



Even better than just in the cloud

Terminology as Linked Open Data

Frederik Pahde (SAP SE)

Christian Lieske (SAP SE)



This presentation was made possible by 

Abstract

Most of today's web-based content lives in isolation. It is hard or even impossible to overcome this. Reasons for the isolation can be found on different levels and range from technical over syntactic to semantic challenges.

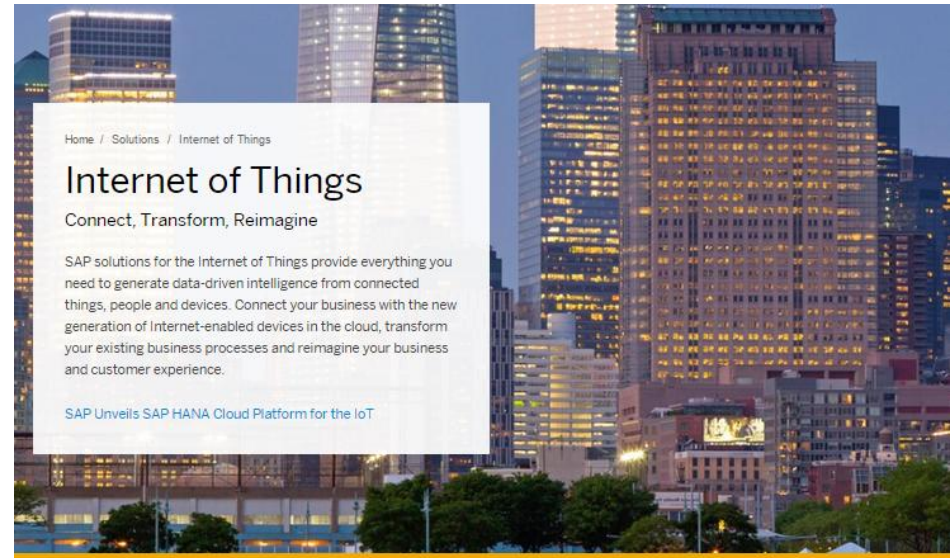
The Semantic Web in general and the Web of Data as one particular flavor provide mechanisms for avoiding challenges that lead to isolated silos of information. The base is a set of technologies that enable machines to interpret information semantically. This enables amongst others to automatically establish relationships between content. In particular in the context of licenses that allow use of content at no cost, interesting new processing opportunities and monetization possibilities arise. Some of them are currently promoted in the keyword "Linked Open Data".

Our presentation explains, how one particular type of data - enterprise terminology - can become a first class citizen of the Semantic Web and possibly even the Linked Open Data movement.

About SAP SE

SAP is the world leader in enterprise applications in terms of software and software-related service revenue. Based on market capitalization, SAP is world's third largest independent software manufacturer.

- More than 291,000 customers in 190 countries
- More than 74,500 employees
- Locations in more than 130 countries [headquarters in Walldorf near Heidelberg]
- A 43-year history of innovation and growth as a true industry leader
- Annual revenue (IFRS) of € 17.56 billion



IOT changes everything

See how technology is now making it possible to connect billions of things to one another to continuously provide more intuitive insights – and help companies revolutionize their business processes.

[Watch the short video](#)

About the Authors

Frederik Pahde

SAP SE



- Finishing his Bachelor of Art (B.A.) in Business Information Systems (focus on Software Engineering) at the Duale Hochschule Baden Württemberg
- Worked in SAP in departments dealing with Liquidity Risk Management, SAP Business ByDemand – Travel OnDemand, PPM Consulting, Custom Development Mobile Design and SAP Language Services

Christian Lieske

SAP SE



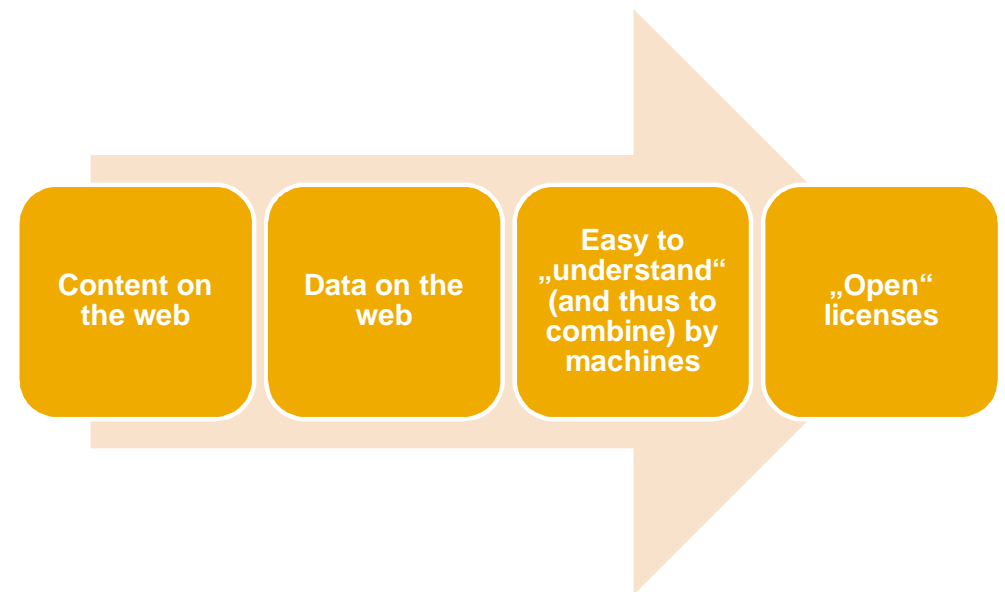
- Knowledge Architect, Globalization Services
- Content engineering and process automation (including evaluation, prototyping and piloting)
- Main fields of interest: internationalization, multilingual content production and Natural Language Processing
- Contributor to standardization at World Wide Web Consortium (W3C; e.g. ITS), OASIS (e.g. XLIFF), Unicode and elsewhere
- Degree in Computer Science with focus on Natural Language Processing and Artificial Intelligence

