Subscribe Now

From RDF to RSS and Atom

Content Syndication with Linked Data

Hypertext 2013, May 1-3, 2013, Paris, France

Alex Stolz and **Martin Hepp**
{firstname.lastname}@unibw.de

E-Business and Web Science Research Group
Universität der Bundeswehr Munich, Germany

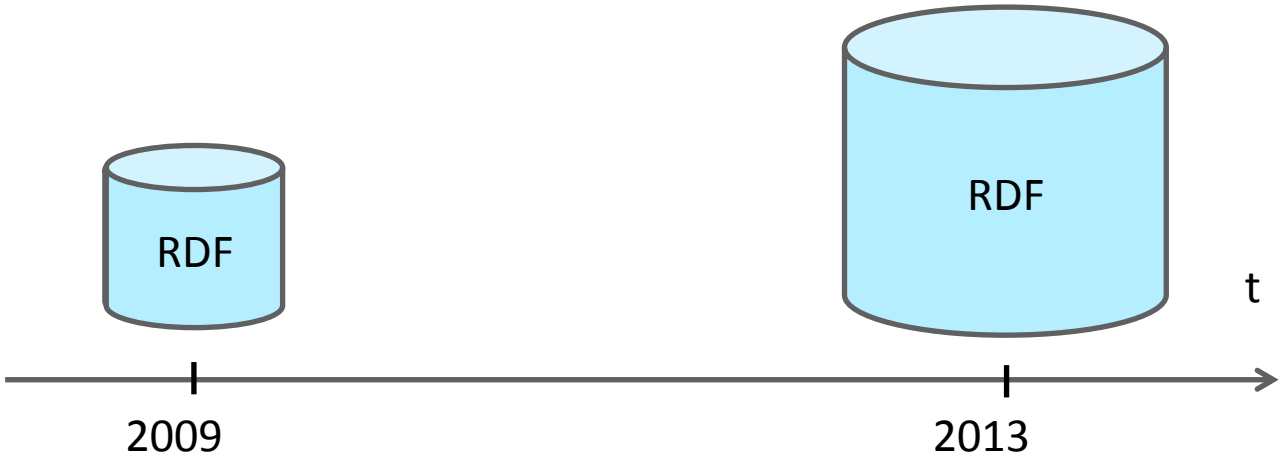
der Bundeswehr
Universität München

Glossary

- **RDF** – Resource Description Framework, a data model consisting of *subject-predicate-object* triples
- **SPARQL endpoint** – Service that provides accessibility to RDF data using SPARQL queries
- **Linked Data** – Interlinked RDF graphs on the Web, built upon URIs, HTTP and RDF
- **Content Syndication** – Publication of page feeds in standardized feed formats, easy to be consumed and integrated by other sites and applications
- **RSS and Atom/ASF** – Popular data feed formats

