SN SciGraph

Building a Linked Open Data Knowledge Graph

Michele Pasin

London Info International 2017
December2017

SPRINGER NATURE

Springer Nature's Metadata Mission Statement

We understand metadata as the gateway to our content.

We provide best quality metadata with state-of-the-art enrichment in all key formats and flavors, available in all relevant delivery models.

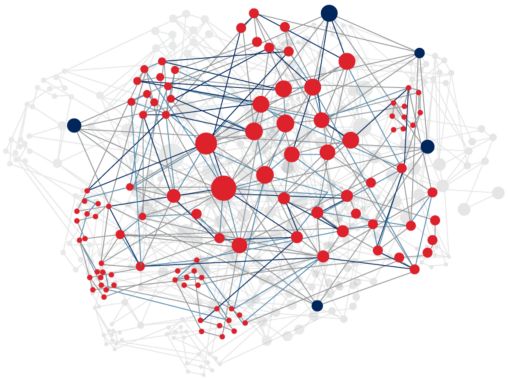
Our bibliographic metadata is free, open and re-usable.

SPRINGER NATURE

Springer Nature SciGraph

A Linked Open Data platform for the scholarly domain

SN SciGraph

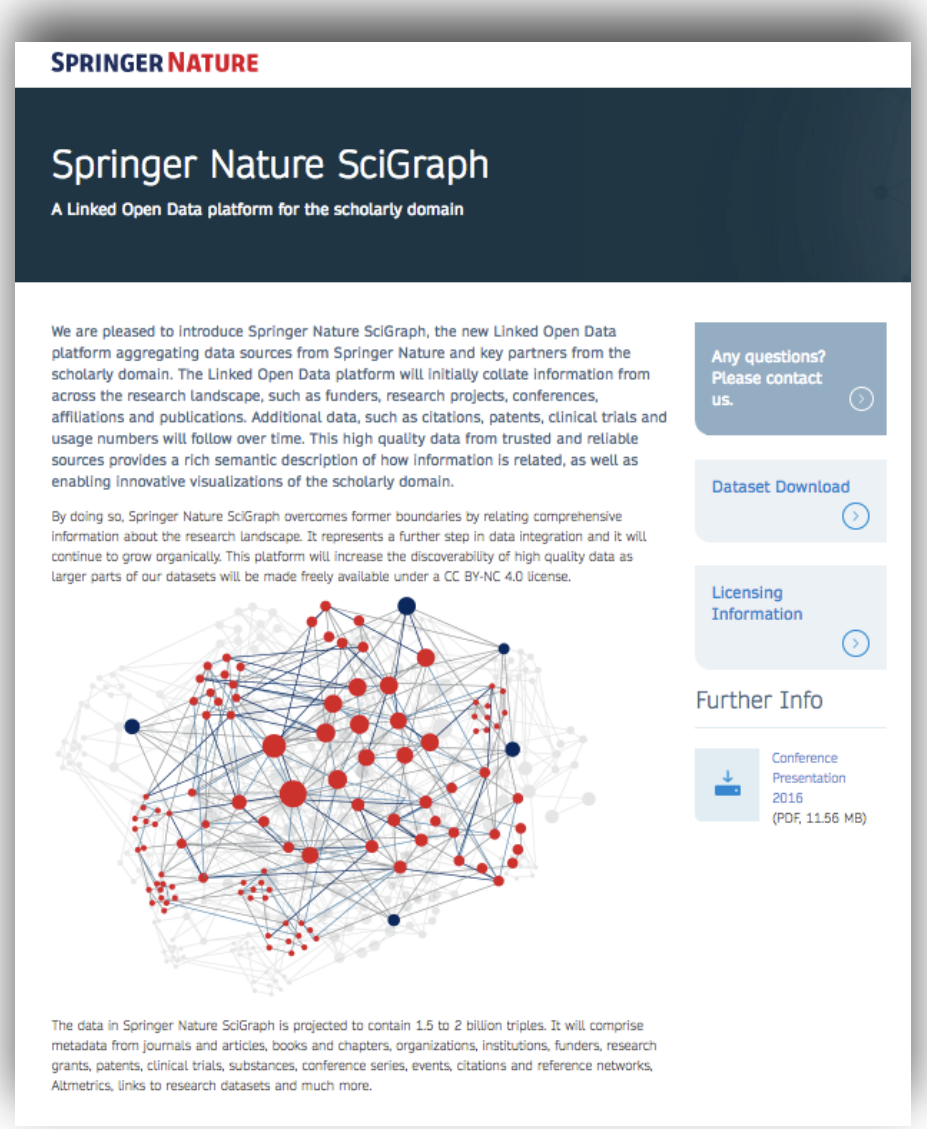


> Collaborative effort between Springer Nature and Digital Science

> Supporting internal use cases, but also contributing to an emerging web of **linked open science data**

> Not just publications data but a wealth of further related information

Linked Open Data Publishing So Far



SPRINGER NATURE

Springer Nature SciGraph

A Linked Open Data platform for the scholarly domain

We are pleased to introduce Springer Nature SciGraph, the new Linked Open Data platform aggregating data sources from Springer Nature and key partners from the scholarly domain. The Linked Open Data platform will initially collate information from across the research landscape, such as funders, research projects, conferences, affiliations and publications. Additional data, such as citations, patents, clinical trials and usage numbers will follow over time. This high quality data from trusted and reliable sources provides a rich semantic description of how information is related, as well as enabling innovative visualizations of the scholarly domain.

By doing so, Springer Nature SciGraph overcomes former boundaries by relating comprehensive information about the research landscape. It represents a further step in data integration and it will continue to grow organically. This platform will increase the discoverability of high quality data as larger parts of our datasets will be made freely available under a CC BY-NC 4.0 license.

Any questions? Please contact us.

