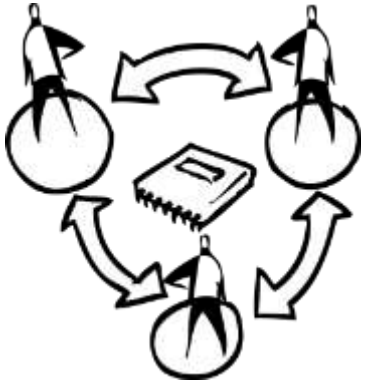
Is Linked Data the future of data integration in the enterprise?

John Walker
Email: john.walker@nxp.com
Twitter: @NXPdata
Pilot Linked Open Data
2013-07-03



ABOUT NXP

- 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

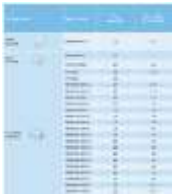
Create Once,
Approve once,
Re-Use Many Times

Flyer (PDF)

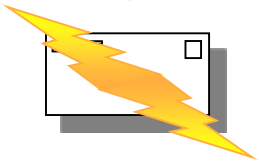
Data sheet (PDF)

Product home page (HTML)

Online Search (HTML)



Selection guide (PDF)



Electronic message

Partner sites

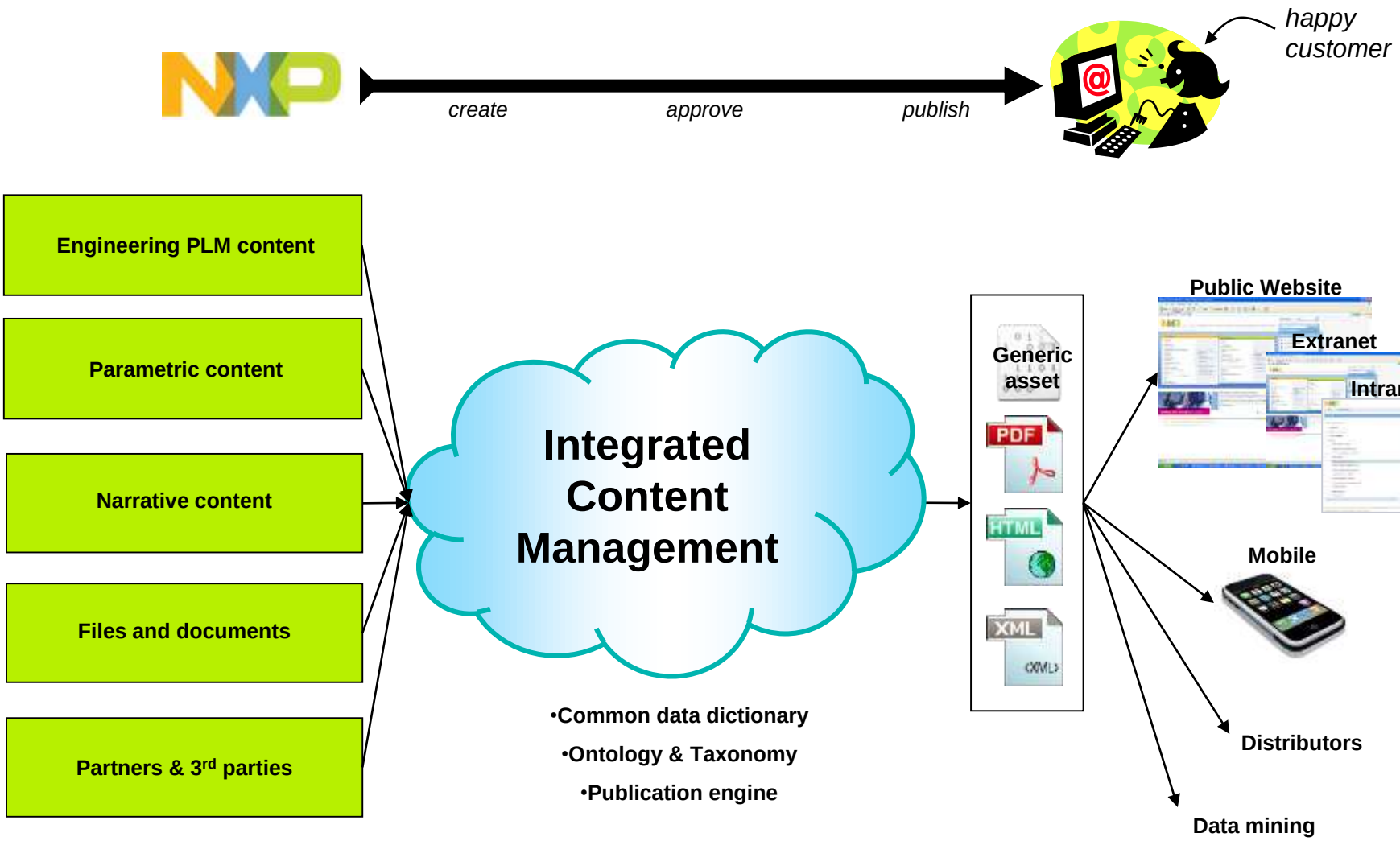
2012 state of play: RDBMS and XML

- 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:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:rdfs="http://www.w3.org/2000/01/rdf-
    schema#" xmlns:skos="http://www.w3.org/2004/02/skos/core#" xmlns:foaf="http://xmlns.com/foaf/0.1/"