Overview/Introduction

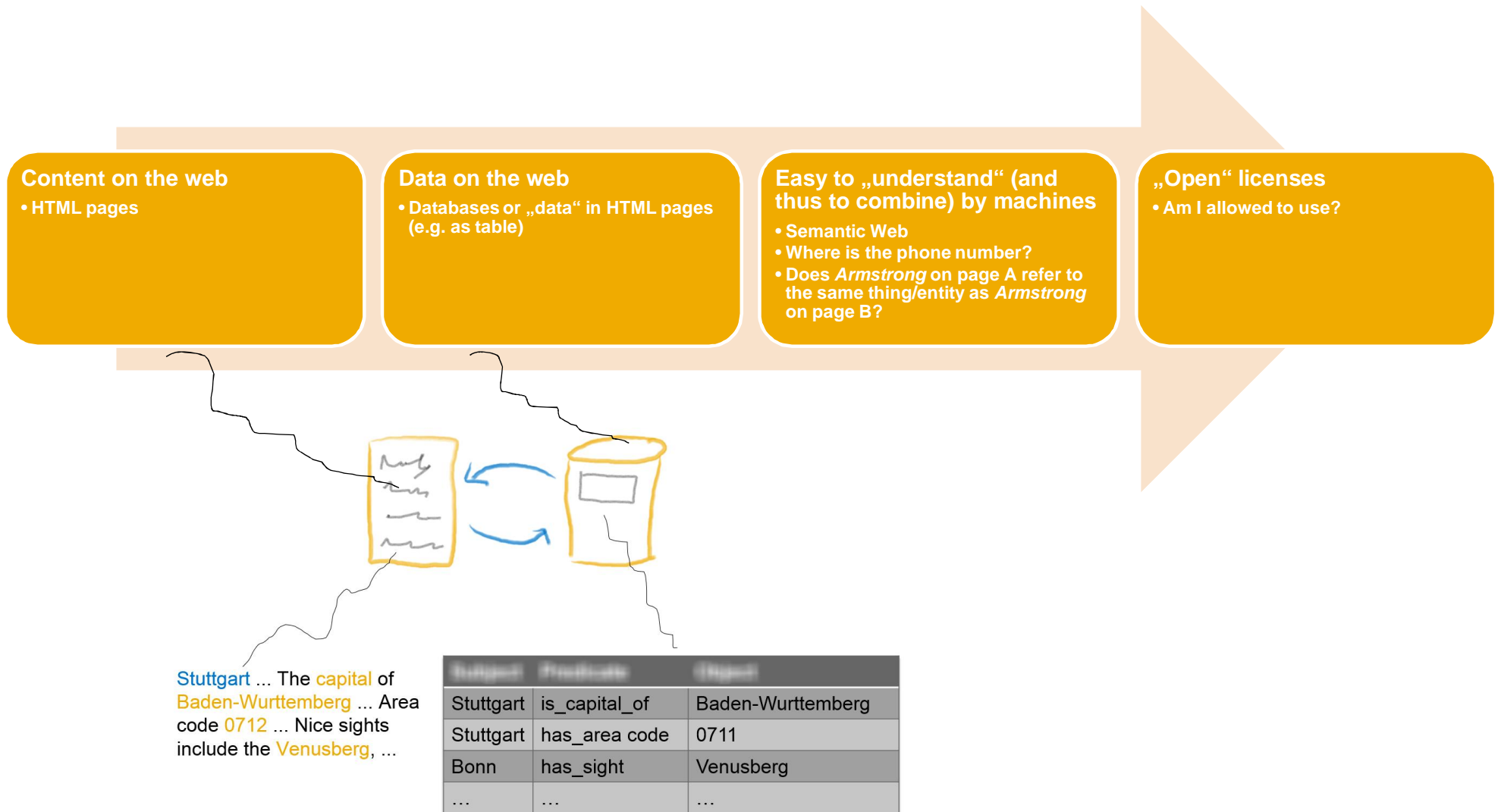
Multilingual Enterprise Assets

- **Terminology databases**
- **Translation memories**
- ...

Linked Open Data



The Evolution of the Web



Data Silos on the Web

Linked Open Data breaks up data silos



Interoperability

Technical

RFC, Webservice, ...

Syntactic

XML, JSON; Part-of-speech, POS,

Semantic

Is A = B?

Sample Opportunities offered by Linked Open Data

Easy creation of resources/assets from scratch

- **All terminology relevant for a subject matter expert**

Automatic enrichment of assets

- **Definitions/explanations for a term**
- **Translation proposal for a term**

Quality control

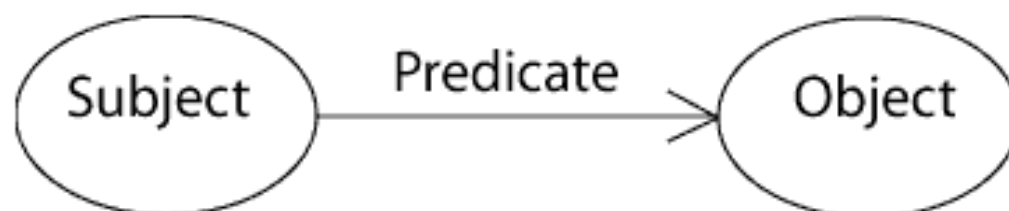
- **Consistency check related to meta data (e.g. part-of-speech)**



Semantic Web technology behind Linked Open Data

Data Representation: Resource Description Format (RDF)

1. Information (about resources) stored as facts/statements



Bob [=subject] is a [=predicate] person [=object]

2. Resources identified via Uniform Resource Identifiers (URIs) – ideally only one URI per resource (i.e. not A in one silo and B in another)

