

Open Science – should
commercial organisations
be banned from
contributing?

Christian Herzog, MD

OPERA Workshop, Copenhagen, March 2019

Introduction and background

Risky
topic...



Part 1: Personal part

Introduction and thoughts on the topic.
Personal views, all mistakes and flaws
are mine, also deliberately over-simplifying.



Part 2: How we as Digital Science try to make 'it' work

Again, not speaking for all colleagues, but trying to
illustrate how Digital Science sees itself in general
in the Open Science / Open Research space,
using the launch of Dimensions as illustration



Commercial companies and Open Research

Really at odds?

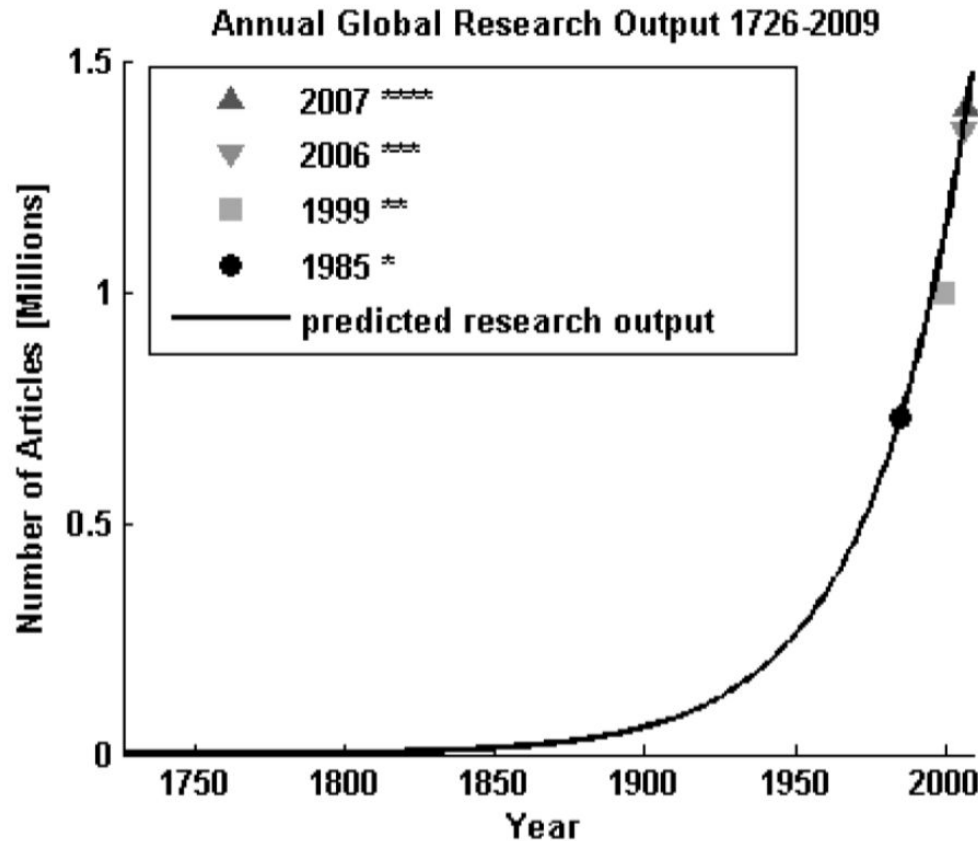


... while the interactions and language sometimes suggest a battlefield...



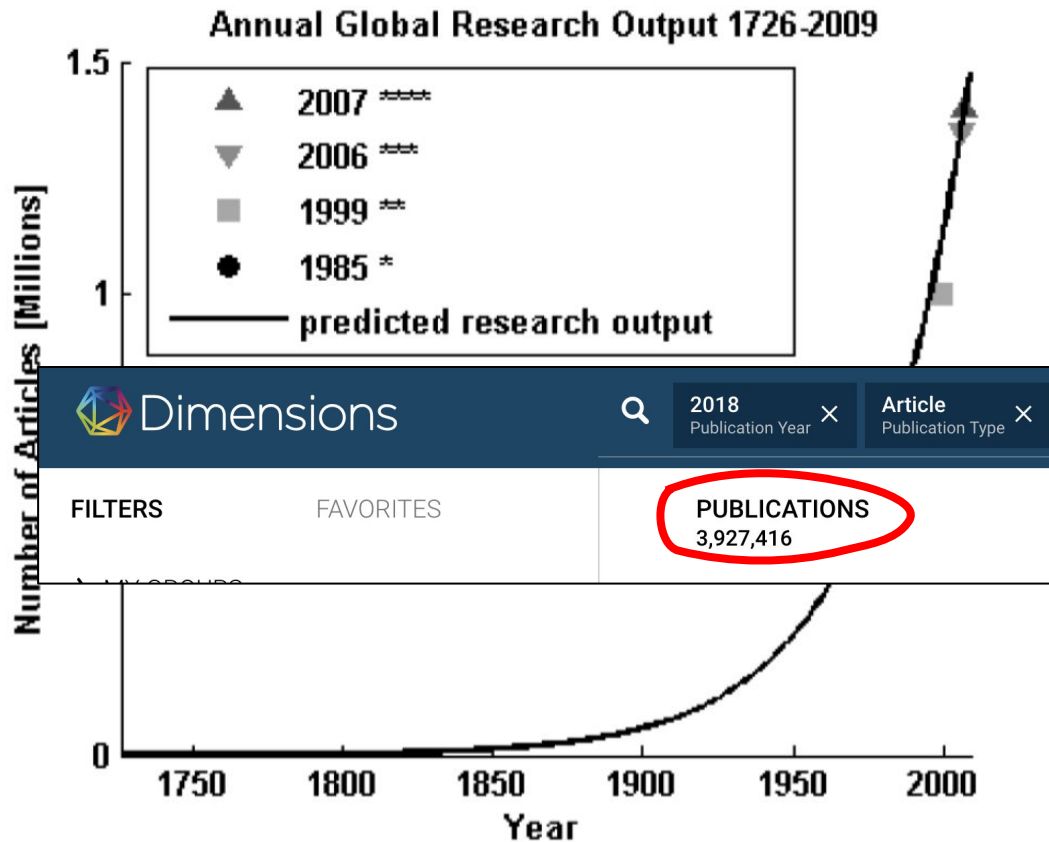
But what is really the core of the conflict
and the changes we are seeing these
days?





