



Leveraging Big Data, Linked (Open) Data and (Multilingual) Language Technology Standards in Global Enterprises



MultilingualWeb-LT Workshop
June 11-13, 2012
Dublin, Ireland

Prof. Dr. Jörg Schütz
bioloom group

Today's Agenda

Overview

- *Application scenario*
- *Needs and Requirements*

Challenges

- *Data, Processes and Workflows*
- *Tools, Curation and Management*

Solutions

- *Standards*
- *Existing Gaps*
- *Proprietary Solution*
- *Future(s)*

Enterprise related Data

RDB
Data
(internal)

Multiple
Data
Streams
(external)

Data and Meaning

Semantics

Metadata

Storage / Tools

Vocabularies

Processing

Needs and Requirements

Social Media

- Identify
- Extract
- Analyze
- Categorize
- Channel

Core Language

- Assure Quality
- Curate Rules, Styles, Meta and Vocabularies
- Monitor and Optimize

Corporate Languages

- Assure Quality
- Curate Rules, Styles, Meta, TMs and Vocabularies
- Monitor and Optimize

 **Dynamic Streams**

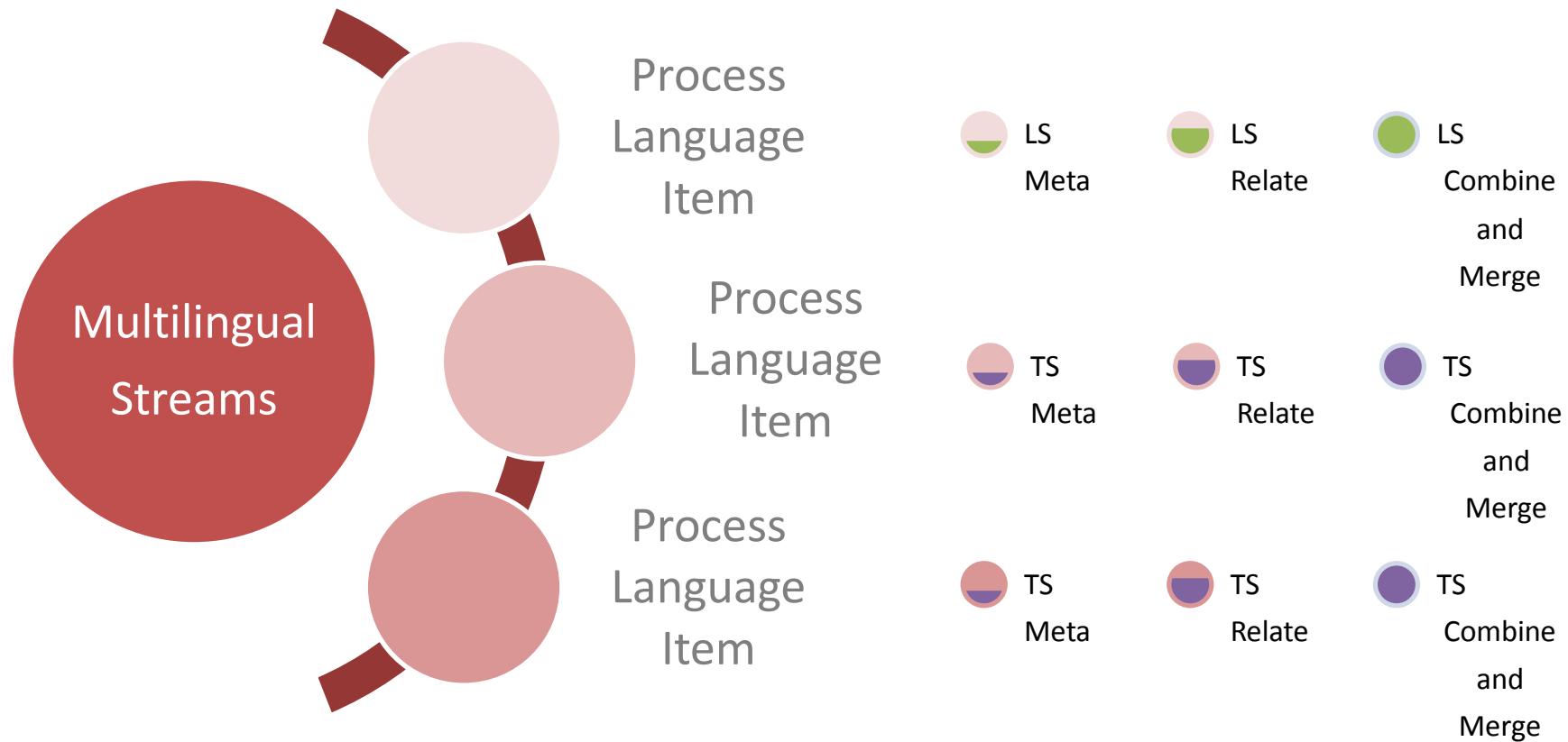
Multiple Languages



Entailed Knowledge

Relate to Curated Repositories

Data, Processes and Workflows



Tools, Curation and Management

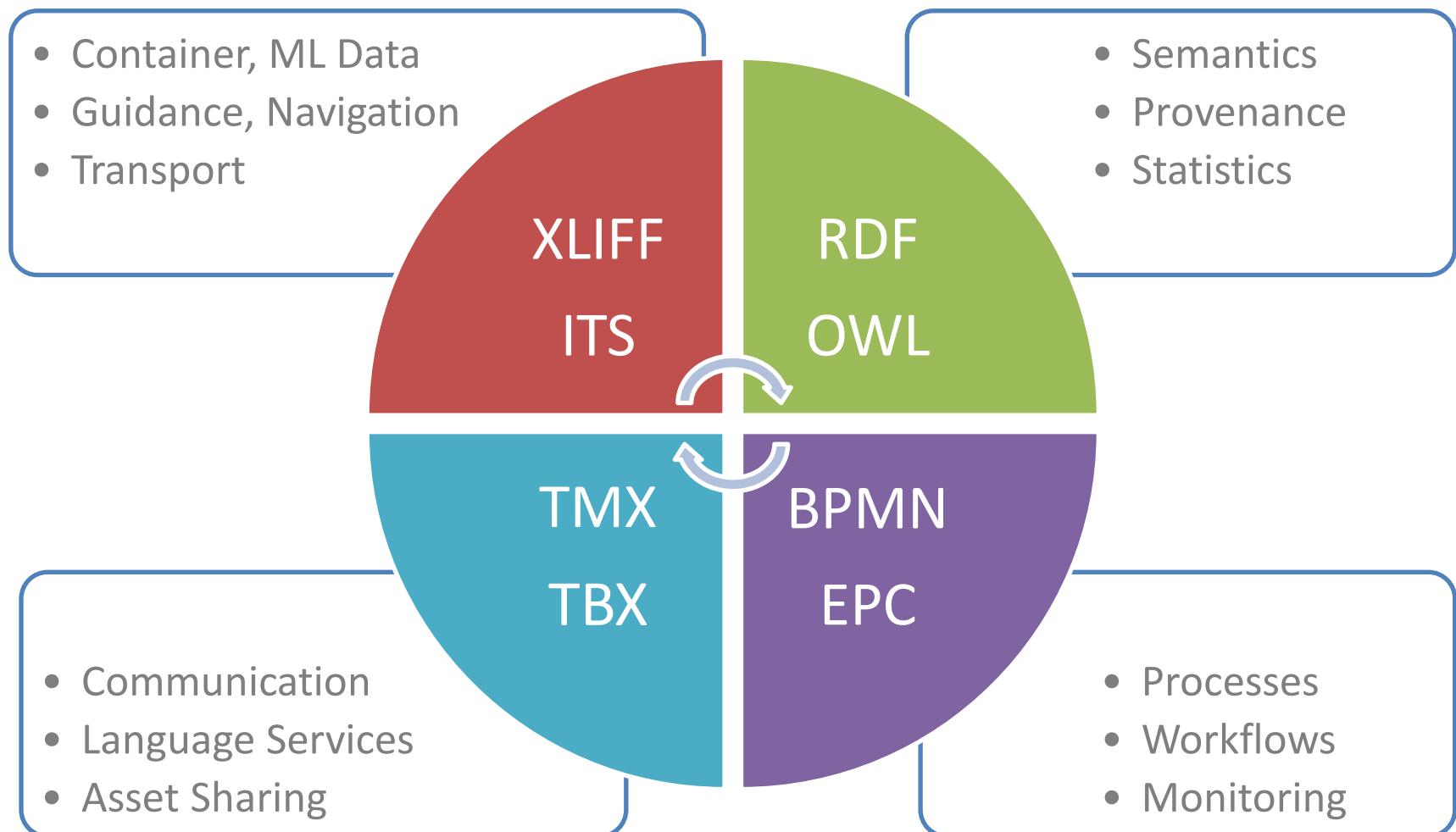
Corporate Communication

- Relational Database Systems
- Content Management
- BPM (BPN, EPC, ...)
- IT Governance and Compliance
- ...