    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="
    xmlns:schema="http://schema.org/" xmlns:dita="http://purl.org/dita/ns#"
    xmlns:frbr="http://purl.org/vocab/frbr/core#">
    <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://
            <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://
            <nxp:publishToWeb>Yes</nxp:publishToWeb>
            <nxp:customerSpecificIndicator>No</nxp:customerSpecific
            <nxp:productSecurityIndicator>No</nxp:productSecurityIn
            <foaf:depiction rdf:resource="http://www.nxp.com/documents/outline_3d/sot402-1_3d.gif"/>
            <nxp:package>SOT402-1</nxp:package>
    </rdf:Description>
</rdf:RDF>
```

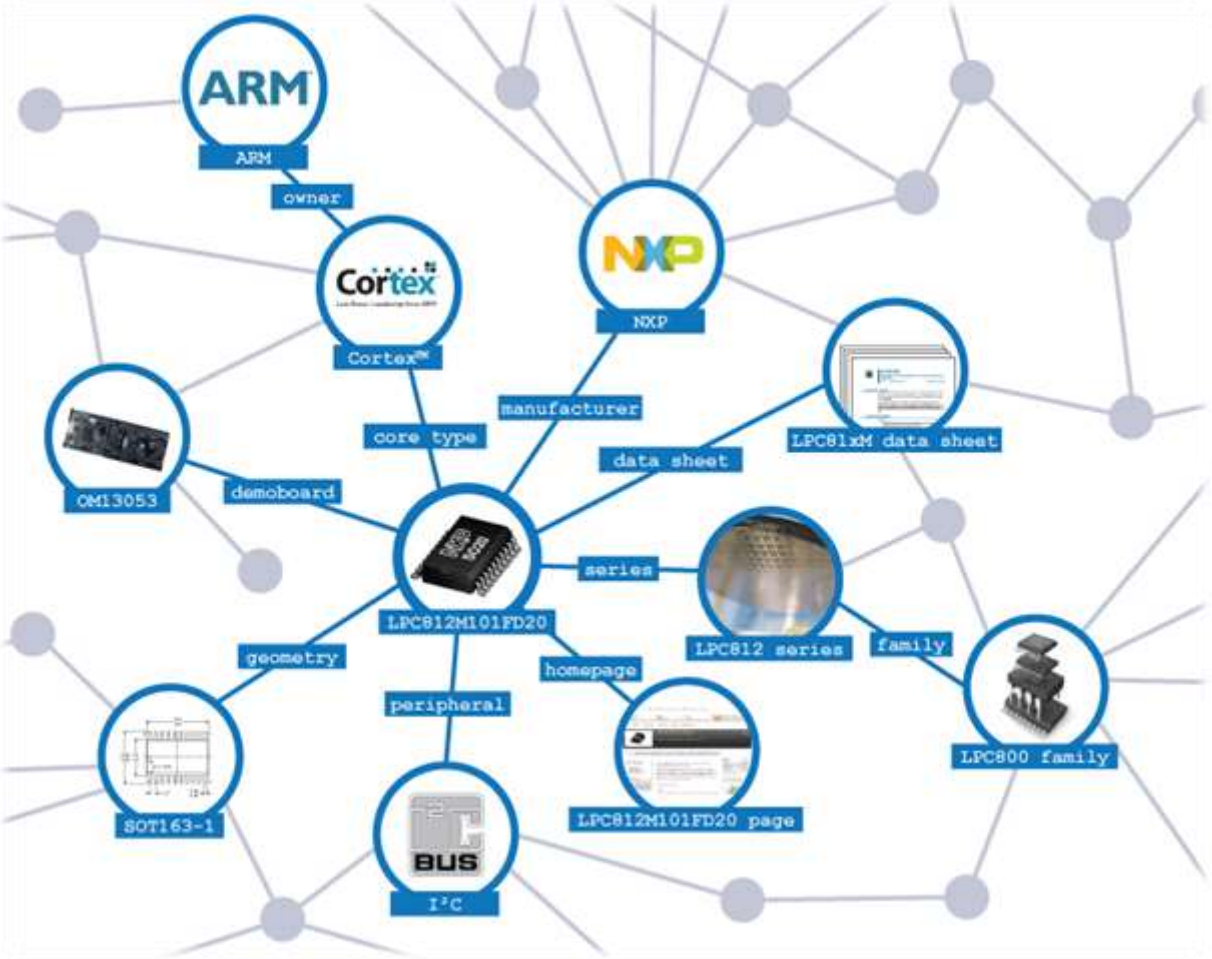
RDF file per type of entity

Description per entity

Literal properties with optional language

Object properties with reference to resource (no nesting)

