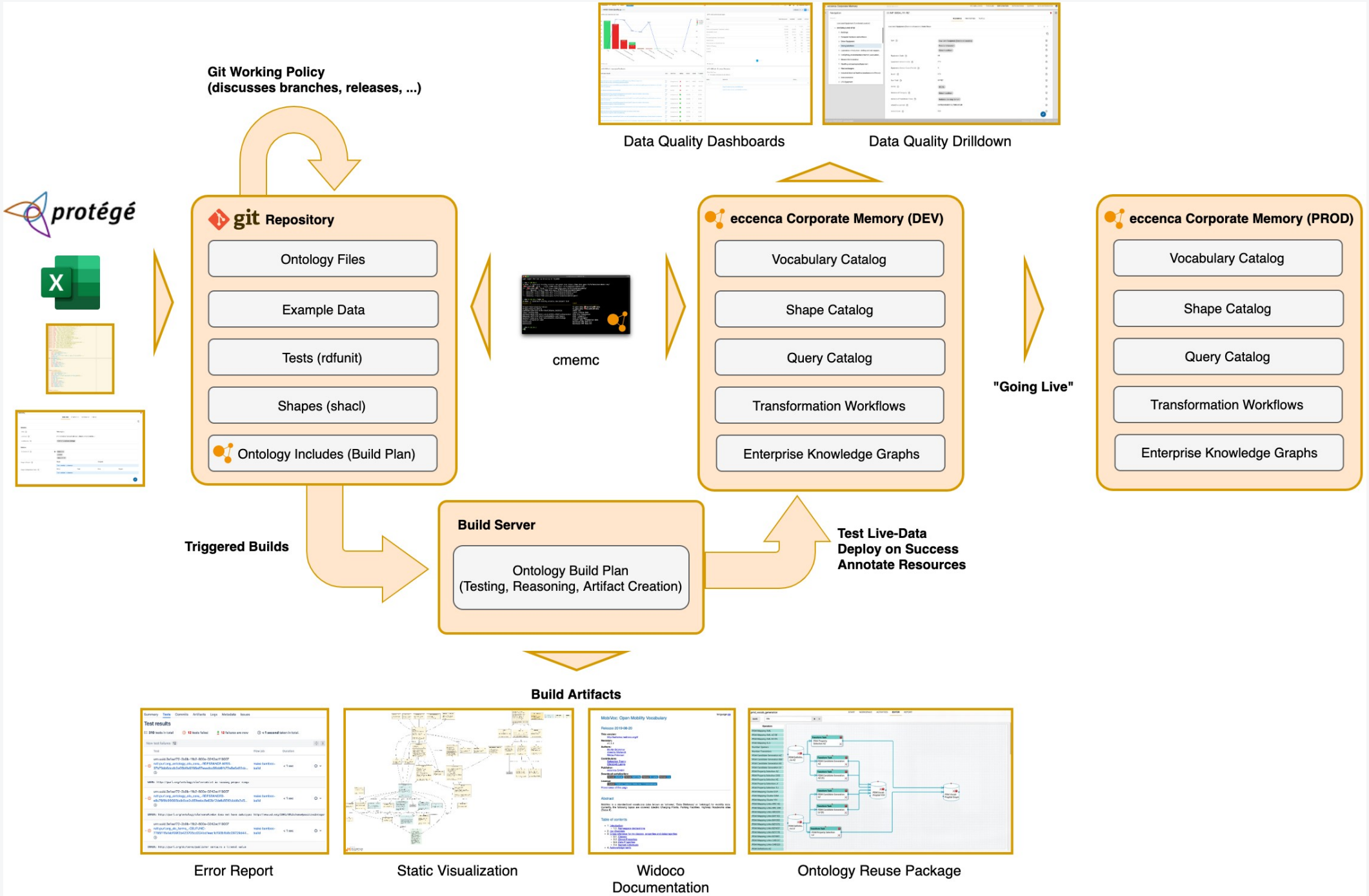
brockmann@eccenca.com
https://eccenca.com