Dataset Download

Licensing Information

Further Info

Conference Presentation 2016 (PDF, 11.56 MB)

The data in Springer Nature SciGraph is projected to contain 1.5 to 2 billion triples. It will comprise metadata from journals and articles, books and chapters, organizations, institutions, funders, research grants, patents, clinical trials, substances, conference series, events, citations and reference networks, Altmetrics, links to research datasets and much more.

At a glance:

- 150 M triples / 32G downloads
- CC-BY-NC license

Metadata about:

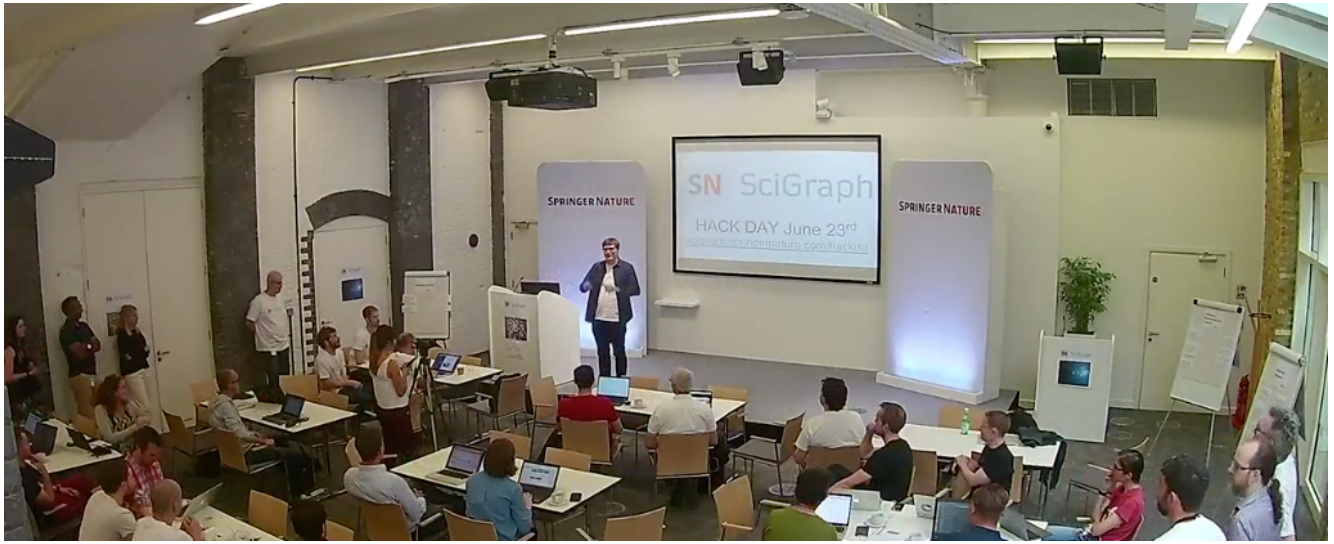
- Articles 2012-2016 (5M) + Abstracts
- Grants (200k)
- Journals (3k)
- Subjects (3k)
- Core Ontology

www.springernature.com/scigraph

Open Data Events: Hack Day June 2017

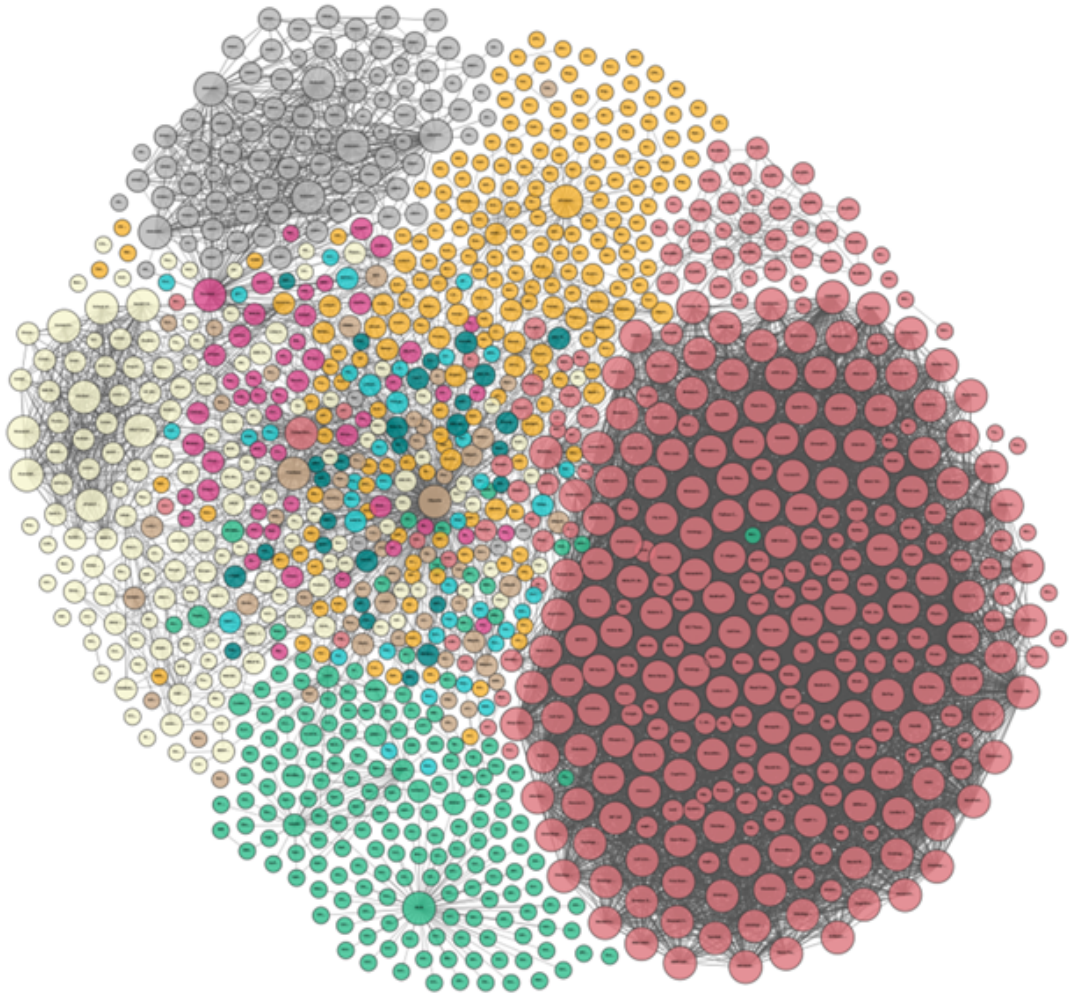
Aims and Scope

- Engagement with Linked Data Researcher Community
- Encourage developers to build tools with our data
- Position ourselves as Open Data research publisher
- Gather first-hand feedback from users of our data