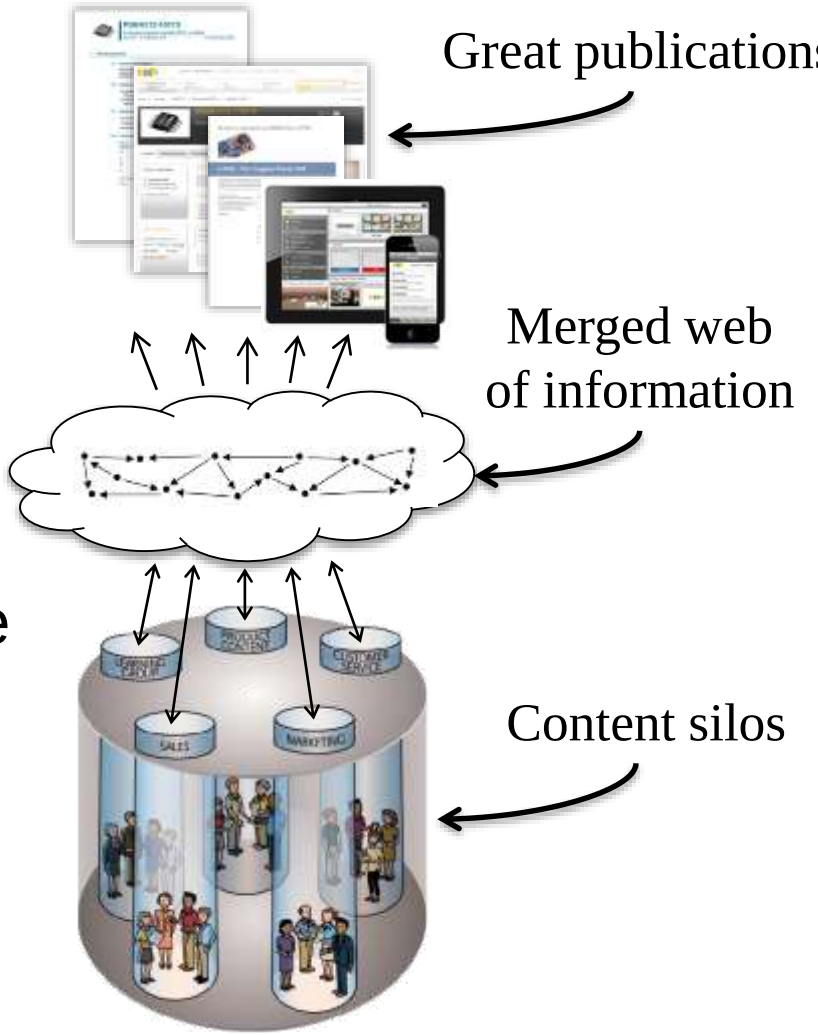
Using existing properties where possible



WHERE WE ARE TODAY

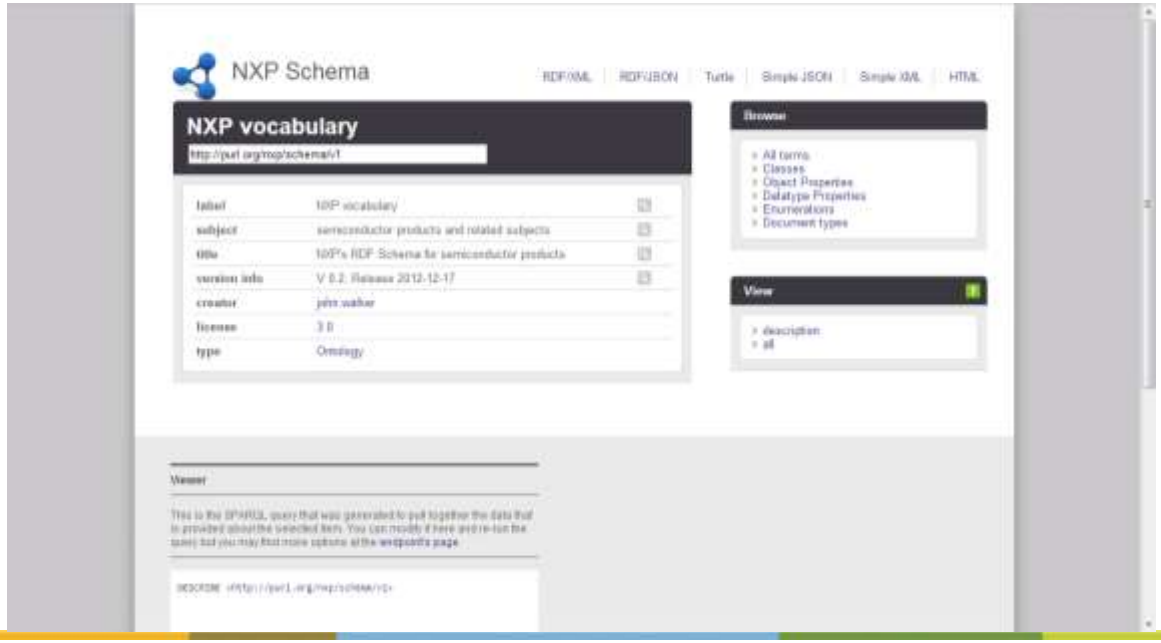
The vision: Unified Content Strategy (v3)

- 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

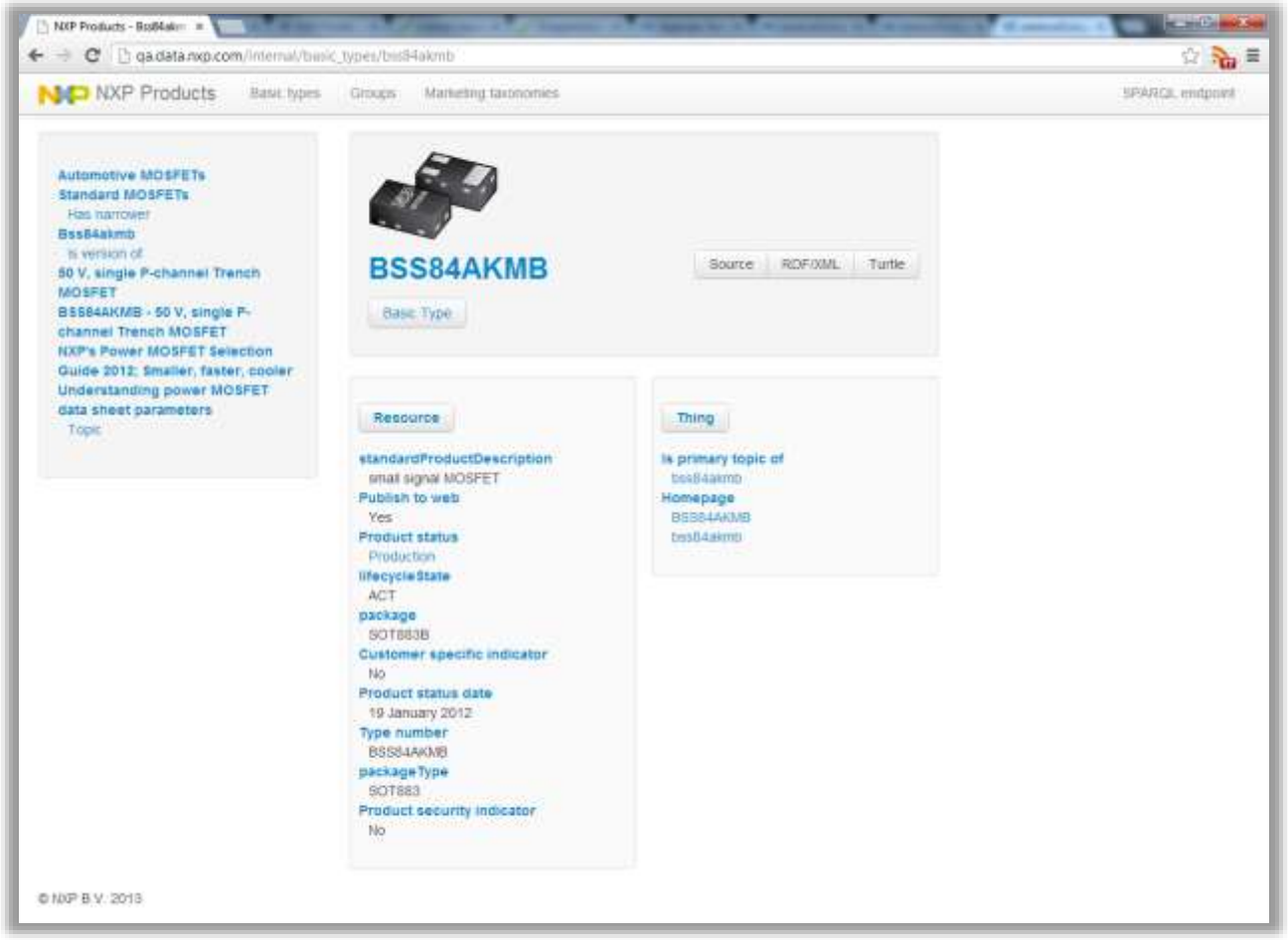
- Maintained in Turtle
- Used purl.org as domain
- Defines NXP-specific classes, properties and enumerations
- http://purl.org/nxp/schema/v1