<http://example.org/bob#me>

3. Different (file) formats/serialization languages (N-Triples, Turtle, RDF/XML, ...)

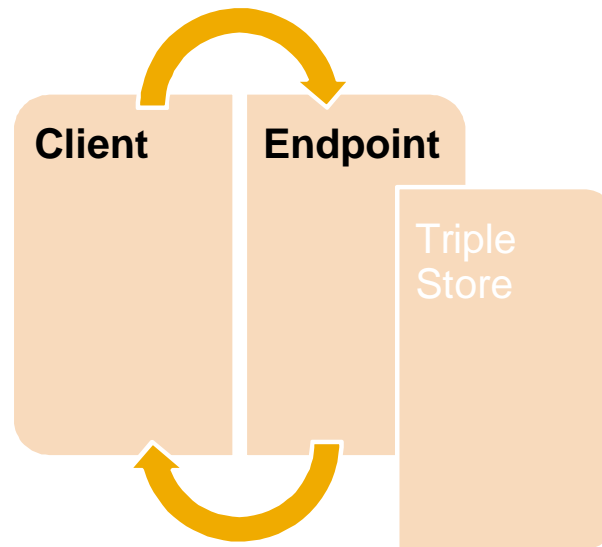
`<http://example.org/bob#me>`

`<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>`

`<http://xmlns.com/foaf/0.1/Person>`

Data Exposure and Access: SPARQL

SPARQL endpoints expose data, triple stores persist data, and SPARQL clients access (or update) data



A special programming/query language is used

```
SELECT ?title WHERE
{ <http://example.org/book/book1>
  <http://purl.org/dc/elements/1.1/title> ?title . }
```



**How to convert
terminological data to RDF?**

TermBase Exchange (TBX)

```
<langSet xml:lang="en">
- <tig>
  <term>progressive lens</term>
  <termNote type="partOfSpeech">noun</termNote>
  <termNote type="administrativeStatus">preferredTerm-a
</tig>
- <tig>
  <term>progressive addition lens</term>
  <termNote type="partOfSpeech">noun</termNote>
  <termNote type="termType">fullForm</termNote>
  <termNote type="administrativeStatus">admittedTerm-a
</tig>
- <tig>
  <term>PAL</term>
  <termNote type="partOfSpeech">noun</termNote>
  <termNote type="termType">acronym</termNote>
  <termNote type="administrativeStatus">deprecatedTerm-admn-sts</termNote>
</tig>
- <tig>
  <term>progressive power lens</term>
  <termNote type="partOfSpeech">noun</termNote>
  <termNote type="administrativeStatus">admittedTerm-admn-sts</termNote>
</tig>
- <tig>
  <term>graduated lens</term>
  <termNote type="partOfSpeech">noun</termNote>
  <termNote type="administrativeStatus">supersededTerm-admn-sts</termNote>
</tig>
</langSet>
```

FI	Financial Accounting
MM	Materials Management
SBO	SAP Business One
SCM-EWM-DLP	Delivery Processing

Verwaltungsinfos	
Originalsprache	DE
Eintragsklasse	Standard Eintrag
Konzept-ID	3526B887AFAB52B9E1000009B38F974
Hinweis zum Eintrag	
Term Kategorie	
Nicht zu übersetzen	Nein
Urheberrechtlich geschützt	Nein
Rechtliche Einschränkung	Nein

Englisch	account
Abkürzung	acct
Abkürzung	ac.
Abkürzung	a/c

Verwaltungsinfos	
Zeitstempel Anlage	27.06.1995 00:00:00
Anleger	ALFORD
Zeitstempel Änderung	01.03.2007 10:37:31
letzter Änderer	HARMAN
Status	3

Definition	
Verwaltungsinfos	

TBX “as/in” RDF

TBX

```
<langSet xml:lang="en">
  <tig>
    <term>progressive lens</term>
    <termNote type="partOfSpeech">noun</termNote>
  </tig>
  <tig>
```

RDF (Turtle)

```
<http://example.com#progressive+power+lens-en>
  a ontolex:LexicalEntry ;
  tbx:partOfSpeech tbx:noun ;
  ontolex:canonicalForm <http://example.com#progressive+power+lens-
en#CanonicalForm> ;
  ontolex:language <http://www.lexvo.org/page/iso639-3/eng> ;
  ontolex:sense <http://example.com#progressive+power+lens-en#Sense> .

<http://example.com#progressive+lens-en#CanonicalForm>
  ontolex:writtenRep "progressive lens"@en .
```



Convert TBX to RDF

The image shows a web browser window on the left and a code editor on the right. The browser window displays the 'TBX2RDF Converter' interface. It has a title bar with the URL 'tbx2rdf.lider-project.eu/converter'. Below the title, there are several input fields: 'TBX Document:' with a 'Choose File' button and 'No file chosen' text; 'Resource URI (where you intend to publish the RDF document):' with the value 'http://example.com#' highlighted; and 'Extra mappings:' with an empty text area. A 'Submit' button is located at the bottom left of the form.

The code editor on the right shows the source code of a TBX file named 'TBXbasic.tbx'. The code is XML-based and includes the following elements:

```
<?xml version='1.0'?>
<martif type="TBX-Basic" xml:lang="en-US">
  <martifHeader>
    <fileDesc>
      <titleStmt>
        <title>TBX-Basic Sample File</title>
        <note>This document is a sample TBX-Basic document instance
          showing various types of
          terminological entries. The entries in this file are for
          demonstration purposes
          only and do not reflect actual terminology data. Any
          references to real
          companies are fabricated for demonstration purposes only.</
          note>
      </titleStmt>
      <sourceDesc>
        <p>This is a sample TBX-Basic file from the LISA Terminology
          Special Interest Group
          (www.lisa.org/term). Address any enquiries to kara@ca.ibm.
          com.</p>
      </sourceDesc>
    </fileDesc>
    <encodingDesc>
      <p type="XCSURI">http://www.lisa.org/fileadmin/standards/tbx_basic/
        TBXBasicXCSV02.xcs
      </p>
    </encodingDesc>
  </martifHeader>
  <text>
    <body>
      <termEntry id="c7">
        <descripGrp>
          <descrip type="definition">corrective lenses used in
            eyeglasses to correct
            presbyopia and other disorders of accommodation,
            characterised by a gradient
            of increasing lens power, added to the wearer's
            correction for the other
            refractive errors.</descrip>
          <admin type="source">Wikipedia</admin>
        </descripGrp>
      </termEntry>
    </body>
  </text>
</martif>
</xml>
```

The code editor shows line numbers from 1 to 31. The status bar at the bottom indicates 'Line 75, Column 25', 'Tab Size: 4', and 'XML'.



