



Linked data, contextualisation and connections - to which end?

Christian Herzog, CEO Dimensions, Digital Science

Linked data
and the future of
information sharing

Part of **DIGITAL**science

CAVEAT AUDITORIUM!

- The term 'Linked Data' is not use in its purest technical form throughout this presentation
- 'Linked data' is used to describe the connections and relations **which are of use to the end user**, not the technical representation of the data (as triplets etc.)
- In essence, 'Linked' is seen from the angle of the consumer of the data, not from the engineer who builds and provides the data infrastructure



The future of linked data is bright and it will empower and simplify information sharing.

- ...getting on the top of the mountain is hard
- ... and jumping requires trust in the cape...

To translate this to our less risky reality:

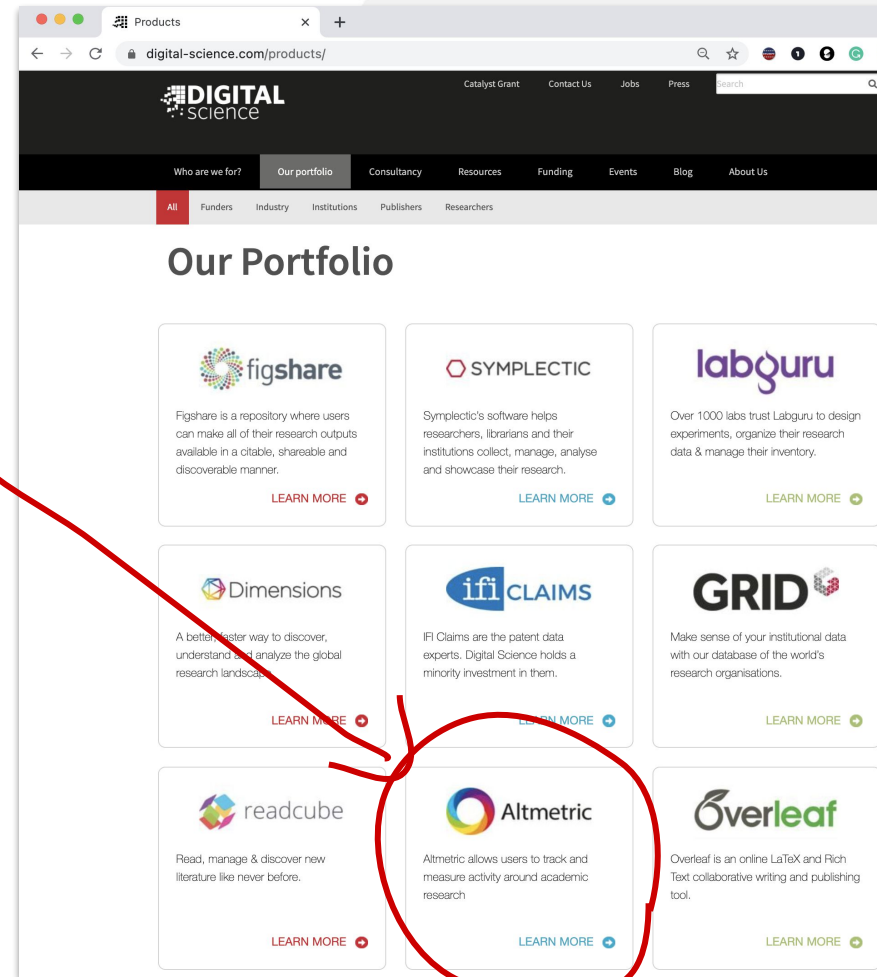
- A lot of players are involved in establishing the linked aspect of the data
- The infrastructure in which is is provided to the beneficiary defines how much value can be derived from it



Digital Science and linked data

'Linking on to it'

- E.g. Altmetric associates 16.7 million research outputs with 158.7 million mentions (on social media, news and blogs)

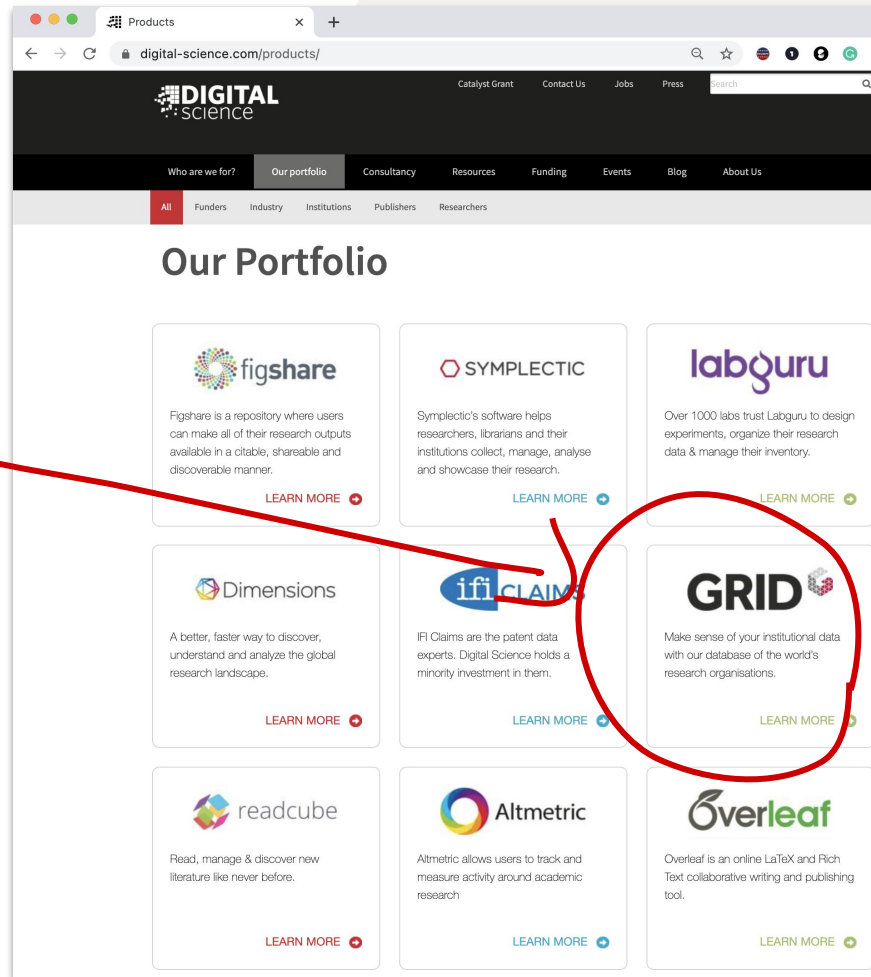


Part of **DIGITAL**science

Digital Science and linked data

'Helping linking it'