© eccenca GmbH 2021

© eccenca GmbH 2021

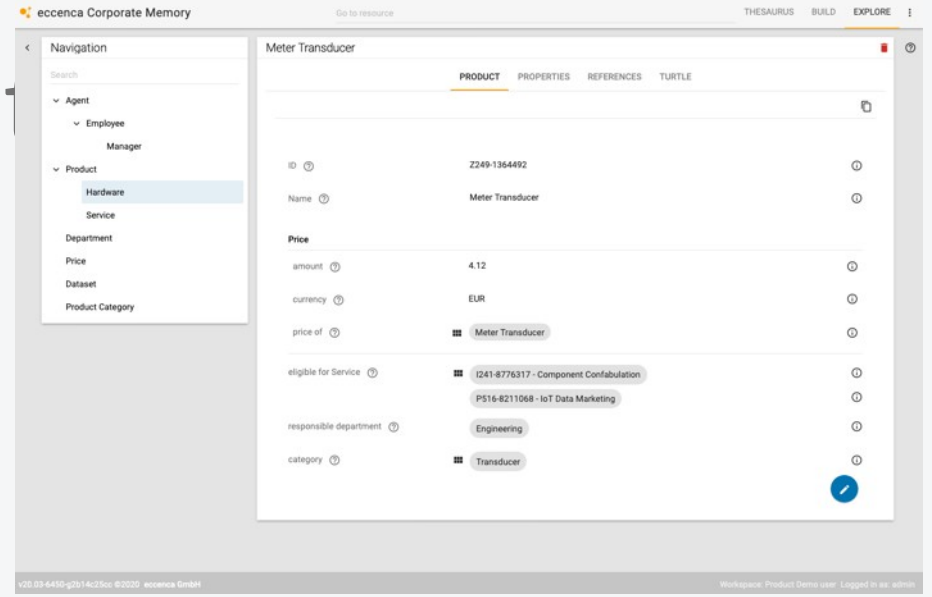
BACKUP

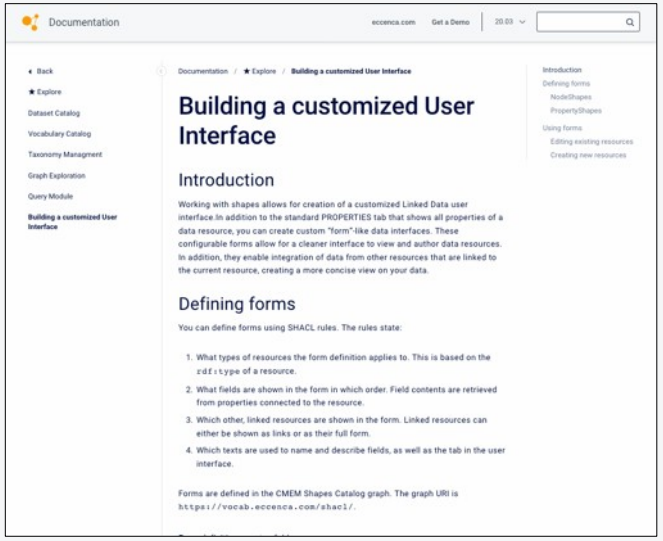
© eccenca GmbH 2021


Git Working Policy
(discusses branches, releases, ...)
Data Quality Dashboards
Data Quality Drilldown
protégé
git Repository
Ontology Files
Example Data
Tests (rdfunit)
Shapes (shacl)
Ontology Includes (Build Plan)
cmemc
eccenca Corporate Memory (DEV)
Vocabulary Catalog
Shape Catalog
Query Catalog
Transformation Workflows
Enterprise Knowledge Graphs
"Going Live"
eccenca Corporate Memory (PROD)
Vocabulary Catalog
Shape Catalog
Query Catalog
Transformation Workflows
Enterprise Knowledge Graphs
Triggered Builds
Build Server
Ontology Build Plan
(Testing, Reasoning, Artifact Creation)
Test Live-Data
Deploy on Success
Annotate Resources
Build Artifacts
Error Report
Static Visualization
Widoco
Documentation
Ontology Reuse Package

