



Is Linked Data the future of data integration in the enterprise?

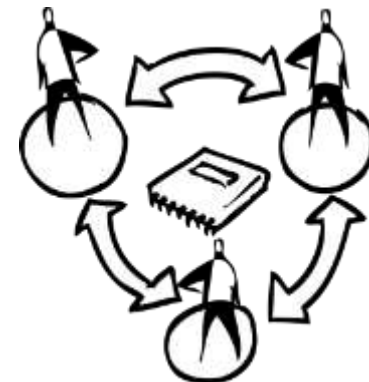
John Walker

Email: john.walker@nxp.com

Twitter: [@NXPdata](https://twitter.com/NXPdata)

Pilot Linked Open Data

2013-07-03



- ▶ NXP is a semiconductor (microchip) manufacturer
- ▶ Established: 2006 (formerly a division of Philips) with 50+ years of experience in semiconductors
- ▶ Headquarters: Eindhoven, The Netherlands
- ▶ Customers include Apple, Bosch, Continental, Delphi, Gemalto, Giesecke/Devrient, Huawei, NSN, Panasonic and Samsung
- ▶ Portfolio of **20,000+** products





THE ROAD TO LINKED DATA

The bad old days

Stone aged attitude to product information

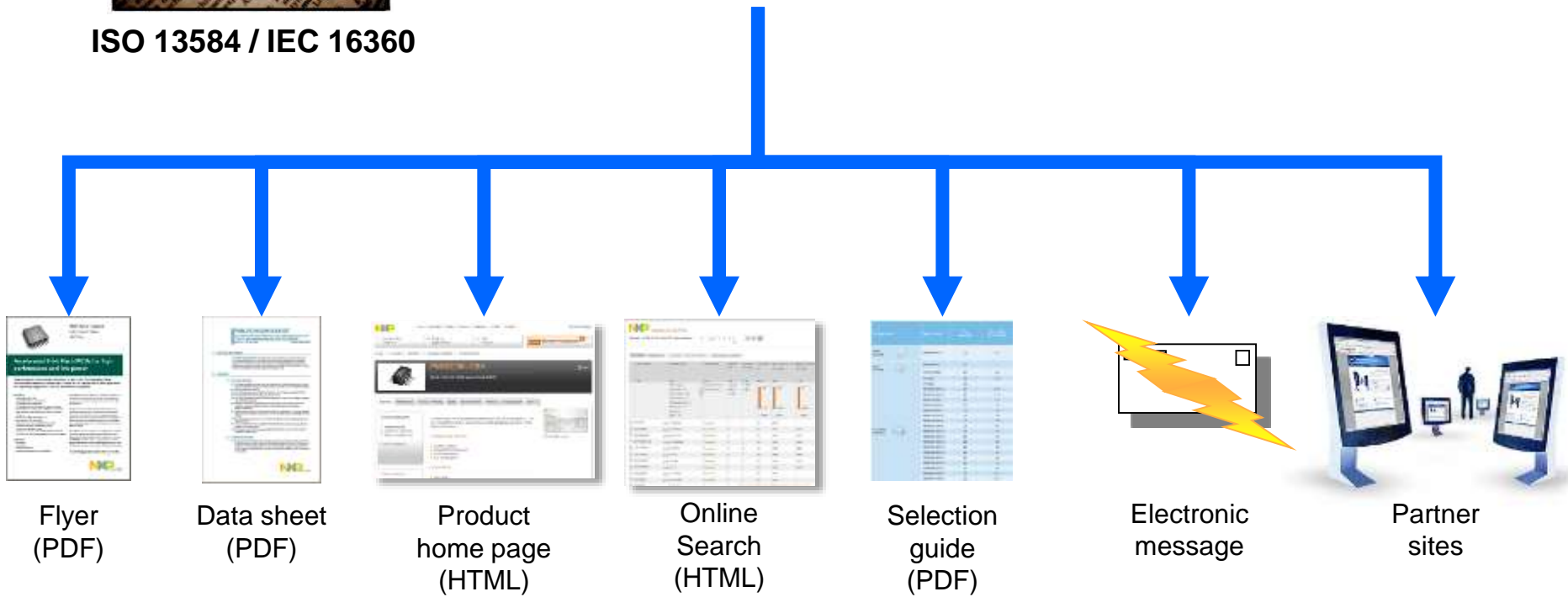
- ▶ Document-centric product information management
 - Multiple separately-maintained content silos
 - Content re-use is manual copy and paste or, worse, re-typing
- ▶ Consequences for NXP:
 - Inconsistent content and uncontrolled publications
 - Duplicated effort and extra time to publish
 - Error prone and costly to maintain
 - Highly-complex process and architecture
- ▶ Consequences for our partners and customers:
 - Unclear what information represents ‘the truth’
 - Manual effort to gather product information
 - Difficult to find all new and updated products