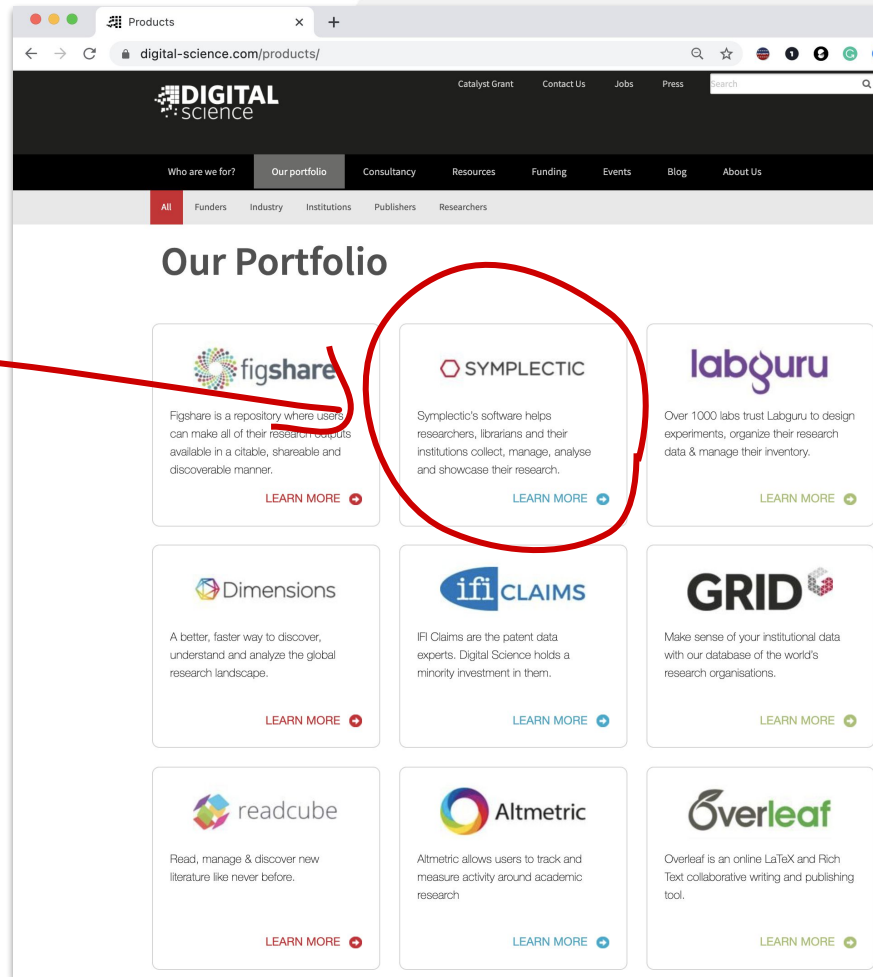
- GRID as persistent identifiers for research organisations
- Currently 99,609 institution records
- Dimensions API includes the matching capabilities to allow to resolve affiliation data
- Currently used as the basis for ROR
- <https://grid.ac>



Digital Science and linked data

‘Making use of the linked data’

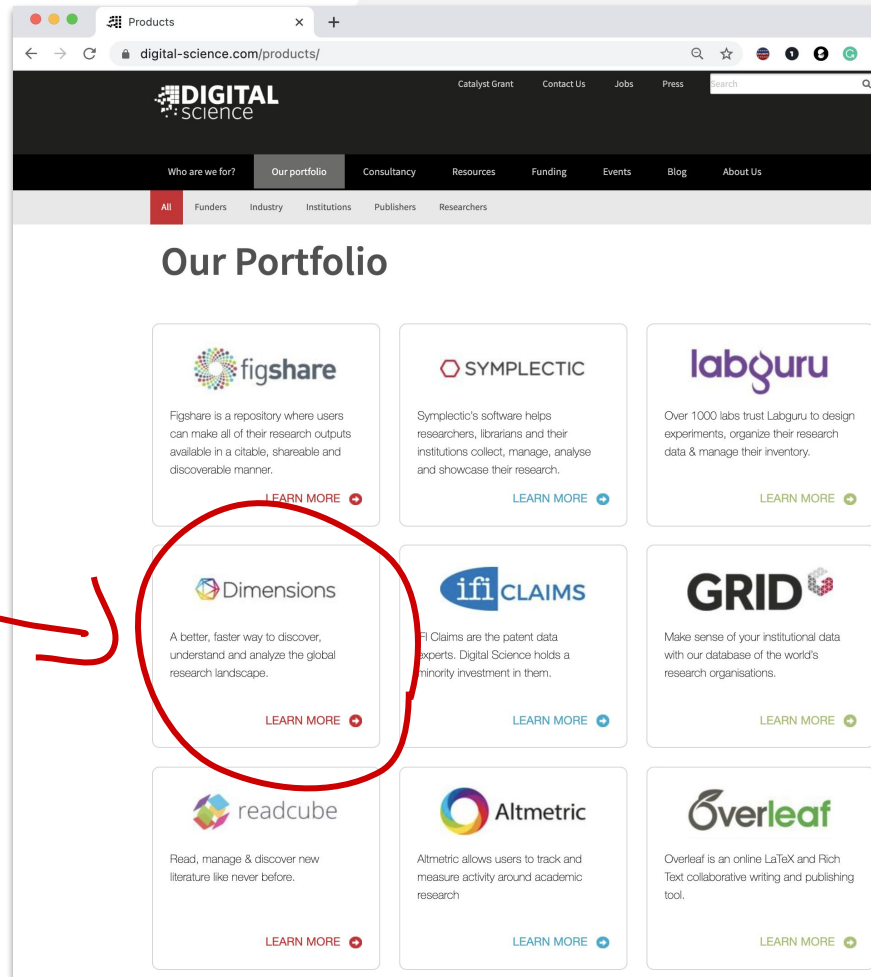
- Applications like Elements make use of linked data to support universities to plan and manage their research activities and outputs



