

Linked Open Data at Springer Nature

Nov 2016

Michele Pasin

Lead Data Architect

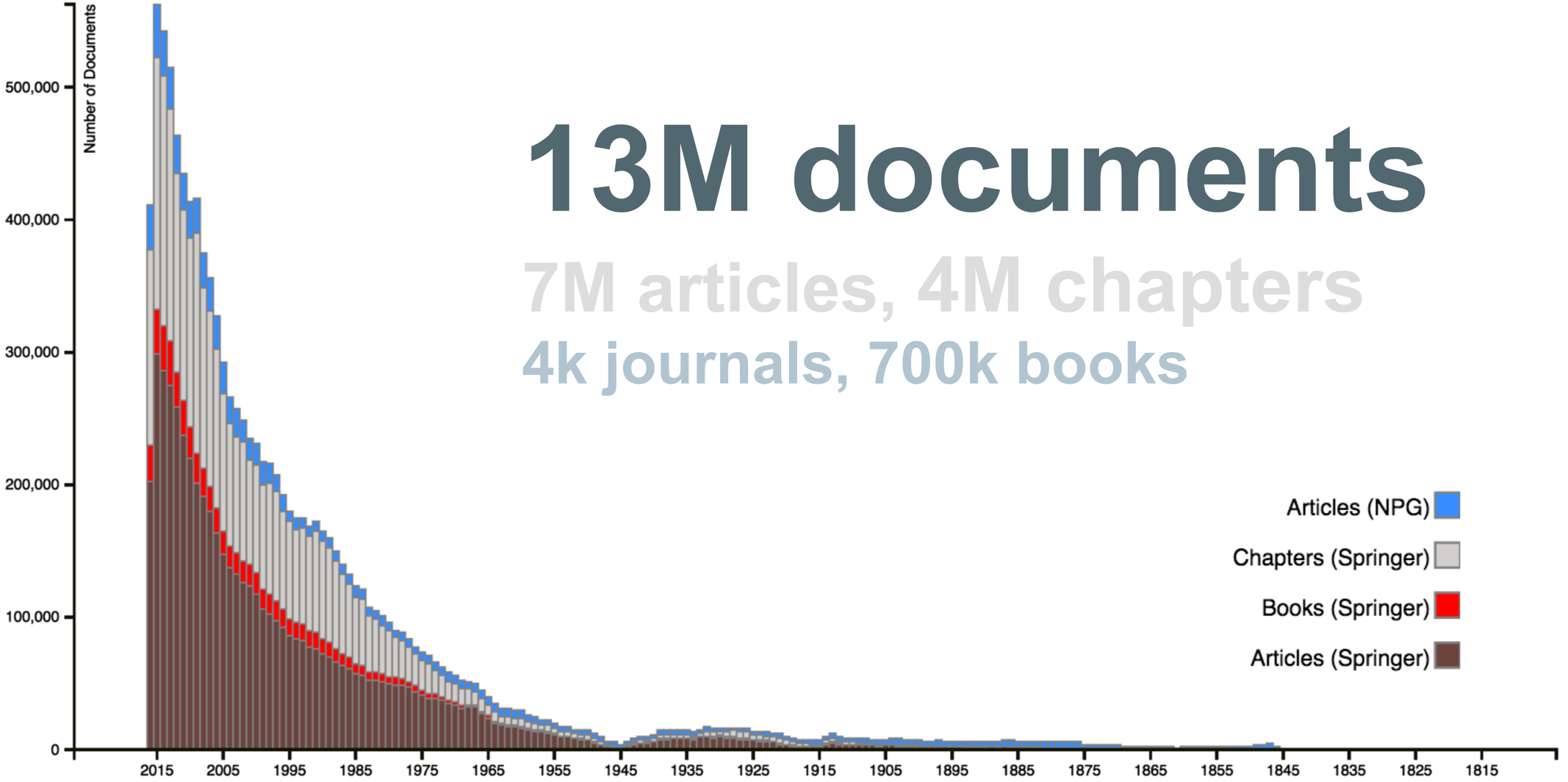
SPRINGER NATURE

SPRINGER NATURE



Formed in **May 2015** through the **merger** of Nature Publishing Group, Palgrave Macmillan, Macmillan Education and Springer Science+Business Media

We publish a lot of science! (since 1815)