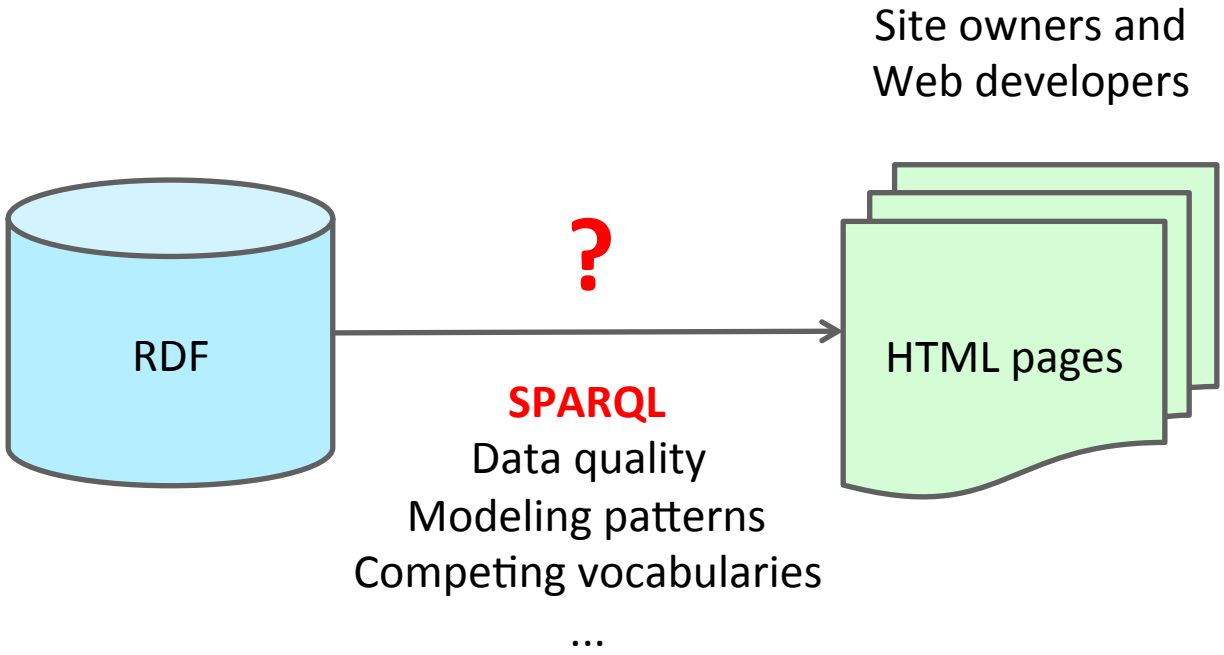
Motivation

1. Recent growth of RDF data



Motivation

2. RDF consumption barrier



Motivation

3. Popularity of Content Syndication on the Web

Provides incentives for both feed publishers and feed subscribers!!

Incentives of Using Data Feeds

<u>Publishers</u>

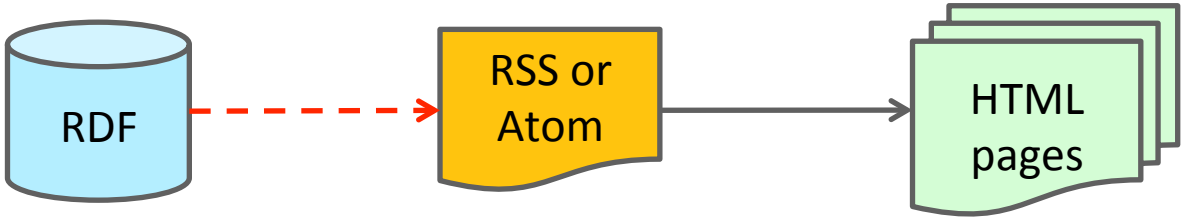
Increase reach of their content

<u>Consumers</u>

Augment their pages by fresh, quality content

Our Contribution

- Use standardized data formats as carrier to leverage Linked Data (RDF data) consumption



- Three important benefits:

  a. Simplicity and broad tooling support of existing data feed formats

  b. Precision of queries against structured data

  c. Ease of integrating content from a large number of Web sites and other sources of RDF in general

Our Approach

Product or Store search

Query and transform

```
<?xml version="1.0" encoding="utf-8"?>
<rss version="2.0">
  <channel>
...
  </channel>
</rss>
```

RSS or Atom feed integration

Affiliate Links?

Sony 4.15MP Handycam High-Definition Camcorder w/120
5 minutes ago
Price: USD 1199.99This high-definition camcorder features x.v. Colo

GoodRelations Product Search :: "camcorder"

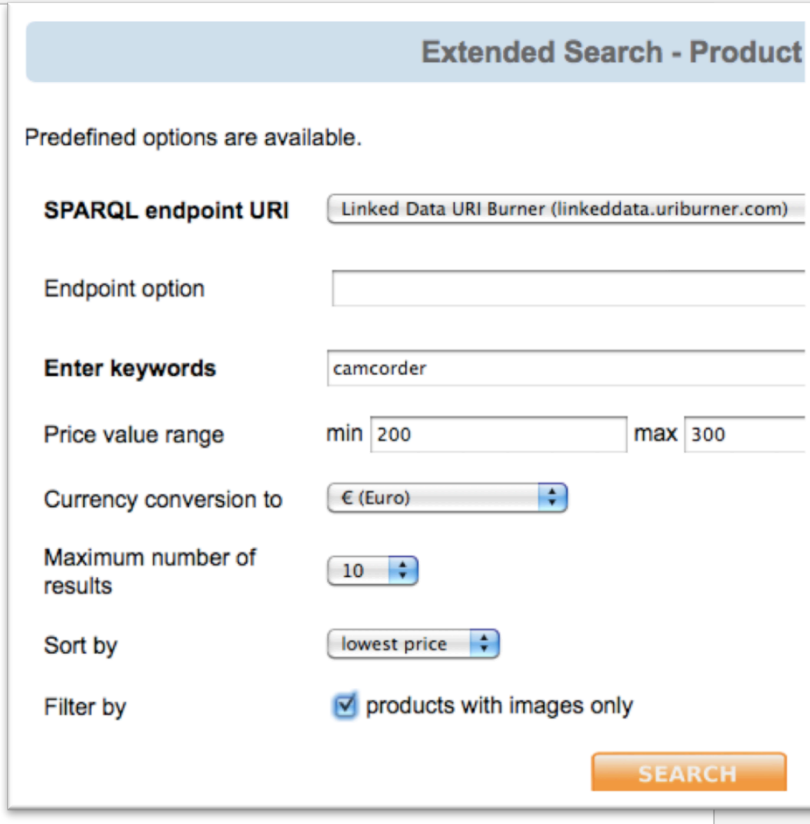
Sony High-Definition MiniDV Handycam Camcorder with 3.2" Widesc
Sony 4.15MP Handycam High-Definition Camcorder w/120GB Hard
JVC Everio X High-Definition Digital Camcorder with 2.8" LCD Monit
Panasonic 2.1MP High-Definition Digital Camcorder w/60GB Hard D

