1

LINKED DATA - THE NEXT 5 YEARS

From Hype to Action

poolparty

SEMANTiCS
Leipzig 2016

Andreas Blumauer
CEO & Managing Partner

Semantic Web Company /
PoolParty Semantic Suite

2

INTRODUCING SEMANTIC WEB COMPANY AND POOLPARTY

Semantic Web Company

- Founded in 2004
- Based in Vienna
- Privately held
- >30 employees, experts in text mining, data science & linked data
- SWC participates in EU-projects with a total funding of over € 17.0 million
- SWC named to KMWorld's 2016 "100 Companies That Matter in Knowledge Management"

PoolParty Semantic Suite

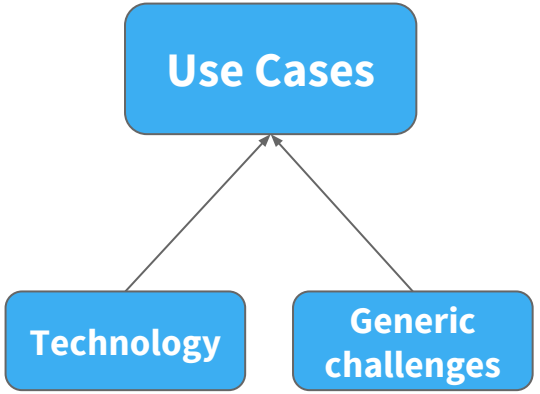
- First release in 2009
- Current version 5.5
- W3C standards compliant
- Over 150 installations world-wide
- 50% of SWC's revenue is reinvested into development of PoolParty
- To be installed on-premises or to be used as cloud service
- KMWorld listed PoolParty as Trend-Setting Product 2015 and 2016

3

Key assumption of this talk

"In the next 5 years
we will see thousands
of enterprises and many other
organisations
making use of linked data and
semantic web technologies,

but only when we are able to
translate
generic challenges and
technical opportunities into
specific, concrete and relevant
use cases."



4

Three generic challenges and hundreds of use cases

- Make informed decisions
- Link data and communicate across boundaries
- Context-aware and adaptive information supply

Nutrition

Food-Drug interactions

Inferring Cuisine - Drug Interactions

Education

Personalised learning pathways

A New Digital Education for Young Australians in the 21st Century

Crisis Management

Lack of data interoperability

Benefits of linked data for interoperability during crisis management

Clean Energy

GHG emissions

Well-founded Policy Making

Law

Complex regulatory systems

Linked Legal Data

Healthcare

Improving health literacy

The Semantic Content Engine of healthdirect Australia

5

How to communicate the technological advantages?

- Linked Data is an agile data integration technology
- RDF stores scale
- Semantic Web is standards-based
- Knowledge graphs can grow over time
- With Linked Data, unstructured and structured information can be handled
- Even complex queries can be executed with SPARQL
- Linked Data goes beyond relational data
- With Linked Data, you can load data, content, and metadata without the pain of upfront data modeling
- ...

healthdirect

Medicines Catalogue

In February 2016, Healthdirect Australia added a major new feature to the healthdirect website – a medicines catalogue. This catalogue provides a searchable database of approved medicines, as registered by the Therapeutic Goods Administration (TGA), available in Australia. People can now search for ***more than 7,000 branded medicine products*** *on the healthdirect website.*

Where is the content sourced from?

A number of trusted sources of medicines information have been relied upon to create the medicines catalogue, including:

- **The Australian Register of Therapeutic Goods (ARTG)** – the reference database of the Therapeutic Goods Administration. It provides information on therapeutic goods that can be supplied in Australia.
- **The Australian Medicines Terminology (AMT)** – the national terminology provides standard naming conventions and terminology to accurately describe medications. The AMT is managed by NeHTA.
- **The Australian Health Thesaurus (AHT)** – managed and maintained by Healthdirect Australia, containing a comprehensive taxonomy of Australian health-related terms and keywords, and their relationships with each other.
- **The Pharmaceutical Benefits Scheme (PBS)** – an Australian Government program that subsidises medicines to make them more affordable.

Why did we create the healthdirect medicines catalogue?

Medicines (both their ingredients and brand names) is a highly searched category online. It is also the most popular topic of calls received by the healthdirect helpline and the after hours GP helpline.

Results for searches on medicines were poor, as medicines content was lost amongst our general content. Research is also showing that health consumers are looking at overseas health and medicines websites, where the regulatory, legal and clinical conditions are different from Australia.

Health consumers are looking to access reliable Australian information about medicines. We developed this section for the website to meet this need, and provide extra value to encourage people to use it and return.

Impact of the healthdirect medicines catalogue

Since the launch of the medicines catalogue there has been a huge increase in traffic to the healthdirect website in just one month. The catalogue has the potential to greatly improve the health literacy of Australians, as it can be freely utilised and is a trusted source of medicines information. The catalogue is regularly updated to ensure accuracy and currency.

Healthdirect medicines sessions

How does the medicines catalogue work?

We created a Terminology Service to manage the data from the different datasets. This tool is used to create and manage translations between vocabularies, terminologies and classification systems.

What content is available in the healthdirect medicines catalogue?

The website has a sophisticated search function which allows you to search by active ingredient or brand name (for example, paracetamol vs Panadol). The search results display a range of information such as medicine descriptions, details of medicine ingredients, and Links to Consumer Medicine Information (CMI) leaflets. Importantly, we also provide product recall information.

We've also published a range of informative articles related to medicines that sit alongside the database, providing information on topics such as children's medicines, generics vs brand names and types of pain killers.

Future developments

We plan to include the integration of additional data sources focussing on;

- ***pregnancy risks,***
- ***adverse events,***
- ***suitability for sport,***
- ***food and drug interactions.***

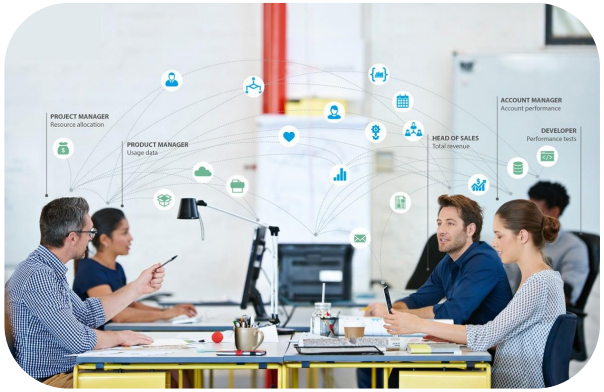
Improved navigation and appropriateness for our audience will also be paramount with any new developments. A medicines glossary will be developed describing the terminology used in a medicines context appropriate for general consumers.