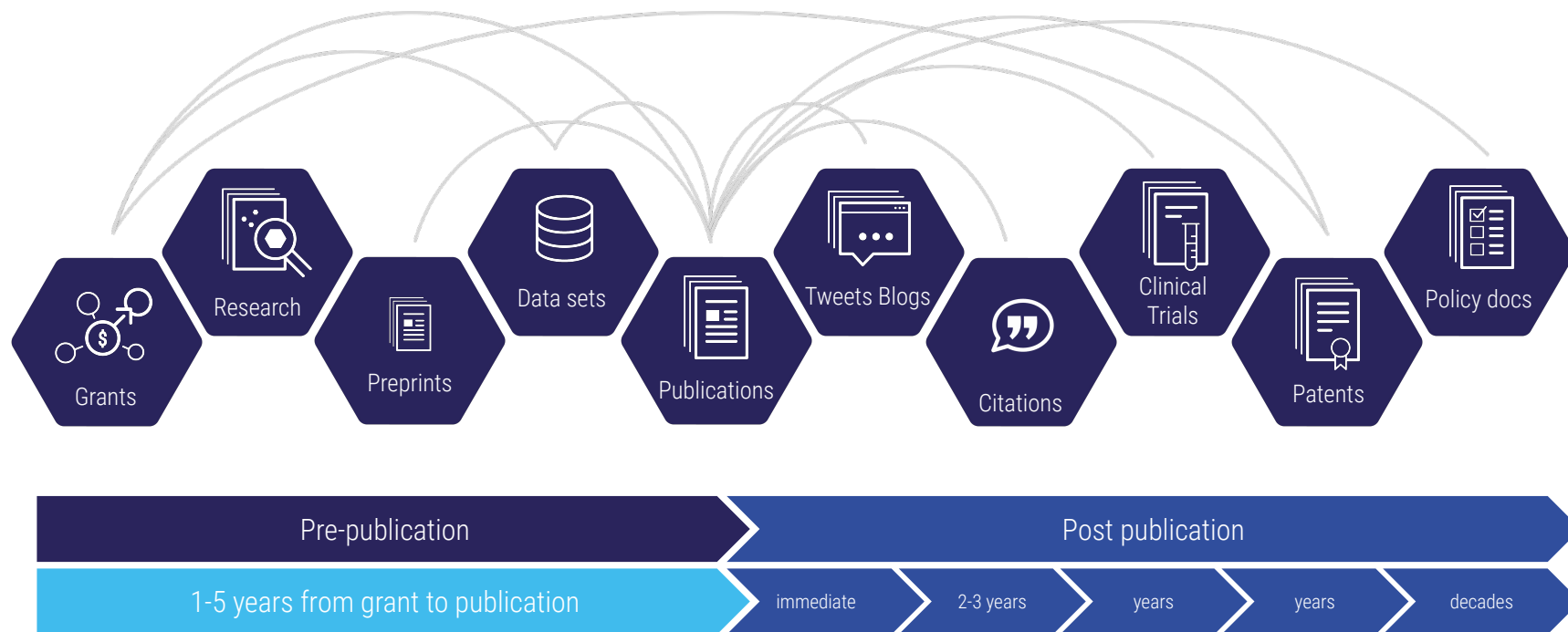
Digital Science and linked data

‘Aggregating and linking it’

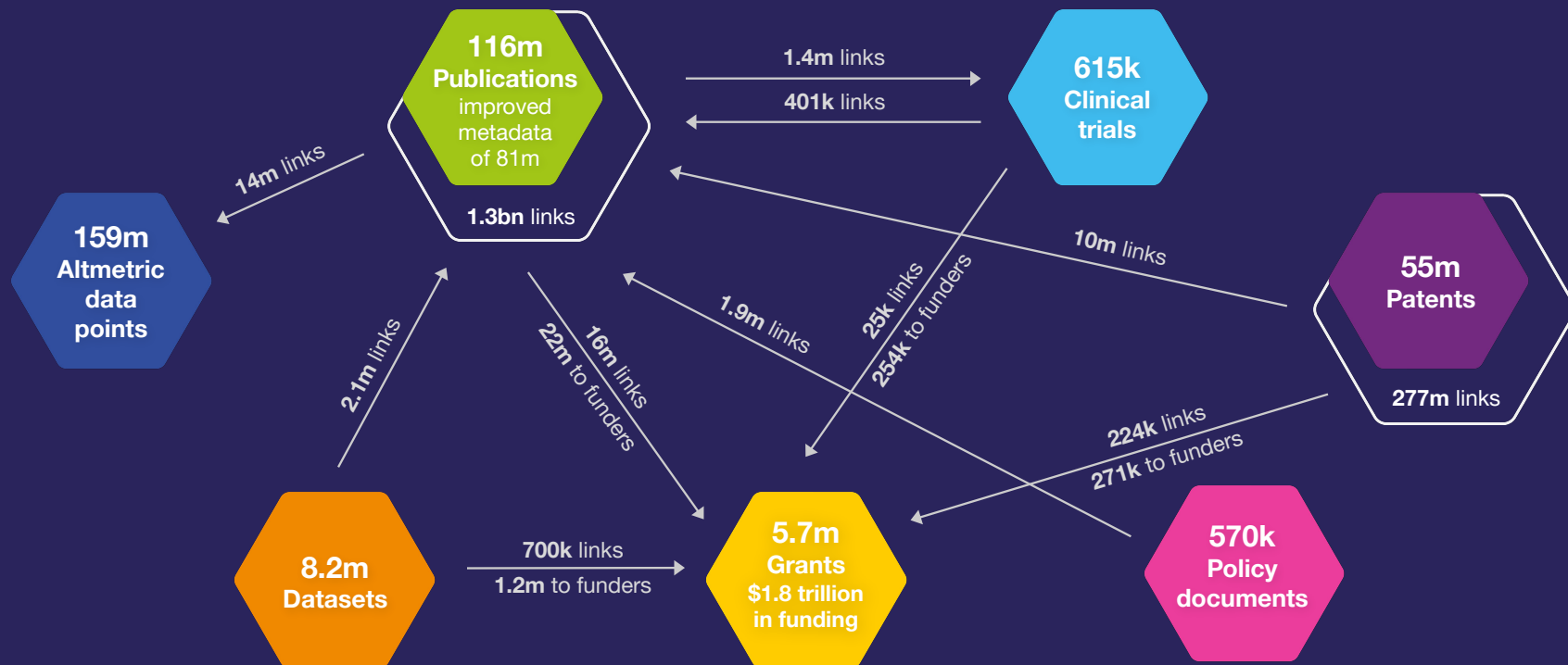
- with Dimensions where we are aggregating, curating and linking the largest research information dataset currently available



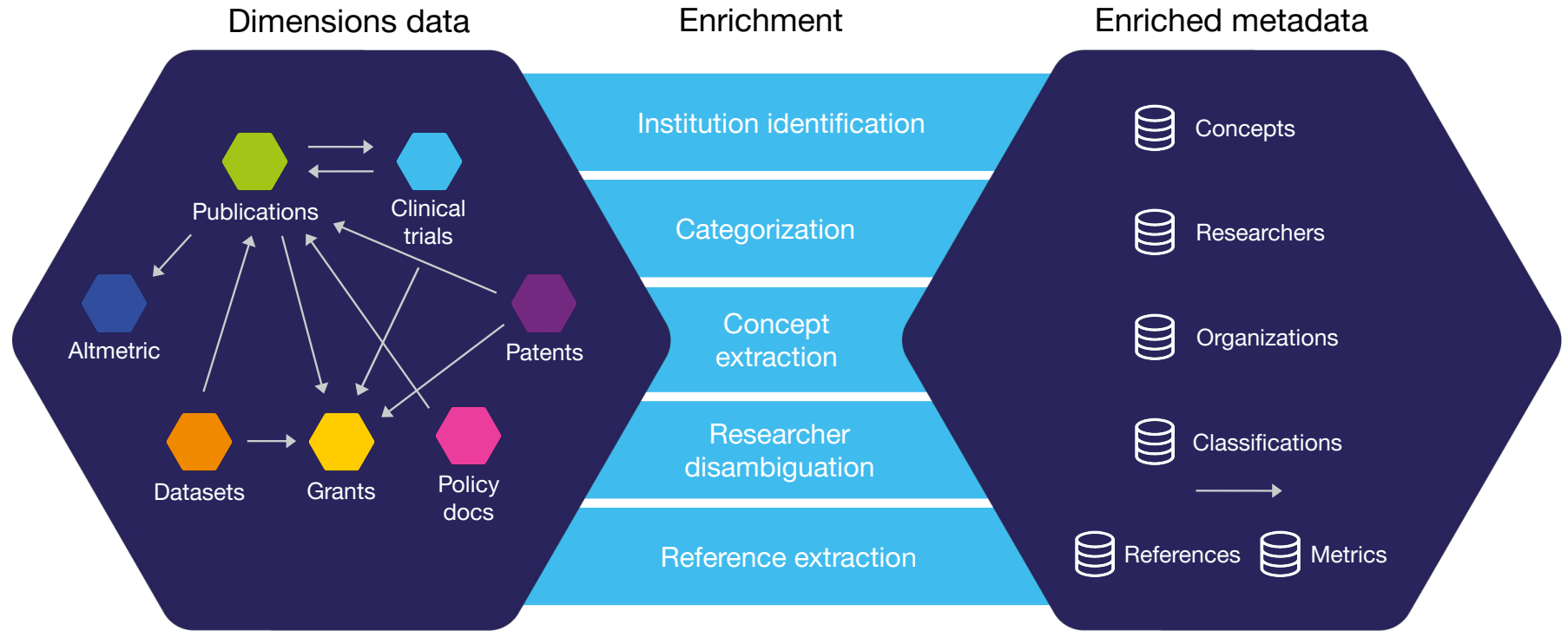
A much broader information landscape is available!