The vision: Unified Content Strategy (v1)



ISO 13584 / IEC 16360

Create Once,
Approve once,
Re-Use Many Times



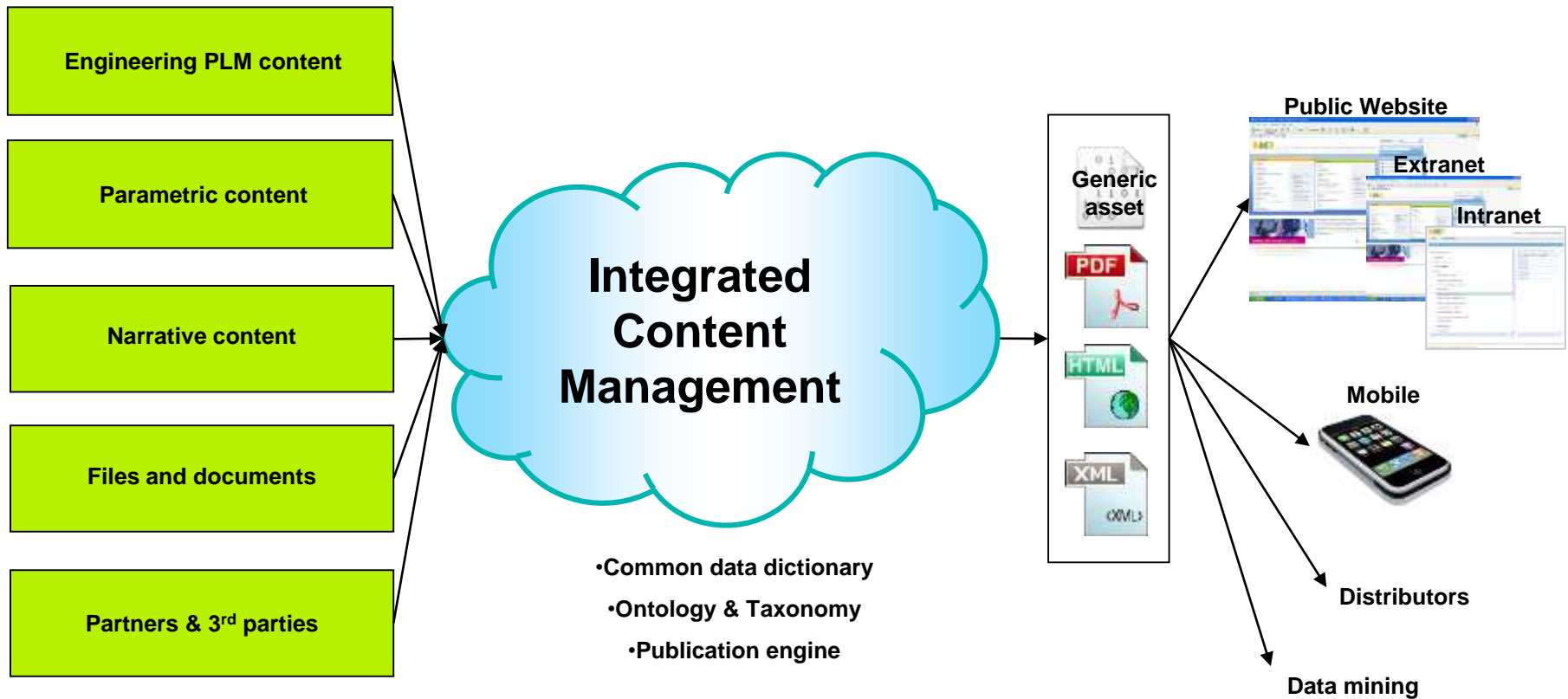
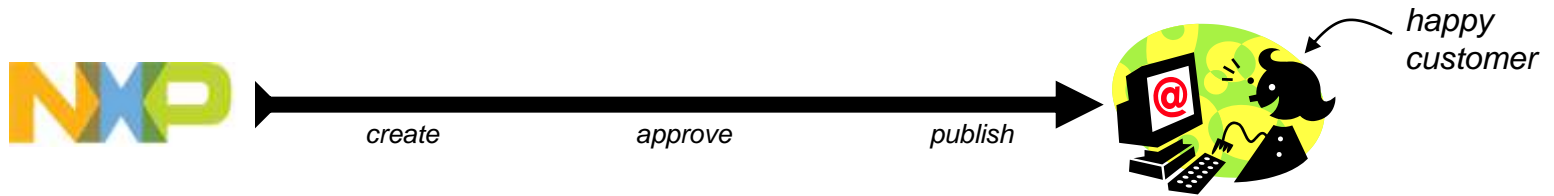
2012 state of play: RDBMS and XML