**How to convert and expose
SAPterm data as RDF?**

SAPterm

The screenshot shows the SAPterm website homepage. At the top, there is a navigation bar with the SAP logo and tagline "The Best-Run Businesses Run SAP". To the right of the logo, there are links for "United States", "Free Trials", "Your SAP.com", and a phone number "+1 800-872-1727". A search bar is also present. Below the navigation bar, there is a main content area with a large heading "Welcome to the SAP terminology database!". Underneath this heading, there is a paragraph of text and a link "Look up SAP terminology >". To the right of the text is a large image of a cable-stayed bridge. Below the main content area, there are four columns of featured content, each with a small image, a title, a short description, and a link.

www.sapterm.com

Apps For quick access, place your bookmarks here on the bookmarks bar. [Import bookmarks now...](#) Other bookm

SAP The Best-Run Businesses Run SAP

United States Free Trials Your SAP.com
+1 800-872-1727 Store

Search

SAPterm

Look up terminology > SAPterm.com Help >
SAP Terminology Community > SAP Help Portal >
SAP Glossary >


SAP.com Terminology@SAP SAPterm

Welcome to the SAP terminology database!

This site gives you access to thousands of terminology entries at SAP.

To begin your search, click on the link below. Enter the term you are searching for in the Search For field. Choose Search to start the search. For more detailed help, please see Additional Resources below.

[Look up SAP terminology >](#)



SAPterm.com Help

Get additional help on how to browse the SAP terminology database.

[View the help now >](#)

SAP Terminology Community
(part of Globalization Services community)

Connect with other terminology enthusiasts to learn, share and network.

[Join the community >](#)

SAP Help Portal

Browse SAP Help Portal to get up-to-date information about our products.

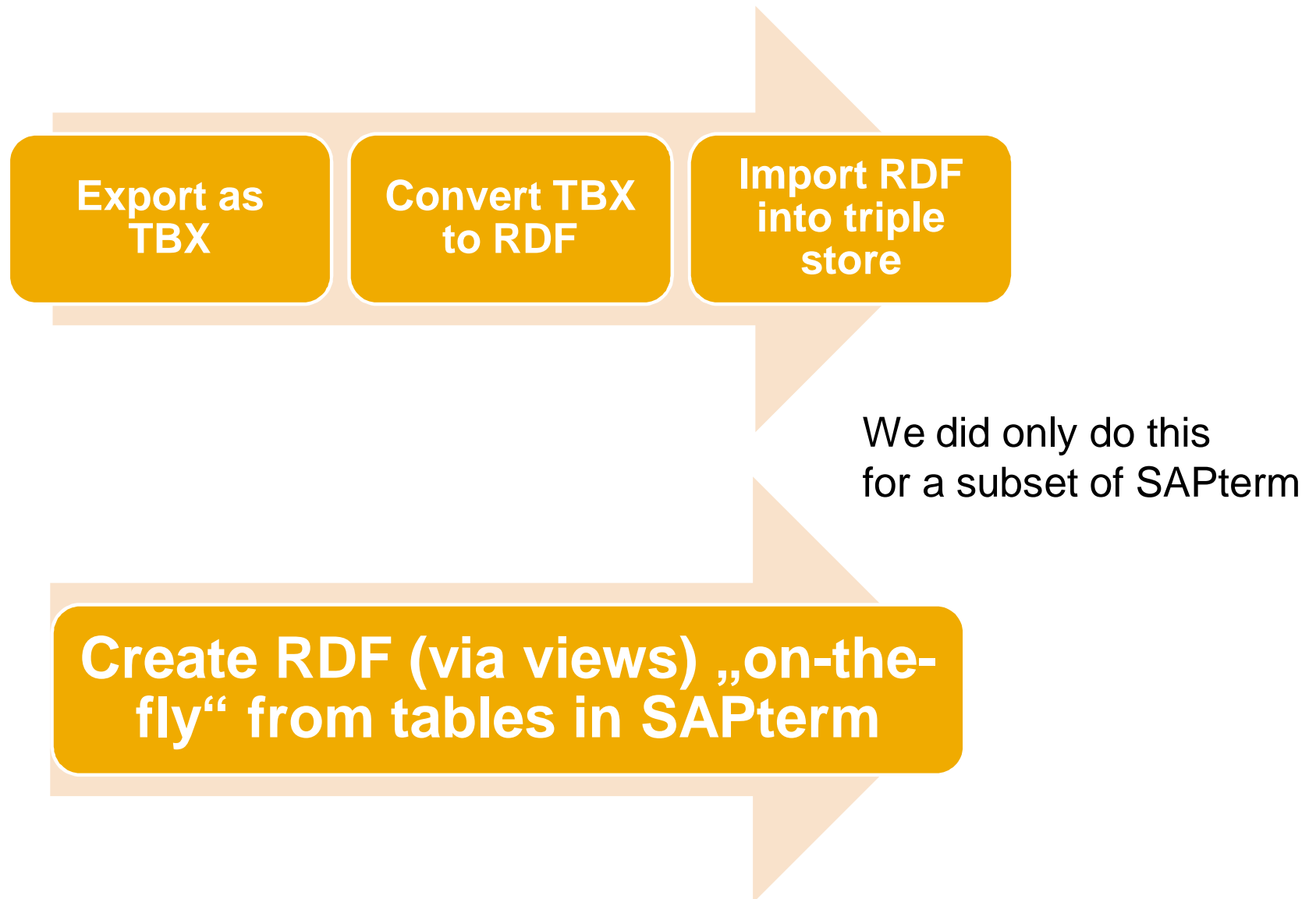
[Browse SAP Help Portal >](#)

