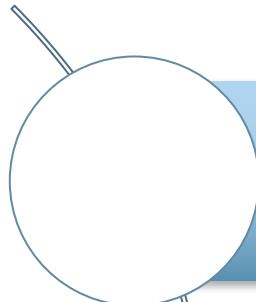


## Building a Linked Open Data Knowledge Graph

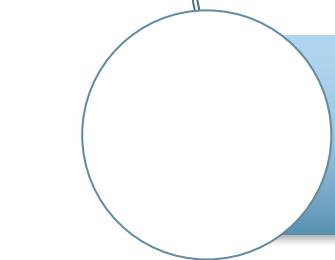
Michele Pasin

London Info International 2017  
December 2017

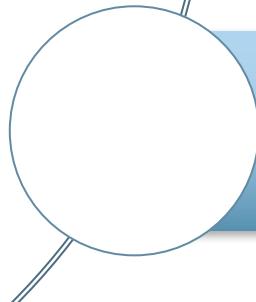
# Springer Nature's Metadata Mission Statement



**We understand metadata as the gateway to our content.**



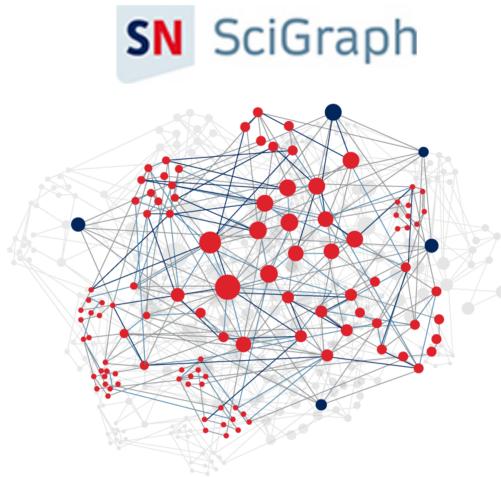
**We provide best quality metadata with state-of-the-art enrichment in all key formats and flavors, available in all relevant delivery models.**



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

# Springer Nature SciGraph

A Linked Open Data platform for the scholarly domain



- > 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

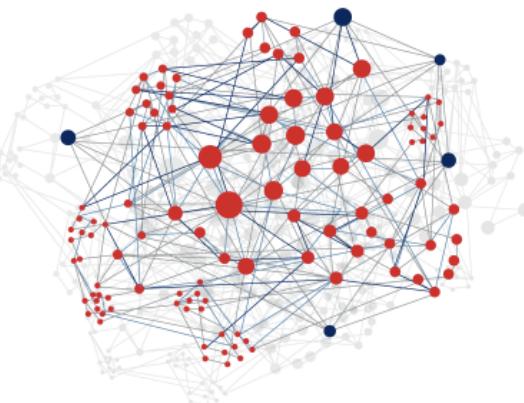
**SPRINGER NATURE**

## 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.



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.

Any questions?  
Please contact us.

Dataset Download

Licensing Information

Further Info

Conference Presentation 2016 (PDF, 11.56 MB)

