CUSTOMER SUCCESS STORY

POOLPARTY SEMANTIC SUITE USED TO INTEGRATE PHARMACEUTICAL GLOBAL RESEARCH DATA

Industry

✓ Pharmaceutical

Challenges

✓ Separate, huge research data not searchable
✓ Missing structure and meta-data on large scale repositories

Solutions

✓ Tag SharePoint documents using PoolParty Powertagging
✓ Transform data into RDF compatible formats with PoolParty UnifiedViews
✓ Integrate multiple platforms and repositories
✓ Prepare and use pharmaceutical standard databses for tagging, such as MeSH, UniProt or ChEMBL
✓ PoolParty GraphSearch used as server and interface solution to search and browse data

Key Benefits

✓ Increased speed of research and target search
✓ Enhancement of collaboration among researchers

Situation: Research Globally Distributed Inaccessible Data

Research data of genetic, protein, cell metabolism, related tests and their outcomes, such as biological activity, was stored in several distributed, partly globally spread repositories and databases. Faulty access to research related data started to seriously hamper new projects and realizations since target search and thus reliable results became increasingly difficult. Several integration systems had been tried but were not producing the necessary results.

Challenge: Develop and Implement an Enterprise Linked Data Strategy

Semantic Web technologies that PoolParty is based on, are particularly apt to solve problems of inaccessible data and inefficient searches. RDF, URIs, data modeling using taxonomies and ontologies and reuse of existing enterprise vocabularies are basic elements to be used. The particular difficulty in such a large organization is that linked data deployment has to be managed across parties, platforms and data security policies, where policies might even be mutually exclusive.

Solution: Transform Data: PoolParty and the ETL Tool UnifiedViews Increase Performance up to 50%

The available data was managed and already partly machine-readable across several platforms and departments. Semantic technologies work with machine-readable data in RDF serialization formats. PoolParty's native solution UnifiedViews is an ETL tool that allows using custom developed plugins to be integrated into processing pipelines. Thus, all transformation processes can be controlled precisely and efficiently, batch processing is possible and output formats and locations can be set individually.

Treat Non-machine-readable Data on Large Scales

The existing data types were not always open to easy transformation or transformation was not possible, such as images or video. Data can also be stored in large relational databases where the amount of data does not allow for transformation in advance. In such cases, the Linked Data Platform Protocol (LDP) and the meta-data added to data types that cannot be transformed to RDF will form an integral part of a setup. Including in this case the use of Apache Spark jobs, that would be executed in regular intervals to query databases on the fly. Thus meta-data would be triplified into the corresponding RDF types and would then be used in search queries based on SPARQL.

Implementation Insights: Connect SharePoint, Medical and Pharmaceutical Vocabularies and use PoolParty GraphSearch

The setup as also displayed in the diagram, ensures that all necessary elements can be integrated and perform reliably and fast. SharePoint as platform to keep and structure documents. Tag them there using PoolParty PowerTagging. Sophisticated search, as well as a term store synchronizing component are made possible with it. The PoolParty thesaurus at its root, together with a custom developed ontology, building a state-of-the-art knowledge graph. The relevant pharmaceutical research databases implemented and installed to use additionally for tagging. PoolParty GraphSearch as server solution with RESTful APIs as integration method and the interface as web-based application to search and browse all data that is part of this project.

Outcome: Data Repositories Around the Globe Accessible Easily and Fast

Regardless of the document or data format, the developed solution ensures now that targets researches for relevant topics are much faster. Search results and the pertaining data of concepts or documents are displayed in place. A quick overview and analytics included in GraphSearch speed up researches as well as collaboration.
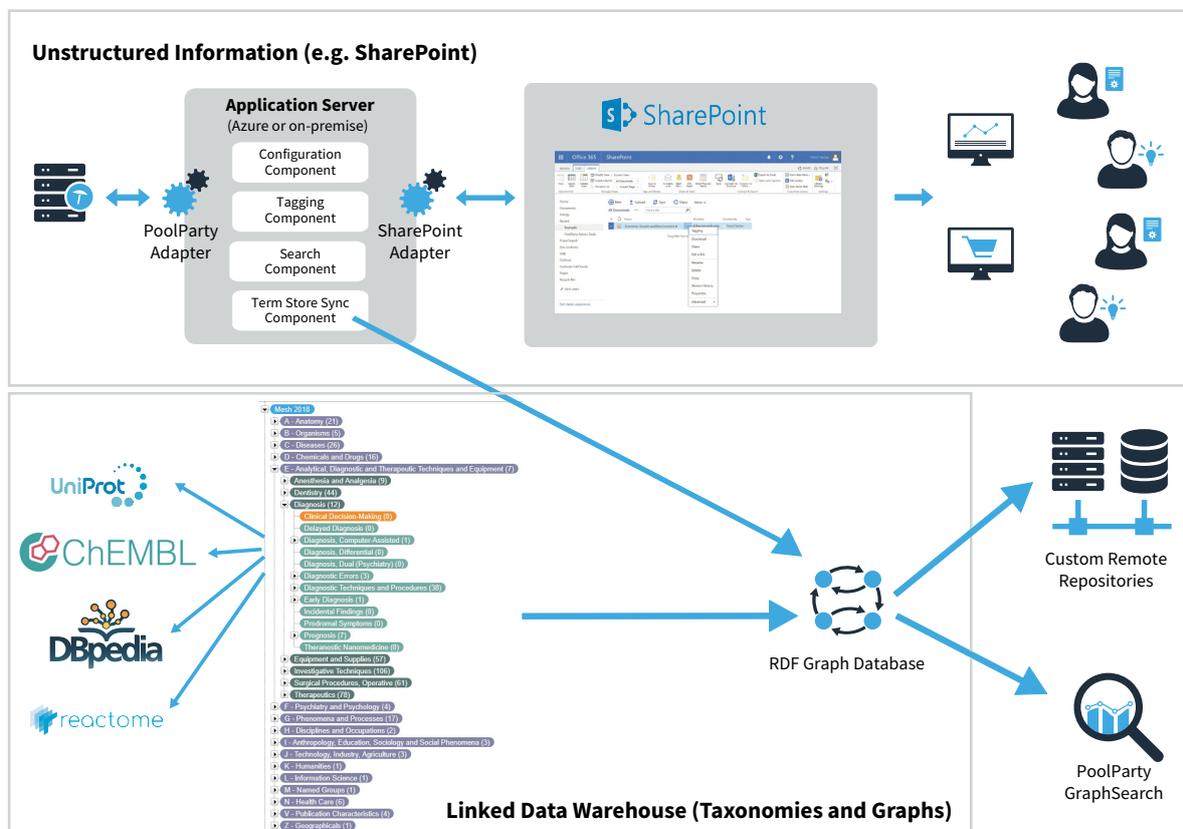
Setup Architecture at Boehringer Ingelheim