SAP Glossary

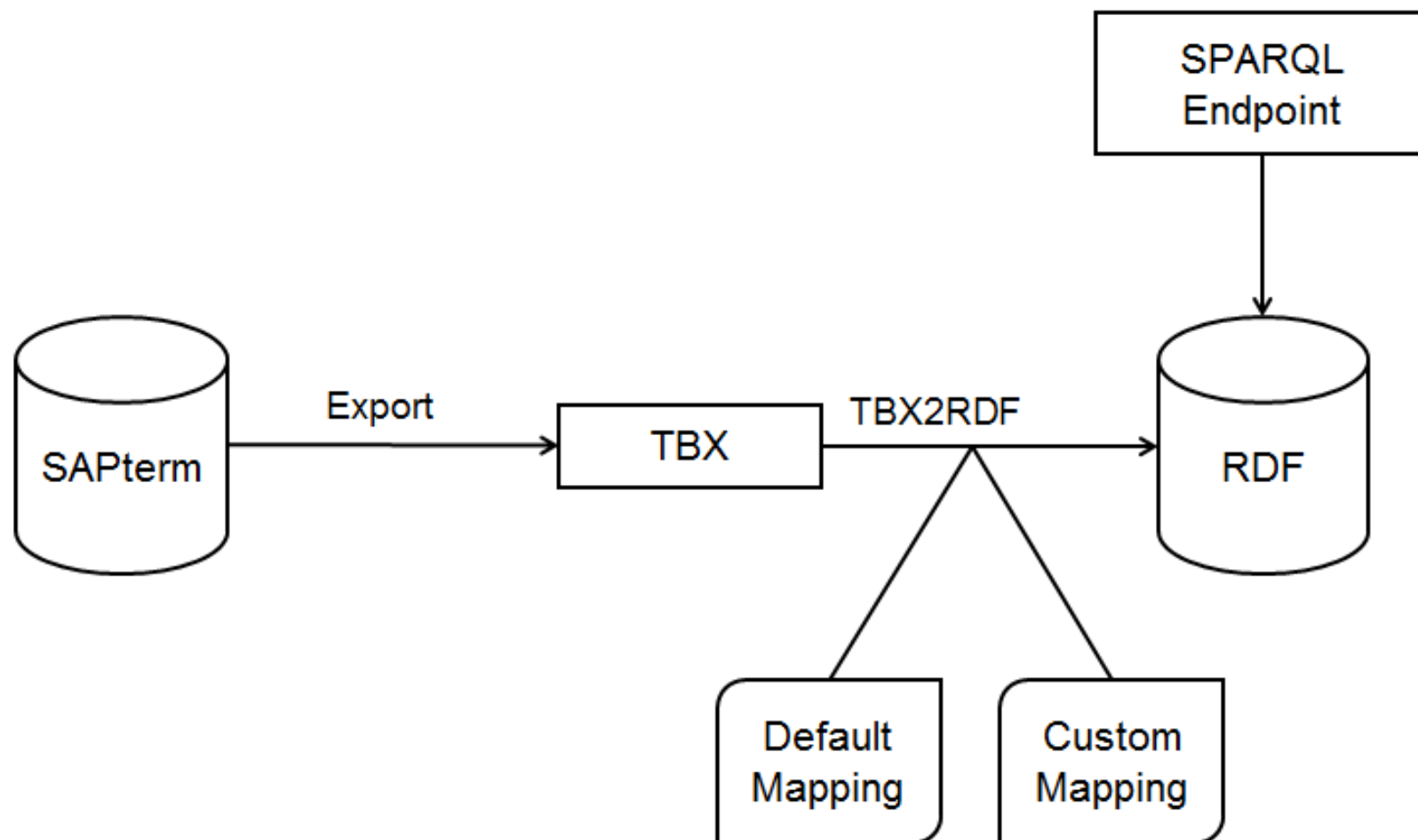
Access a glossary with SAP-specific technical and business terms and their definitions in English and German.

[View SAP Glossary >](#)

Two Approaches



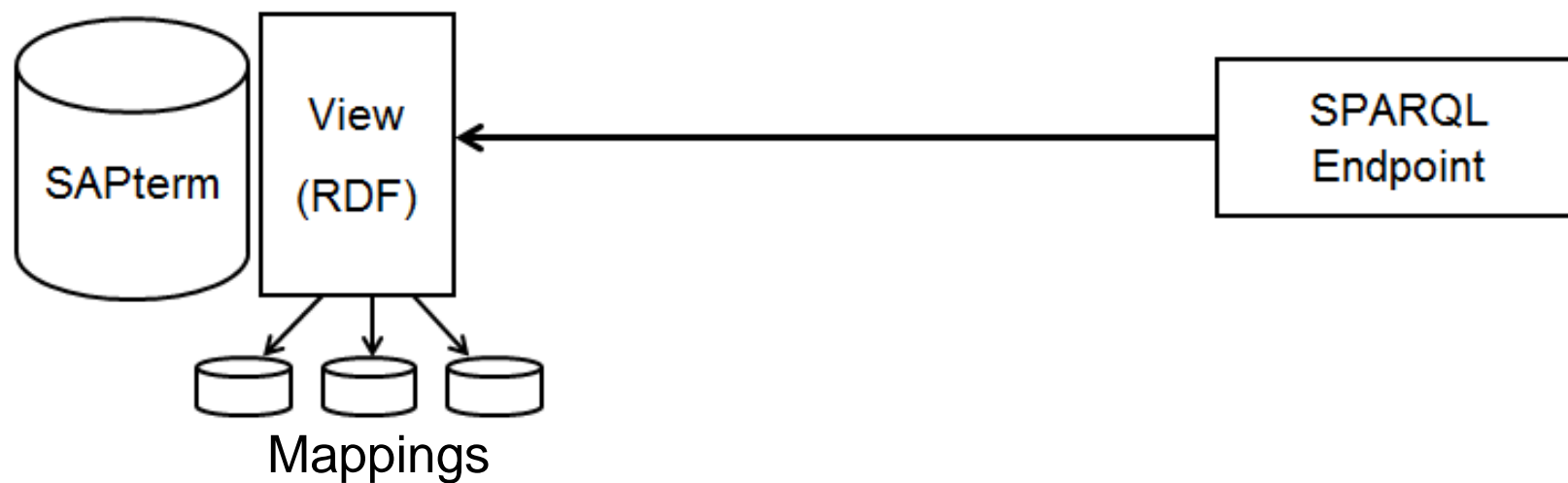
Approach 1: Conversion via TBX (1/2)