... and how that looks in Dimensions



Harmonised and curated data - see results without having to align the data first





e.g. plastic AND instrument

Save / Export

PUBLICATIONS
115,816,709

DATASETS
8,225,401

GRANTS
5,713,768

PATENTS
54,961,345

CLINICAL TRIALS
616,288

POLICY DOCUMENTS
568,871

Example of publication detail record in Dimensions

- Largest amount of links realised by indexing of 81M full-text records with noun phrase detection
- E.g. the detected noun phrase "stable chronic obstructive pulmonary disease patients" links 825 publications, 1 grant and 6 clinical trials and 6 policy documents together

Dimensions e.g. plastic AND instrument

Publication - Article

Persistent Systemic Inflammation is Associated with Poor Clinical Outcomes in COPD: A Novel Phenotype

PLoS ONE, 7(5), e37483 - May 2012
<https://doi.org/10.1371/journal.pone.0037483>

Authors

Alvar Agustí - Thorax Institute, Hospital Clinic, Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), University of Barcelona and Centro de Investigación en red de enfermedades respiratorias (CIBERES), Barcelona, Spain; Fundación Investigación Sanitaria Illes Balears (FISB), Palma de Mallorca, Spain

Lisa D Edwards - Gaudium Clinic (United States)

Stephen I. Rennett - University of Nebraska Medical Center

15 more

Abstract

BACKGROUND: Because chronic obstructive pulmonary disease (COPD) is a heterogeneous condition, the identification of specific clinical phenotypes is key to developing more effective therapies. To explore if the persistence of systemic inflammation is associated with poor clinical outcomes in COPD we assessed patients recruited to the well-characterized ECLIPSE cohort (NCT00292552). METHODS AND FINDINGS: Six inflammatory biomarkers in peripheral blood (white blood cells (WBC) count and CRP, IL-6, IL-8, ferritinogen and TNF-α levels) were quantified in 1,755 COPD patients, 297 smokers with normal spirometry and 202 non-smoker controls that were followed-up for three years. We found that, at baseline, 30% of COPD patients did not show evidence of systemic inflammation whereas 16% had persistent systemic inflammation.

More

Acknowledgements

Authors thank all participants for their willingness to contribute to this study and all field personnel for their commitment and quality of their work. Principal investigators and centers participating in ECLIPSE (NCT00292552, SC0104645) Bulgaria: Y Ivanov, P K Kostov, Sofia, Canada: J Bourbeau, Montreal; M Prasad, Vancouver; P Hernandez, Halifax; K Killian, Hamilton; R Levy, Vancouver; F Mallat, Paris; Maltina, Montreal; D O'Donnell, Kingston, Czech Republic: J Kopecká, Praha, Denmark: J Vestbo, Hvidovre, The Netherlands: E Wouters, Horn, New Zealand: D Quinn, Wellington, Norway: P Bakke, Bergen, Slovenia: M Kosnik, Golnik, Spain: A Agustí, Jaume Sureda, Palma de Mallorca, Ukraine: Y Pechenko, Kiev, V Gavrilov, Kiev, L Yashina, W Machee, Edinburgh; D Singh, Manchester, J Wedzicha, London, USA: A Accurso, San Antonio, TX, S

More

Supplemental data

Persistent Systemic Inflammation is Associated with Poor Clinical Outcomes in COPD: A Novel Phenotype

Showing 1/15 Figure_S1.tif

Group	0	1	2	3	4	5+
COPD (n=1755)	30%	31%	19%	12%	6%	1%
Smokers (n=297)	42%	35%	16%	5%	2%	0%
Non smokers (n=202)	77%	18%	2%	2%	0%	0%

1 / 15

Publication references - 44 [Show all](#)

Systems medicine and integrated care to combat chronic noncommunicable diseases

Juan Boagard, Josep M Arto, Peter J Sher, Ian M Adcock, Kian Fan Chung, Josep Roca, Alvar Agustí, Chris Brightling, Anne Carbon-Thomson, Alfredo Cesaric, Sonia Abdellhak, Stylianos E Antonarak...

2011, Genome Medicine - Article

[CiteSpace](#) [155](#) [PubMed](#) [17](#) [View PDF](#) [Add to Library](#)

Identification and prospective validation of clinically relevant chronic obstructive pulmonary disease (COPD) subtypes

Judith Garcia-Americh, Federico P Gómez, María Benet, Eva Fariñas, Xavier Basagaña, Àngel Gayet, Carlos Paré, Xavier Fieira, Jaume Ferrer, Antoni Ferrer, Josep Roca, Juan R Galdiz, Jaume Sureda...

2010, Thorax - Article

[CiteSpace](#) [234](#) [PubMed](#) [1](#) [Open Access](#)

Network medicine: a network-based approach to human disease

Albert-László Barabási, Natal Gulbahce, Joseph Loscalzo

2010, Nature Reviews Genetics - Article

[CiteSpace](#) [238](#) [PubMed](#) [71](#) [View PDF](#)

Supporting grants - 2 [Show all](#)

Sorted by: [Start Date](#) [Funding amount \(USD\)](#) [Funding period](#)

University of Edinburgh/MRC Centre for Inflammation Research

Medical Research Council
to John Innes

3,286,616

2011 - 2017

Pathobiology of alpha-1-antitrypsin deficiency and the serpinopathies

Medical Research Council
to David Arthur Lomas

4,712,147

2011 - 2016

Clinical trial references - 1 [Show all](#)

Sorted by: [Date](#) [Trial period](#)

A Multicentre 3 Year Longitudinal Prospective Study to Identify Novel Endpoints and Compare These With Forced Expiratory Volume in 1 Second (FEV1) for Their Ability to Measure and Predict COPD Severity and Its Progression Over Time

(GlaxoSmithKline (United Kingdom))

2005 - 2010

Publication citations - 504 [Show all](#)

Sorted by: [Date](#)

Chronic lung diseases: prospects for regeneration and repair.

Peter J Barnes, Gary P Anderson, Main Fagerås, Maria G Selvali

2021, European Respiratory Review - Article

[PubMed](#) [18](#) [Open Access](#) [Add to Library](#)

Combination of Systemic Inflammatory Biomarkers in Assessment of Chronic Obstructive Pulmonary Disease: Diagnostic Performance and Identification of Networks and Clusters

Iva Hladik, Daniela Belamark, Martina Bosnar, Domagoj Kifer, Andrea Vukic Dugac, Lada Rumora

2020, Diagnostics - Article

[PubMed](#) [1](#) [View PDF](#) [Add to Library](#)

Association of Plasma Heat Shock Protein 70 with Disease Severity, Smoking and Lung Function of Patients with Chronic Obstructive Pulmonary Disease

Iva Hladik, Andrea Vidina Tomaskovic, Marija Grdic Rajkovic, Sanja Popovic-Gile, Andrea Vukic Dugac, Lada Rumora

2020, Journal of Clinical Medicine - Article

[CiteSpace](#) [1](#) [View PDF](#) [Add to Library](#)

Patent citations - 6 [Show all](#)

Sorted by: [Date](#)

Anti-TNF- α /CXCL10 Double-Targeting Antibody and Use Thereof

Seoul National University Hospital - KANG HEUN-SOO, PARK SO-HYUN, SONG YEONG WOOK, SHIN KI CHUL, LEE ELUN YOUNG, LEE ELUN BONG, PARK YOUNG WOO, PARK BUM-CHAN, LEE DONG HEE, KI...