GoodRelations Product Search :: "lcd tv"

Audiovox 9" Class LCD Flat-Panel TV VE927
Sanus VisionMount Full-Motion Mount for 23" - 40" TVs - Extends 9-
Altra Oval TV Stand for Flat-Panel TVs Up to 47" 1126096
Dynex™ LCD Screen Cleaning System DX-CLNKITCL

Query Builder

- Two search types:
  - *Products*
  - *Stores*
- Three search levels:
  - *Basic*: Keyword search
  - *Extended*: **Query builder form**
  - *Expert*: SPARQL query with predefined variables



Query builder form for product searches

Feed Delivery

RSS feed:

```
<item>
    <title>Camcorder with Widescreen LCD</title>
    <description>The best camcorder...</description>
    <link>http://www.example.com/camcorder</link>
    <pubDate>Tue, 04 Dec 2012 23:33:07 +0100</pubDate>
    <guid isPermaLink="false">...omitted...</guid>
</item>
```

Atom feed:

```
<entry>
    <title>Camcorder with Widescreen LCD</title>
    <content>The best camcorder...</content>
    <link rel="alternate">http://www.example.com/camcorder</link>
    <updated>2012-12-04T23:33:07+01:00</updated>
    <id>...omitted...</id>
</entry>
```

Caveats of Feed Integration

- No preservation of valuable meta-data found in the original content
- No links back to the original data source (backward links)
- No RDF triple that links the original locations with the URIs on which content is republished (forward links)

“Viral RDFa”

RSS feed:

```
<item>
    <title>Camcorder with Widescreen LCD</title>
    <description>The best camcorder...</description>
    <link>http://www.example.com/camcorder</link>
    <pubDate>Tue, 04 Dec 2012 23:33:07 +0100</pubDate>
    <guid isPermaLink="false">...omitted...</guid>
</item>
```

(Invisible) RDFa snippet:

backward link

```
<span typeof="gr:Offering" about="http://www.example.com/camcorder#product">
    <span property="gr:name" content="Camcorder with WideScreen LCD"></span>
    <span property="gr:description">The best camcorder...</span>
    <span rel="foaf:page" resource=""></span>
</span>
```