- ▶ Multiple RDBMS and XML:DB with XML messages for data exchange
- ▶ Every application has it's own database
- ▶ Every RDBMS has it's own interpretation of the data
- ▶ Data synchronization issues are regular occurrence
- ▶ Links between objects are implicit
- ▶ Same data is sometimes replicated in multiple XML documents
- ▶ Each system is essentially closed with 'throw it over the wall' attitude to publishing data
- ▶ Proliferation of APIs for specific purposes without any overall standards

What we were searching for

- ▶ More open access to query and use data
- ▶ More formal semantics (ontology)
- ▶ Shared model and key identifiers
- ▶ Explicit relations between conceptual entities
- ▶ Ability to easily query from multiple view points
- ▶ Ability to evolve and extend schema

The vision: Unified Content Strategy (v2)





- ▶ Stumbled across RDF and Linked Data via DITA Concept Scheme and SKOS
- ▶ Immediately recognized potential

How we got started

Just do it!

- ▶ Get hands on
 - Started transforming XML to RDF/XML with XSLT
 - Used Kasabi as data store and SPARQL endpoint
 - Used Puelia as front end
- ▶ Sell the idea
 - Break through resistance to ‘new/unproven’ technology
 - Emphasize benefits over traditional approach
- ▶ Don't mention RDF
 - RDF/XML is still XML after all ;-)

Find the right nail



- ▶ Problems about relationships between things are ideal
- ▶ BI / BW issue
 - Reporting on user behavior across several systems
 - Unable to reconcile data from various sources
- ▶ We were able to provide the data that linked everything together
- ▶ Agreed upon principled use of RDF/XML that can be validated as XML
- ▶ Easy to use in traditional BW star schema

- ▶ We solved the immediate issue and got several RDF dumps to experiment with
- ▶ Based on this we got the OK to invest further time/resources

```
<?xml version="1.0" encoding="UTF-8"?>
```

RDF file per type of entity

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  schema#" xmlns:skos="http://www.w3.org/2004/02/skos/core#"
  xmlns:nxp="http://purl.org/nxp/schema/v1/" xmlns:dcterms="http://purl.org/dc/terms/"
  xmlns:nie="http://www.semanticdesktop.org/ontologies/2007/01/
  xmlns:nfo="http://www.semanticdesktop.org/ontologies/2007/03
  xmlns:xs="http://www.w3.org/2001/XMLSchema#" xmlns:owl="http://www.w3.org/2002/
  xmlns:schema="http://schema.org/" xmlns:dita="http://purl.org/dita/ns#"
  xmlns:frbr="http://purl.org/vocab/frbr/core#">
```

Description per entity