Application US 20160108118-A1 - Published 2016-04-21 Filed 2014-05-22 Priority 2013-05-22

Family 11 [5](#)

Anti-TNF- α /CXCL10 double targeting antibody and use thereof

Seoul National University Hospital - KANG HEUN-SOO, PARK SO-HYUN, SONG YEONG WOOK, SHIN KI CHUL, LEE ELUN YOUNG, LEE ELUN BONG, PARK YOUNG WOO, PARK BUM-CHAN, LEE DONG HEE, KI...

Grant US 10087248-B2 - Published 2018-10-02 Granted 2018-10-02 Priority 2013-05-22

Family 11

Policy document citations - 1 [Show all](#)

Sorted by: [Date](#)

Chronic obstructive pulmonary disease in over 16s: diagnosis and management : E: Predicting and preventing exacerbations

2018, National Institute for Health and Care Excellence

- Cited publications with indicators

- Supporting grants linked to the publication

- References to clinical trials

- Citing publications

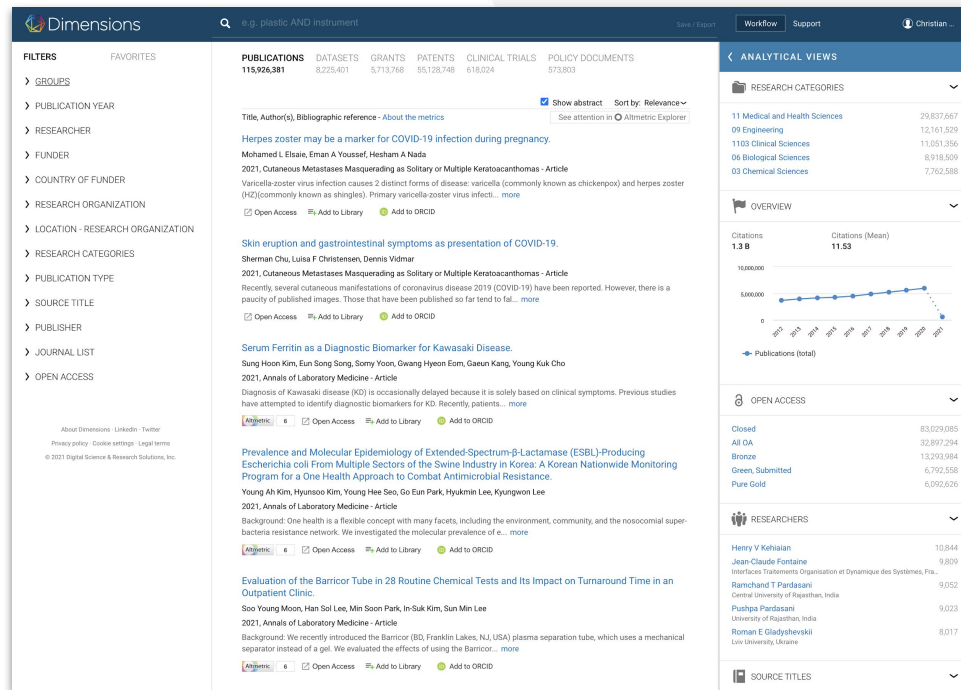
- Citing patents

- Related Policy Documents

But how to serve the linked data up to users with different use cases, different levels of technical skills?

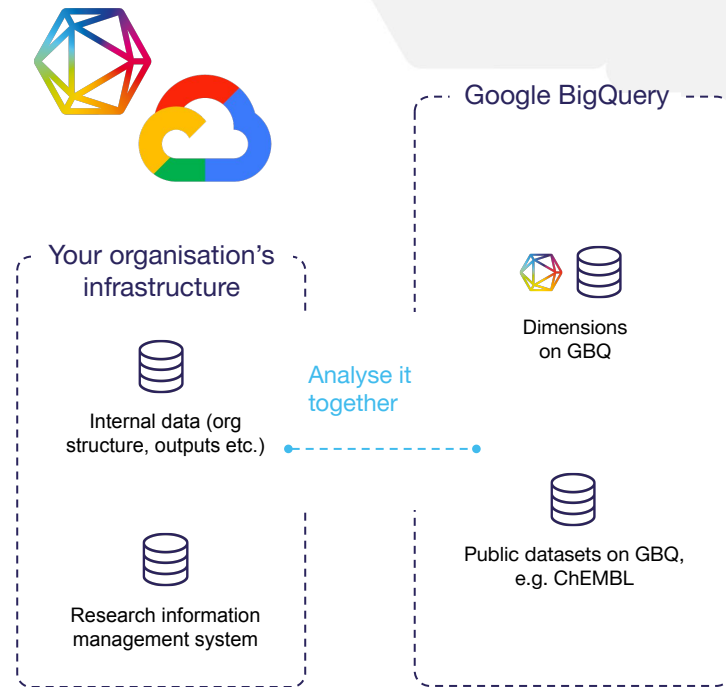
Discovery use case - the easy to use web application

- Linked information is an incremental value, not a game changer
- Needs to be provided as additional context
- Relies on the user to 'connect the dots'
- But no context and data in silos is not acceptable anymore!
- Dimensions with 116M publications and datasets and all the context on a record level is openly available



Easy access to all the links - Dimensions on Google BigQuery

- Access to the linked data matters
- Ability and skills to work with the raw data have increased tremendously with new technology advancing!
- With Dimensions, the underlying data is available together with the computing infrastructure - start within seconds!
- No new skills required, use SQL to explore the links or standard BI tools
- ... and add your own linked data into the analysis!