Further Information

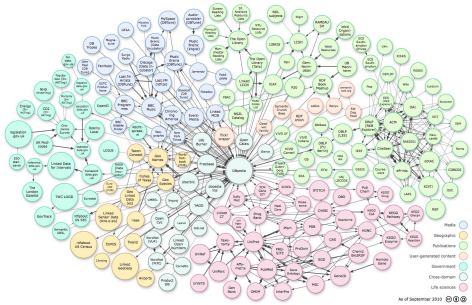
The medicines section can be viewed at ***www.healthdirect.gov.au/medicines***

6

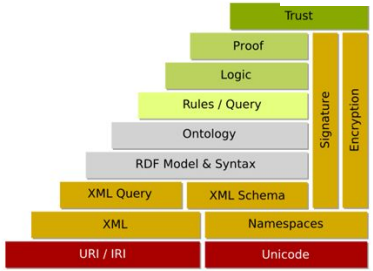
From technologies over data to applications



2017



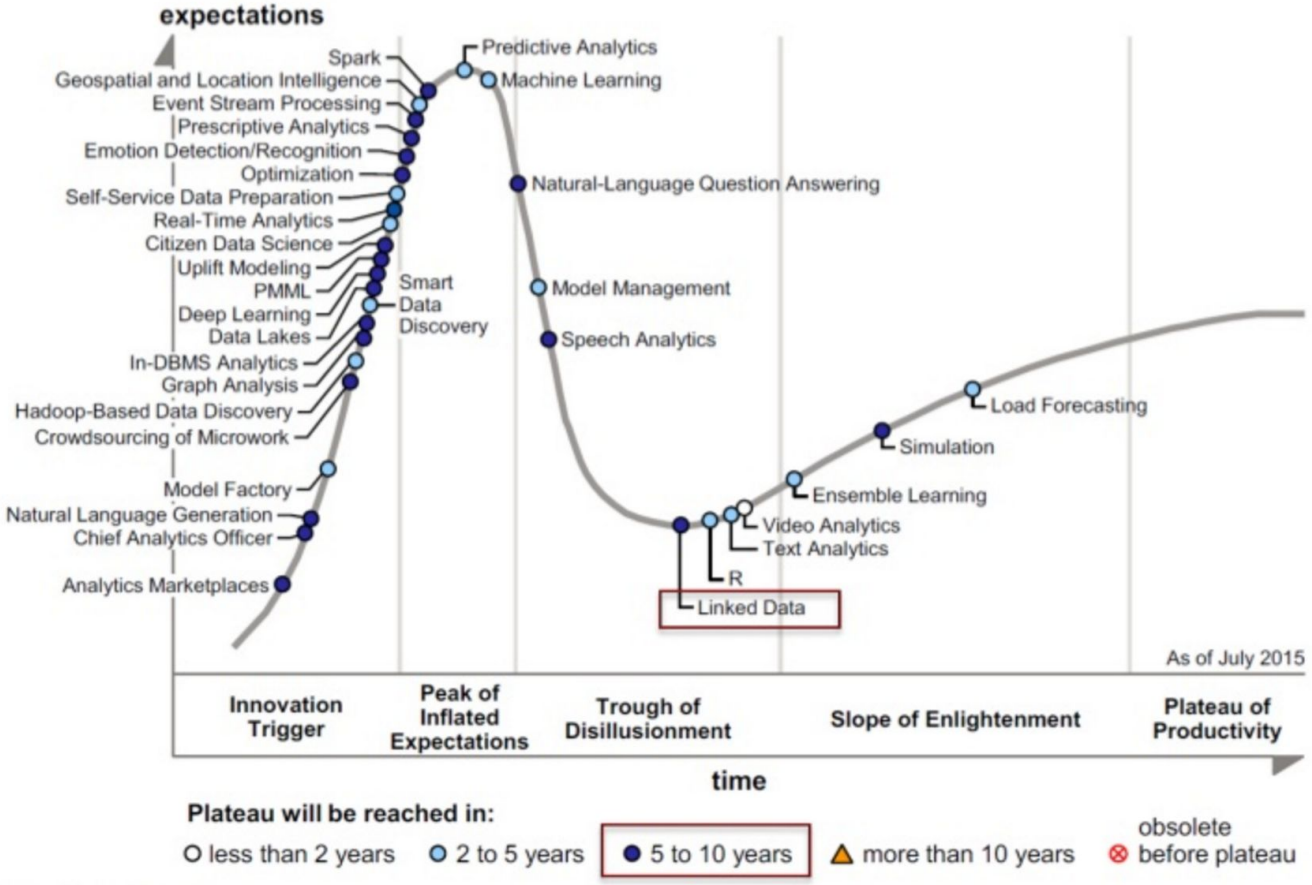
2013



2006

7

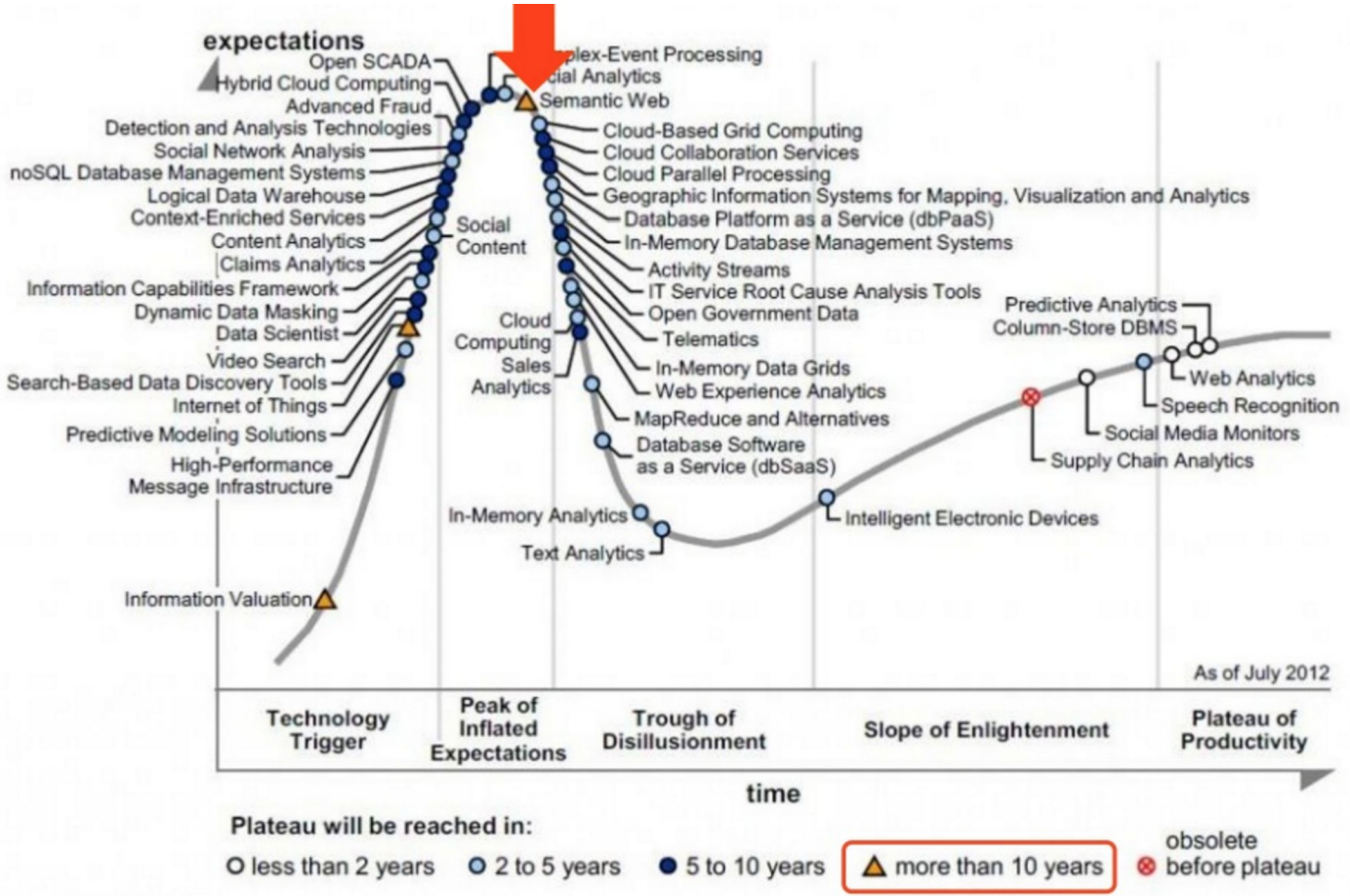
Gartner's Hype Cycle for Advanced Analytics and Data Science, 2015



Source: Gartner (July 2015)

8

Gartner's Hype Cycle for Big Data, 2012

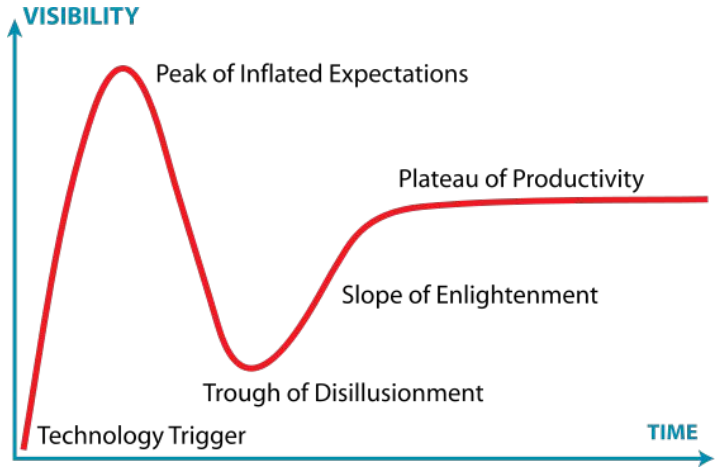


Source: Gartner (July 2012)

9

Escaping from the Trough of Disillusionment

Climbing the Slope of Enlightenment



Slope of Enlightenment:

More instances of how the technology can benefit the enterprise start to crystallize and become more widely understood.

Second- and third-generation products appear from technology providers.

More enterprises fund pilots.

Conservative companies remain cautious.

From: https://en.wikipedia.org/wiki/Hype_cycle

10

Climbing the Slope of Enlightenment:

From Hype to Action - Some ingredients

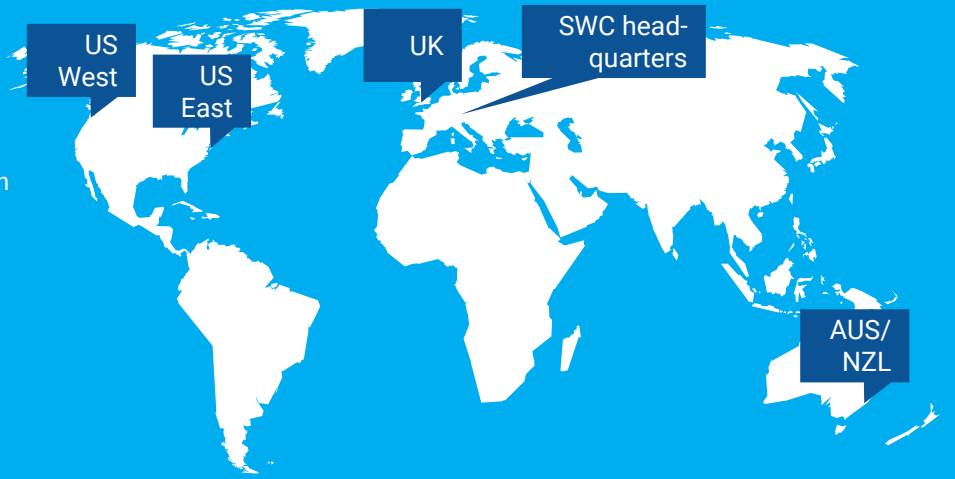
- **Learning from the past**
  Main obstacles for potential users
- **Roll-out scenarios**
  Best practices for implementing linked data on a larger scale
- **Planning a semantic project**
  'Low-hanging fruits' and a concrete action plan
- **Semantic Web eco-systems**
  Data and members, you would like to have linked
- **Complementary technologies**
  Other technologies as enablers for enterprise linked data
- **Integration & Network effects**
  How to engage end-users and the community

11

SELECTED POOLPARTY CUSTOMER REFERENCES AND PARTNERS

Some of our Customer References

- Credit Suisse
- Boehringer Ingelheim
- Roche
- adidas
- The Pokémon Company
- Canadian Broadcasting Corporation
- SwissRe
- Wolters Kluwer
- Bank of America
- HealthStream
- TC Media
- Techtarget
- BMJ Publishing Group
- CafePress
- Pearson - Always Learning
- Education Services Australia
- American Physical Society
- Healthdirect Australia
- World Bank Group
- Inter-American Development Bank
- Renewable Energy Partnership
- Wood MacKenzie
- Oxford University Press
- International Atomic Energy Agency
- Norwegian Directorate of Immigration
- Ministry of Finance (AT)
- Council of the E.U.
- Australian National Data Service