- Exponential growth starting after 1945, almost a ‘Scientific Revolution’, or the ‘Industrialisation of Science’
- Not a technical trigger - printing has been around for a while
- Purely growth in size, resulting in diversification of disciplines, no. of journals





- Exponential growth starting after 1945, almost a ‘Scientific Revolution’, or the ‘Industrialisation of Science’
- Not a technical trigger - printing has been around for a while
- Purely growth in size, resulting in diversification of disciplines, no. of journals



Scholarly societies
facilitated research
communications,
community aligned
and driven

Large publishing
industry, meeting the
needs of the growing
research eco-system



Resulted in 'mostly happy' for a while....



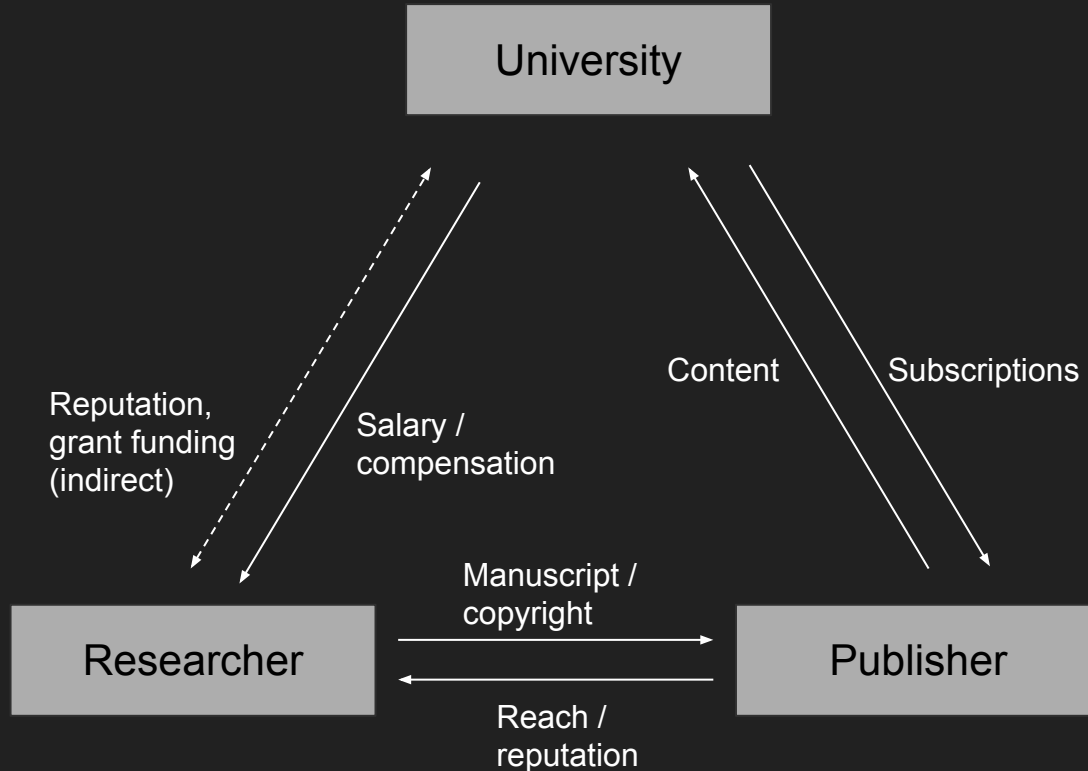


Growth results in increased competition

- Reputation on a journal level - nothing new
- But reputation management went 'industrial' and commercial with a 'replacement currency'
- e.g. JIF - a metric to measure and rank and run journals - used for very different tasks
- High dependency for tenure & review process, national assessment framework etc.