## 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](http://www.springernature.com/scigraph)

# Open Data Events: Hack Day June 2017

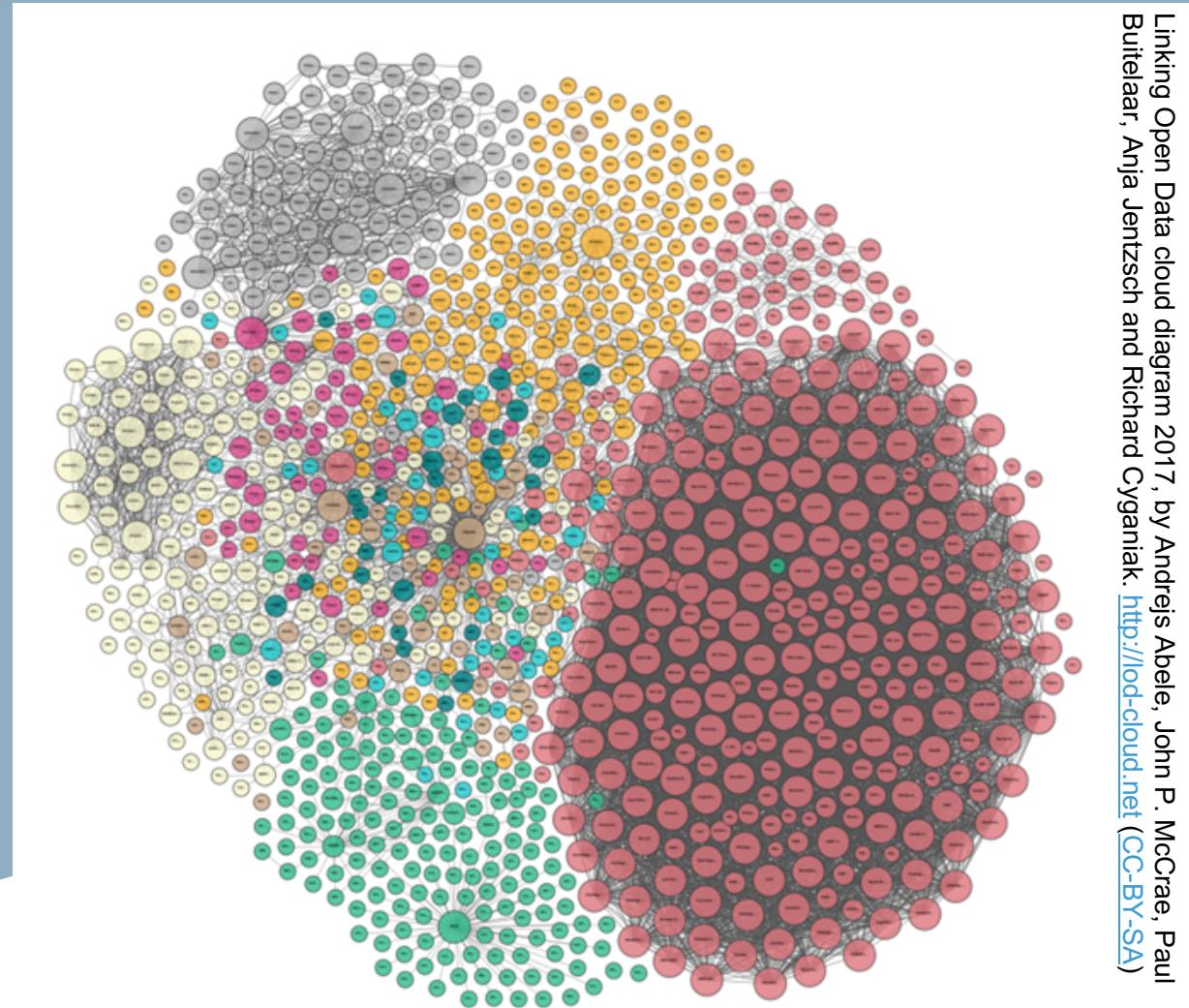
## Aims and Scope

- 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)

# Libraries using Linked Data

 LIBRARY OF CONGRESS

ASK A LIBRARIAN DIGITAL COLLECTIONS LIBRARY CATALOGS

Search Search Local

The Library of Congress > Linked Data Service

**LIBRARY OF CONGRESS  
LINKED DATA SERVICE**

**LC Linked Data Service**  
Authorities and Vocabularies



Search

Enter Keyword or Phrase

All  
LC Subject Headings  
LC Name Authority File  
LC Classification  
LC Children's Subject Headings

Search

**Available Datasets**

The Linked Data Service provides access to commonly found standards and vocabularies promulgated by the Library of Congress. This includes data values and the controlled vocabularies that house them. The following are currently available:

> <a href="#">LC Subject Headings</a>	> <a href="#">MARC Relators</a>
> <a href="#">LC Name Authority File</a>	> <a href="#">MARC Countries</a>
> <a href="#">LC Classification</a>	> <a href="#">MARC Geographic Areas</a>
> <a href="#">LC Children's Subject Headings</a>	> <a href="#">MARC Languages</a>
> <a href="#">LC Genre/Form Terms</a>	> <a href="#">MARC Genre Terms</a>
> <a href="#">LC Medium of Performance Thesaurus for Music</a>	> <a href="#">ISO639-1 Languages</a>
	> <a href="#">ISO639-2 Languages</a>

Schemes

- > [Identifiers](#)
- > [Carriers](#)
- > [Content Types](#)
- > [Media Types](#)
- > [Resource Types](#)
- > [Description Conventions](#)