Data defined UI Configurat

- 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

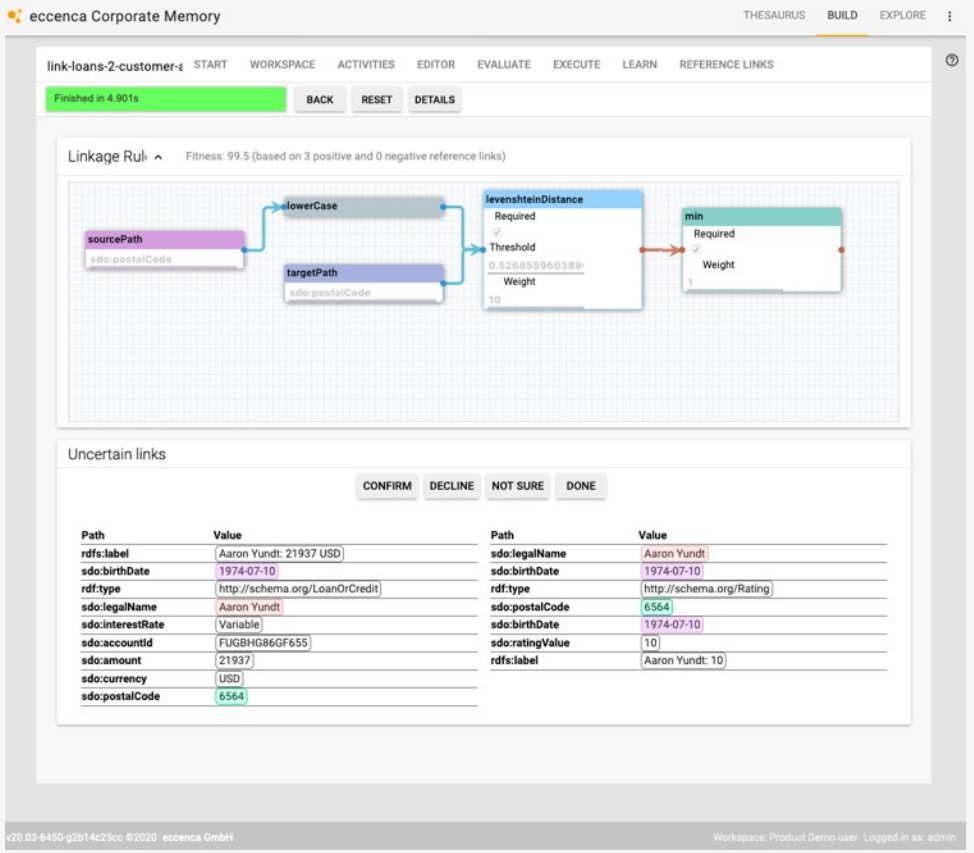


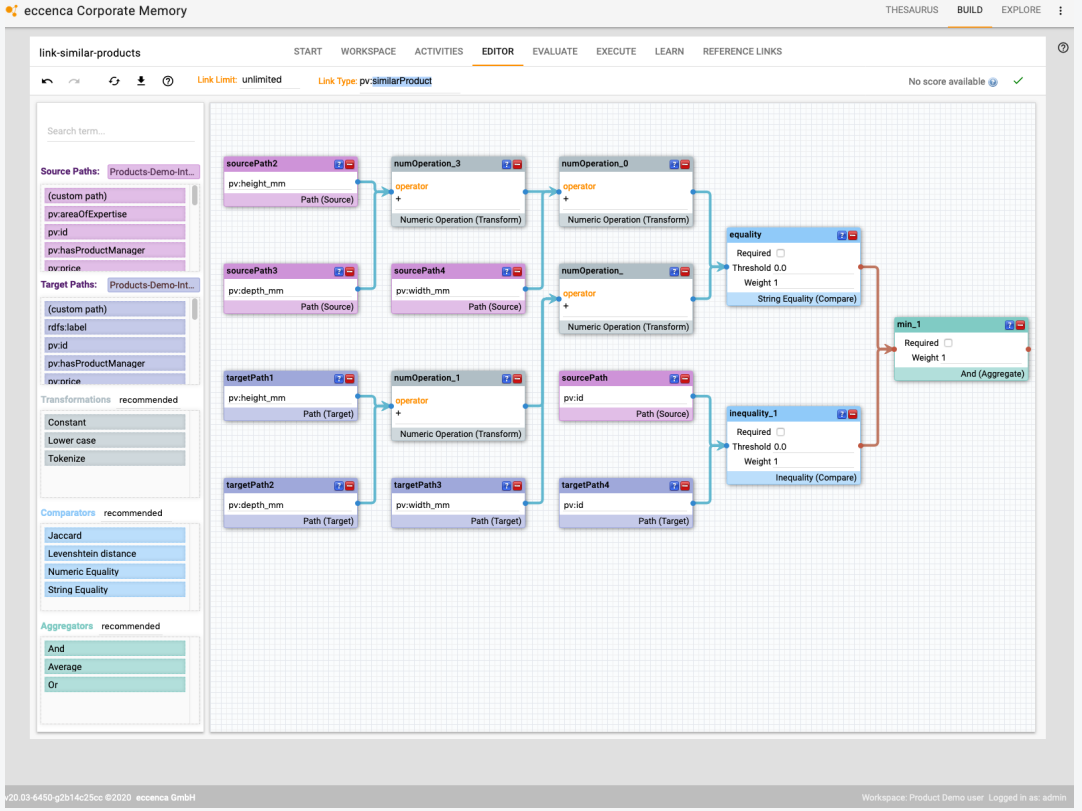


https://documentation.eccenca.com/latest/explore/building-a-customized-user-interface