```
<rdf:Description rdf:about="http://qa.data.nxp.com/id/basic_types/74lvt00pw">
  <rdf:type rdf:resource="http://purl.org/nxp/schema/v1/BasicType"/>
  <nxp:productStatusDate rdf:datatype="http://www.w3.org/2001/XMLSchema#date">2011-10-
  20</nxp:productStatusDate>
  <skos:prefLabel xml:lang="en-US">74LVT00PW - 3.3 V Quad 2-input NAND gate</skos:prefLabel>
  <nxp:productStatus rdf:resource="http://www.nxp.com/documents/outline_3d/sot402-1_3d.gif"/>
  <foaf:homepage rdf:resource="http://en.qa.beta.nxp.trim">
  <nxp:typeNumber>74LVT00PW</nxp:typeNumber>
  <foaf:isPrimaryTopicOf rdf:resource="http://qa.data.nxp.com/internal/basic_types/74lvt00pw"/>
  <foaf:isPrimaryTopicOf rdf:resource="http://www.nxp.com">
  <nxp:publishToWeb>Yes</nxp:publishToWeb>
  <nxp:customerSpecificIndicator>No</nxp:customerSpecificIndicator>
  <nxp:productSecurityIndicator>No</nxp:productSecurityIndicator>
  <foaf:depiction rdf:resource="http://www.nxp.com/documents/outline_3d/sot402-1_3d.gif"/>
  <nxp:package>SOT402-1</nxp:package>
```

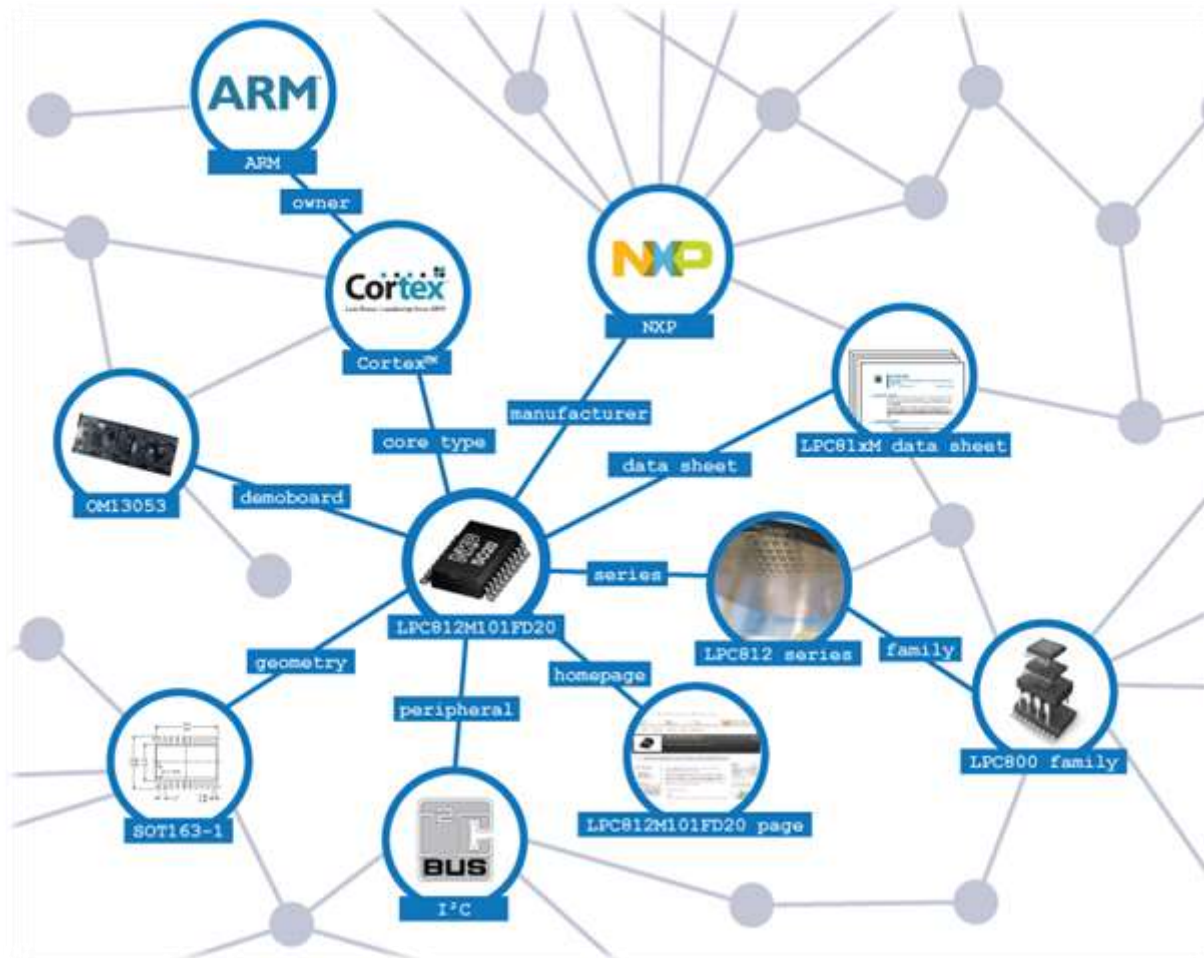
Literal properties with optional language

Object properties with reference to resource (no nesting)

Using existing properties where possible