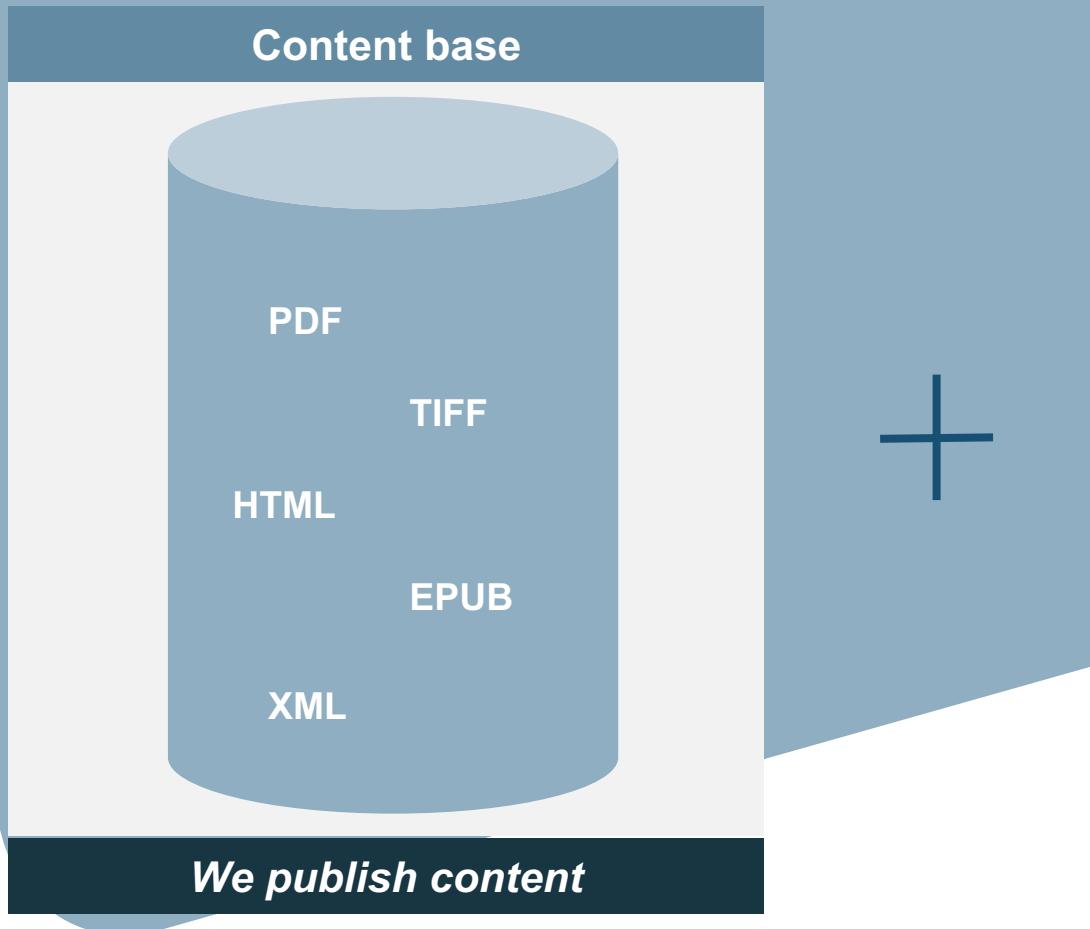
## Library of Congress Linked Data Service (2009)

- 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:

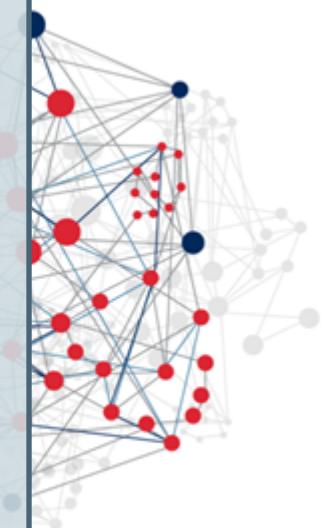
- British Library (**BL**)
- French National Library (**BNF**)
- German National Library (**DNB**)
- National Library of Spain (**BNE**)
- National Library of Sweden (**LIBRIS**)
- Hungarian National Library (**NSL**)