13M documents

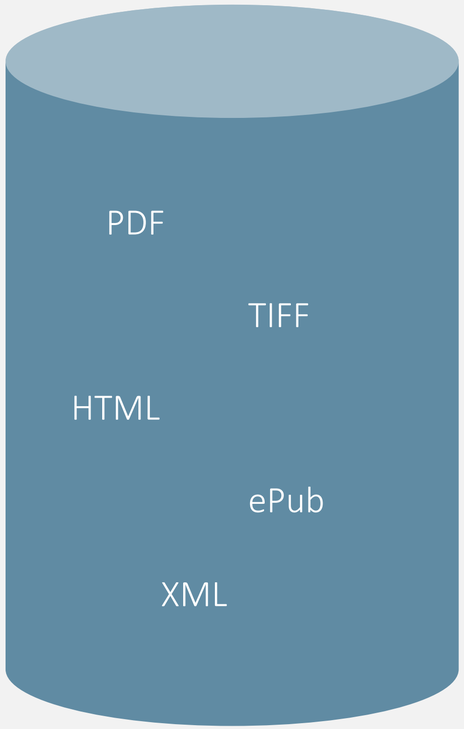
7M articles, 4M chapters

4k journals, 700k books

SPRINGER NATURE

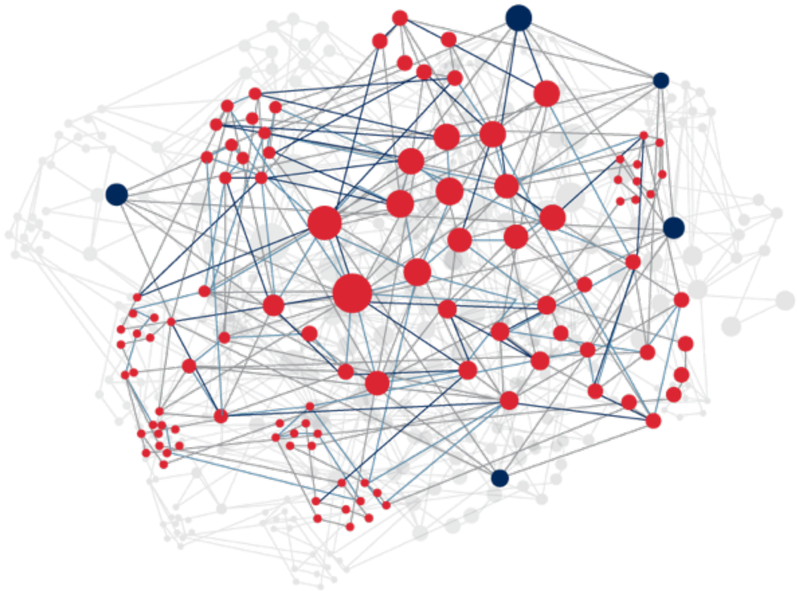
Vision

Today: Content base



We publish science

Tomorrow: Knowledge Graph



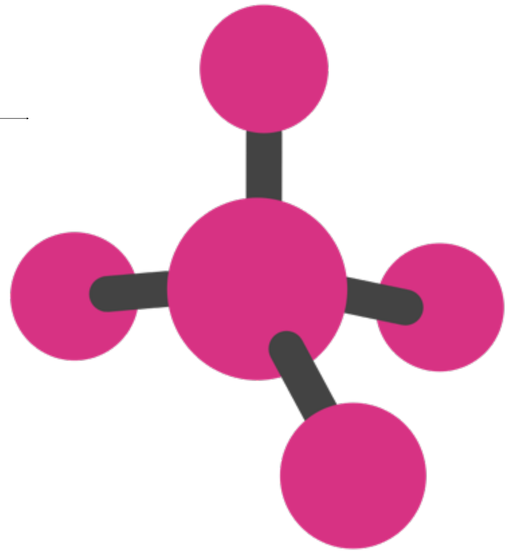
We manage knowledge

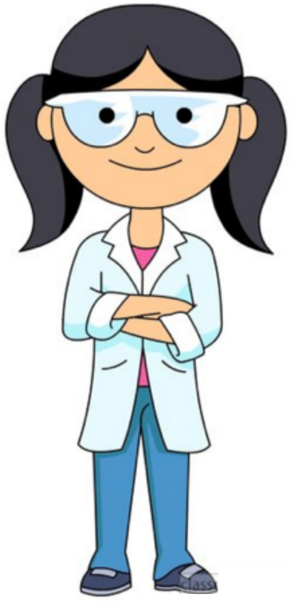
SPRINGER NATURE

Three areas of knowledge we care about

is about



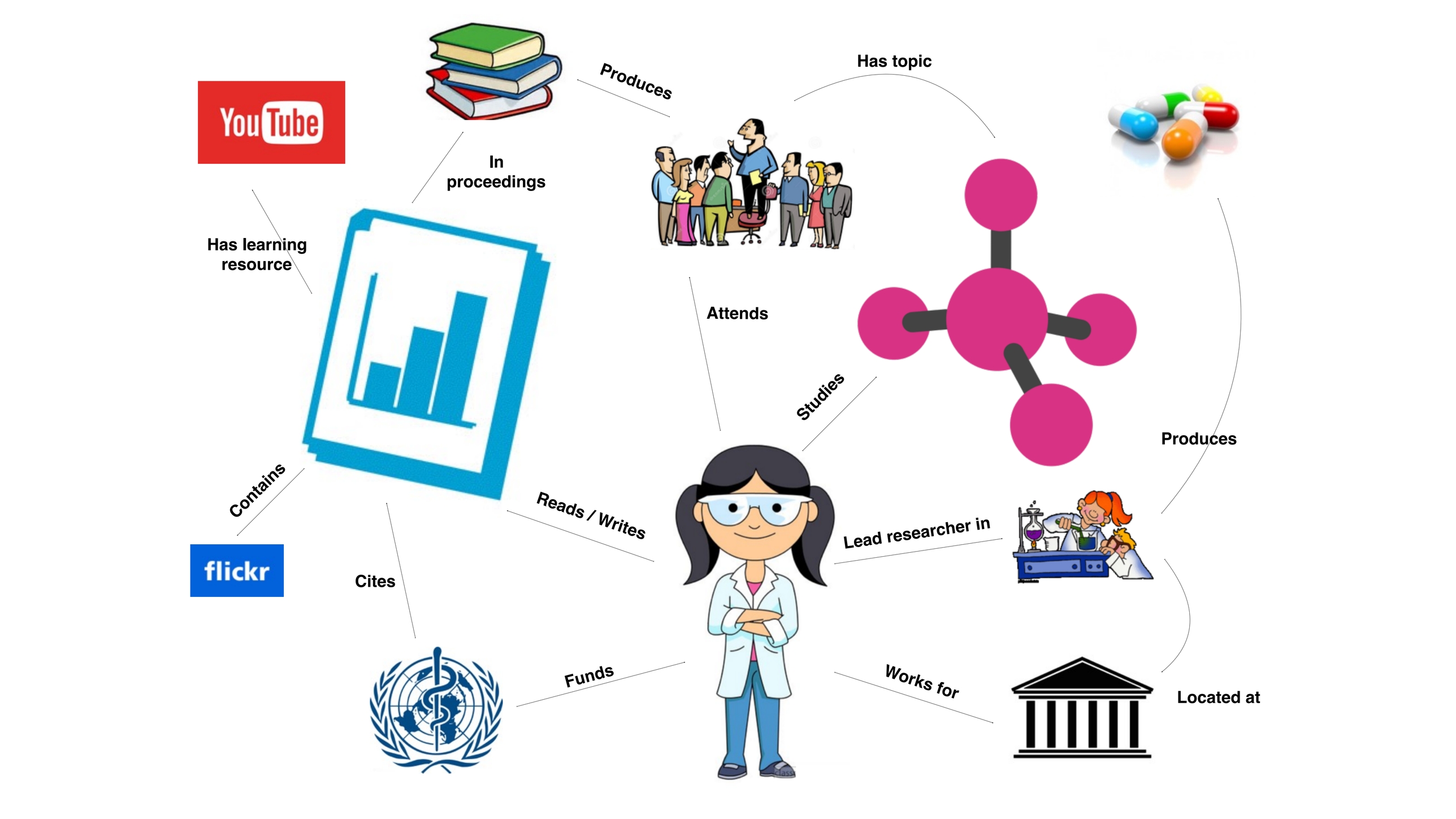




reads / writes

interested in

SPRINGER NATURE


Produces
Has topic
In proceedings
Has learning resource
Attends
Studies
Contains
Reads / Writes
Lead researcher in
Produces
Cites
Funds
Works for
Located at

We've come a long way...

2012

2013

2014

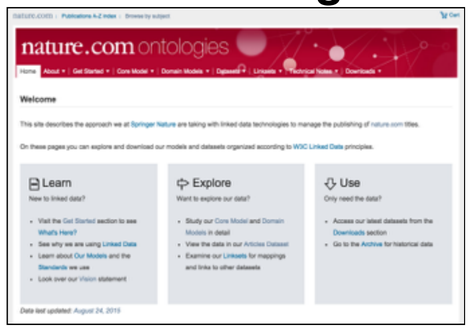
2015

2016

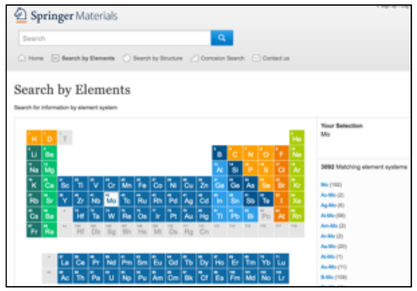
NPG Linked Data Platform



Nature Ontologies



Springer Materials



Springer Conferences



Scigraph.com



..and many more!

Why Linked Data

- **Enterprise metadata** management
  - addressing the data fragmentation problem
  - promoting a culture of data governance and reuse
- Content **discoverability**
  - search: handling morphological variations, synonyms, generalisations etc…
  - browse: more detailed and interesting discovery pathways
  - dynamic semantic publishing: repurposing content based on context and metadata
- We believe in **open science**
  - science is a *global* collaborative effort that we want to support!
  - *Open Access* for readers, *Open Peer Review* & *Open Choice* for authors
  - Linked Open Data publishing (since 2012)

Challenges

- **Developers**: lack of interest for RDF
  - too ‘academic’, too many namespaces
  - solution: focus on data integration, **JSON-LD**



- **Scalability** & performance
  - semantic graph databases are less mature than conventional technologies
  - solution: **GraphDB** + **Elasticsearch**

GraphDB by ontotext

- **Collaboration** among publishers
  - business concerns: ‘sharing our data, we’re losing competitive advantage’
  - solution: keep healthy **balance** between publishing Linked Open Data and private Linked Data
  - solution: importance of neutral, third party **mediators**, e.g. CrossRef with conference identifiers

Crossref
Open Funder Registry



Thanks

email :

michele.pasin@nature.com