```
</rdf:Description>
```

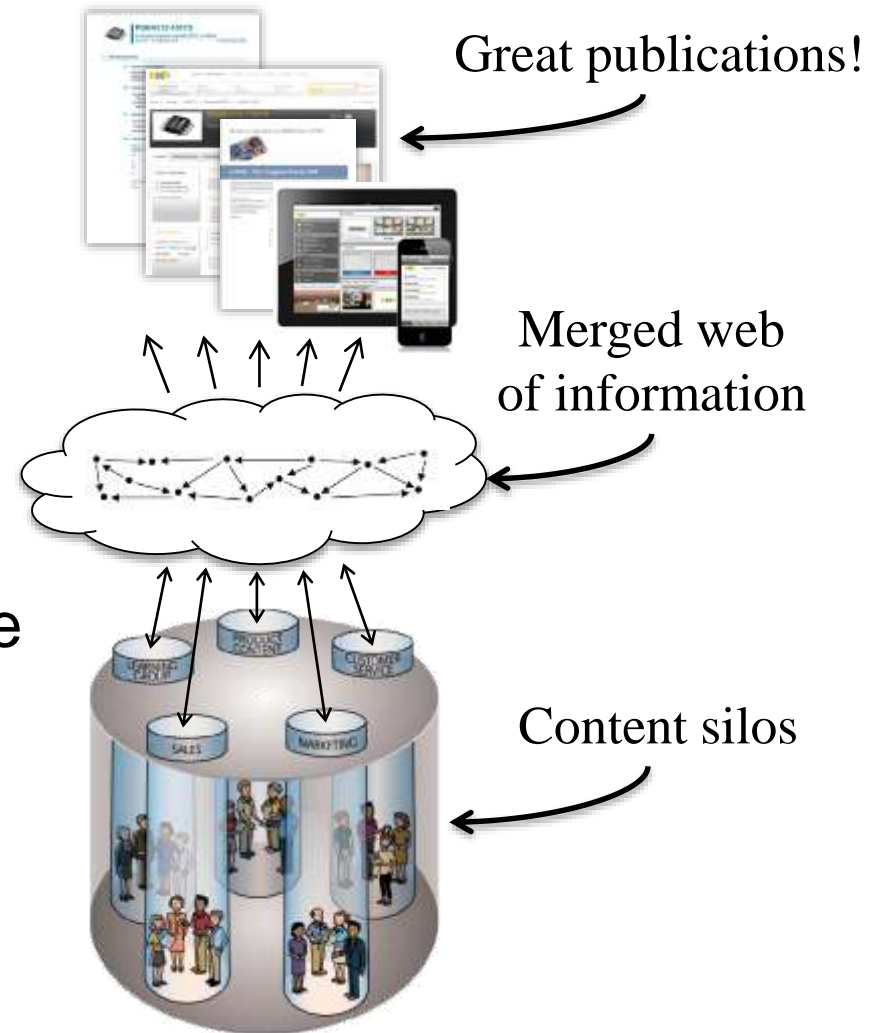
```
</rdf:RDF>
```



WHERE WE ARE TODAY

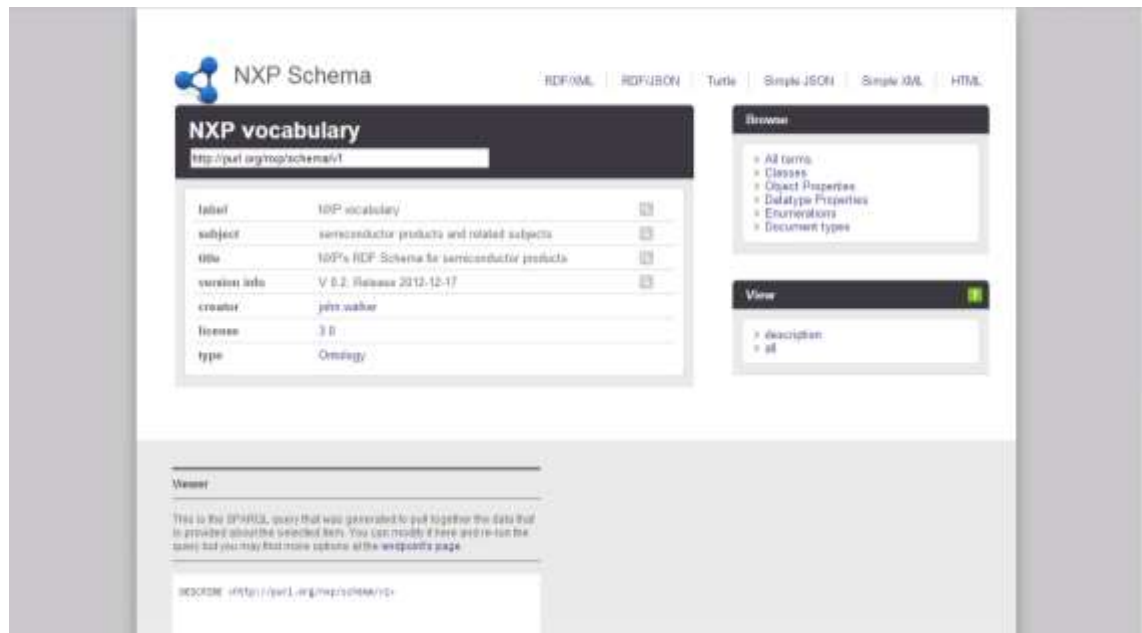
The vision: Unified Content Strategy (v3)

- ▶ **One** integrated trusted source
- ▶ **One** process
- ▶ **Reduced** number of tools
- ▶ **High** quality
- ▶ **Richly** structured
- ▶ **Easy** to create, update & (re)use
- ▶ **High** speed
- ▶ **Publish** to multiple channels



Minimal NXP vocabulary

- ▶ Maintained in Turtle
- ▶ Used purl.org as domain
- ▶ Defines NXP-specific classes, properties and enumerations
- ▶ <http://purl.org/nxp/schema/v1>



The screenshot displays the NXP Schema website interface. At the top, there is a navigation bar with the NXP logo and the text "NXP Schema". Below this, there are tabs for different data formats: RDF/XML, RDF/JSON, Turtle, Simple JSON, Simple XML, and HTML. The main content area is titled "NXP vocabulary" and shows the URL "http://purl.org/nxp/schema/v1". A table lists the following metadata:

label	NXP vocabulary
subject	semiconductor products and related subjects
title	10P's RDF Schema for semiconductor products
version info	V 0.2, Release 2012-12-17
creator	jens.nahler
license	3.0
type	Ontology

On the right side, there is a "Browse" section with a tree view showing categories: All terms, Classes, Object Properties, Datatype Properties, Enumerations, and Document types. Below that is a "View" section with a tree view showing: description and all. At the bottom, there is a "Viewer" section with a warning message: "This is the SPARQL query that was generated to put together the data that is provided about the selected item. You can modify it here and re-run the query and you may find more options at the wikipedia page." Below the warning is a code block containing the SPARQL query: "DESCRIBE <http://purl.org/nxp/schema/v1>".

Triple store

- ▶ Using Dydra (<http://dydra.com>) as triple store
- ▶ Cloud based
 - Reduces need for in-house knowledge
- ▶ SPARQL 1.1 endpoint
- ▶ NXP-specific instance
- ▶ Stored queries



DYDRA Home About Docs Blog Signup Login

Networks Made Friendly

Dydra is a powerful graph database in the cloud, allowing your business to make the most of highly connected data, such as social networks. It's fast, easy to use and affordable. [Learn more >](#)

Take Dydra for a spin
Fill the ground running, it's simple and takes just a minute of your time.

1 Create **5 dydra create foaf**
Report
Query
Database: /foaf
Created: <http://dydra.com/foaf/>

Hassle-free operations
We take care of system administration and maintenance. You just use it.

Reliability comes standard
We focus on data integrity and availability. You focus on your business.

Always in balance
Dydra grows with your data. Worrying about scaling is a thing of the past.

Saves you money
The basic plan is free. We'll offer more powerful paid tiers after the beta period.

Join the private beta
Enter your email address and we'll let you know when a slot is available.