# Vision: From Content to Data



# Vision: From Content to Data

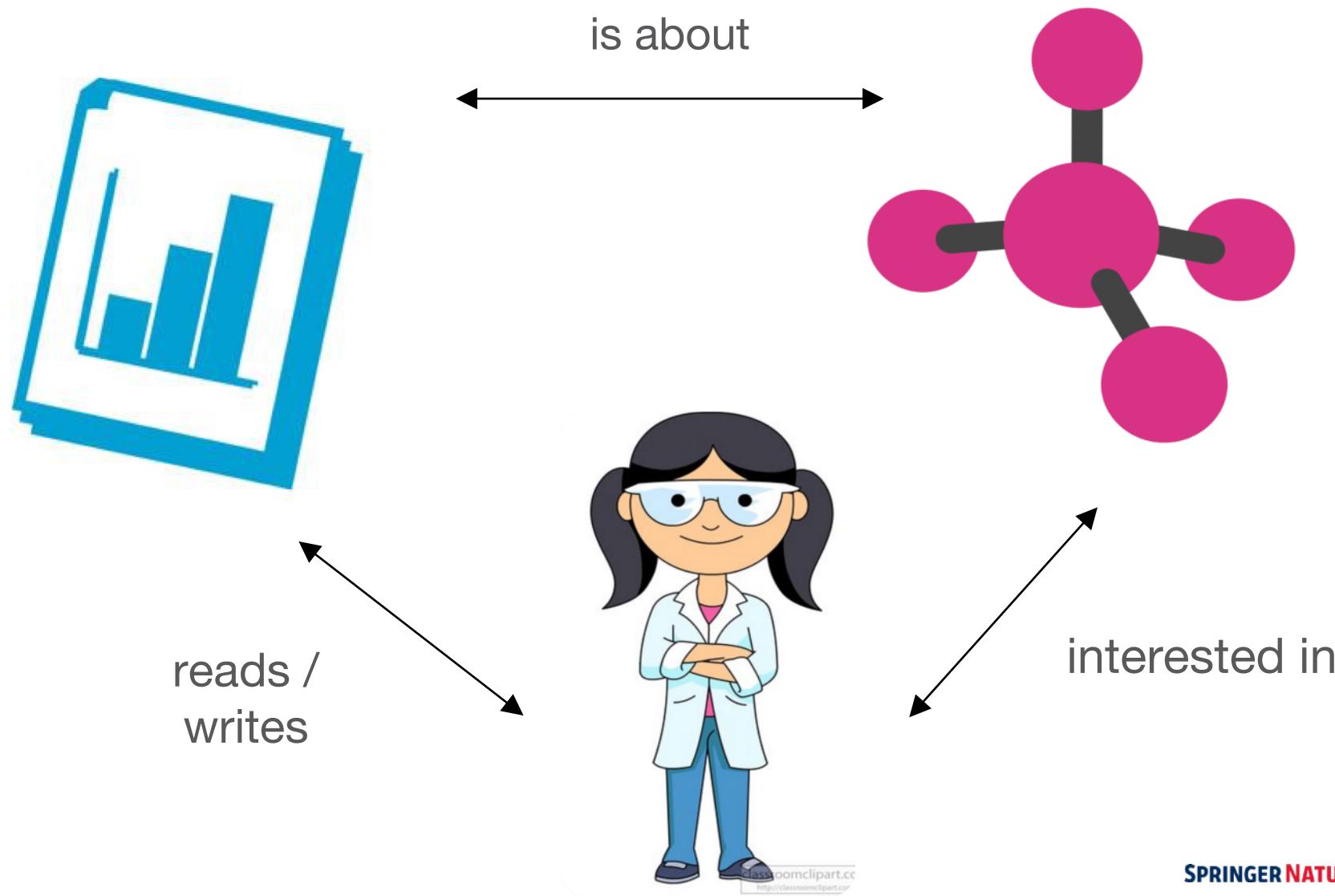
- We create the largest state-of-the-art linked open data aggregation platform for the scholarly domain.
- In doing so, we increase content discoverability and provide data tools and services for researchers, authors, editors, librarians, data scientists, funders, conference organizers, and many others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