forward link

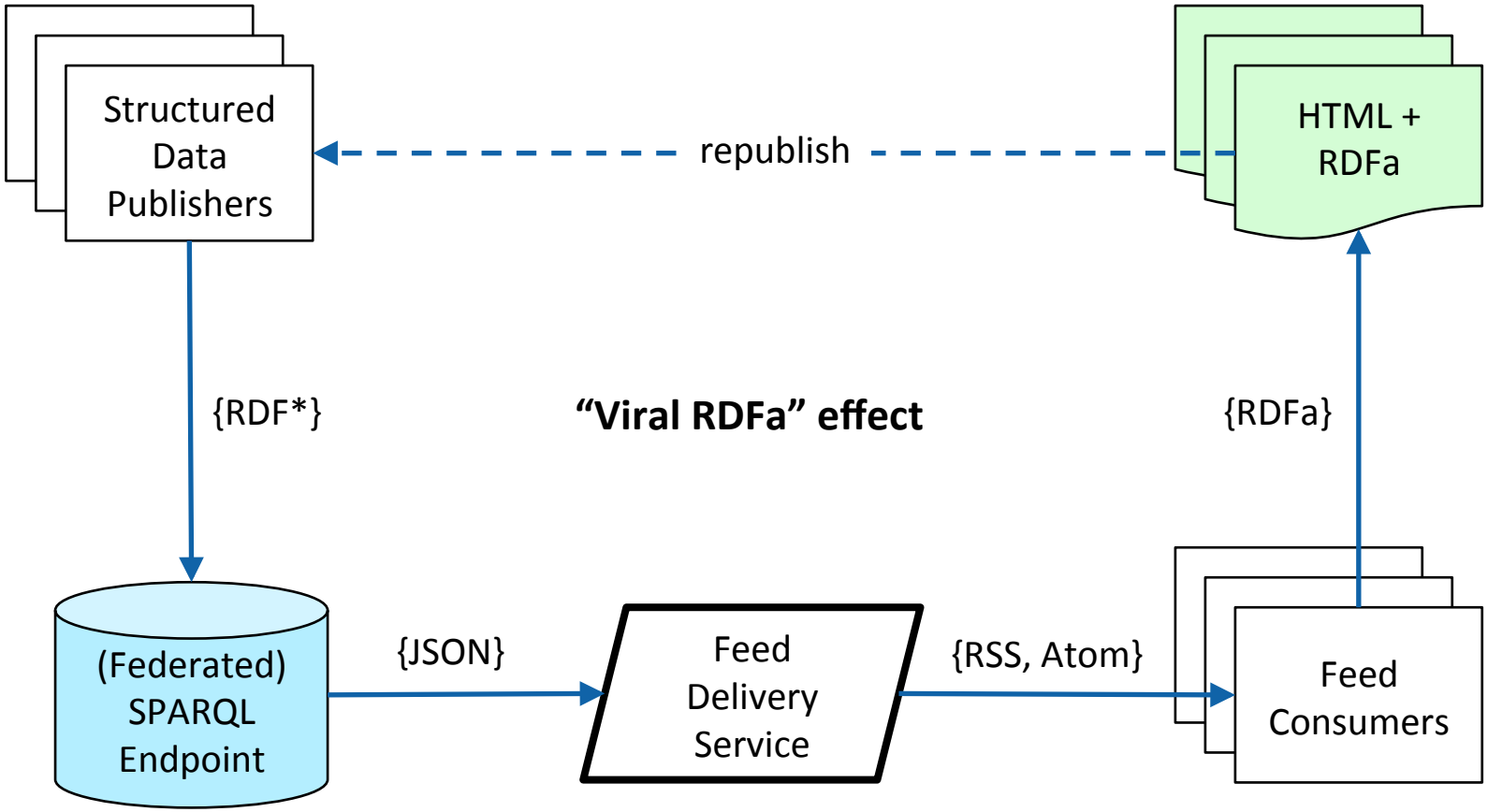
“Viral RDFa”

“Viral RDFa” technique:

```
<item>
    <title>Camcorder with Widescreen LCD</title>
    <description>&lt;span typeof=&quot;gr:Offering&quot;
about=&quot;http://www.example.com/camcorder#product&quot;&gt;
    &lt;span property=&quot;gr:name&quot; content=&quot;Camcorder
with WideScreen LCD&quot;&gt;&lt;/span&gt;
    &lt;span property=&quot;gr:description&quot;&gt;The best
camcorder...&lt;/span&gt;
    &lt;span rel=&quot;foaf:page&quot;
resource=&quot;&quot;&gt;&lt;/span&gt;
&lt;/span&gt;</description>
    <link>http://www.example.com/camcorder</link>
    <pubDate>Tue, 04 Dec 2012 23:33:07 +0100</pubDate>
    <guid isPermaLink="false">...omitted...</guid>
</item>
```