Approach 1: Conversion via TBX (2/2)

```
<?xml version="1.0" encoding="utf-8" standalone="no"?><martif type="TBX" xml:lang="en">
  <martifHeader>
    <fileDesc>
      <titleStmt>
        <title>SAPterm</title>
      </titleStmt>
      <publicationStmt>
        <p>Distributor=SAP Language Services contact=SLS-TERM Company=SAP AG
          URL=http://www.sap.com/company/legal/impressum.epx Region=world
          PubStatus=unknown DateValue=March 2010</p>
      </publicationStmt>
      <sourceDesc>
        <p>CreaTool=SAPterm CreaToolVersion= 740 OrigFormat=R3-internal
          AdminLang=en CreaDate= 20150211073900 CreaId= LIESKE</p>
      </sourceDesc>
    </fileDesc>
    <encodingDesc>
      <p type="DCSName">sapTEXXCSV02.xcs</p>
    </encodingDesc>
  </martifHeader>
  <text>
    <body>
      <termEntry id="tid_db6_014D420D507ED411B1360060B03C6BFB">
        <descripGrp>
          <descrip type="definition"/>
          <transacGrp>
            <transac type="transactionType">origination</transac>
          </transacGrp>
          <transacGrp>
            <transac type="transactionType">modification</transac>
          </transacGrp>
        </descripGrp>
        <transacGrp>
          <transac type="transactionType">origination</transac>
          <transacNote type="responsibility">https://people.wdf.sap.corp/#/?query=sapname:S5_NGUYENK</transacNote>
          <date>20080307024400</date>
        </transacGrp>
        <transacGrp>
          <transac type="transactionType">modification</transac>
          <transacNote type="responsibility">https://people.wdf.sap.corp/#/?query=sapname:S5_PHAM</transacNote>
          <date>20080407020448</date>
        </transacGrp>
        <admin type="conceptOrigin">de</admin>
        <descrip type="characteristic">standardEntry</descrip>
        <admin type="productSubset">BC</admin>
      </termEntry>
    </body>
  </text>
</martif>
```

Approach 2: On-the-fly conversion via Views (1/2)



Concept 2: On-the-fly conversion via Views (2/2)

Snorql: Exploring <http://qe-dev:6789/D059348/sparql>

SPARQL:

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX dc: <http://purl.org/dc/elements/1.1/>
PREFIX dcterms: <http://purl.org/dc/terms/>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX sim: <http://purl.org/ontology/similarity/>
PREFIX mo: <http://purl.org/ontology/mo/>
PREFIX ov: <http://open.vocab.org/terms/>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
```