Web of Data

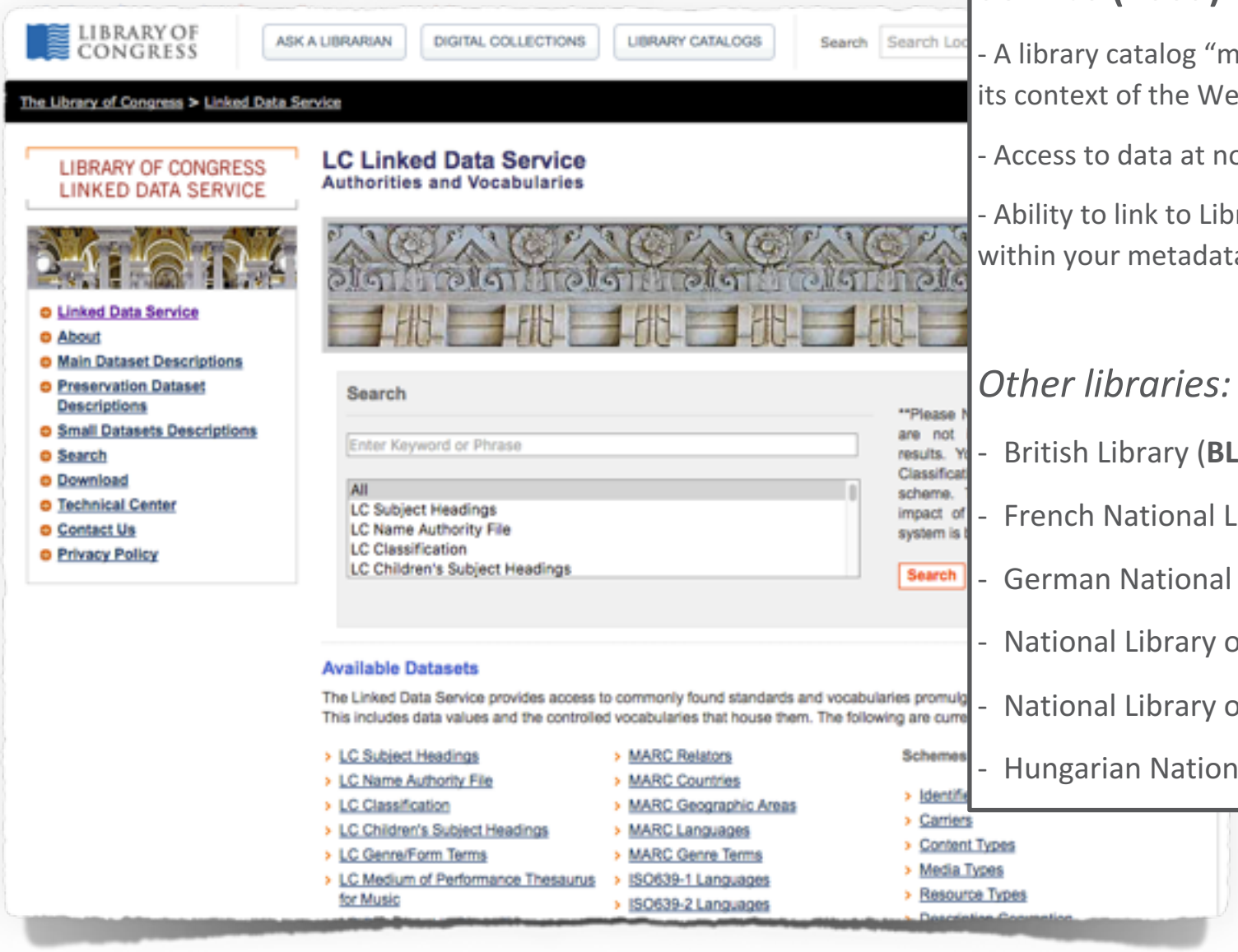
- Be part of the LOD cloud!
- The future library and trade catalog is the Internet!



Linking Open Data cloud diagram 2017, by Andrejs Abele, John P. McCrae, Paul Buitelaar, Anja Jentzsch and Richard Cyganiak. http://lod-cloud.net (CC-BY-SA)

SPRINGER NATURE

Libraries using Linked Data



Library of Congress Linked Data Service (2009)

- A library catalog “must be designed by considering its context of the Web”

- Access to data at no cost.

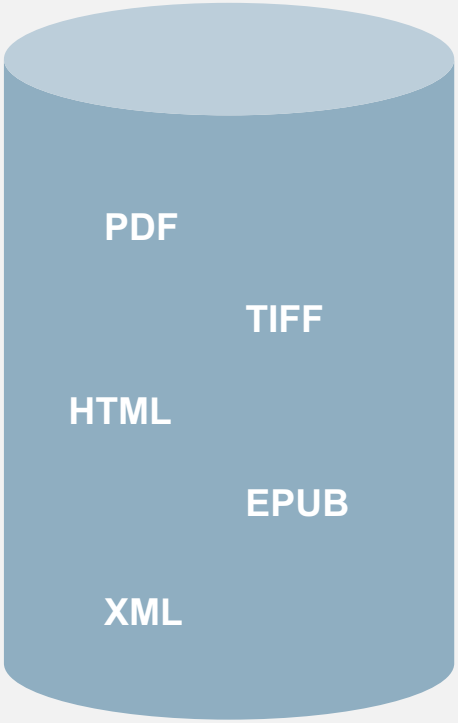
- Ability to link to Library of Congress data values within your metadata via Linked Data.