© eccenca GmbH 2022

Linking as general rules mechanism



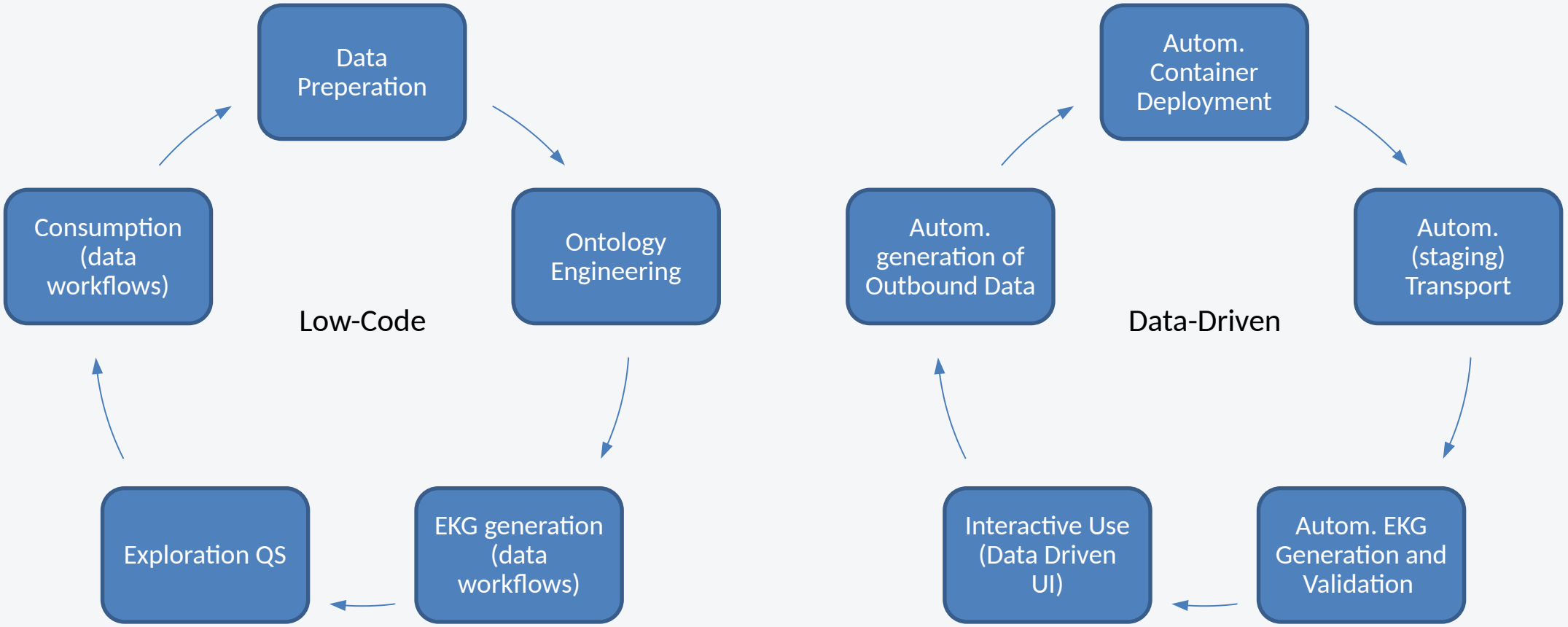


DevOps based Ontology Engineering Pipeline

- 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

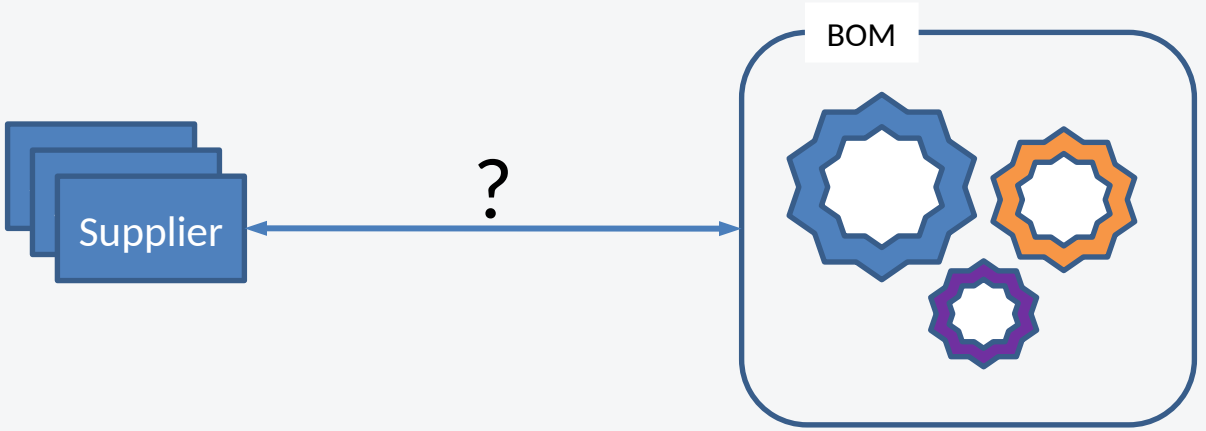


© eccenca GmbH 2021

SAMPLE USE CASES

© eccenca GmbH 2021

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



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

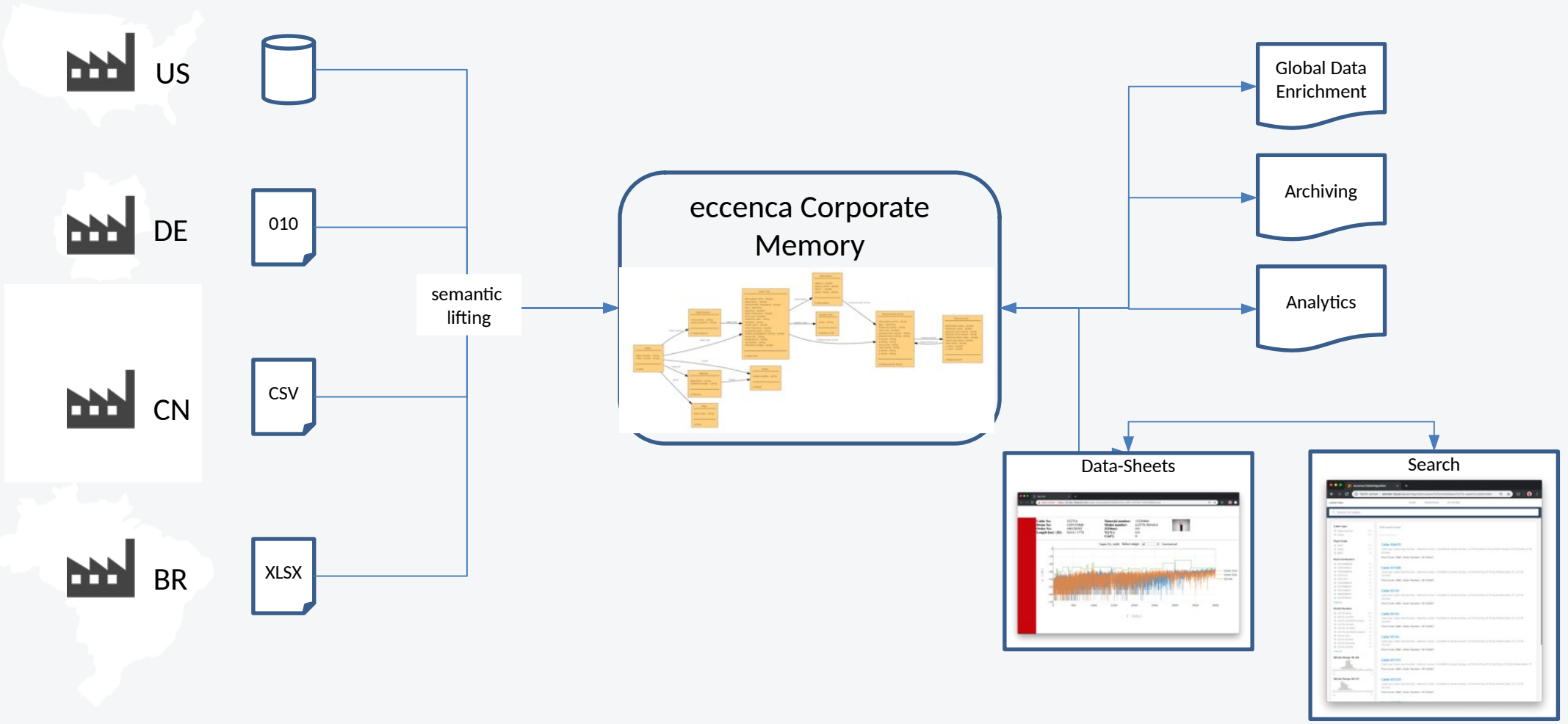
MPN	**Supplier**
293D105X9016A2##E3	VISHAY
B45196E3105K10	KEMET
T491A105K016AS	KEMET
T491A105K016AT	KEMET
TAJA105K016RNJ	AVX