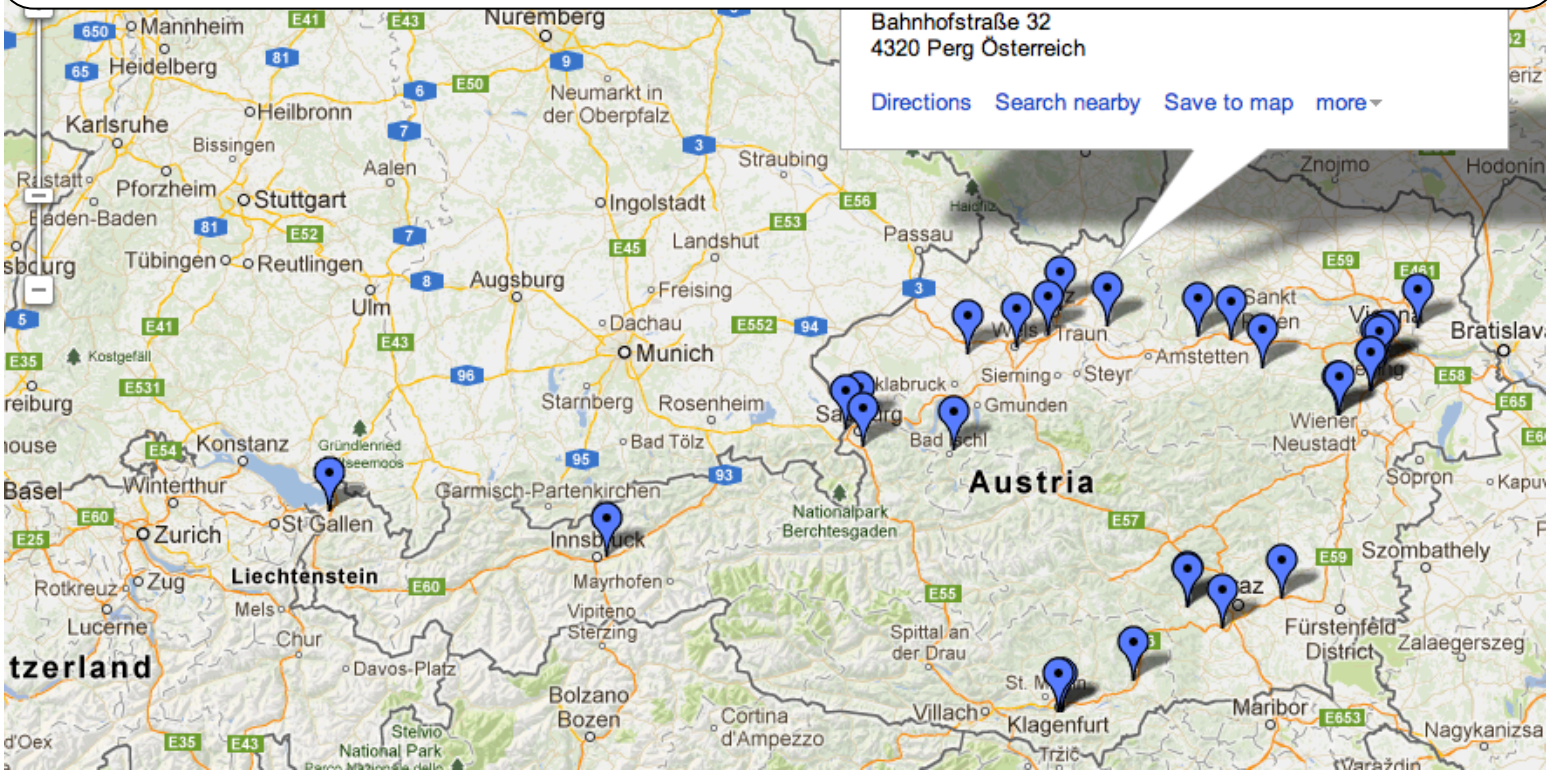
HTML entity-encoded RDFa

“Viral RDFa”


Structured Data Publishers
republish
HTML + RDFa
{RDF*}
“Viral RDFa” effect
{RDFa}
(Federated) SPARQL Endpoint
{JSON}
Feed Delivery Service
{RSS, Atom}
Feed Consumers

Geo-Location Information

```
<georss:point>48.248212 14.627774</georss:point>
<geo:lat>46.786</geo:lat>
<geo:long>11.2279</geo:long>
```