Other libraries:

- British Library (**BL**)
- French National Library (**BNF**)
- German National Library (**DNB**)
- National Library of Spain (**BNE**)
- National Library of Sweden (**LIBRIS**)
- Hungarian National Library (**NSL**)

Vision: From Content to Data

Content base



We publish content

Knowledge Graph

We manage knowledge

SPRINGER NATURE

Vision: From Content to Data

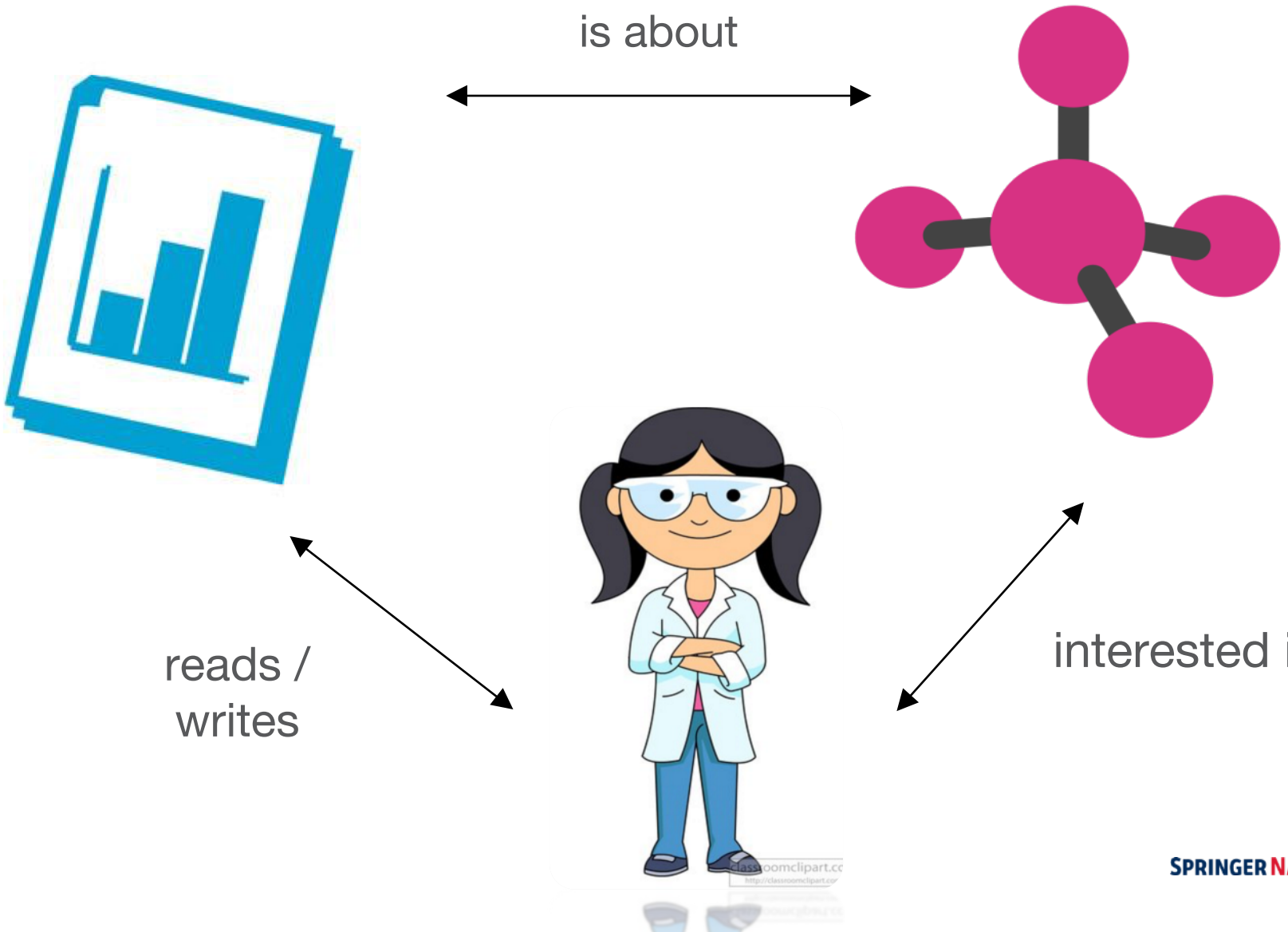
- We create the largest state-of-the-art linked open data aggregation platform for the scholarly domain.
- In doing so, we increase content discoverability and provide data tools and services for researchers, authors, editors, librarians, data scientists, funders, conference organizers, and many others by adding value across all content types.