Language Technologies

- Terminology Database Systems
- Translation Memories
- Checking Tools
- Crawler, Parser, ..., Machine Translation
- ...

Combining Standards



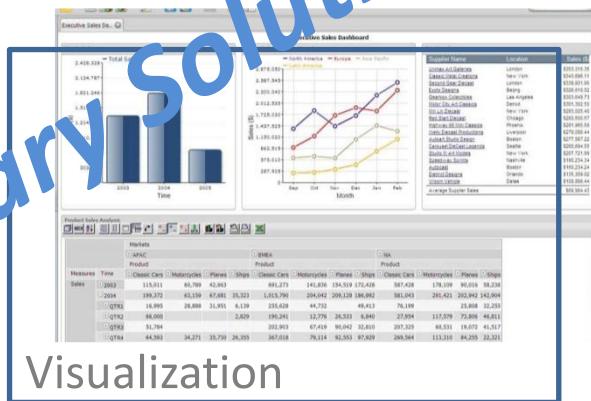
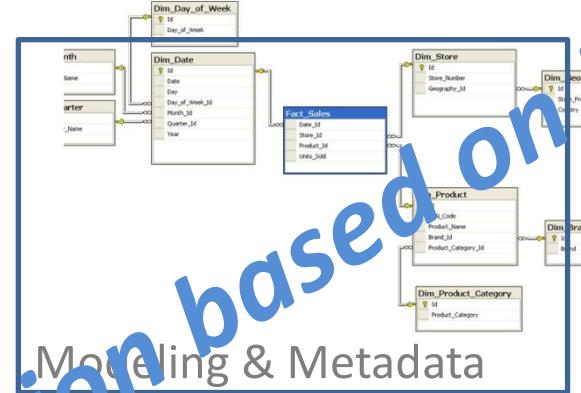
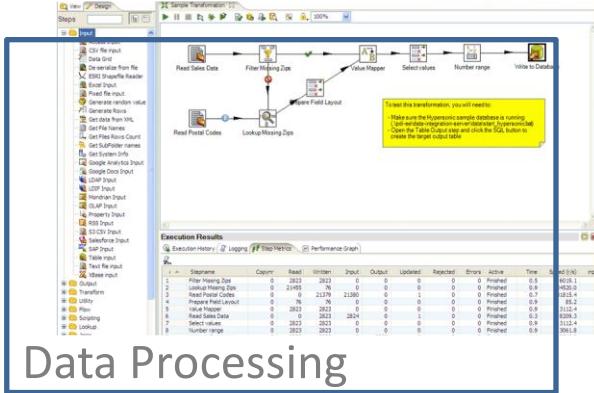
Existing Gaps ...

Too complex, not intuitive, not fitting, ...

Standards (partly) immature, too leaky, too flexible,
too error-prone, too non-existent, ...