Some of our Partners

- Accenture
- EPAM Systems
- Tellura
- Term Management
- Taxonomy Strategies
- MarkLogic
- Solnet Solutions
- Wolters Kluwer
- Mekon
- Ontotext

12

Emerging technologies:

Supporting the Pioneers

Top-down approach
“Big Bang”
→ Lack of down-to-earth approach

Revolutionists
“Make all things new”
→ Lack of implementation expertise

Sustainers
“Not invented here”
→ Lack of innovative capacity

Bottom-up approach
“Quick win”
→ Lack of strategic relevancy

13

How to disarm Killer Phrases

Top-down approach
- Application scenarios are unclear
- It cannot be integrated
- It cannot be managed

Revolutionists
- It doesn't work with R
- It doesn't support cognitive computing
- Doesn't support NLQA

Sustainers
- It doesn't scale
- It isn't secure
- It isn't well supported

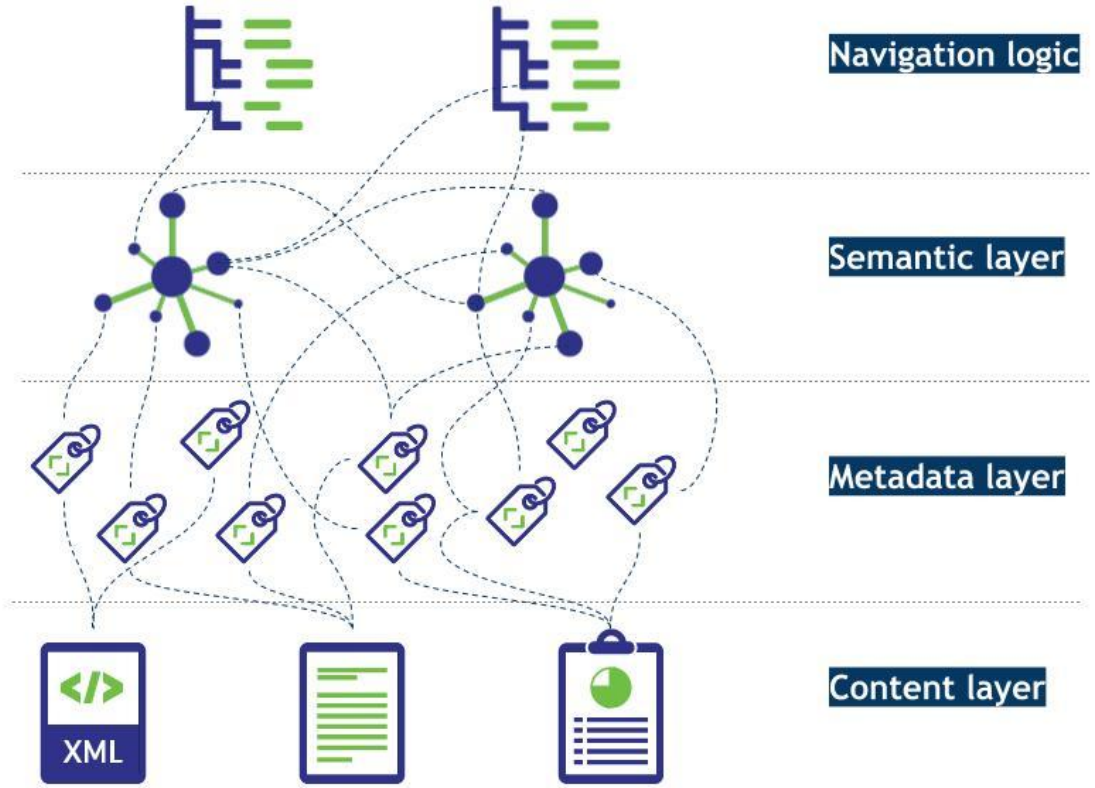
- It's too expensive
- It's too academic
- It cannot be sustained

Bottom-up approach

14

What is ‘it’ really?

Understand and emphasize the fundamental differences



‘Things’ but not Strings:
Using a ‘Semantic Knowledge Graph’

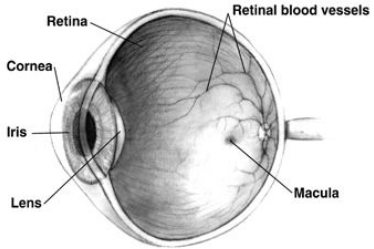
Retina

prefLabel

http://www.my.com/
taxonomy/62346723

image

http://www.my.com/
images/90546089



Home → Medical Encyclopedia → Retina

Retina

The retina is the light-sensitive layer of tissue at the back of the eyeball. Images that come through the eye's lens are focused on the retina. The retina then converts these images to electric signals and sends them along the optic nerve to the brain.

The retina usually looks red or orange because there are many blood vessels right behind it. An ophthalmoscope allows a health care provider to see through your pupil and lens to the retina. Sometimes photos or special scans of the retina can show things that the provider cannot see just by looking at the retina through the ophthalmoscope. If other eye problems block the provider's view of the retina, ultrasound can be used.

Anyone who experiences these vision problems should get a retinal examination:

- Changes in sharpness of vision
- Loss of color perception
- Flashes of light or floaters
- Distorted vision (straight lines look wavy)

Watch this video about:
Retina

Funduscope

prefLabel

http://www.my.com/
taxonomy/
97345854

Ophthalmoscope

altLabel

has broader

http://www.mycom.com
/taxonomy/4543567

prefLabel

Diagnostic Equipment

16

Climbing the Slope of Enlightenment:

Some ingredients

poolparty

- **Learning from the past**
  Main obstacles for potential users
- **Roll-out scenarios**
  Best practices for implementing linked data on a larger scale
- **Planning a semantic project**
  'Low-hanging fruits' and a concrete action plan
- **Semantic Web eco-systems**
  Data and members, you would like to have linked
- **Complementary technologies**
  Other technologies as enablers for enterprise linked data
- **Integration & Network effects**
  How to engage end-users and the community

17

Build capacities first

Top-down approach
- Define benefits and KPIs
- Specify a target architecture
- Define roles and responsibilities

Revolutionists
- Realistic targets
- Semantics as a ‘horizontal’ technology
- Envision something ‘really game-changing’