We publish content

We manage knowledge

SPRINGER NATURE

Three areas of knowledge we care about

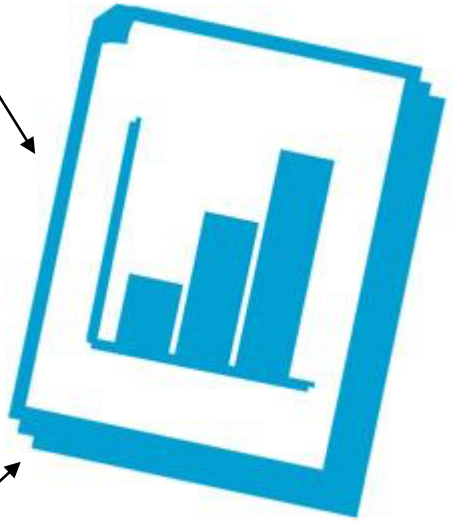


Produces

Has topic

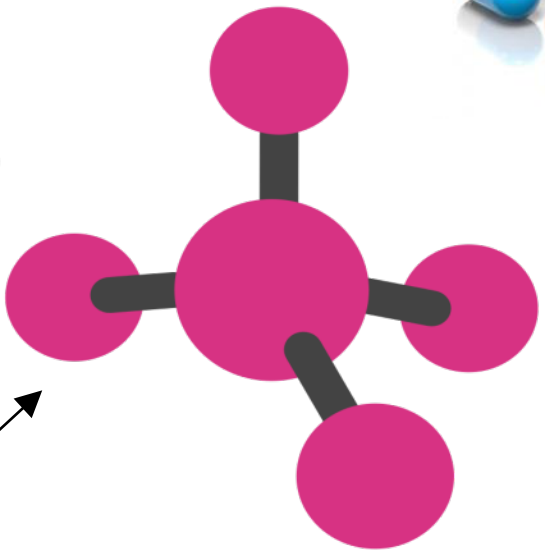
In proceedings

Has learning resource



Attends

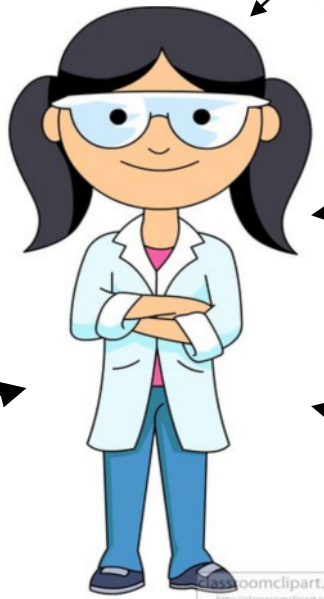
Studies



Produces

Contains

Reads / Writes



Lead researcher in

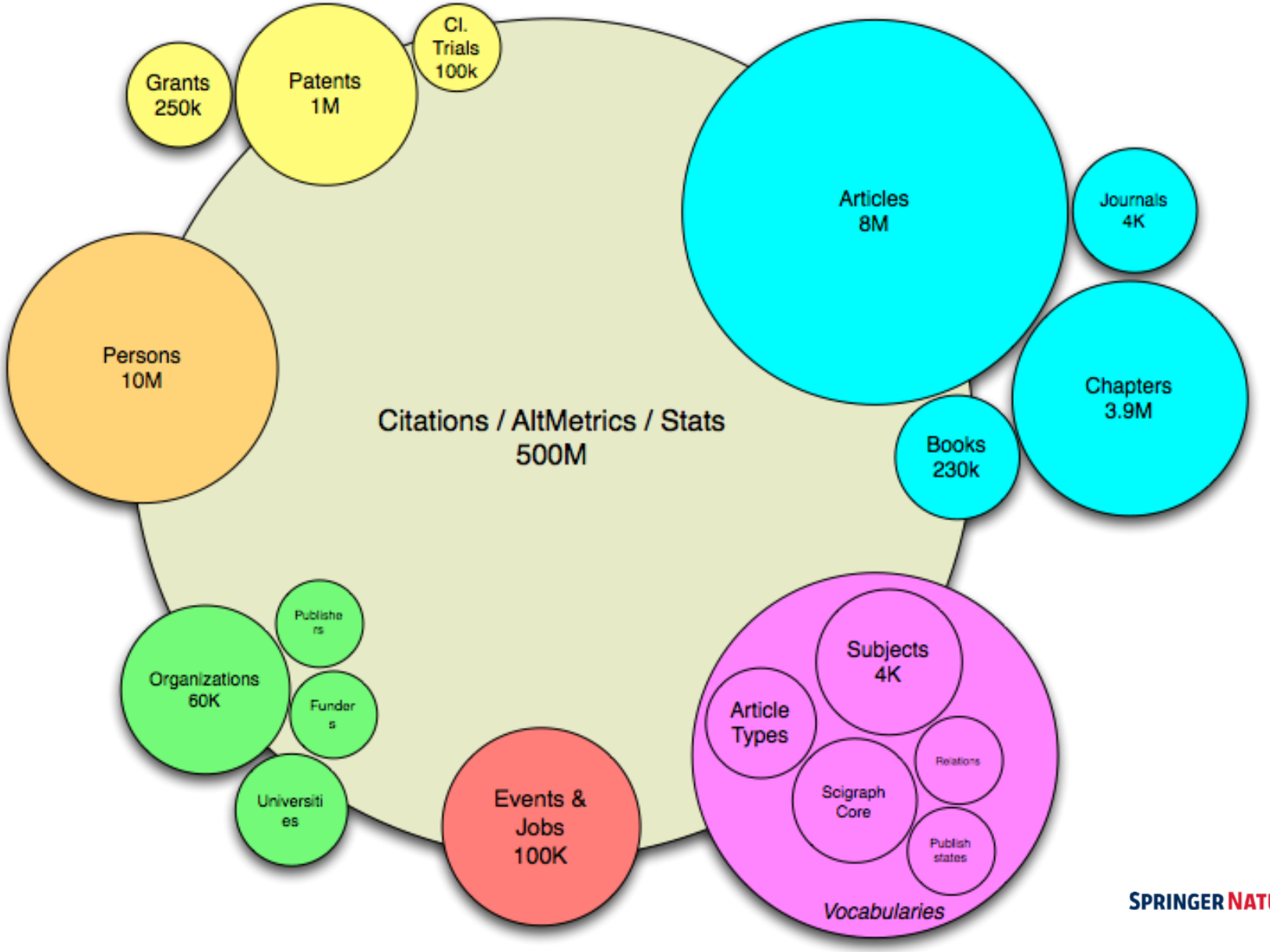
Cites

Funds

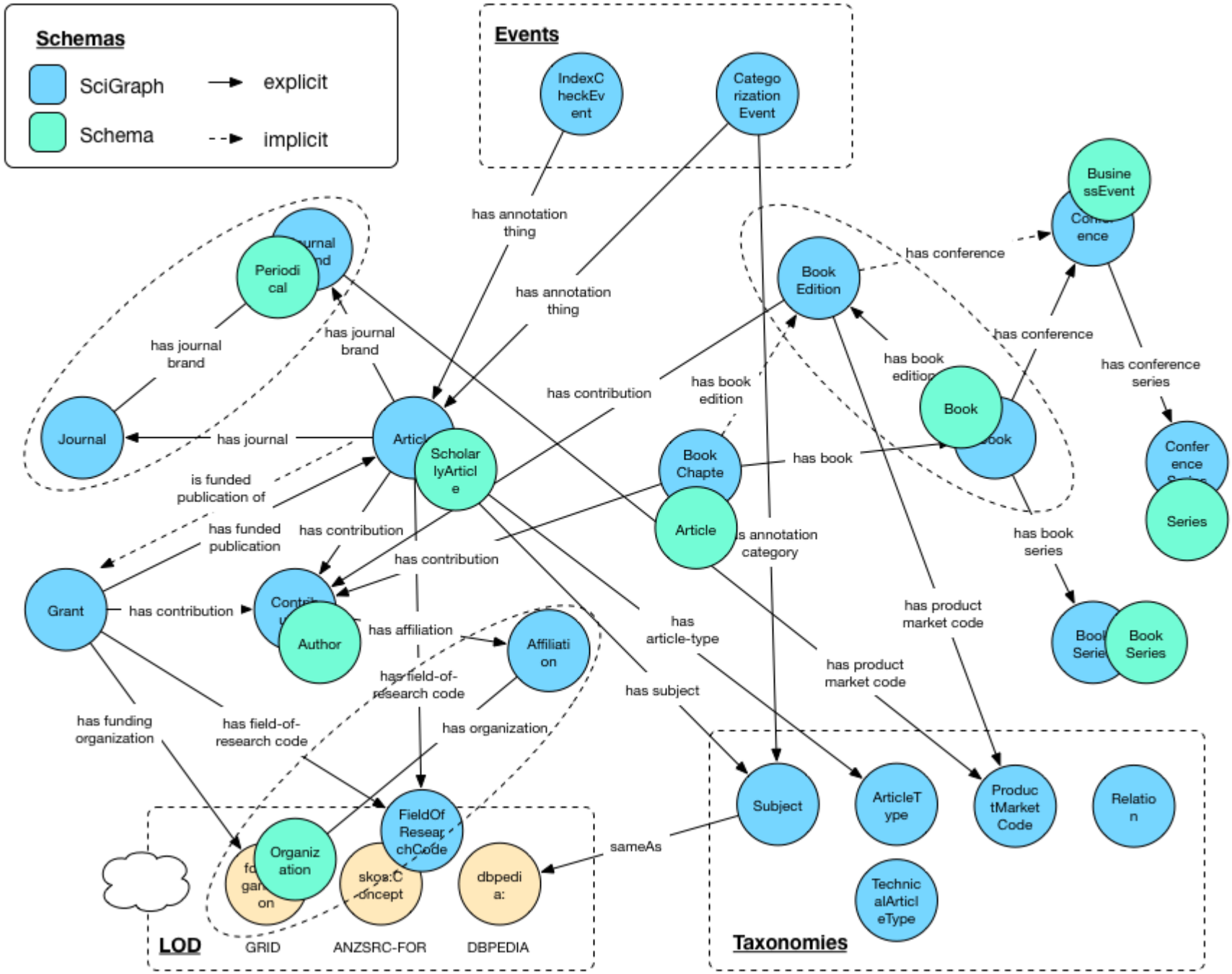
Works for

Located at

Springer Nature SciGraph Data Landscape



SPRINGER NATURE


Schemas
SciGraph
Schema
explicit
implicit
Events
IndexCheckEvent
Categorization Event
BusinessEvent
Conference
has annotation thing
has annotation thing
has conference
has conference
has conference series
Periodical
has journal brand
has journal brand
Book Edition
has book edition
has book edition
Book
Journal
has journal
has contribution
Article
ScholarlyArticle
Book Chapter
has book
Conference Series
is funded publication of
has funded publication
has contribution
has contribution
has contribution
Article
has annotation category
has book series
Grant
has contribution
Author
has affiliation
Affiliation
has field-of-research code
has organization
has article-type
has product market code
has product market code
Book Series
has funding organization
has field-of-research code
has subject
Subject
ArticleType
ProductMarketCode
Relation
Organization
FieldOfResearchCode
sameAs
TechnicalArticleType
LOD
GRID
ANZSRC-FOR
DBPEDIA
Taxonomies

Springer Nature SciGraph Applications

Springer Nature Scigraph: Applications

Linked Open Data Publishing

- **Researchers** can analyze/build upon our data
- Contributing to Open Research

Business Intelligence and Analytics

- Dashboards for understanding the research landscape
- **Editors**, **Sales**, **Marketing** etc.