Triple store

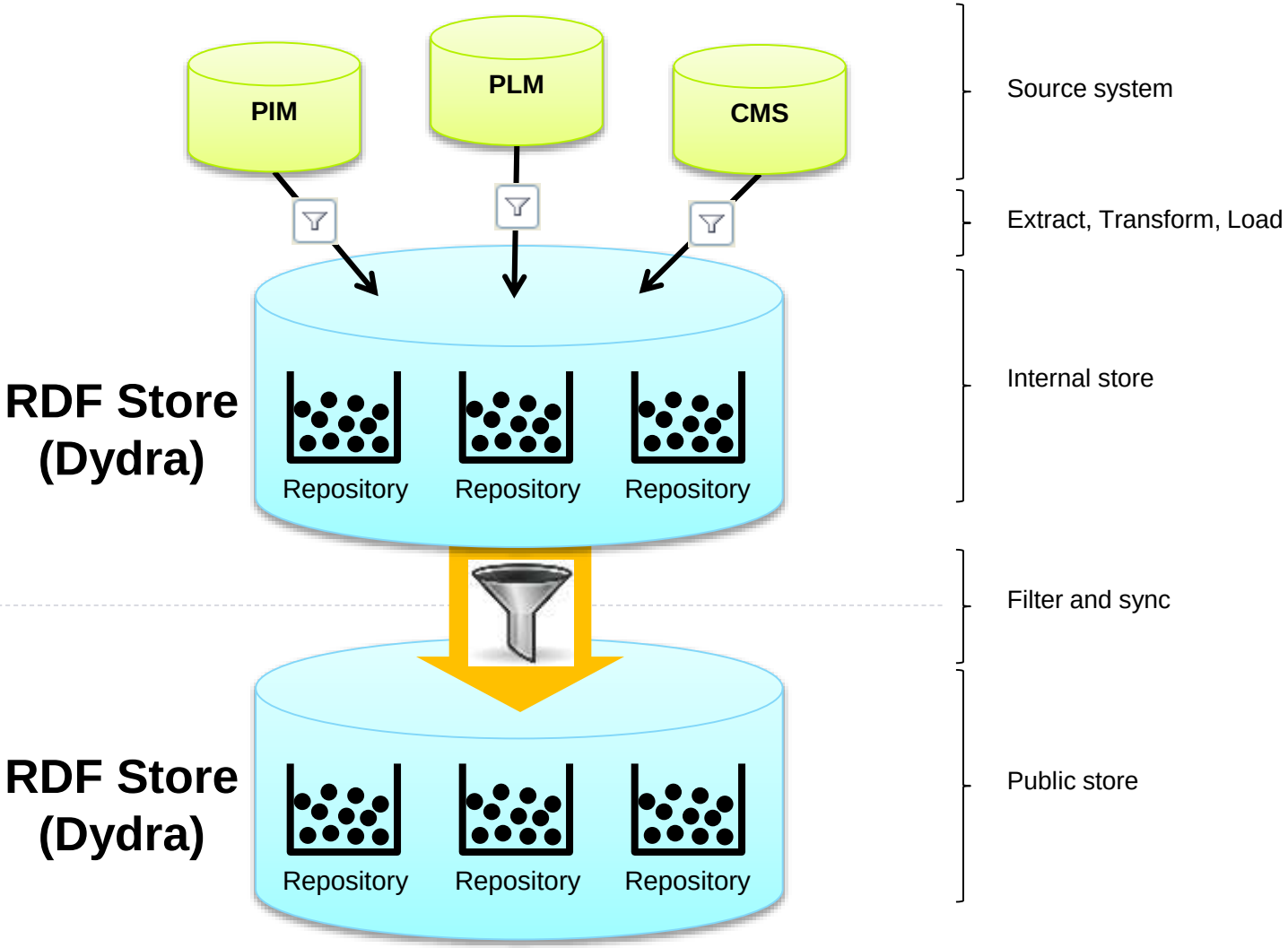
- Using Dydra (http://dydra.com) as triple store
- Cloud based
  - Reduces need for in-house knowledge
- SPARQL 1.1 endpoint
- NXP-specific instance
- Stored queries

NXP Linked Product Data



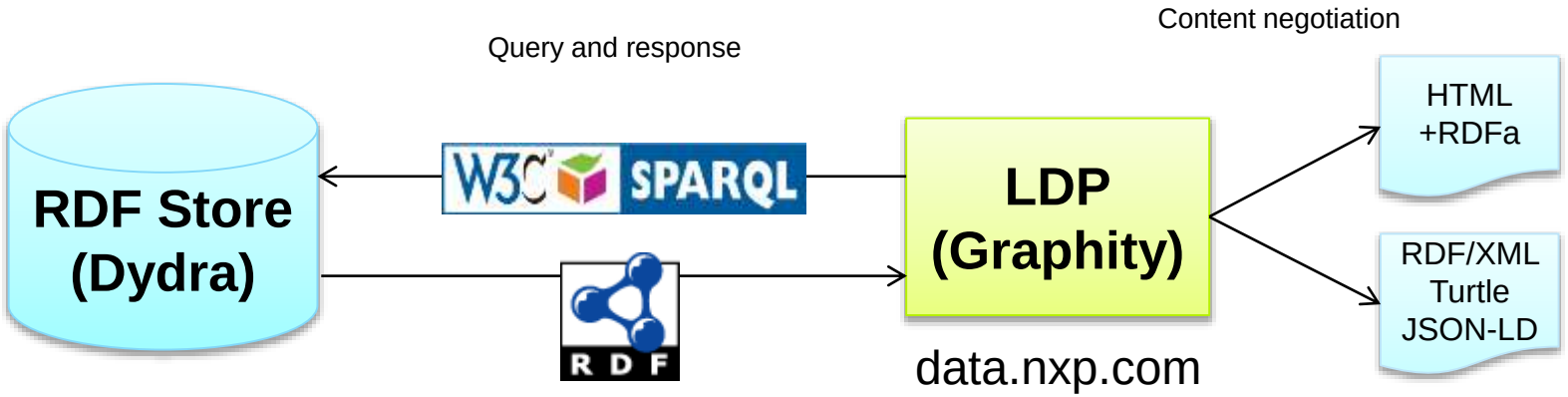
Using Graphity as publication framework: http://graphityhq.com/

Current back-end architecture



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