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
…	

© eccenca GmbH 2021

Cross Site Semantic Data Harmonization



© eccenca GmbH 2021

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

© eccenca GmbH 2021

About eccenca GmbH

brox IT-Solutions GmbH

brox information excellence

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

AKSW – Universität Leipzig & Fraunhofer IAIS

AKSW

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

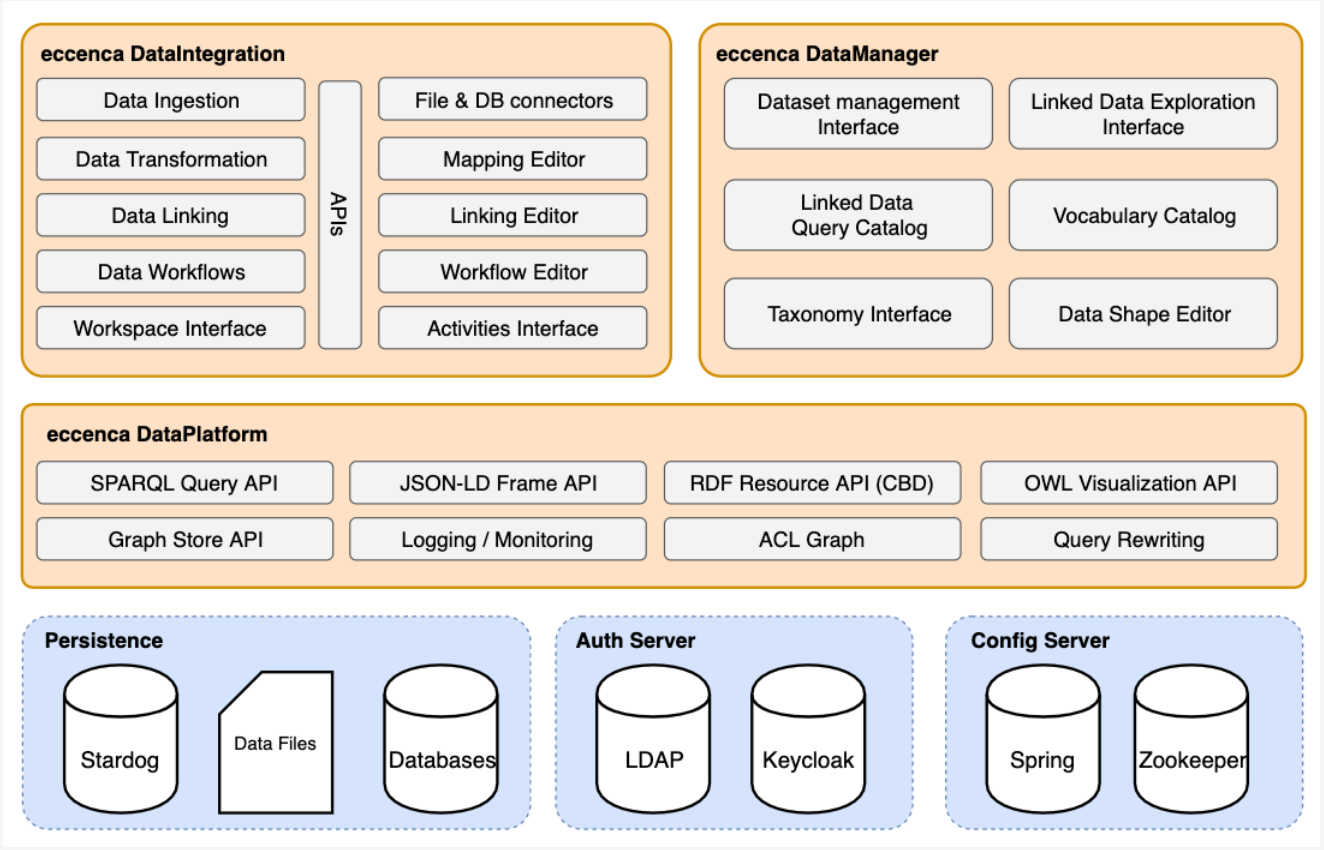
eccenca GmbH

eccenca Command your Data!

