The Coypu Platform
Data Integration Workflow

Dr. Natanael Arndt
Senior Linked Data Expert @ eccenca GmbH
The Coypu-Plattform

Datenfluss/Verarbeitungsprozess

Artefakte
- Vokabulare
- Unternehmensdaten
- Hintergrundwissen

Nutzer-Interaktion
- Konvertrierung
- Quality Assurance
- Wissens-Evolution
- Kuratierung
- Synchronisierung

Cognitive Knowledge Graph

Cognitive Economy Intelligence (KI)

KI-Artefakte
- Muster
- Konzepte
- Analysen
- Protokolle

Interpretation
- Handlungs-empfehlungen
- Prognosen
- Analysen
- Visualisierung

Funktionen
- Rollen
  - Unternehmen
  - Steuerberater
  - Wirtschaftsberater
  - Daten
  - Wissenschaftler
  - Interessengruppen

- Regeln
  - Rechtliche Rahmen
  - Datenschutz
  - Krisendefinitionen
  - Risikoparameter
  - Zugangskontrolle

- Aggregationen
  - Wertschöpfungskette
  - Wertschöpfungsnetz
  - Marktspezifisch
  - Gesamtwirtschaft

- Extraktion, Publikation
  - Taxonomien
  - Klassifikationen
  - Prognosen
  - Propagierung von Aktualisierungen
  - Erklärbare KI
The Coypu-Plattform

Datenfluss/Verarbeitungsprozess

Vokabulare
- Nutzer-Interaktion
  - Konvertierung
  - Quality Assurance
  - Wissens-Evolution
  - Kuratierung
  - Synchronisierung

Cognitive Knowledge Graph
Cognitive Economy Intelligence (KI)

KI-Artefakte
- Muster
- Konzepte
- Analysen
- Protokolle

Interpretation
- Handlungsempfehlungen
- Prognosen
- Analysen
- Visualisierung

Unternehmensdaten

Hintergrundwissen

Artefakte

Funktionen

Rollen
- Unternehmen
- Steuerberater
- Wirtschaftsberater
- Daten
- Wissenschaftler
- Interessengruppen

Regeln
- Rechtliche Rahmen
- Datenschutz
- Krisendefinitionen
- Risikoparameter
- Zugangskontrolle

Aggregationen
- Wertschöpfungskette
- Wertschöpfungsnetz
- Marktpezifisch
- Gesamtwirtschaft

Extraktion, Publikation
- Taxonomien
- Klassifikationen
- Prognosen
- Propagierung von Aktualisierungen
- Erklärbare KI
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 ...

1. Gather example data and documentation for each source and target
2. Create core semantic model bottom-up
3. Review model with domain experts
4. Map first data source, review data with domain expert
5. Iteratively expand model in further mappings

© eccenca GmbH 2022
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
Transformation of the Datasources
Transformation of the Datasources

Mapping editor

Search term...

Transformations

© eccenca GmbH 2022
Explore the Data

<table>
<thead>
<tr>
<th>Properties</th>
<th>Turtle</th>
</tr>
</thead>
<tbody>
<tr>
<td>rdf:type</td>
<td>Country</td>
</tr>
<tr>
<td></td>
<td>SHOW IN LIST ADD</td>
</tr>
<tr>
<td>owl:sameAs</td>
<td>geoscheme_Western_Europe</td>
</tr>
<tr>
<td></td>
<td>2.DE</td>
</tr>
<tr>
<td></td>
<td>lana_de</td>
</tr>
<tr>
<td></td>
<td>num_276</td>
</tr>
<tr>
<td></td>
<td>code_DEU</td>
</tr>
<tr>
<td></td>
<td>SHOW ALL 8 IN LIST ADD</td>
</tr>
<tr>
<td>label</td>
<td>Germany</td>
</tr>
<tr>
<td></td>
<td>SHOW IN LIST ADD</td>
</tr>
<tr>
<td>coy:ref:code-lana</td>
<td>de</td>
</tr>
<tr>
<td></td>
<td>SHOW IN LIST ADD</td>
</tr>
<tr>
<td>coy:code-3166-1-alpha-2</td>
<td>DE</td>
</tr>
<tr>
<td></td>
<td>SHOW IN LIST ADD</td>
</tr>
<tr>
<td>coy:code-3166-1-num</td>
<td>276</td>
</tr>
<tr>
<td></td>
<td>SHOW IN LIST ADD</td>
</tr>
<tr>
<td>coy:authority-code</td>
<td>DEU</td>
</tr>
<tr>
<td></td>
<td>SHOW IN LIST ADD</td>
</tr>
<tr>
<td>coy:code-3166-1-alpha-3</td>
<td>DEU</td>
</tr>
<tr>
<td></td>
<td>SHOW IN LIST ADD</td>
</tr>
<tr>
<td>coy:UNSD-geoscheme</td>
<td>Western Europe</td>
</tr>
<tr>
<td></td>
<td>SHOW IN LIST ADD</td>
</tr>
</tbody>
</table>
Explore the Data
The Mapping Workflow
User Journey and Functional Areas