Content Discoverability

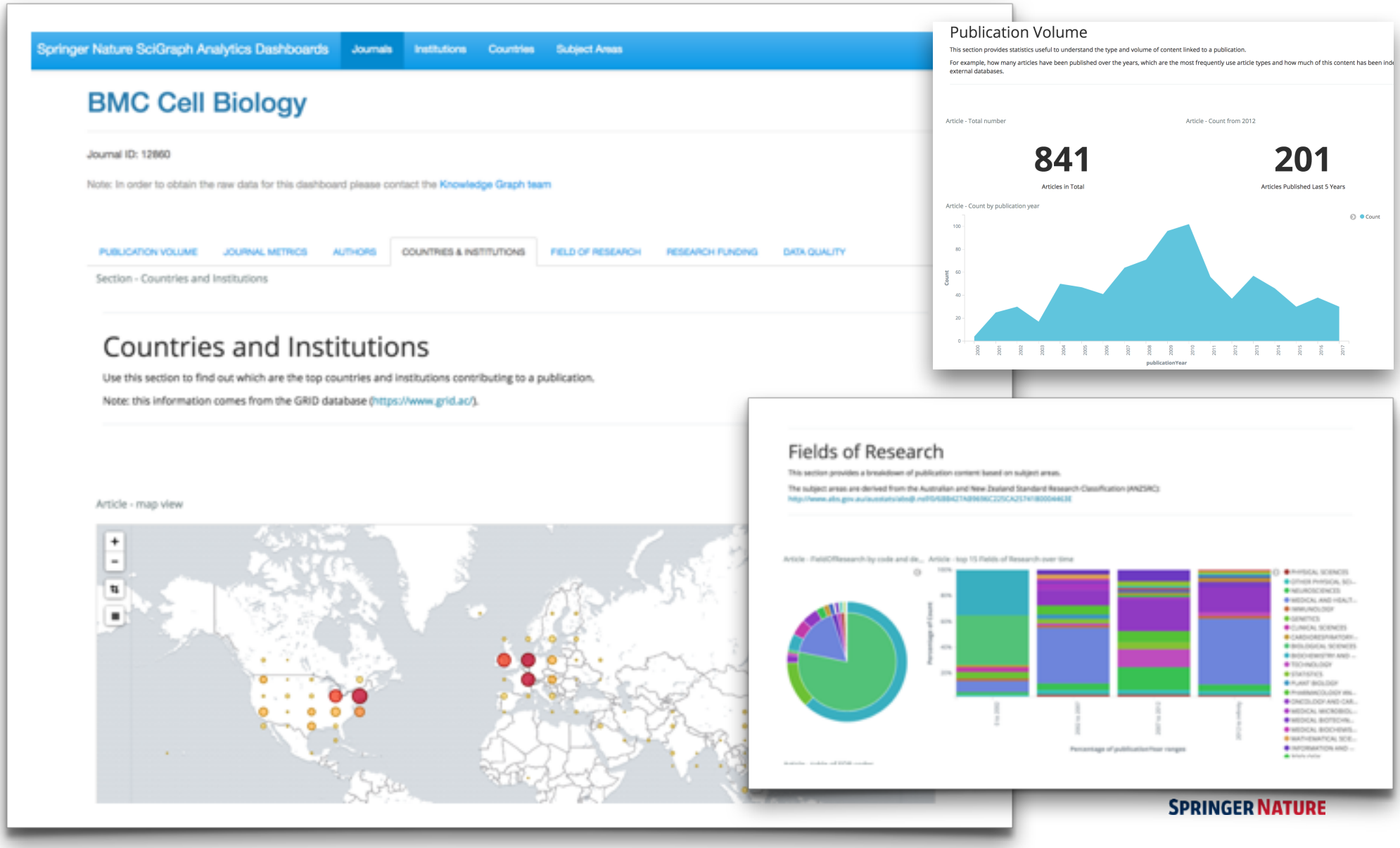
- SN SciGraph Data Explorer
- APIs for better **end user** applications