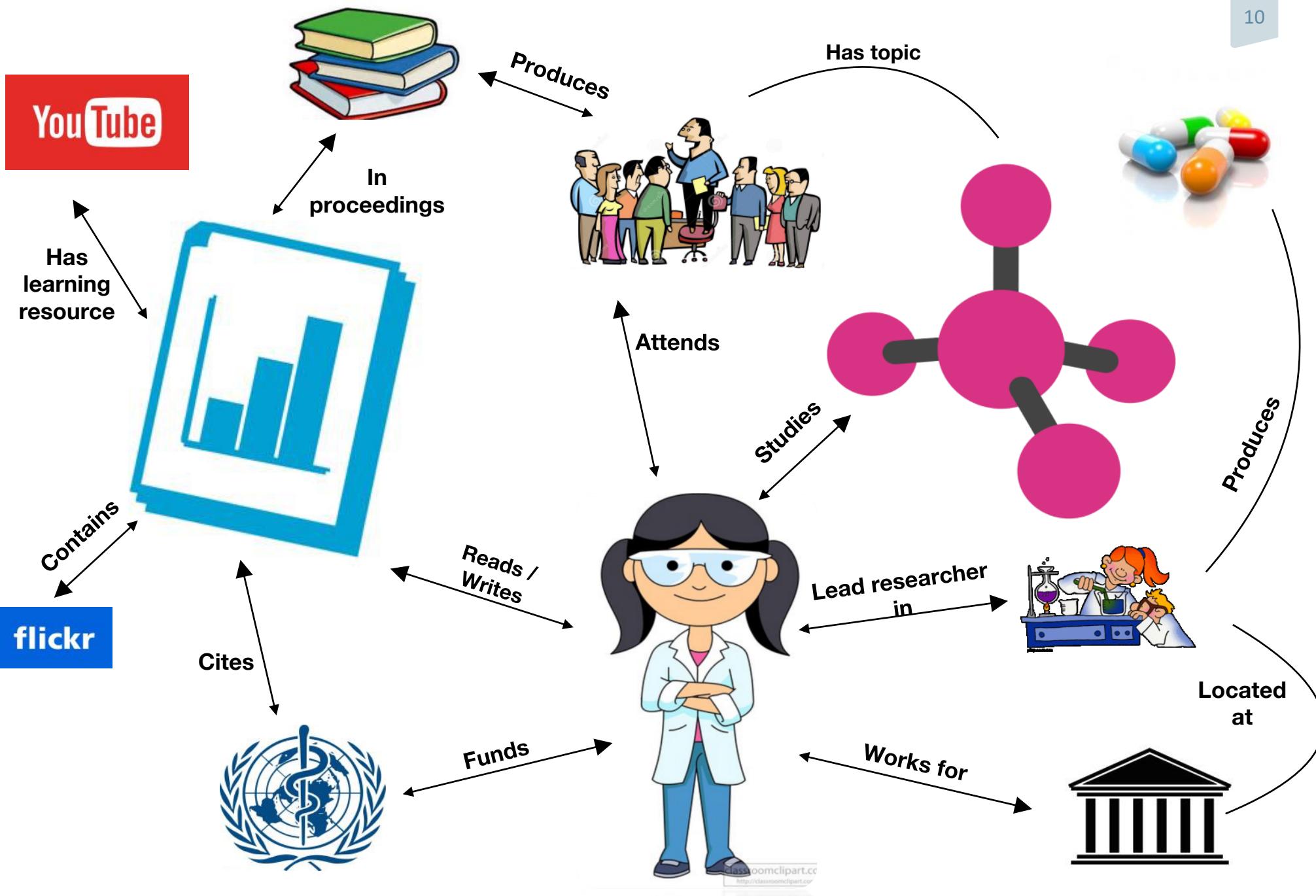


*We publish content*

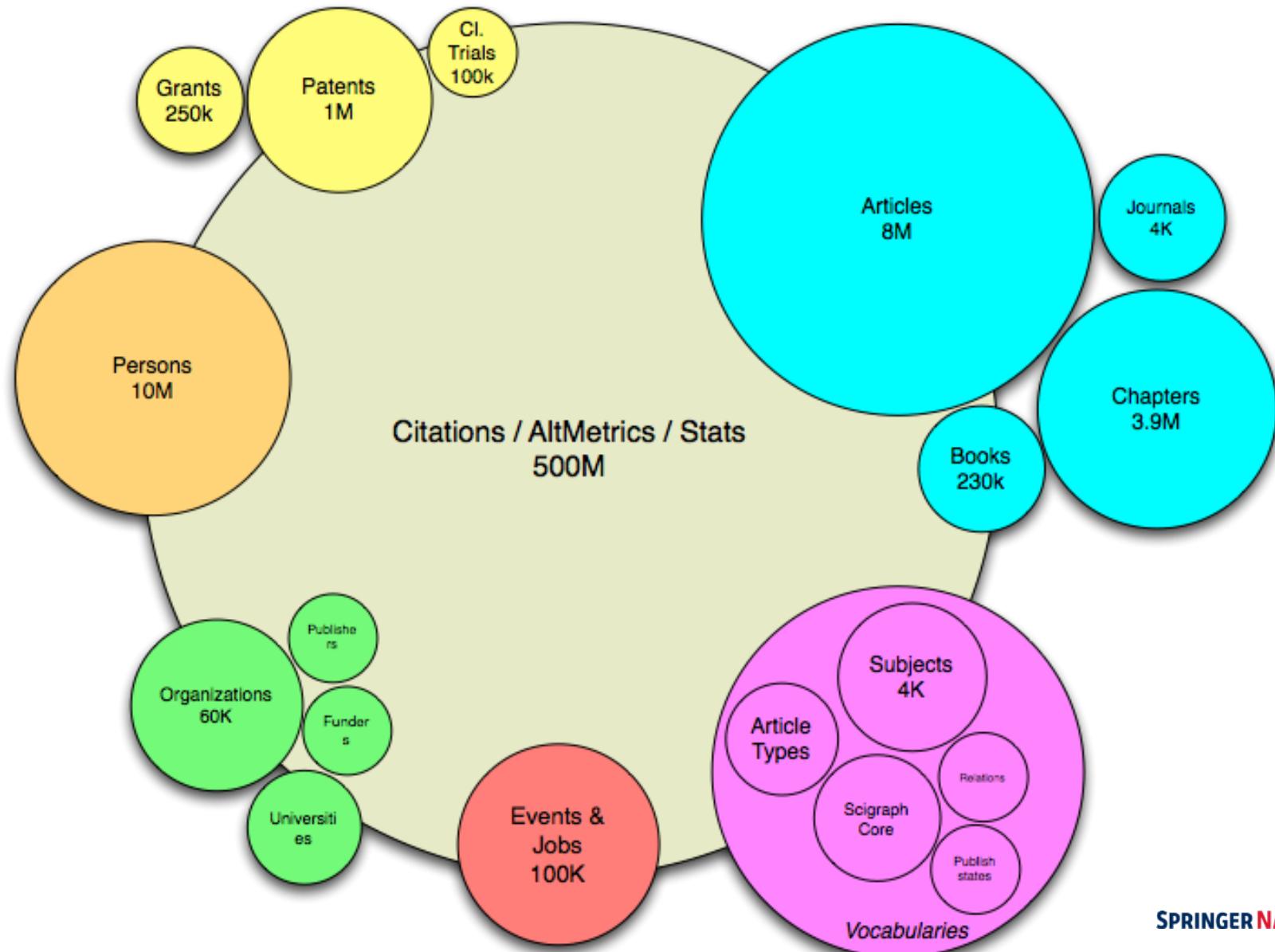