Evaluation

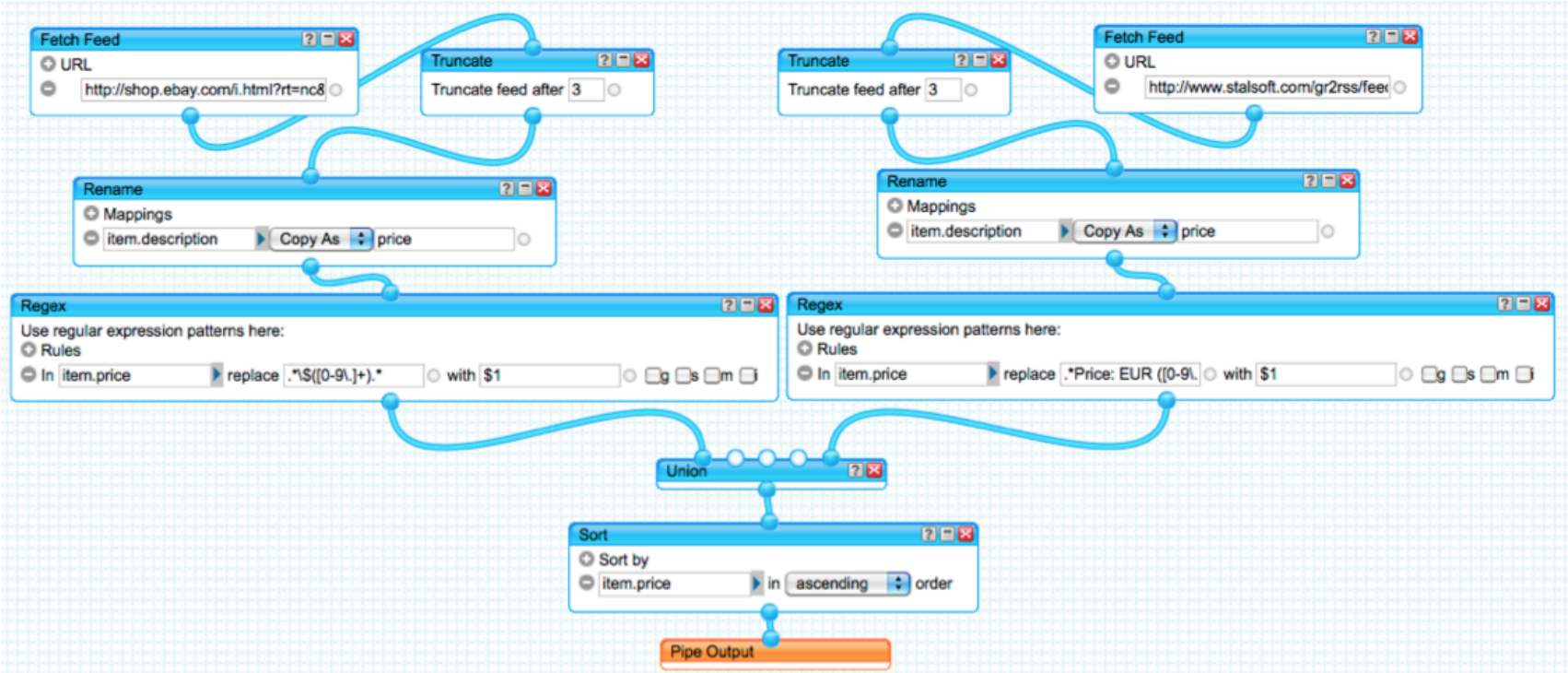
1. Qualitative comparison with available alternatives: DERI Pipes, Yahoo Pipes, eBay RSS feed generator, Amazon RSS feeds, manual integration in PHP

- Is it possible to integrate different data from multiple sources?
- Is it possible to integrate RDF data from SPARQL endpoints?
- Can the results be made available in standard feed syntaxes?
- Does the technology provide a user interface for specifying the desired content?
- Can the query address individual data elements from multiple sources (e.g. can it sort the union of offers from multiple vendors by price)?
- Are the canonical identifiers (URIs, product identifiers, etc.) of entities preserved so that clients can link the information with external data easily?

→ *integration task tedious (if not impossible)*

Evaluation

2. Integration experiments with DERI Pipes and Yahoo Pipes: Failed to integrate endpoints, but succeeded in integrating two RSS feeds using Yahoo Pipes



Conclusion

- Leverage Linked Data consumption by
  - providing a convenient way of querying Linked Data
  - exploiting widely standardized data feed formats
- Take away the barriers of integrating data sources
- Create incentives for both ~~feed publishers~~ publishers of RDF data and feed subscribers that integrate RSS or Atom feeds into their Web pages

Acknowledgments

The work described in this paper was partly sponsored by the German Federal Ministry of Education and Research (BMBF) by a grant under the KMU Innovativ program as part of the Intelligent Match project (FKZ 01IS10022B).

GEFÖRDERT VOM

Bundesministerium
für Bildung
und Forschung

Thank you!

Do you have any questions?

Alex Stolz
alex.stolz@unibw.de

Project page:
http://www.stalsoft.com/gr2rss/