2,212,574,878 of the most basic relations

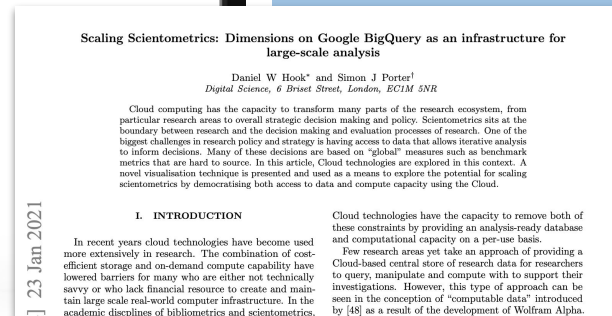
- Only taking into account:
 - Person/researcher
 - Citing relationship
 - Affiliation/organisation
 - Funded by
- None of the classification or indexing related relations included
- None of the relations which could be derived using ontologies mining the fulltext

From	Relation	To	Count
Clinical Trials	Author	Researchers	323,063
Clinical Trials	Cites	Publications	400,755
Clinical Trials	Funder	Organisations	253,853
Clinical Trials	Affiliations	Organisations	1,677,585
Clinical Trials	Is funded	Grants	25,204
Datasets	Author	Researchers	11,539,400
Datasets	Cited by	Publications	2,000,147
Datasets	Funder	Organisations	1,708,718
Datasets	Affiliations	Organisations	4,381,033
Datasets	Is funded	Grants	1,709,400
Grants	Grantee	Researchers	7,792,978
Grants	Affiliations	Organisations	6,095,142
Grants	Funder	Organisations	5,714,237
Patents	Researchers	Researchers	2,219,748
Patents	Cites	Publications	9,928,789
Patents	Cited by	Patents	301,091,117
Patents	Current assignees	Organisations	44,142,419
Patents	Funder	Organisations	282,392
Patents	Assignees	Organisations	28,080,358
Patents	Original assignees	Organisations	67,576,475
Patents	Funded by	Grants	220,394
Publications	Author	Researchers	255,412,952
Publications	Cites	Publications	1,335,819,988
Publications	Affiliations	Organisations	85,190,166
Publications	Funder	Organisations	21,664,752
Publications	Funded by	Grants	15,923,666
Publications	Cites	Clinical Trials	1,400,147

A few example use cases

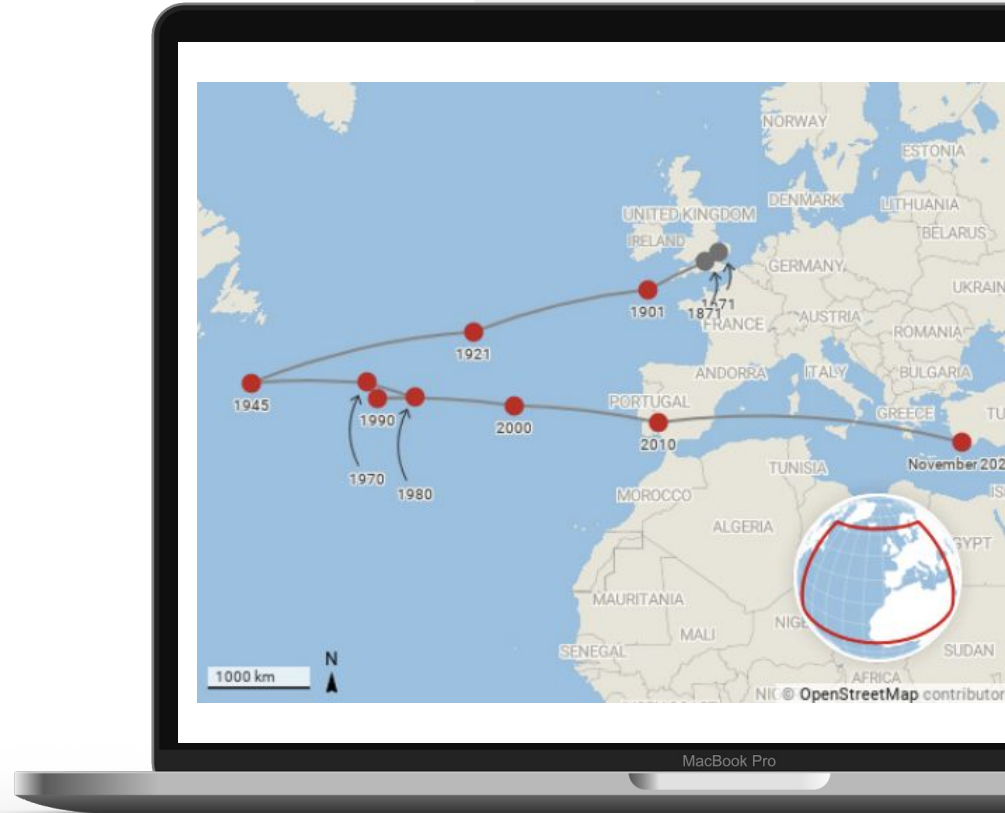
Research use case - the centre mass of research shifting through the centuries

- Motion of the centre of mass of research production from 1671 to present day.
- The centre of mass calculation is unweighted by citations or other measures and is based solely on the **outputs of papers by institutions** that appear in the GRID database
- <https://arxiv.org/pdf/2101.09567.pdf>



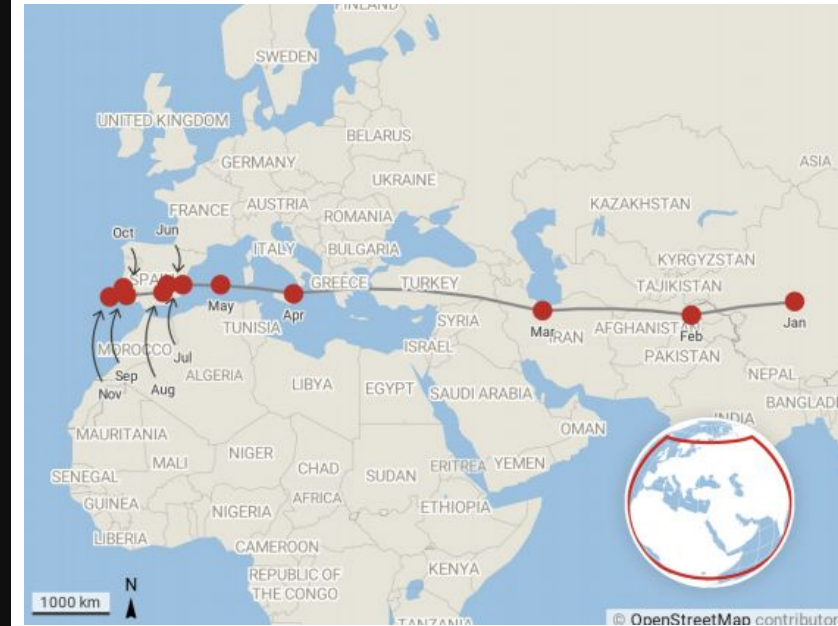
Research use case - the centre mass of research shifting through the centuries

- Motion of the centre of mass of research production from 1671 to present day.
- The centre of mass calculation is based to outputs of papers by institutions that appear in the GRID database and **weighted by citations**



COVID-19 research - centre of mass movement in recent months

- Centre of mass for COVID-19 related research activities from January 2020 to Nov 2020
- The direction 'East to West' will most likely be reversed in the coming 24 months
- On a more serious note, this shows that a linked research information infrastructure allows 'realtime bibliometrics'



Accessibility of linked data matters to many matters!

The analysis were tapping into more than 1.4 billion relations - and only took a 30 seconds on a dataset with 116 million records

Can be replicated within seconds by everyone - without technical barriers, without tedious data preparation!

output in any year, we average the spatial location of all the papers produced in that year. We can think of this position as the "average centre of global research production" or the "centre of mass (gravity of global research output)".