*We manage knowledge*

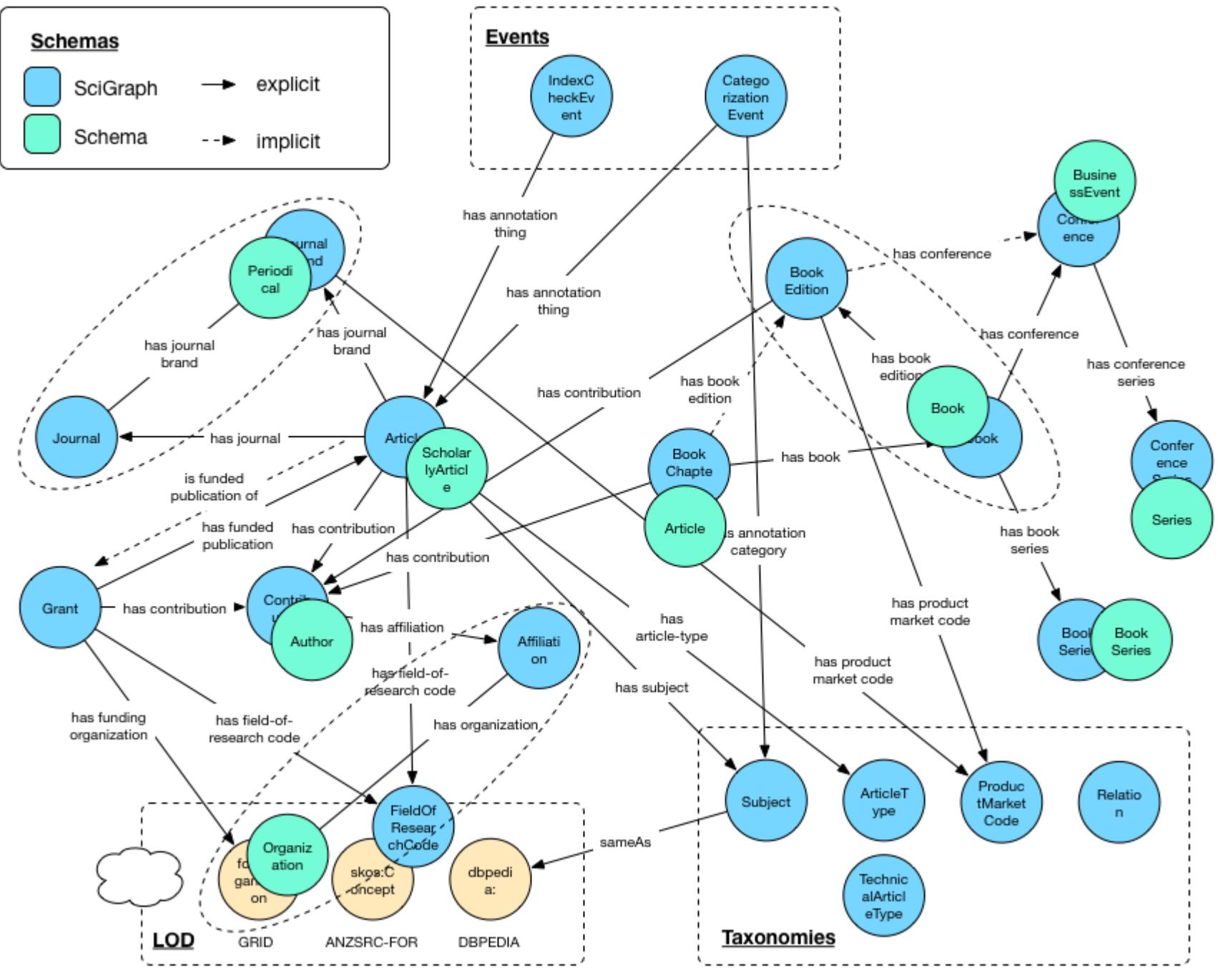
# Three areas of knowledge we care about