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

© eccenca GmbH 2021

Technical Architecture



© eccenca GmbH 2021

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 GmbH 2021

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 GmbH 2021

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)

© eccenca GmbH 2021

Semantic Data Management
Data in Context

What

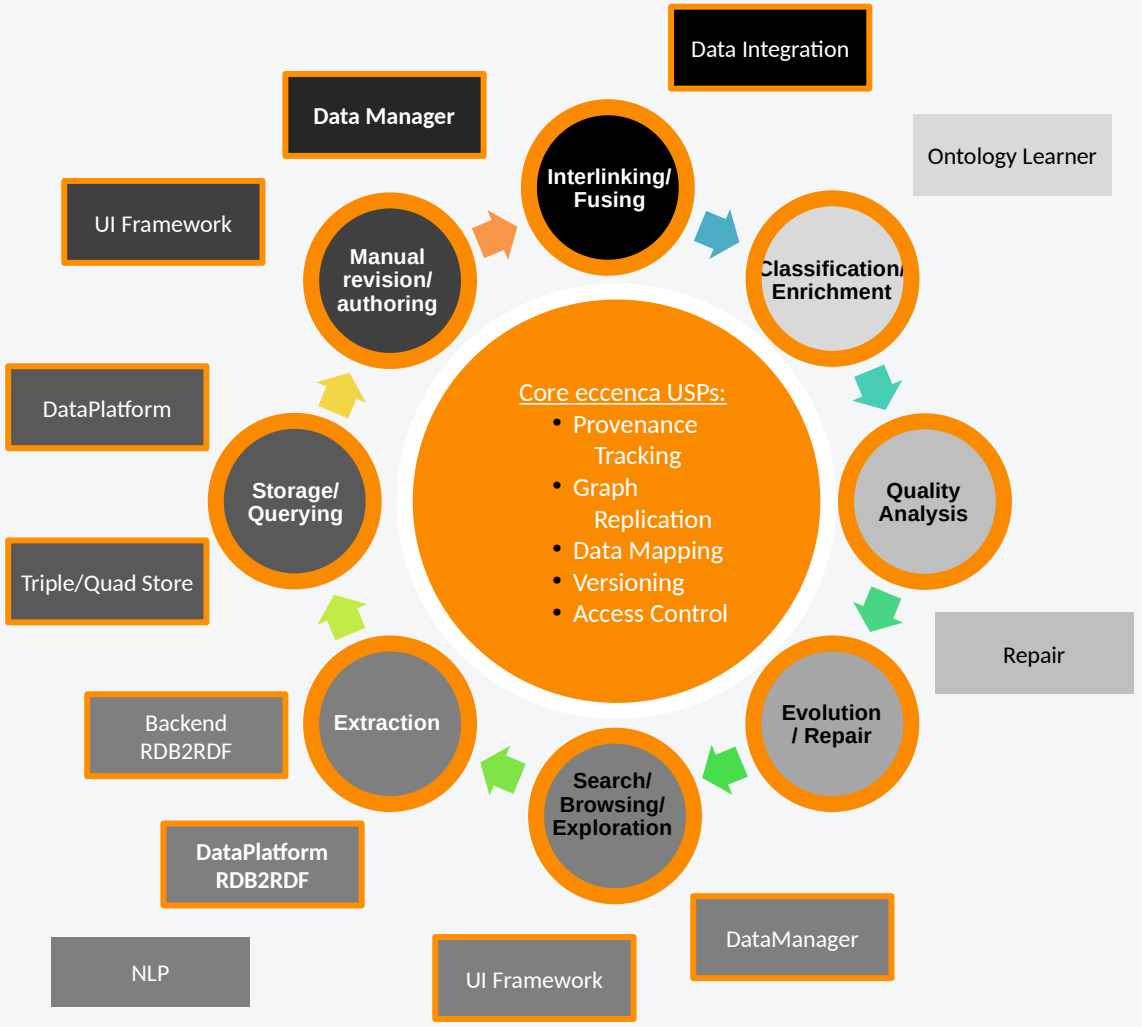
- Create knowledge graphs by connecting datasets and metadata to logical models
- Physical data models unchanged!
-
  - Explore metadata & structures
  - Query & Access data via models
  - Integrate data on model level
  - Share data on model level