```
SELECT ?s ?o
from <http://www.sap.com/dataset/sapterm/Term>
WHERE {
  ?s <http://www.w3.org/ns/lemon/ontolex#WrittenRep> ?o
}
LIMIT 100
```

Results:

Browse:

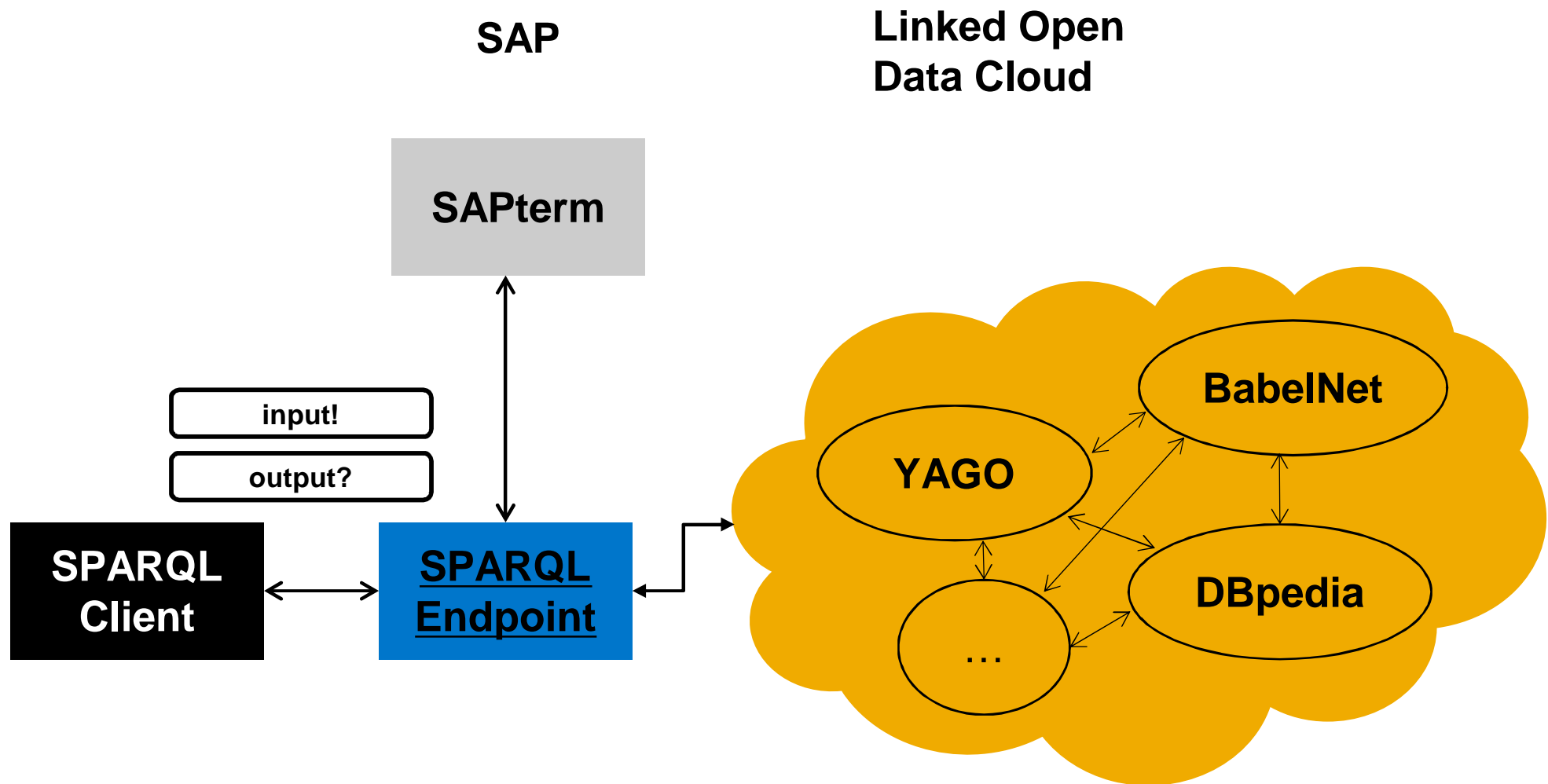
- [Super Classes](#)
- [Classes](#)
- [Properties](#)

Powered by [DOR Saver](#)

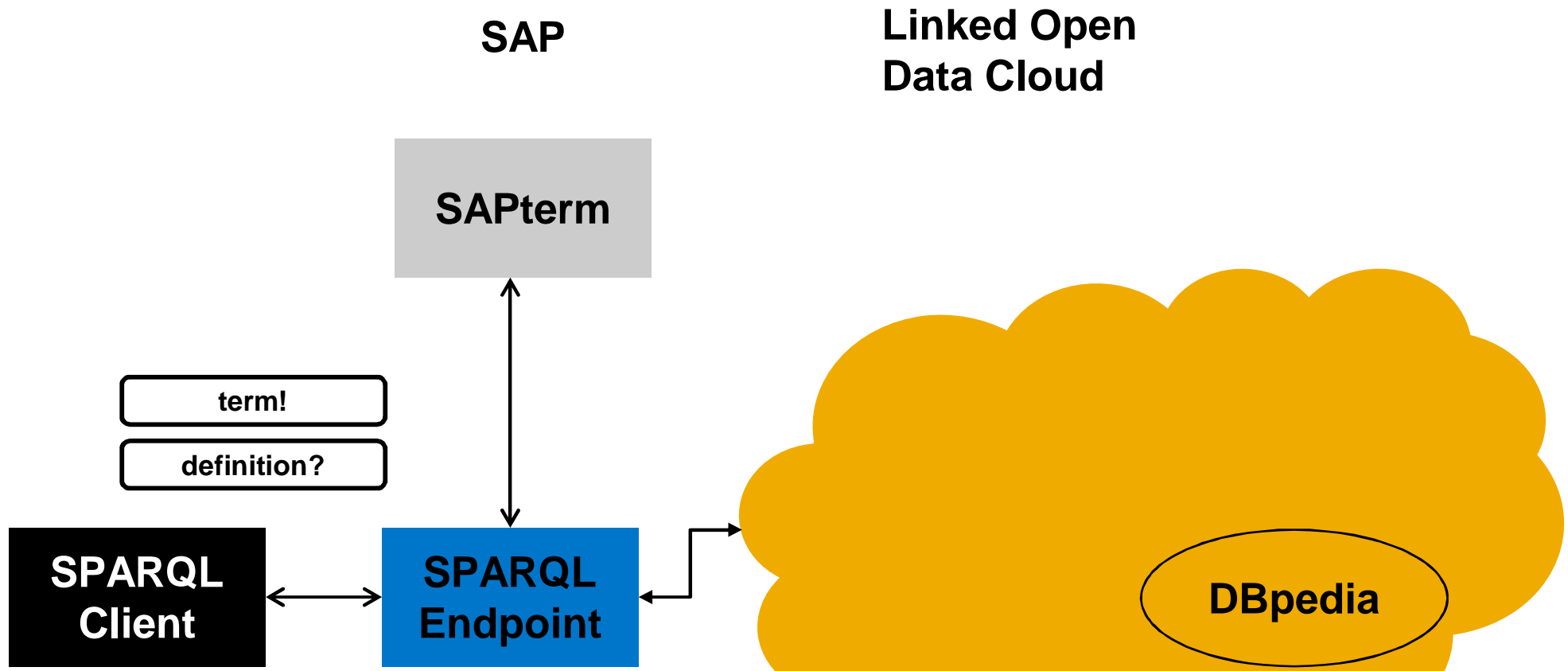


Show me the cool stuff

Queries to Interconnected Knowledge Sources



Example 1: Enrichment (1/2)



writtenRep	definition
"Datenerfassung"@de	"Datenerfassung bezeichnet alle Methoden der zeitgleichen oder zeitfolgerichtigen Messungen und Zählungen, gegebenenfalls einschließlichem Zeitstempel für messbare oder zählbare Daten und Gruppen von zusammenhängenden Daten. Charakteristisch ist die unmittelbare und direkte Zugänglichkeit der Messgröße oder einer damit physikalisch eindeutig verbundenen Ersatzgröße. [..]"@de

Example 1: Enrichment (2/2)

Virtuoso SPARQL Query Editor

[About](#) | [Namespace Prefixes](#) | [Inference rules](#)

Default Data Set Name (Graph IRI)

Query Text

```
SELECT ?writtenRep, ?definition WHERE {
  ?term <http://www.w3.org/ns/lemon/ontolex#writtenRep> "Datenerfassung"@de.
  ?term <http://www.w3.org/ns/lemon/ontolex#writtenRep> ?writtenRep.