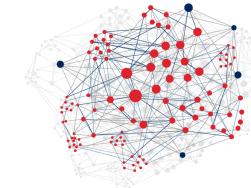


# Springer Nature SciGraph Data Landscape





# 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

# Springer Nature SciGraph Analytics Dashboards (internal only)

Springer Nature SciGraph Analytics Dashboards Journals Institutions Countries Subject Areas

## BMC Cell Biology

Journal ID: 12860

Note: In order to obtain the raw data for this dashboard please contact the [Knowledge Graph team](#)

PUBLICATION VOLUME JOURNAL METRICS AUTHORS COUNTRIES & INSTITUTIONS FIELD OF RESEARCH RESEARCH FUNDING DATA QUALITY

Section - Countries and Institutions

### Countries and Institutions

Use this section to find out which are the top countries and institutions contributing to a publication.

Note: this information comes from the GRID database (<https://www.grid.ac/>).

Article - map view

Publication Volume

This section provides statistics useful to understand the type and volume of content linked to a publication. For example, how many articles have been published over the years, which are the most frequently used article types and how much of this content has been indexed in external databases.

Article - Total number Article - Count from 2012

**841** **201**

Articles in Total Articles Published Last 5 Years

Article - Count by publication year

Count publicationYear

Fields of Research

This section provides a breakdown of publication content based on subject areas. The subject areas are derived from the Australian and New Zealand Standard Research Classification (ANZSRC). <http://www.abs.gov.au/ausstats/abs@.nsf/088B427A8988C205CA2741800044E3E>

Article - FieldOfResearch-by code and description Article - top 15 Fields of Research over time