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

© eccenca GmbH 2021

Linked Data Life Cycle



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

© eccenca GmbH 2021

Changing the Data-Collaboration Paradigm

… by turning STRINGS into THINGS

Schema defines Semantics

API
EDI
XML

TODAY

Local Standardization
Slow Innovation
NxM cost
Changes break the contract

RDF

TOMORROW

Low Standardization
Agile Innovation
N+1 cost
Changes amend the contract

Standardized Schema (SPO)
Vocabularies define
Semantics

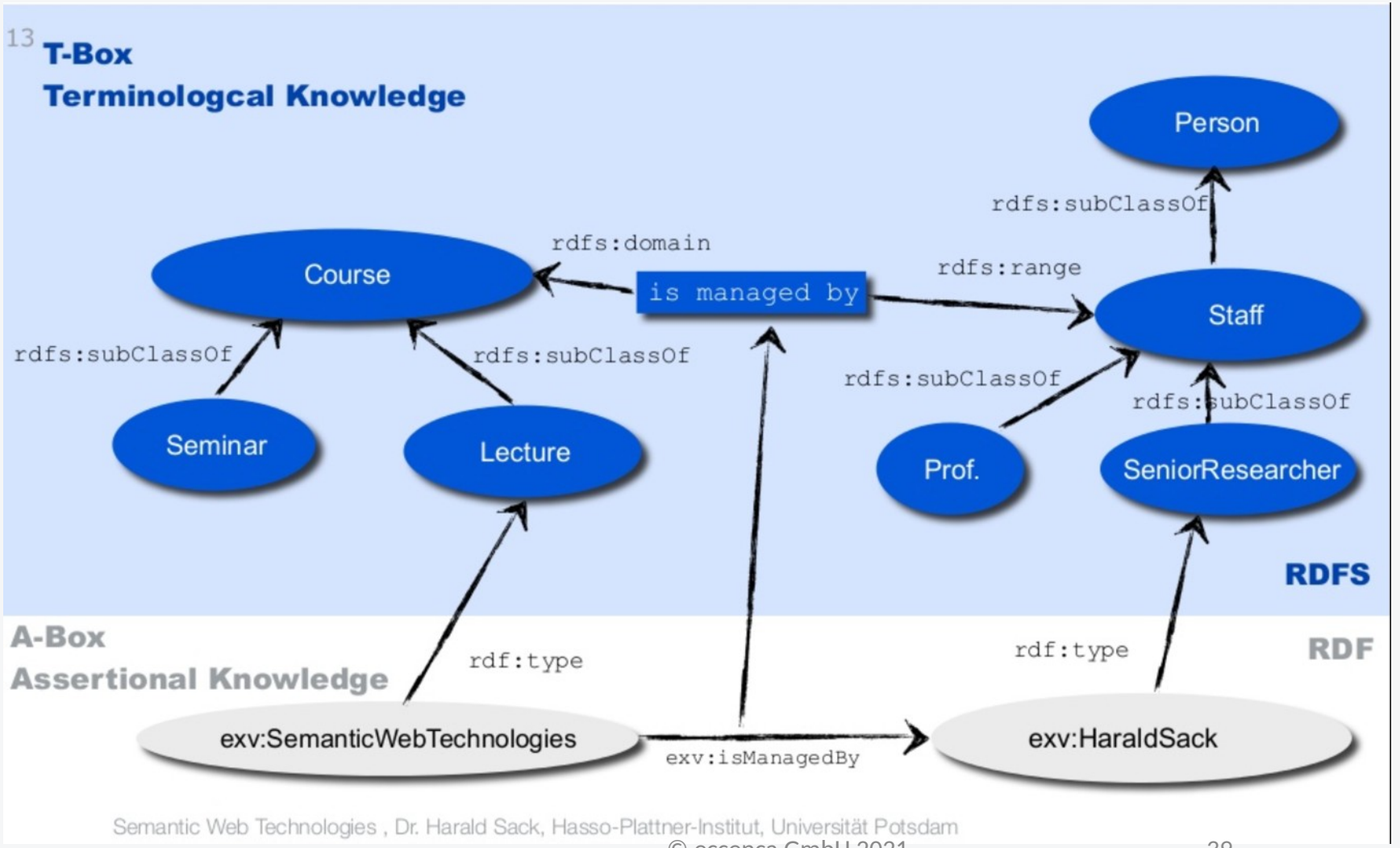
© eccenca GmbH 2021

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.