  SERVICE <http://dbpedia.org/sparql> {
    ?dbpediaEntry <http://www.w3.org/2000/01/rdf-schema#label> ?writtenRep.
    ?dbpediaEntry <http://dbpedia.org/ontology/abstract> ?definition.
    FILTER(langMatches(lang(?definition), "de"))
  }
}
```

Sponging:

Results Format:

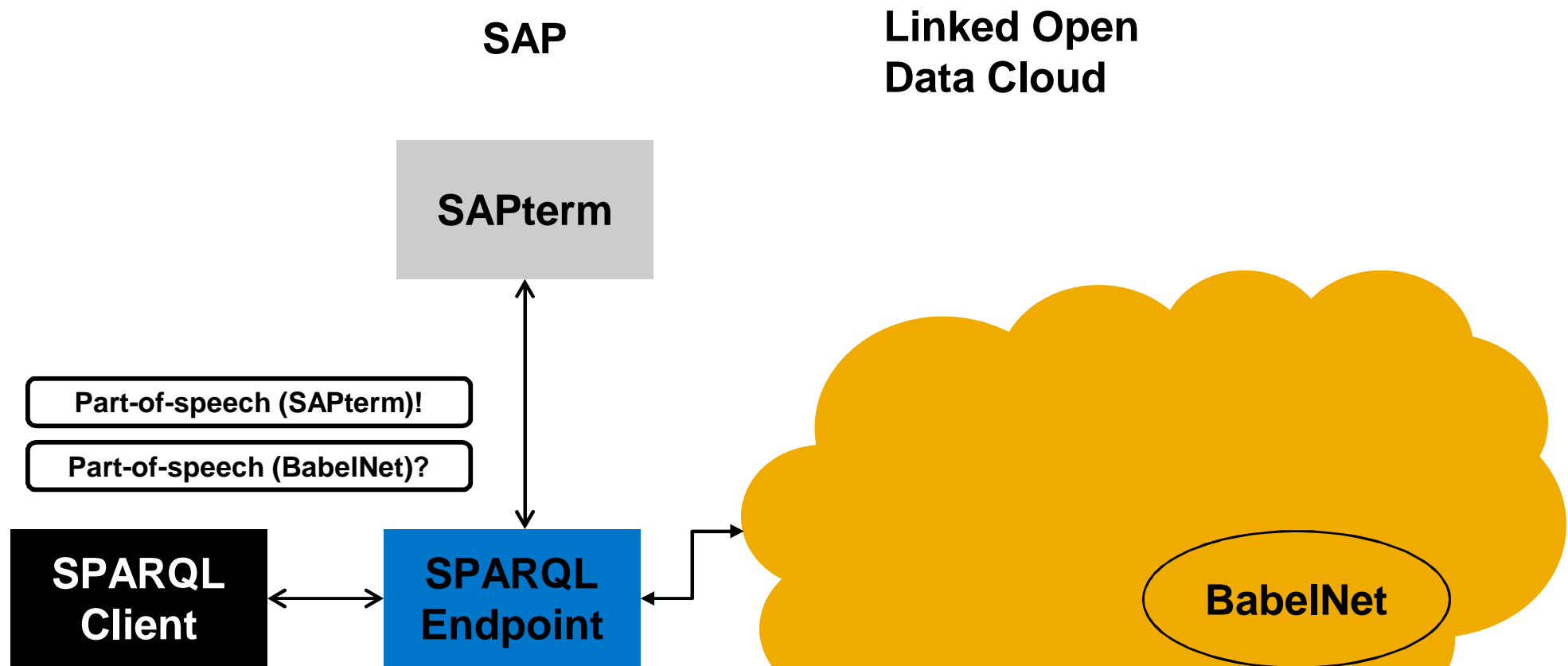
Execution timeout: milliseconds (values less than 1000 are ignored)

Options: Strict checking of void variables

(The result can only be sent back to browser, not saved on the server, see [details](#))

Copyright © 2015 [OpenLink Software](#)
Virtuoso version 07.10.3211 on Win64 (x86_64-generic-win-64), Single Server Edition

Example 2: Quality Control



writtenRep	posSAPterm	posBabelNet
"Datenerfassung"@de	http://tbx2rdf.lider-project.eu/tbx#noun	http://www.lexinfo.net/ontology/2.0/lexinfo#noun

Uncool: Deviation from „Only one URI“ Principle

writtenRep	posSApTerm	posBabelNet
"Datenerfassung"@de	http://tbx2rdf.lider-project.eu/tbx#noun	http://www.lexinfo.net/ontology/2.0/lexinfo#noun

```
SELECT ?writtenRep, ?definition WHERE {
  ?term <http://www.w3.org/ns/lemon/ontolex#writtenRep> "Datenerfassung"@de
  .
  ?term <http://www.w3.org/ns/lemon/ontolex#writtenRep> ?writtenRep.

  SERVICE <http://dbpedia.org/sparql> {
    ?dbpediaEntry <http://www.w3.org/2000/01/rdf-schema#label> ?writtenRep.
    ?dbpediaEntry <http://dbpedia.org/ontology/abstract> ?definition.
    FILTER(langMatches(lang(?definition), "de"))
  }
}
```

Summary/Conclusion

New possibilities for multilingual assets such as terminology with Linked Open Data

- Enrichment
- Quality Control
- ...

Projects such as LIDER do nice work

We can get terminology into the Linked Open Data world

- TBX is a good starting point

Some stuff is still uncool

- Maturity of prototypes
- Deviation from „Only one URI“ Principle

Interesting Touchpoints

W3C LD4LT (Linked Data for Language Technology) Community Group (CG)

- <http://www.w3.org/community/ld4lt/>
- Anyone can join 😊

W3C Best Practices for Multilingual Linked Open Data CG

- <https://www.w3.org/community/bpmlod/>
- Anyone can ... 😊

W3C Ontology-Lexica CG

- <https://www.w3.org/community/ontolex/>
- Anyone ... 😊

Linked Data as an enabler of cross-media and multilingual content analytics for enterprises across Europe (Lider)

- <http://www.lider-project.eu/>

mlider

Open framework of e-services for multilingual and semantic enrichment of digital content (FREME)

- <http://www.freme-project.eu/>

FREME



Thank you

Contact information:

frederik.pahde@sap.com

christian.lieske@sap.com