Your email address:

NXP Linked Product Data

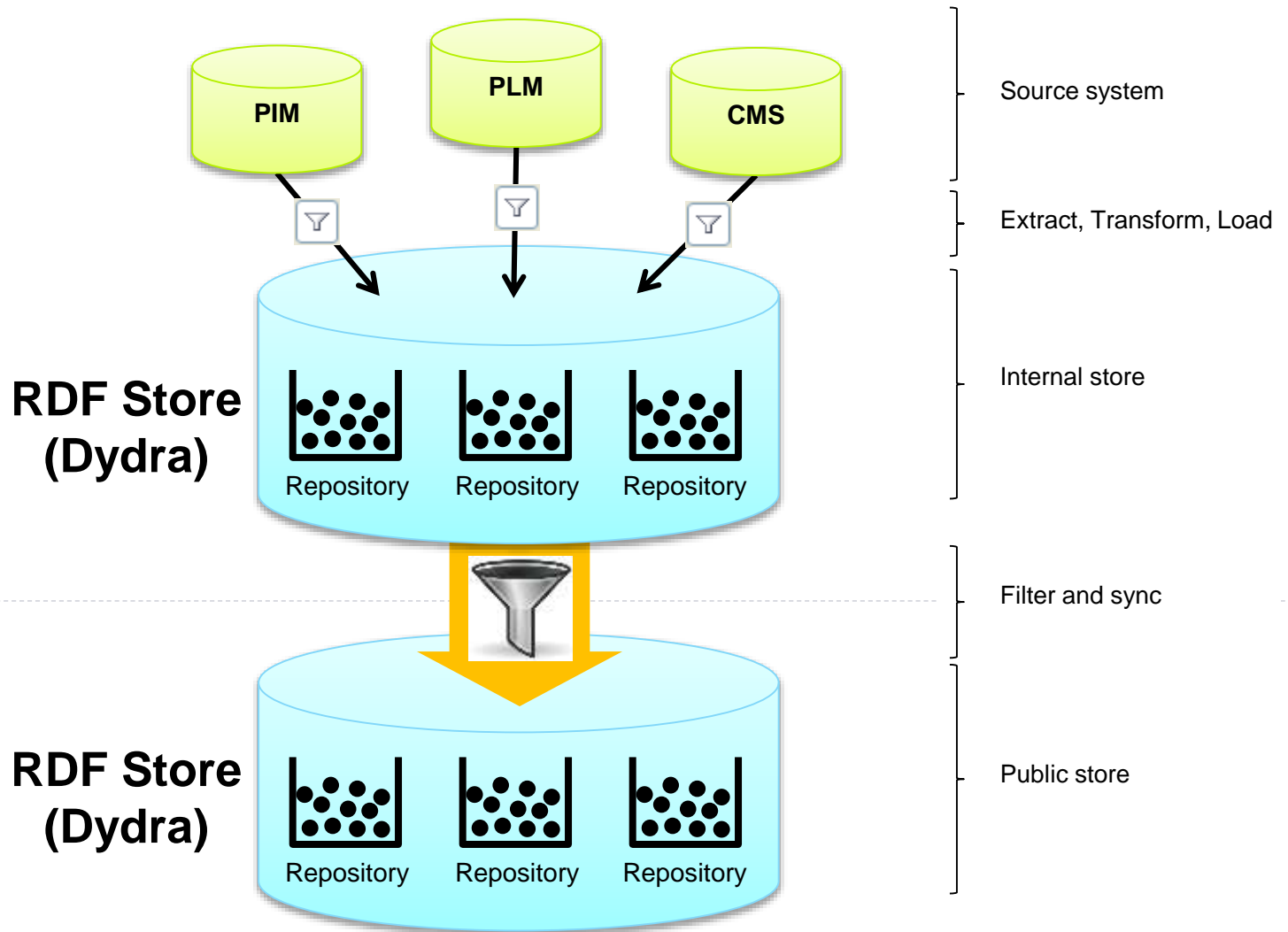
The screenshot displays a web browser window with the URL `qa.data.nxp.com/internal/basic_types/bss84akmb`. The page features the NXP logo and navigation links for 'Basic types', 'Groups', and 'Marketing taxonomies'. The main content area is divided into several sections:

- Automotive MOSFETs**: A list of related products including 'Standard MOSFETs', 'Bss84akmb', and '50 V, single P-channel Trench MOSFET'. It also includes a link to 'NXP's Power MOSFET Selection Guide 2012: Smaller, faster, cooler' and 'Understanding power MOSFET data sheet parameters'.
- BSS84AKMB**: A central product card showing an image of the component, the part number 'BSS84AKMB', and a 'Basic Type' button. It also includes 'Source', 'RDF/XML', and 'Turtle' buttons.
- Resource**: A section containing metadata such as 'standardProductDescription' (small signal MOSFET), 'Publish to web' (Yes), 'Product status' (Production), 'lifecycleState' (ACT), 'package' (SOT883B), 'Customer specific indicator' (No), 'Product status date' (19 January 2012), 'Type number' (BSS84AKMB), 'package#Type' (SOT883), and 'Product security indicator' (No).
- Thing**: A section indicating that the product is the 'primary topic of' a 'Homepage' for 'BSS84AKMB'.

At the bottom left of the page, the copyright notice reads '© NXP B.V. 2013'.

Using Graphity as publication framework: <http://graphityhq.com/>

Current back-end architecture