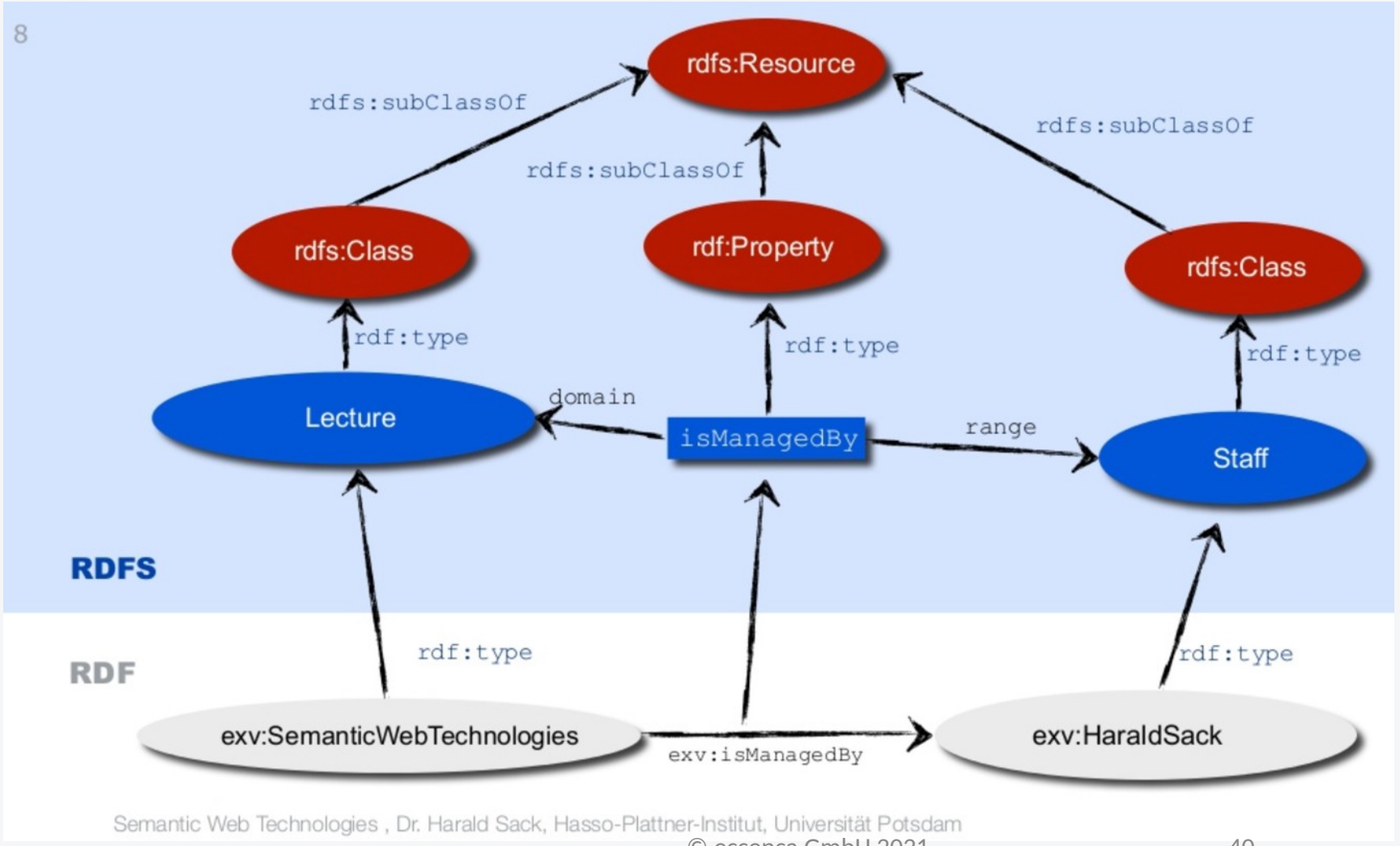
© eccenca GmbH 2021

RDF +RDFS



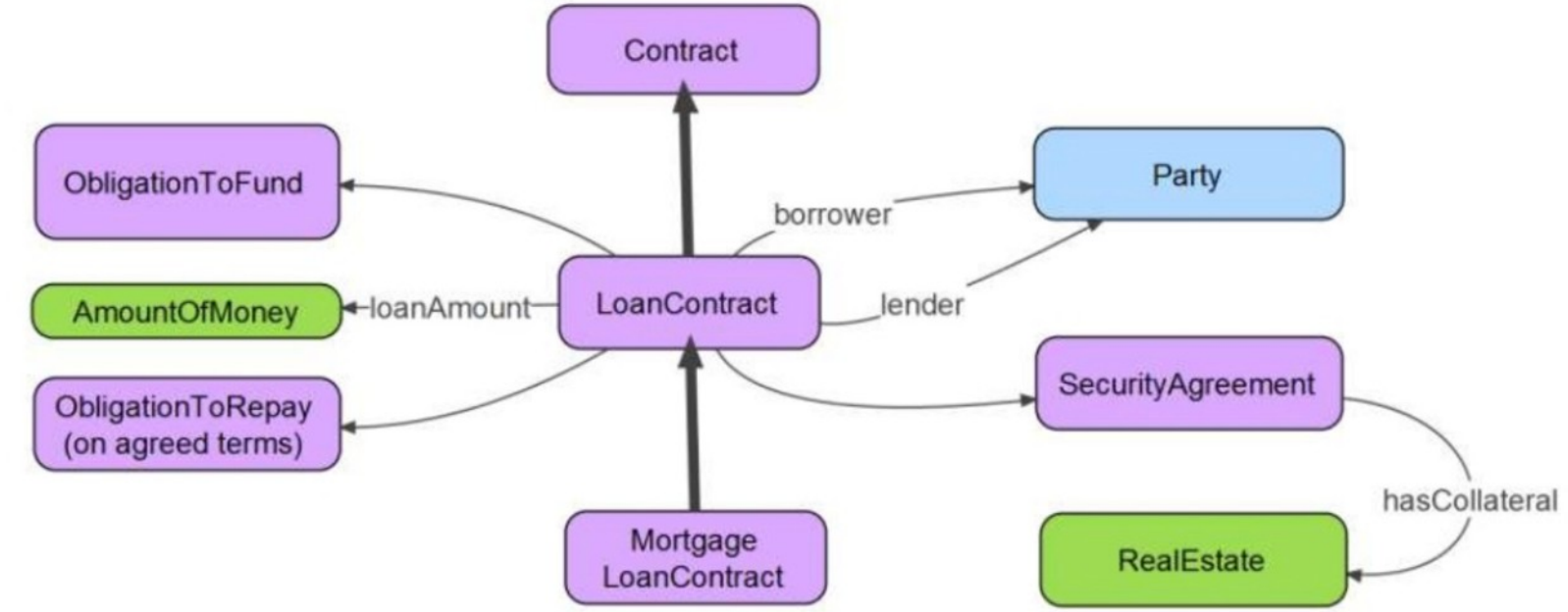
© eccenca GmbH 2021

Classes, Properties, Instances



© eccenca GmbH 2021

OWL: Inferencing classification.



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

© eccenca GmbH 2021

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 .
```

☹