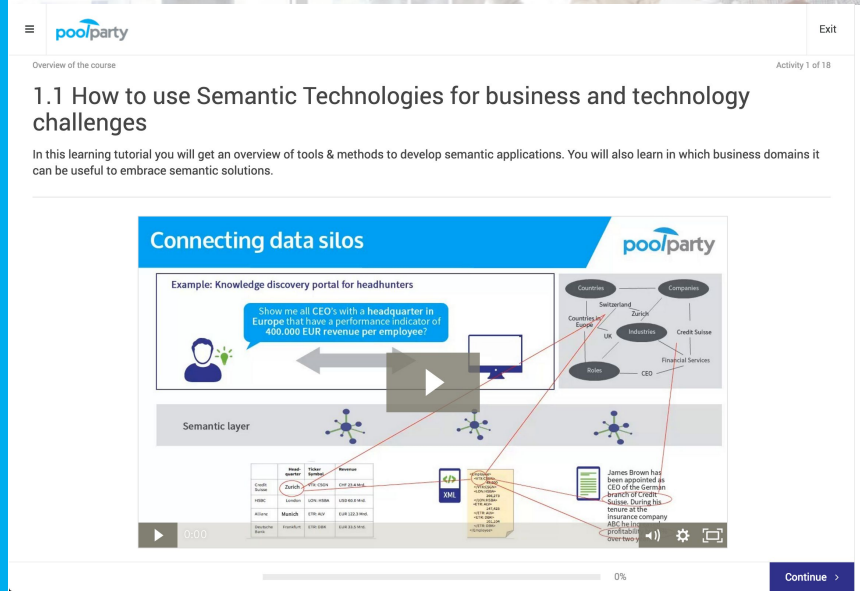
Sustainers
- Software evaluation
- Benchmark tests
- Make a PoC

- Develop a Linked Data Strategy
- Understand the benefits of standards
- See the network effects, overcome departmental thinking

Bottom-up approach

18

PoolParty Academy



- Become productive quickly
- Three e-learning tracks
- Certification program
- Complemented by tailored workshops

- https://www.poolparty.biz/academy/

19

Climbing the Slope of Enlightenment:

Some ingredients

- **Learning from the past**
  Main obstacles for potential users
- **Roll-out scenarios**
  Best practices for implementing linked data on a larger scale
- **Planning a semantic project**
  ‘Low-hanging fruits’ and a concrete action plan
- **Semantic Web eco-systems**
  Data and members, you would like to have linked
- **Complementary technologies**
  Other technologies as enablers for enterprise linked data
- **Integration & Network effects**
  How to engage end-users and the community

Planning a Semantic Project:

Some building blocks in a multi-stakeholder environment

poolparty

Describe the use case / business case

Specify target applications and KPIs

Develop and implement in an agile way

Software Development

Analyse available and relevant data sources

Reuse / build taxonomies and ontologies

Make data available, focus on data quality

Make architectural decisions

Develop a URI strategy

Transform, map and link data

Data Engineering

Transfer knowledge to pioneers / key people

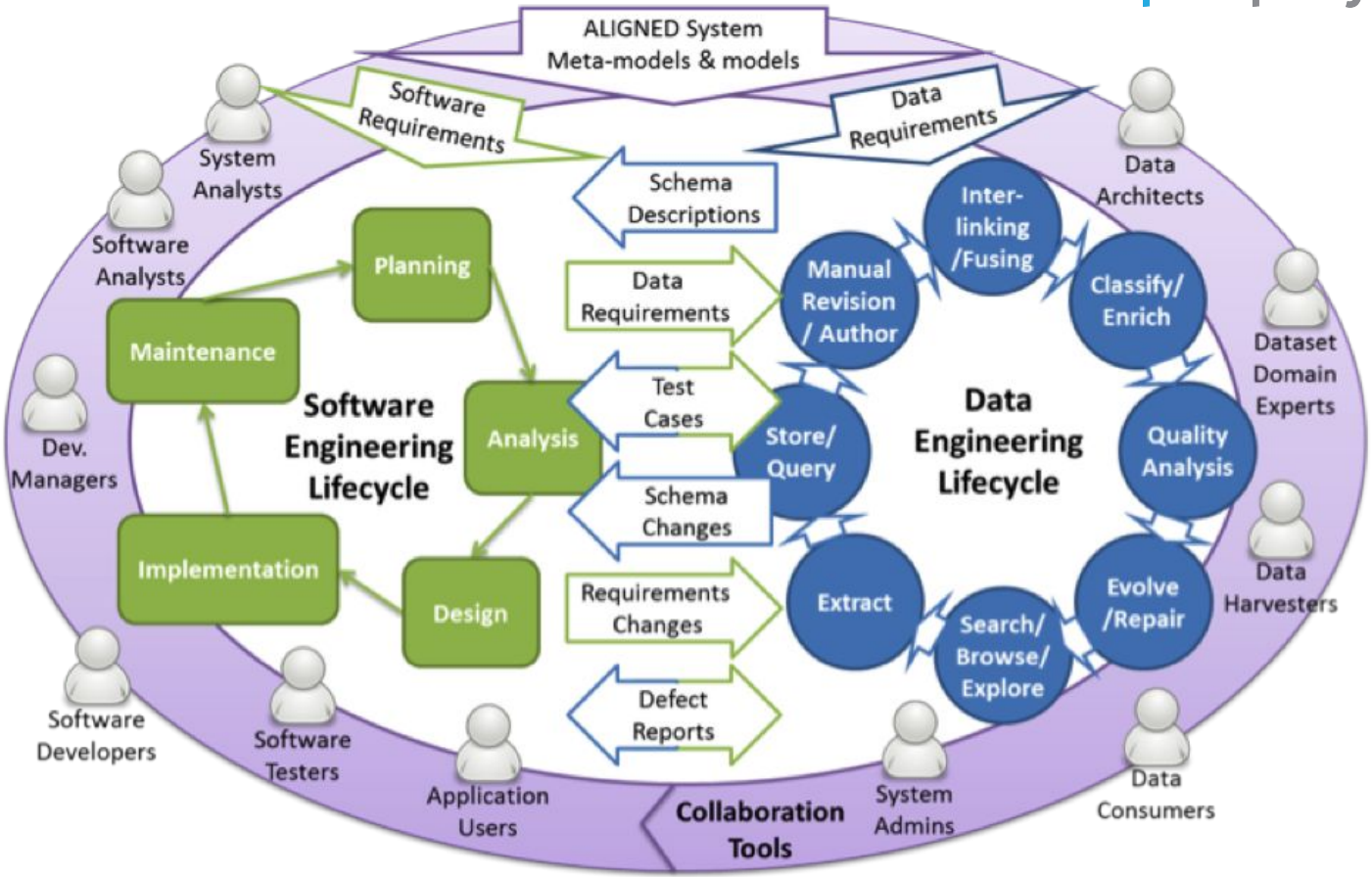
Make workshops / provide e-learning to project team

Explain benefit to the end-users, engage the community

Capacity Building

21

Aligned Software & Data Development Life Cycle



http://aligned-project.eu/

22

Climbing the Slope of Enlightenment:

Some ingredients

- **Learning from the past**
  Main obstacles for potential users
- **Roll-out scenarios**
  Best practices for implementing linked data on a larger scale
- **Planning a semantic project**
  'Low-hanging fruits' and a concrete action plan
- **Semantic Web eco-systems**
  Data and members, you would like to have linked
- **Complementary technologies**
  Other technologies as enablers for enterprise linked data
- **Integration & Network effects**
  How to engage end-users and the community

23

Identify Data Sets and Ontologies

Example for a Linked Data Portfolio

24

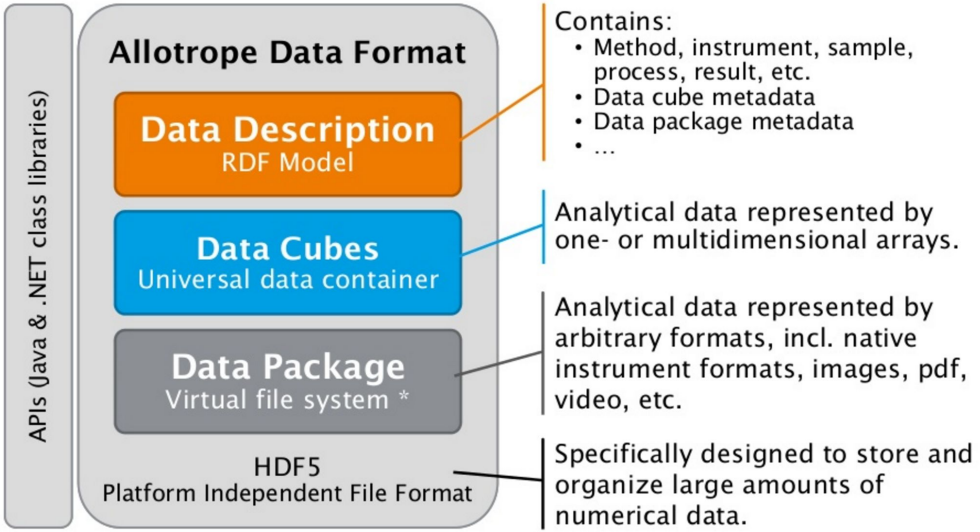
Best practice: Allotrope Foundation

Allotrope Foundation
Building an Open Framework for Laboratory Data

Current Allotrope Foundation members include:

- AbbVie
- Amgen
- Baxter
- Bayer
- Biogen
- Boehringer Ingelheim
- Bristol-Myers Squibb
- Eli Lilly
- Genentech/Roche
- GlaxoSmithKline
- Merck & Co.
- Pfizer

Utilize, implement, and integrate data standards into analytical laboratory workflows.



http://www.allotrope.org/

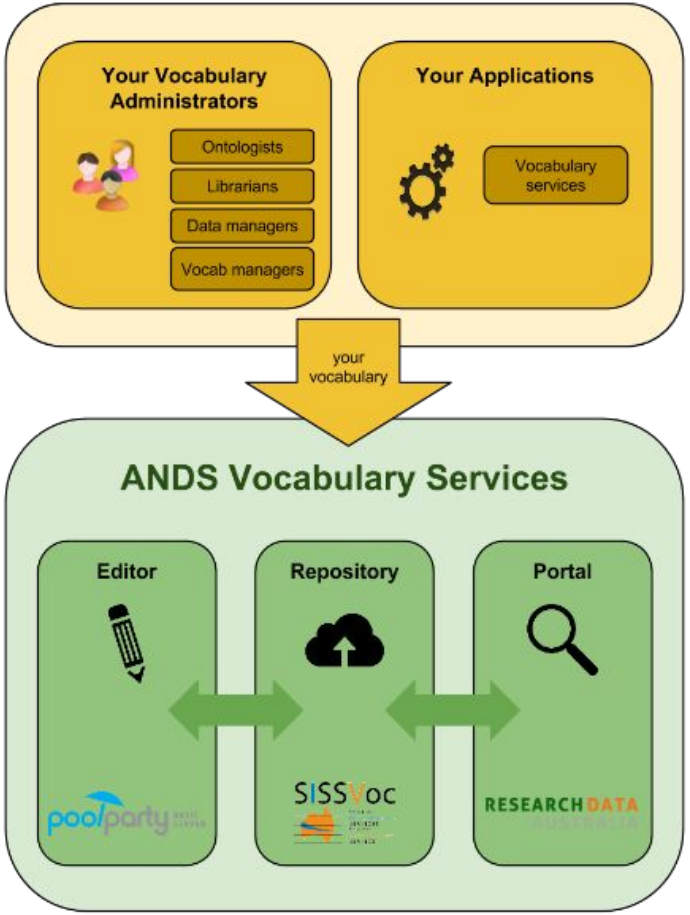
25

Best practice: Australian National Data Service (ANDS)

Current ANDS partners include:

- Over 40 research institutions
- 12 NCRIS data intensive partners
- 6 high performance computing partners
- 17 partners from public sector
- 5 partners from cultural heritage

When research communities agree to use common language for the concepts in datasets, then the discovery, linking, understanding and reuse of research data are improved.



26

Climbing the Slope of Enlightenment:

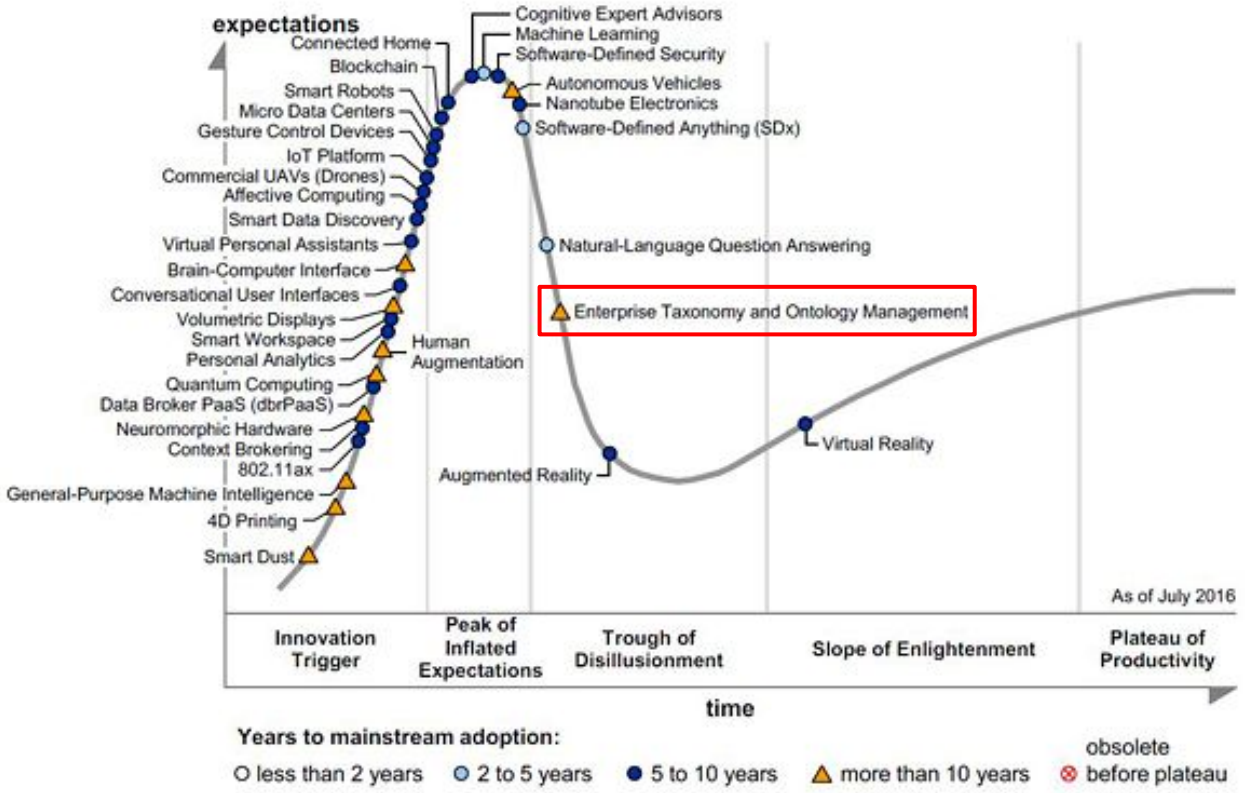
Some ingredients

- **Learning from the past**
  Main obstacles for potential users
- **Roll-out scenarios**
  Best practices for implementing linked data on a larger scale
- **Planning a semantic project**
  'Low-hanging fruits' and a concrete action plan
- **Semantic Web eco-systems**
  Data and members, you would like to have linked
- **Complementary technologies**
  Other technologies as enablers for enterprise linked data
- **Integration & Network effects**
  How to engage end-users and the community

27

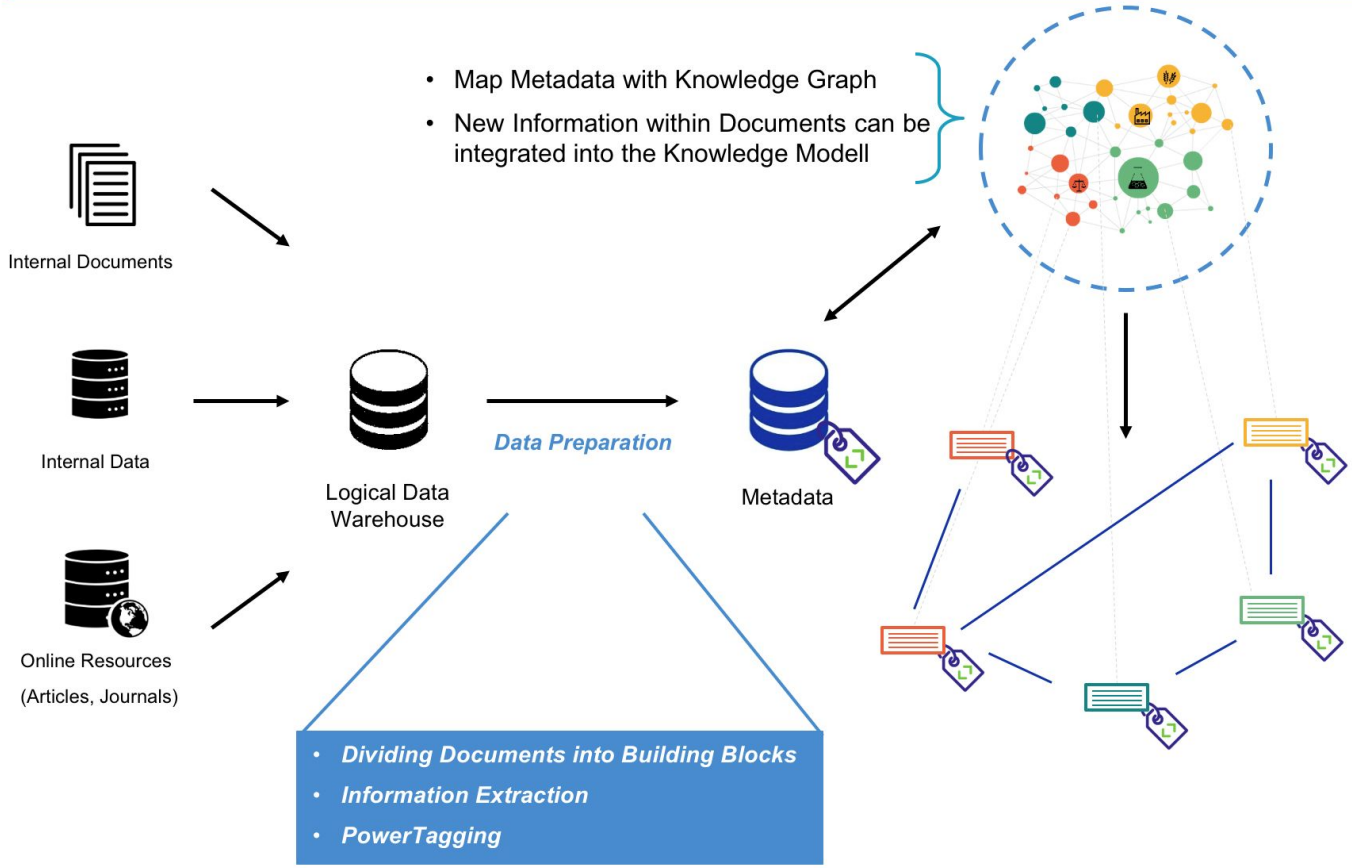
Change is the only constant:

Gartner's Hype Cycle for Emerging Technologies, 2016



Source: Gartner (July 2016)