Applications

Analytics Dashboards

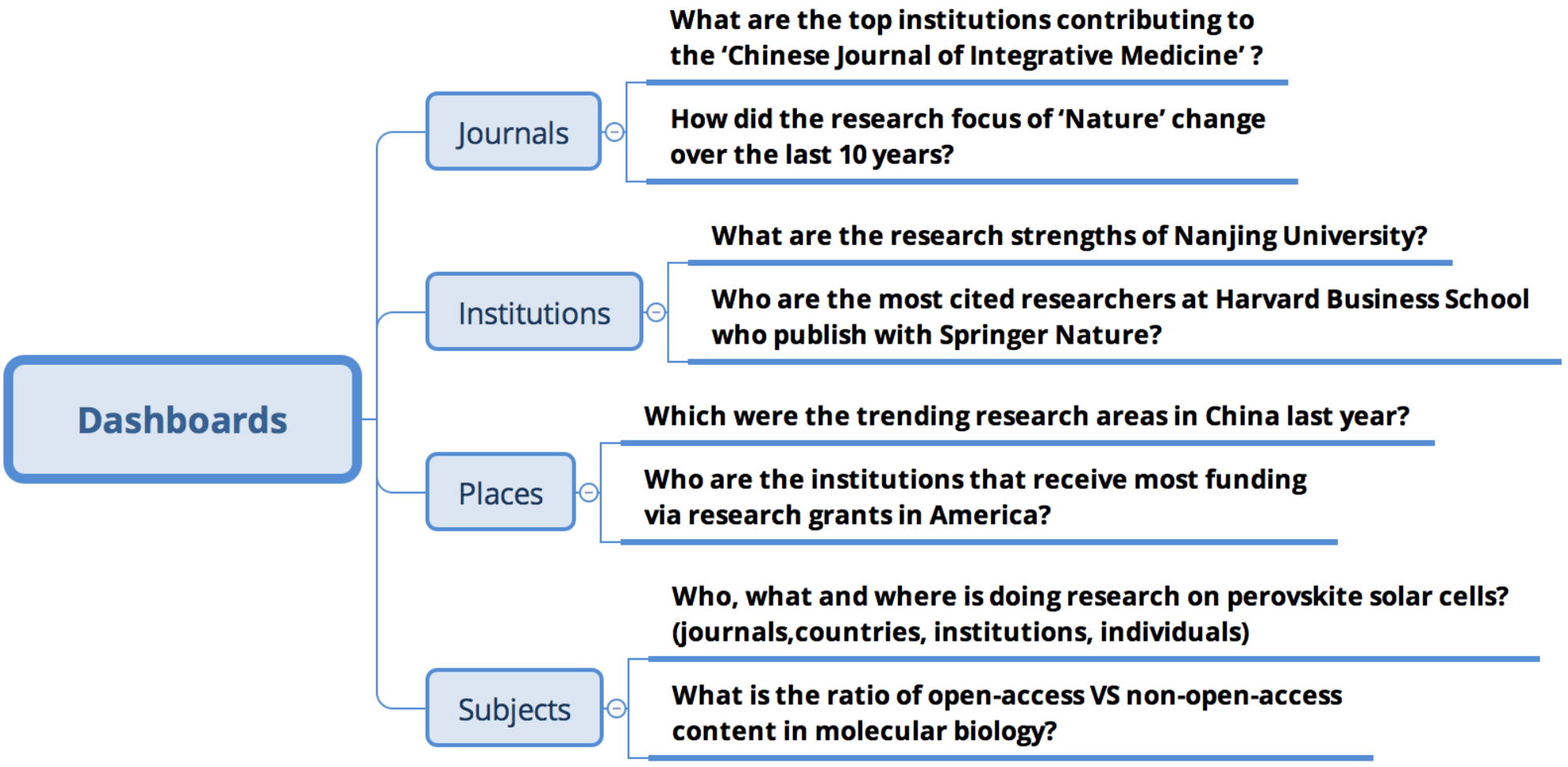
Springet Nature SciGraph

Springer Nature SciGraph Analytics Dashboards (internal only)



SPRINGER NATURE

Springer Nature SciGraph Analytics: Supporting Data Driven Decisions

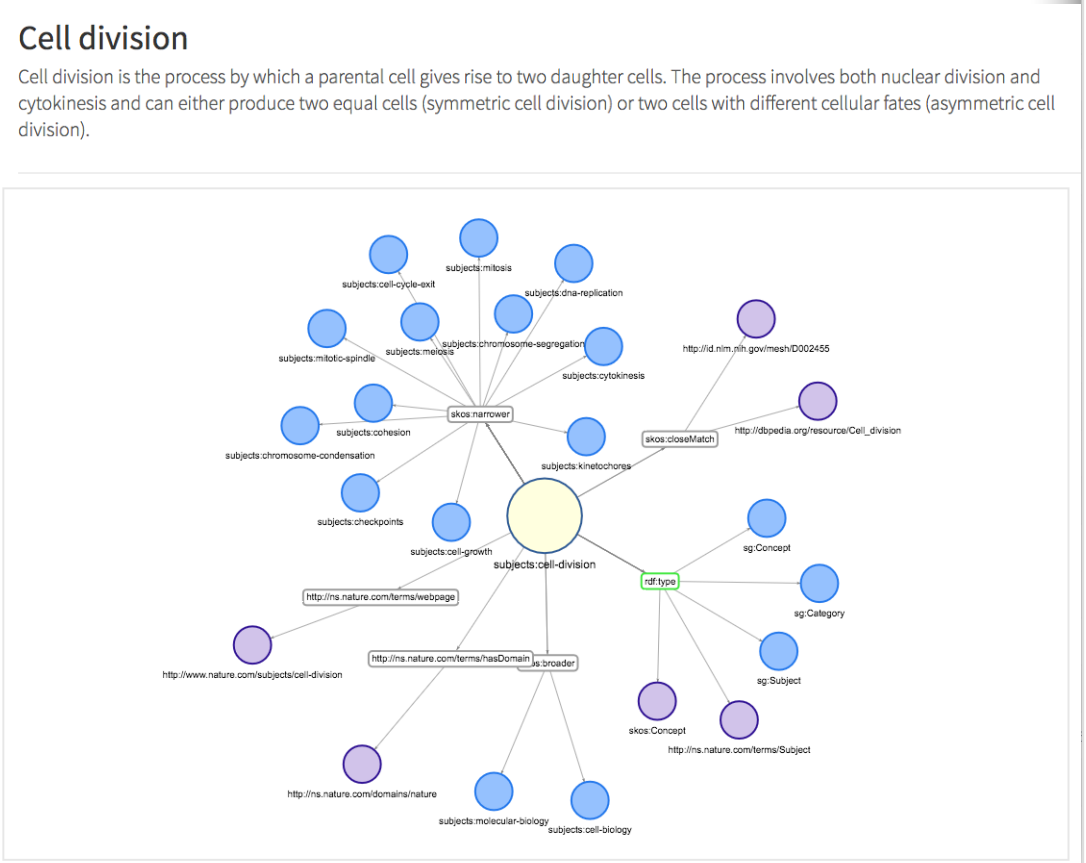


Applications

Linked Data Explorer

Springet Nature SciGraph

Linked Data Explorer (public)



Purpose

- simple UI for exploring graph contents interactively
- internal version: unrestricted access to data and statistics
- external version: only public data, allows linked data ‘dereferencing’

Features

- text & graphical UI on top of graph database
- one page per URI, permits to download machine readable descriptions of data

TRY IT OUT: http://scigraph.springernature.com

SPRINGER NATURE

Linked Data Explorer (public)

Purpose

- 1BL Triples
- 150+ years of content
- Dual License (CC-BY, CC-BY-NC)
- Linked Data Pages (API)

Summary, Next Steps and Public Outreach

Summary

- **Key Facts**
  - Springer Nature Scigraph is our LOD platform: Focus on data re-use, integration and discoverability
  - Collaboration between Springer Nature and Digital Science (and other partners)
  - Data publishing: ~1B triples released as of today, including complete archive, updated CC license
  - Data browser: Visualization, interactive exploration of the graph (Linked Data Browser),
  - Internal use cases: Ontology management, analytics dashboards, semantic publishing

Looking Ahead

- **Next Steps**
  - Data publishing: more releases including citations data
  - Analytics dashboards publicly acessible in particular to support *Sales, Marketing and Editorial*
  - Exploring metadata deliveries to third parties (Google, libraries)
  - SN Taxonomy Manager: Beta going live this month
  - Going Pan-publisher

- **Public Outreach**
  - Collaboration with DBpedia: Internship in London/Leipzig
  - Hack Day: Research data Publishing hack day (London, November 17, 2017)
  - Webinar for more discussions about SciGraph (Dec 2017)

Thank you

How to get in touch:

- E-mail
  michele.pasin@springernature.com
- Portal
  http://www.springernature.com/scigraph
- E-Mail
  scigraph@springernature.com
- Twitter
  #scigraph