We also introduce a citation-weighted version of this calculation which may be interpreted as a measure of centrality of global research attention to research output. Formally, we define the centre of mass of a set of research objects to be the spatial average (or centroid) of the affiliations of the co-authors of the output. On a paper with n co-authors, each co-author is associated with $1/n$ of the paper. If a given co-author is affiliated with n_i institutions, then each institution will have a weight of $1/n$ of that co-author's part of the paper, and $1/n$ of the overall paper. Thus, each author-institution pairing has a weight $w_{i,n}$ where

$$\sum_{i,n} w_{i,n} = 1. \quad (1)$$

We do not need to explicitly sum over authors to get the overall contribution of a specific institution nor do we need to worry about repetition of institutions since, in our calculation, we reduce an institution to its longitude and latitude of its principal campus. Hence, there is a natural accumulation of weight to a geographical location.

This reduction to longitude and latitude is made possible through the use of GRID. The longitude and latitude of research institutions is not held natively within the Dimensions dataset. However, each institution in Dimensions is associated with a persistent unique identifier that allows us to connect to other resources. In the case of Dimensions the institution identifier is the GRID identifier. GRID not only includes some helpful data about institutions such as the longitude and latitude that we use here but also acts as a gateway to numerous such as ROR, the Research Organization Registry that will in turn facilitate access to other pieces of information.

The maps that we use simply calculate the average longitude, latitude and latitude lat of a single research output using:

$$\bar{lon} = \frac{1}{n} \sum_{i=1}^n lon_i$$
$$\bar{lat} = \frac{1}{n} \sum_{i=1}^n lat_i$$
$$\bar{lon}_{avg} = \frac{1}{n} \sum_{i=1}^n lon_{avg,i}$$
$$\bar{lat}_{avg} = \frac{1}{n} \sum_{i=1}^n lat_{avg,i} \quad (2)$$

where T is the total number of publications.

We can then extend this to a group of outputs by introducing an index, k , that ranges over each output in the relevant set to create the average longitude \bar{lon}_{avg} and average latitude \bar{lat}_{avg} of the whole set.

$$\bar{lon} = \frac{1}{T} \sum_{k=1}^T \sum_{i=1}^n lon_i$$
$$\bar{lat} = \frac{1}{T} \sum_{k=1}^T \sum_{i=1}^n lat_i$$
$$\bar{lon}_{avg} = \frac{1}{T} \sum_{k=1}^T \sum_{i=1}^n lon_{avg,i}$$
$$\bar{lat}_{avg} = \frac{1}{T} \sum_{k=1}^T \sum_{i=1}^n lat_{avg,i} \quad (3)$$

where T_i is the total number of institutional affiliations on the i^{th} paper in the average.

Longitude and latitude are defined as angles on the surface of a sphere with longitude in the range $[-90, 90]$ and latitude in the range $[-180, 180]$. The conventions in Eqs. 2 guarantee that the final results of these calculations are also in these ranges.

Further weighting factors can also be added to the calculation to highlight lines of particular interest. For example, if we were to consider an example using research publications and we wished to calculate not just the centre of the output but also the centre of the combination of output weighted by the attention given to that output, then we might introduce a weighting by the number of citations received for each paper.

In that case (3) would need to be updated and the form for the centroid would be:

$$\bar{lon} = \frac{1}{T} \sum_{k=1}^T \sum_{i=1}^n \sum_{j=1}^{T_i} lon_{i,j}$$
$$\bar{lat} = \frac{1}{T} \sum_{k=1}^T \sum_{i=1}^n \sum_{j=1}^{T_i} lat_{i,j} \quad (4)$$

where C_i is the number of citations of i^{th} paper and C is the sum of all citations across papers in the set.

Indeed, if we were interested in the level of non-scholarly attention we might replace citations by some relevant alternative data.

The code snippet below is the implementation of Eqs. 4 using Google's implementation of RGL in the Dimensions dataset. In addition to the calculation explained above, the code below takes into account cases where research may use an affiliation by missing that information, as detailed in the use of null data.

NOTE: pub_id, first_name, last_name, and institution are stored in the use of null data.

```
def pub_avg_longitude_longitude(pub_id):  
    p_year =  
        a_first_name,  
        a_last_name,  
        a_institution,  
        a_institution_id  
    # count the distinct number  
    # of institutions per author  
    COUNT(DISTINCT a_institution_id) num_insts,  
    # sum of all the GRID IDs  
    SUM(a_institution_id) grid_id,  
    # count the number of authors  
    COUNT(a_institution_id) num_authors  
    # count the number of authors  
    # affiliation to GRID  
    # overpartition by  
    pub_id authors
```

NOTE: 'a_institution_id', data_analytic.
pub_id, first_name, last_name, and institution are stored in the use of null data.

```
SELECT  
    SUM(a_institution_id) num_insts,  
    SUM(a_institution_id) grid_id,  
    COUNT(a_institution_id) num_authors  
    FROM dimensions.  
    data_analytic pub_id
```

```
GROUP BY  
    pub_id,  
    p_year,  
    a_first_name,  
    a_last_name,  
    a_institution
```

```
),  
/* get the location for each GRID. Each  
row that is being summed here  
represents a single author. If  
they have more than one affiliation  
then the contribution of the  
author is split equally. */  
pub_center_year AS  
(  
    SELECT pub_id,  
        COORDS(address, latitude/pr.  
num_insts/pr authors) latitude,  
        COORDS(address, longitude/pr.  
num_insts/pr authors) longitude  
    FROM pub_avg_longitude_longitude  
    ORDER BY grid_id
```

```
),  
/* sum the centre mass for all  
publications, replacing p_year.  
time_cited in the respective to an  
explicit value of "1" (recovers a  
negative time from the calculation)  
data_analytic pub_id  
GROUP BY pub_id
```

```
/* sum the centre mass for all  
publications, replacing p_year.  
time_cited in the respective to an  
explicit value of "1" (recovers a  
negative time from the calculation)  
data_analytic pub_id  
GROUP BY pub_id
```

```
/* sum the centre mass for all  
publications, replacing p_year.  
time_cited in the respective to an  
explicit value of "1" (recovers a  
negative time from the calculation)  
data_analytic pub_id  
GROUP BY pub_id
```

```
/* sum the centre mass for all  
publications, replacing p_year.  
time_cited in the respective to an  
explicit value of "1" (recovers a  
negative time from the calculation)  
data_analytic pub_id  
GROUP BY pub_id
```

```
/* sum the centre mass for all  
publications, replacing p_year.  
time_cited in the respective to an  
explicit value of "1" (recovers a  
negative time from the calculation)  
data_analytic pub_id  
GROUP BY pub_id
```

```
/* sum the centre mass for all  
publications, replacing p_year.  
time_cited in the respective to an  
explicit value of "1" (recovers a  
negative time from the calculation)  
data_analytic pub_id  
GROUP BY pub_id
```

negative bias (i.e. only publications of the major year are averaged together), however, this does introduce an explicit bias in the analysis in that a citation bias may have a counterweight effect. A further analysis could be performed that only considered the citations in a n -year window following the date of publication of the paper. Of course, introducing such a parameter also makes a value judgement about the lifetime of a piece of research.

In Sec. III we use this method to showcase three analyses:

- 1) a standard unweighted calculation of the centre of mass of research output from 1871 to present day; 2) a calculation of the centre of mass of research weighted by citation attention over the same time period; 3) a calculation of the citation-weighted centre of mass of research based just on data from the freely available COVID-19 dataset that is available on the Google BigQuery interface.

C. Data specificity

The details of the high-level data within Dimensions, including information about coverage and the treatment of single identifiers is described in several recent publications, for example [25, 26].