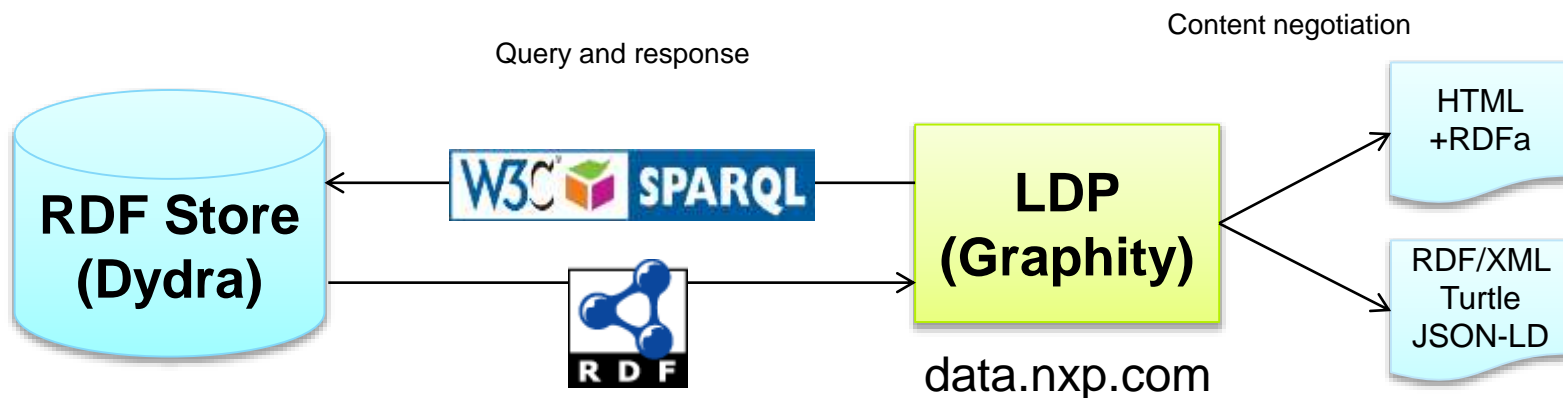


WWW



Current publication architecture

- ▶ Use W3C web standards RDF, SPARQL for portable solution
- ▶ Work with any RDF store with only minor configuration



Count of resources (top 10)

type	count
nxp:SalesItem	56389
nxp:ProductType	43680
nxp:BasicType	26426
nfo:FileDataObject	25596
nfo:PlainTextDocument	25596
nxp:ValueProposition	19922
<http://purl.org/dita/ns#Topicref>	19807
nxp:PTopicDita	19728
nxp:FinancialClassification	14496
nfo:FileDataObject	12050

In total 268,678 entities with over 2.4M triples

```
@prefix nxp: <http://purl.org/nxp/schema/v1/> .
```

```
@prefix nfo: <http://www.semanticdesktop.org/ontologies/nfo/#> .
```



Successes so far

- ▶ Canonical data source for master data
 - Provide the linking data
 - Unambiguous references
- ▶ Using stored SPARQL SELECT queries to expose REST APIs
 - Results in tabular XML, JSON or CSV/TSV format
 - Easy to manage queries
 - Extremely quick to set up new APIs (15 – 30 mins)
- ▶ Able to answer previously unanswerable questions
- ▶ Easy to integrate new sources
- ▶ Minimal investment compared to traditional BW/BI projects

NEXT STEPS



What's cooking

- ▶ More data
 - Full parametric product data
 - More sources integrated
 - Full blown ontologies
 - Use RDF as persistent source
- ▶ More consumers
 - Customer facing Linked Data applications
 - Faceted and guided search
 - Data source for nxp.com
- ▶ Linked Open Data
 - Publish data under open license (where applicable)

Broader online ecosystem