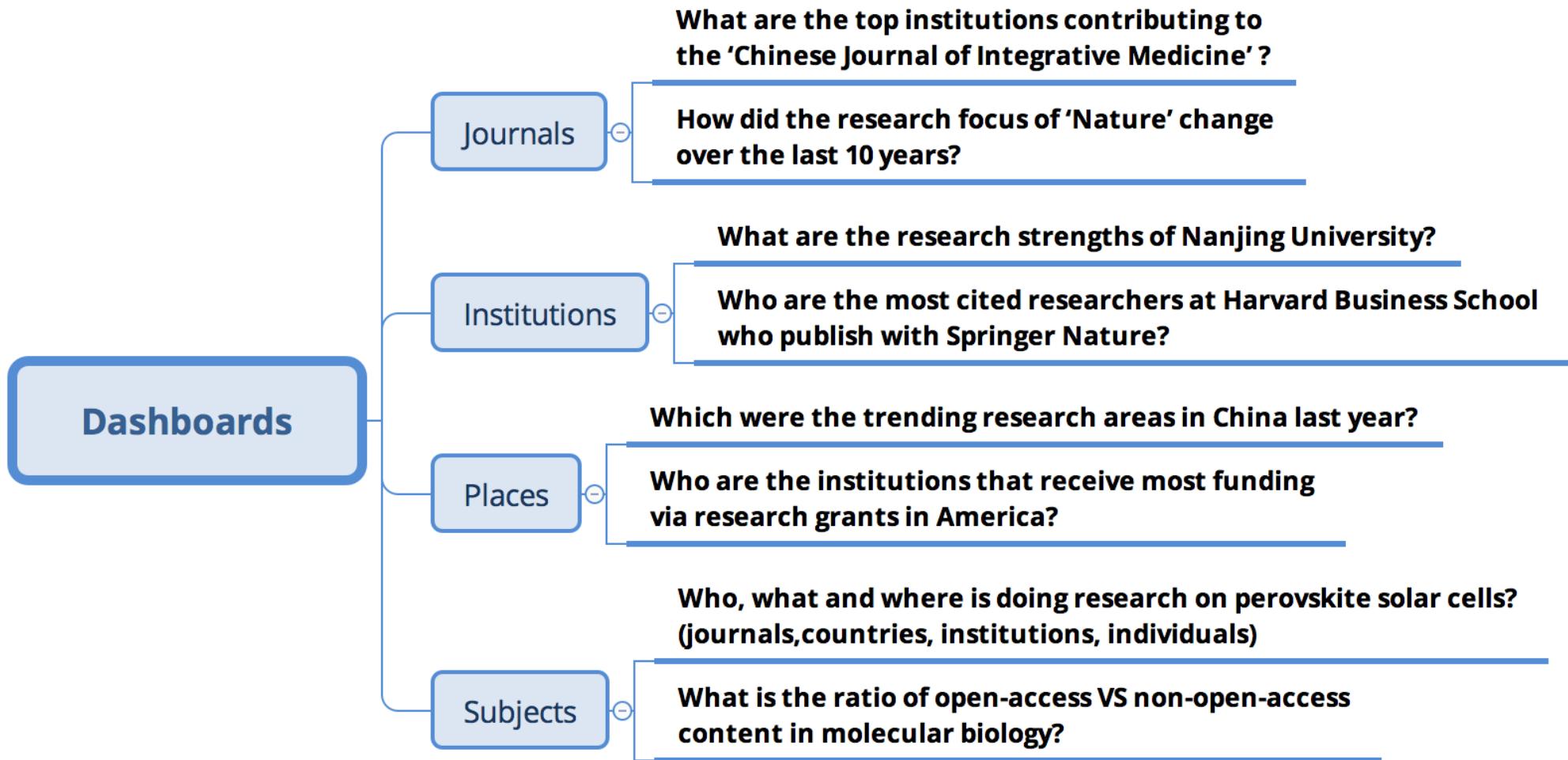
Percentage of publications

Percentage of publication from range

- PHYSICAL SCIENCES
- OTHER PHYSICAL SCIENCES
- NEUROSCIENCES
- MEDICAL AND HEALTH SCIENCES
- IMMUNOLOGY
- GENETICS
- CHEMICAL SCIENCES
- CAMPAGNE FOR RESEARCH
- BIOLOGICAL SCIENCES
- BIOTECHNOLOGY AND BIOPROCESS
- TECHNOLOGY
- STATISTICS
- PLANT BIOLOGY
- PHYSIOLOGY AND PHARMACOLOGY
- ONCOLOGY AND CARCINOGENESIS
- MEDICAL MICROBIOLOGY
- MEDICAL BIOTECHNOLOGY
- MEDICAL BIOCHEMISTRY
- MATHEMATICAL SCIENCES
- INFORMATION AND COMPUTER SCIENCE
- OTHER SUBJECTS

SPRINGER NATURE

# Springer Nature SciGraph Analytics: Supporting Data Driven Decisions

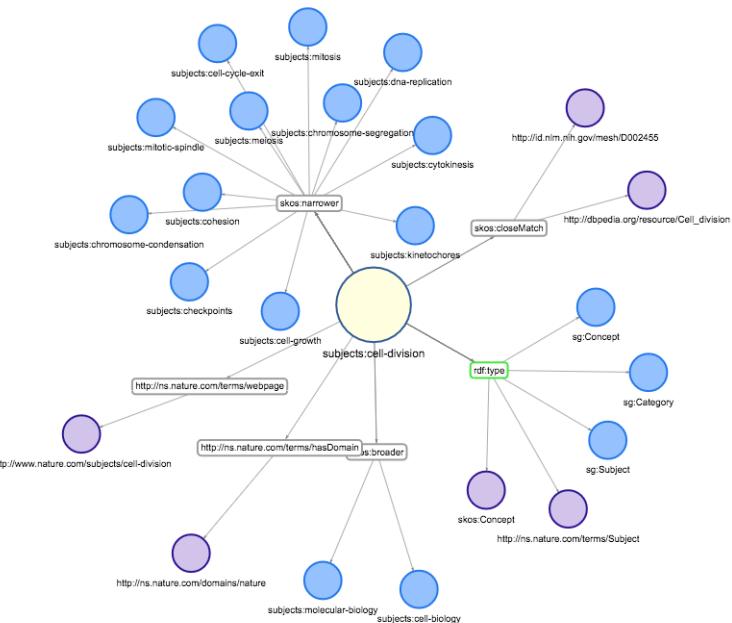


# Applications Linked Data Explorer

# Linked Data Explorer (public)

## Cell division

Cell division is the process by which a parental cell gives rise to two daughter cells. The process involves both nuclear division and cytokinesis and can either produce two equal cells (symmetric cell division) or two cells with different cellular fates (asymmetric cell division).



## 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 accessible 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](mailto:michele.pasin@springernature.com)
- Portal  
<http://www.springernature.com/scigraph>
- E-Mail  
[scigraph@springernature.com](mailto:scigraph@springernature.com)
- Twitter  
[#scigraph](#)