Once the data are prepared from a script such as the one above they were downloaded from the interface and are initially analysed in Mathematica. The graphics shown in Sec. III are produced using Mathematica.

At the Mathematica analysis stage, we plotted every year of data from the system. However, this gave us a satisfactory picture as the data are quite noisy. In the early years of the dataset (approximately from 1871-1900) the number of publications with a GRID-based institution number is the single digit. A consequence of noise contribute to this picture: (i) the low number of annual publications; (ii) the low level of annual academic affiliations of authors in early work; (iii) affiliations to institutions that are not part of GRID. Figure 1 shows the number of publications with at least one recognizable GRID-recognized affiliation in each year in the Dimensions dataset.

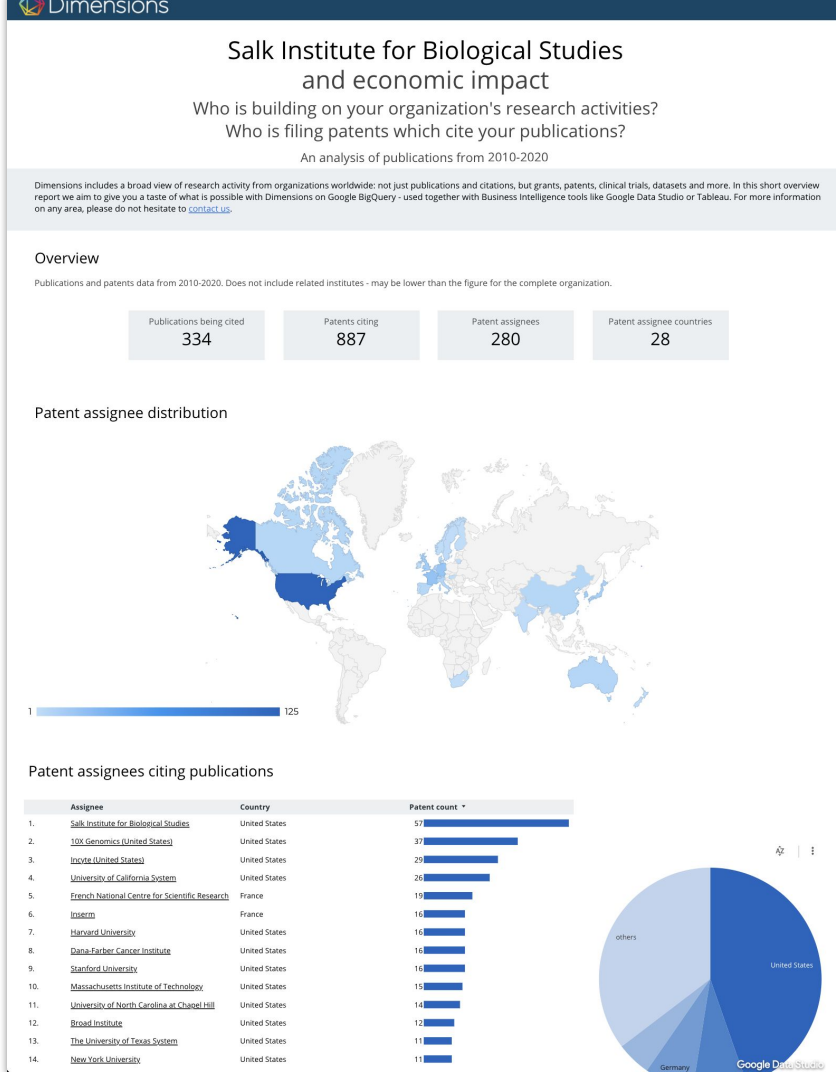
From 1900, the data begins to settle as it begins to be appropriate to treat it statistically in the context of a statistical calculation such as the one outlined in Sec. IIB. Between 1900 and 1970, the year-on-year variability of the data decreases, and from the 1970s the data describes a fairly consistent path with less significant deviations.

Listing 1: Listing to produce a citation-weighted centre of mass year-by-year using RGL on Google BigQuery with Dimensions data.

In the final analysis presented, we focus on the COVID-19 dataset. The data is available in the Dimensions dataset. In this situation, we are again in the low of relatively small numbers where we have had to be careful about analysis. In fact, however, the COVID-19 dataset has grown quickly during 2020 with a low level of noise. However, given to several thousand papers per month in November

Economic impact based on patent citation analysis

- One of the links established: which publications are cited by patents
- If the publications of one organisation are taking into account, the organisation can quickly see which companies and other institutions are 'standing on their shoulders'
- Without 'ready-made' links, this is out of reach for most organisations
- With a linked data infrastructure, it takes seconds
 - As a simple SQL query
 - ... or even as a dashboard summarising all the information



Supporting the publishing workflow - possible with a ready to go linked research information resource

Manuscript processing

Context for authors via ORCID or name matching

Context for research topic at host institutions

Funded grants and funders for APC eligibility

Journals with semantically similar publications

Review process

Identify reviewers based semantic matching

Provide conflicts of interest for potential review candidates (co-author and institutional affiliation)

Post publication

Monitor reception and indicators (citations, altmetrics)

Build up dynamically fueled author network

Monitor preprint activity of your authors

Monitor funding of your authors

All possible with a few persistent identifiers and linked data - e.g. provided as Dimensions on Google BigQuery!

Creating real 'Linked Data' is possible...

- ... but more a question of copyrights and interests!
- Digital Science realized for Springer Nature a Linked Data graph with appr. 2 billion triples using Dimensions as the basis



The screenshot shows the SN SciGraph website. At the top, the Springer Nature logo is on the left, and search and language options are on the right. Below the header is a dark blue banner with a network graph background. The banner contains the text "SN SciGraph" in large white letters, followed by "A Linked Open Data platform for the scholarly domain" and a "Watch video" button. Below the banner, the main content area has a welcome message, additional data details, and a paragraph about the platform's goals. To the right of the main text is a vertical sidebar with several buttons: "Any questions? Please contact us.", "Latest Datasets Download", "Data Explorer", "Licensing Information", and "Results: First SN SciGraph Hack Day". At the bottom, there is a section titled "Further Info" with a PDF download link for "SN SciGraph at a glance".

SPRINGER NATURE

Search EN

Menu

SN SciGraph

A Linked Open Data platform for the scholarly domain

Watch video

Welcome to SN SciGraph, our Linked Open Data offering which aggregates data sources from Springer Nature and key partners from the scholarly domain. The Linked Open Data platform collates information from across the research landscape, for example 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, SN 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. Our aim is to increase the discoverability of high quality data as larger parts of our datasets are being made available under CC-BY licensing.

SciGraph

Any questions?
Please contact us.

Latest Datasets
Download

Data Explorer

Licensing
Information

Results: First SN
SciGraph Hack
Day

Further Info

SN SciGraph at a glance
(PDF, 979.37 KB)

The data in SN SciGraph is projected to contain 1.5 to 2 billion triples. We are constantly iterating on adding more 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.

- ... and while we cannot provide with Dimensions a Linked (Open) Data set in the pure technical sense (yet)
- ... the team is focusing on providing a versatile linked data 'cape' which suits many, which is safe and is easy to wear!
- ... to make linked data speak so that we can understand the complex processes better and facilitate information sharing!



Thank you!

Christian Herzog
CEO Dimensions

c.herzog@digital-science.com

+49-171-5678917