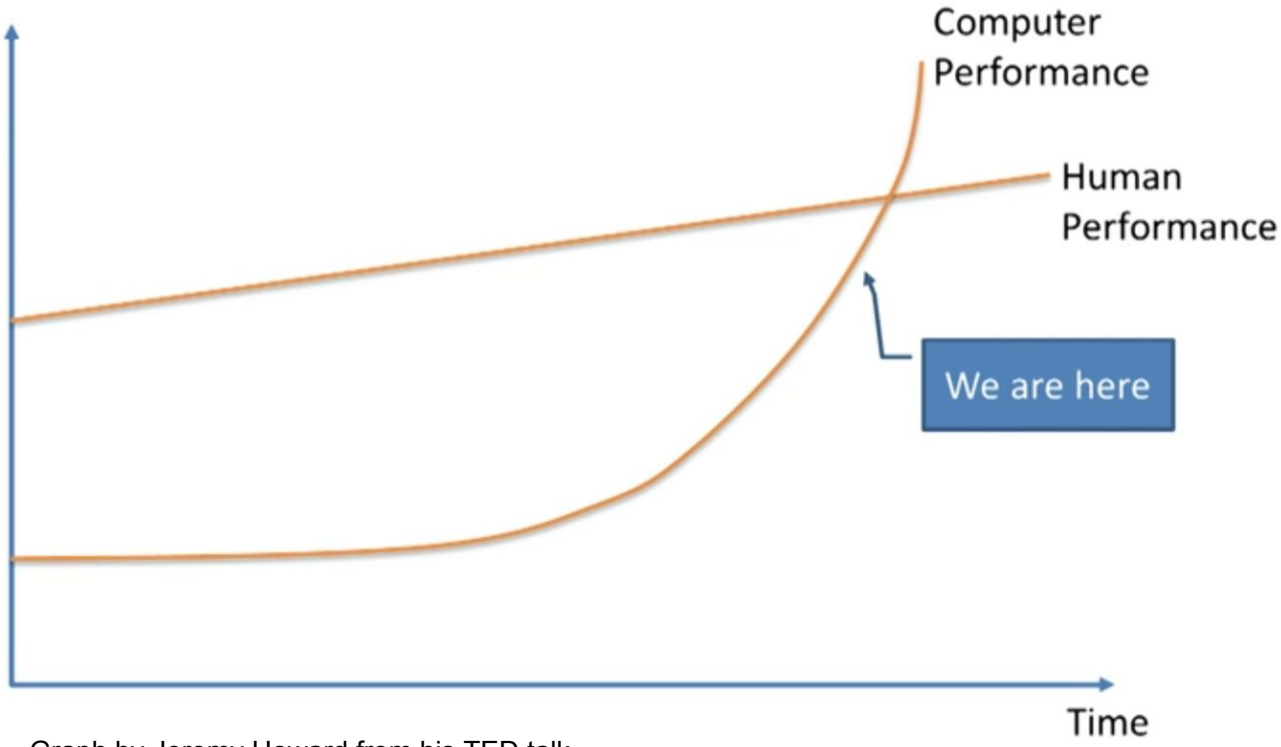
Data Integration & Mapping of Knowledge: Basis for a Cognitive Approach



29

Change the way how you think that change will happen:

When is the AI revolution coming?



Graph by Jeremy Howard from his TED talk "The wonderful and terrifying implications of computers that can learn."

30

Climbing the Slope of Enlightenment:

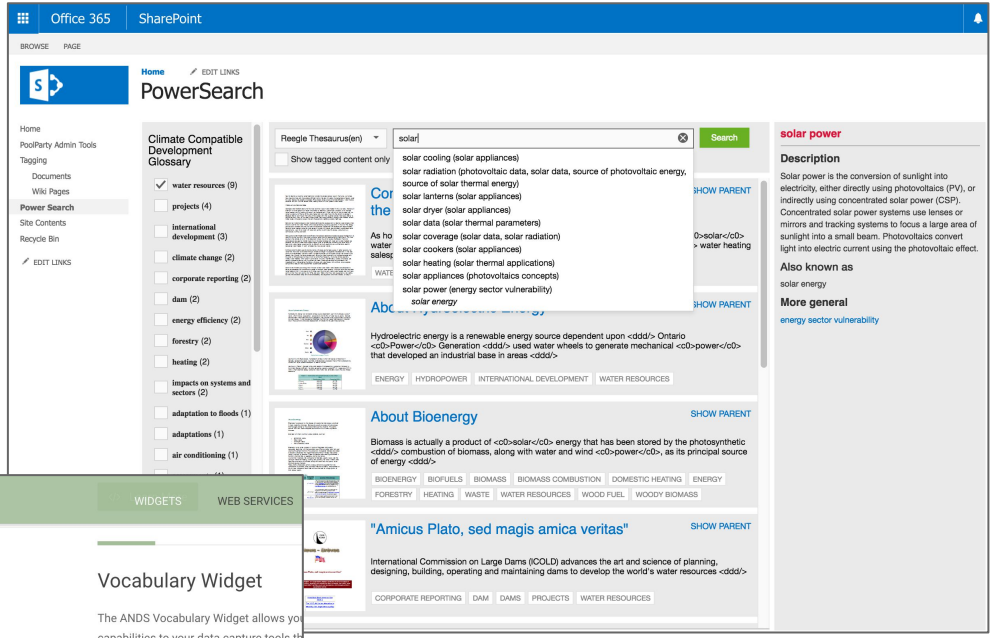
Some ingredients

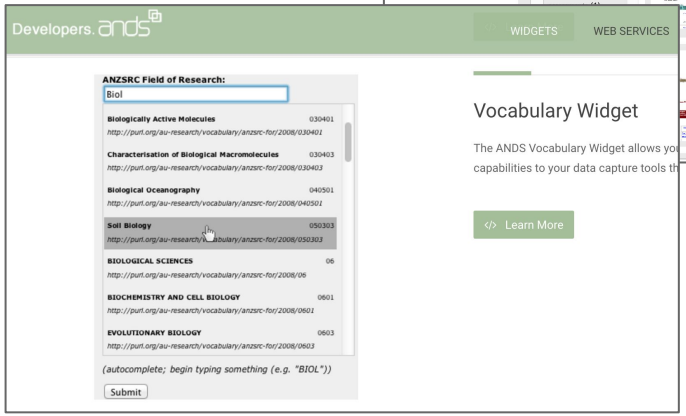
poolparty

- **Learning from the past**
  Main obstacles for potential users
- **Roll-out scenarios**
  Best practices for implementing linked data on a larger scale
- **Planning a semantic project**
  'Low-hanging fruits' and a concrete action plan
- **Semantic Web eco-systems**
  Data and members, you would like to have linked
- **Complementary technologies**
  Other technologies as enablers for enterprise linked data
- **Integration & Network effects**
  How to engage end-users and the community

31

Make services easily accessible





32

CONNECT

Andreas Blumauer

CEO, Semantic Web Company

- a.blumauer@semantic-web.at
- http://at.linkedin.com/in/andreasblumauer
- https://twitter.com/semwebcompany
- https://ablvienna.wordpress.com/