General lack of interoperability

How everything fits together

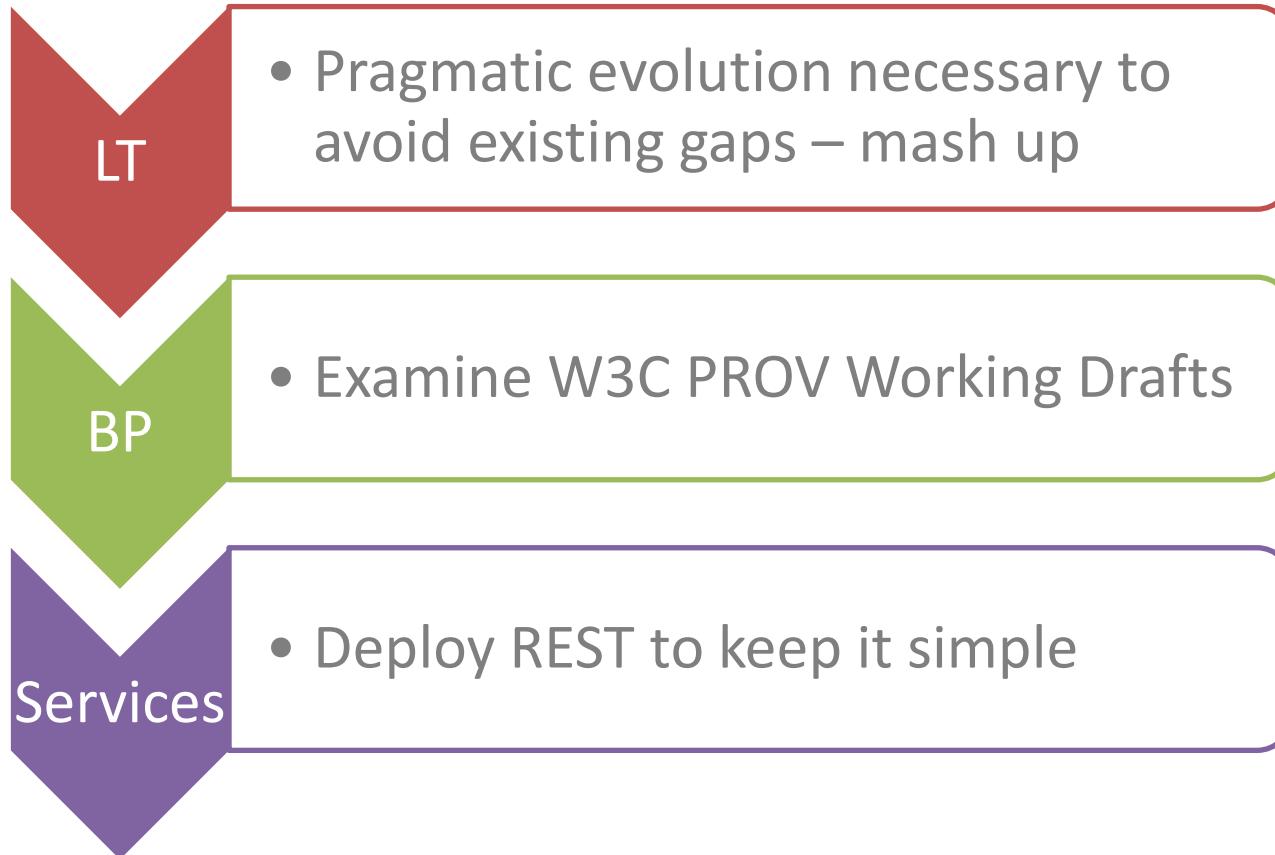


The image displays three distinct software interfaces arranged in a grid, each representing a different aspect of a data solution:

- Data Processing:** A screenshot of a data transformation tool's interface. It shows a flowchart with various steps: "Read Sales Data", "Filter by Date", "Value Mapper", "Selected values", "Number range", "Table to Database", "Read Postal Codes", and "Lookup/Missing Zip". A yellow box contains instructions: "Execute this transformation, you will need: Make sure the Hyperion's sample database is running. Open the Table Catalog step and click the SQL button to choose the sample catalog." Below the flowchart is a table titled "Execution Results" showing the progress of 8 steps.
- Modeling & Metadata:** A screenshot of a data modeling tool showing a star schema. The central fact table is "Fact_Sales" with dimensions "Dim_Date", "Dim_Store", "Dim_Prod", and "Dim_PCat". Other dimension tables include "Dim_Day_of_Week", "Dim_Market", "Dim_Geo", "Dim_Brand", and "Dim_Prod_Cat". Relationships between tables are indicated by arrows.
- Visualization:** A screenshot of a dashboard titled "Executive Sales Dashboard". It features three main components: a bar chart showing "Total Sales" over "Time" (2003, 2004, 2005), a line chart showing "Sales" over "Month" (Jan, Feb, Mar, Apr, May, Jun, Jul, Aug, Sep, Oct, Nov, Dec, Jan, Feb, Mar), and a table titled "Customer Sales" listing sales data for various regions and products.

A large blue watermark reading "Proprietary Solution based on RDF" is diagonally overlaid across the entire image.

Extend / Refit or Mash up ?



Enterprise Data Web

RDB
Data
(internal)

Multiple
Data
Streams
(external)

Linked Enterprise Data

RDF / RDFS

Provenance

Triple Stores

Vocabularies
SKOS, OWL

SPARQL, Inferencing

Recommendations

KISS Principle

Re-use and Combine

Show how it works

Word Wide Web

Enterprise Data Web

RDB
Data
(internal)

Multiple
Data
Streams
(external)

Linked Enterprise Data

RDF / RDFS

Provenance

Triple Stores

Vocabularies
SKOS, OWL

SPARQL, Inferencing

HOW STANDARDS PROLIFERATE: (SEE: A/C CHARGERS, CHARACTER ENCODINGS, INSTANT MESSAGING, ETC)

SITUATION:
THERE ARE
14 COMPETING
STANDARDS.

14?! RIDICULOUS!
WE NEED TO DEVELOP
ONE UNIVERSAL STANDARD
THAT COVERS EVERYONE'S
USE CASES.



SOON:

SITUATION:
THERE ARE
15 COMPETING
STANDARDS.

Thanks for listening!

Additional Info: joerg@bioloom.de