Build
- Data Ingestion
- File & DB Connectors
- Dataset Management
- Workspace Interface
- Mapping Editor
- Linking Editor
- Workflow Editor

Knowledge Graph

Explore
- Vocabulary Catalog
- Taxonomy Management
- Graph Exploration
- Data Shape Editor
- Query Catalog
- OWL Visualization
- Graph Visualization

Data Shopping

Consume
- Target File Types
- SPARQL Endpoint
- SQL Endpoint
- Graph Store API
- RDF Resource API (CBD)
- JSON-LD Frame API
- Custom Endpoints (Template based)

Govern
- Corporate Memory Control (CLI)
- APIs
- ACL Graph
- Query Rewriting
- Logging / Monitoring

Integrations
- re: dash
- Power BI
- Apache Zeppelin
- Qlik
- Tableau
- Jupyter

© eccenca GmbH 2022
Integration with the Databus

Corporate Memory (CMEM)
- Data Integration
- Data Manager
- Data Platform
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
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

Data Preparation

Ontology Engineering

Autom. Container Deployment

Autom. EKG Generation and Validation

Autom. generation of Outbound Data

Interactive Use (Data Driven UI)

Autom. (staging) Transport

Low-Code

Consumption (data workflows)

EKG generation (data workflows)

Exploration QS

Data-Driven

© eccenca GmbH 2021
SAMPLE USE CASES
Apply Machine Learning to perform “Form Fit Functional” Material Linking

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

<table>
<thead>
<tr>
<th>MPN</th>
<th>Supplier</th>
</tr>
</thead>
<tbody>
<tr>
<td>293D105X9016A2##E3</td>
<td>VISHAY</td>
</tr>
<tr>
<td>B45196E3105K10</td>
<td>KEMET</td>
</tr>
<tr>
<td>T491A105K016AS</td>
<td>KEMET</td>
</tr>
<tr>
<td>T491A105K016AT</td>
<td>KEMET</td>
</tr>
<tr>
<td>TAJA105K016RNJ</td>
<td>AVX</td>
</tr>
</tbody>
</table>

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 Corporate Memory

Semantic lifting

010

CSV

XLSX

Global Data Enrichment

Archiving

Analytics

Data-Sheets

Search
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: Initierung/Leitung eclipse.org/SMILA
eccenca Enterprise Search
Key Accounts: Volkswagen, Audi, Skoda, MAN, Telekom, Daimler, Bosch, Siemens, Continental

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

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

© 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)
• 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!
- 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

Core eccenca USPs:
- Provenance Tracking
- Graph Replication
- Data Mapping
- Versioning
- Access Control

DataPlatform
- RDB2RDF
- DataManager
- NLP
- Backend RDB2RDF
- Triple/Quad Store
- UI Framework
- Data Integration
- Interlinking / Fusing
- Classification / Enrichment
- Quality / Evolution
- Search / Browse / Explore
- Repair
- Manual revision / authoring
- Extraction / Mapping
- Storage / Querying
- Manual Revision / Authoring
- Linking / Fusion
- Classification / Enrichment
- Search / Browse / Explore

© eccenca GmbH 2021
Changing the Data-Collaboration Paradigm

... by turning STRINGS into THINGS

Today:
- Local Standardization
- Slow Innovation
- NxM cost
- Changes break the contract

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

API
EDI
XML

RDF

Schema defines Semantics

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.
RDF + RDFS
Classes, Properties, Instances

![Diagram showing relationships between classes, properties, and instances in RDFS and RDF standards.](image-url)
**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

<table>
<thead>
<tr>
<th>Constraint</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>node*</td>
<td>All values of a given property must have a given shape. Recursion is not allowed in current SHACL.</td>
</tr>
</tbody>
</table>

:User a sh:NodeShape, rdfs:Class ;

:Company a sh:Shape ;

:alice a :User;
  schema:worksFor :OurCompany .

:bob a :User;
  schema:worksFor :Another .

:OurCompany
  schema:name "OurCompany" .

:Another
  schema:name 23 .