A growing research system and asymmetrical relations

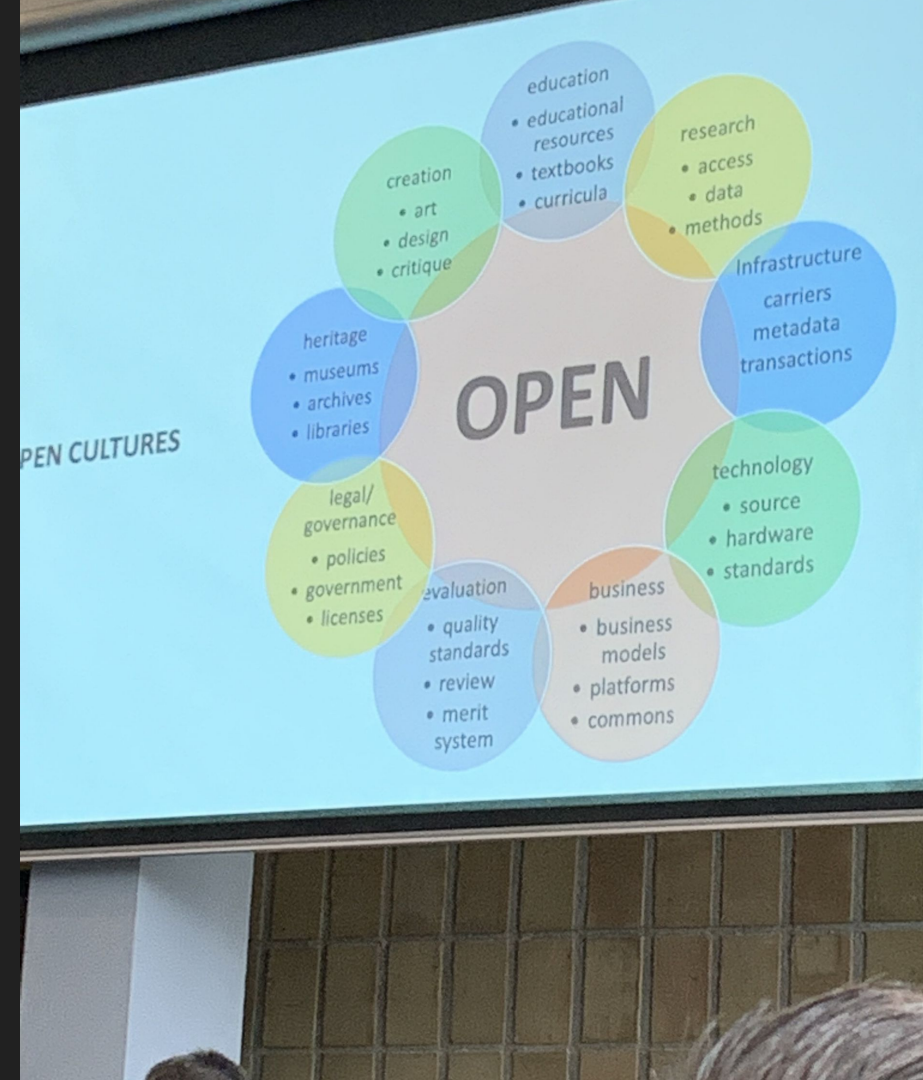


Commercial overdrive kicked in...

- Externally: increasing subscription costs, bulk licensing only, dependency on content
- Increasing: dependency on the citation / journal level reputation

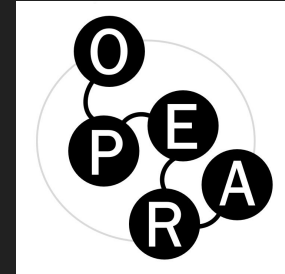


Open Research -
the concept is not
really new... but
also multi-faceted



Two aspects to Open Research

- Publications, data sets etc. openly available and accessible for collaborative research activities
- Changing the value / reputation management system required as well





Open
Research



Commercial
companies



Commercially organised players



A way to organise a group of people, focused on an objective

- Ideally, private resources are invested to create a product / services which solves an issue, a problem shared by many
- The costs for the individual / organisational user for the services / products is significantly lower than building it and the value exceeds the costs
- And the client / user gets to decide, whether the product is worth the price
- If not, the entrepreneur / company has lost its investment / private money
- The key question: what are the objectives / values of the group of people?





Startup vs. big corporate player



Digital Science (a commercially organised company) and the launch of Dimensions

Digital Science - a digital native / unusual corporate animal

- Organised as a commercial entity, started about 10 years ago, with the clear mission to shape change which was about to come
- Purpose: attract, support and develop small innovative software companies which focus on improving a step / challenge in the research process
- Track record of early movers in their space: Altmetric, Figshare, Symplectic, ÜberResearch and more...
- Nothing to defend - no legacy business
- Committed to open research
- A shareholder committed to create value in terms of impact



Sustainable startup genes and constantly evolving

- Portfolio structure -small focused groups working with partners and clients - empowered teams
- Close to the research community
- Not a 'fire it up and look for the exit' mentality, most funders are still around - 10 years after DS started
- Founders often move into new roles
- Bottom up integration - portfolio companies grow up together as siblings
- Starting to integrate and share more and more...
- Overall: highly motivated team, opinionated, value driven, striving to create a positive impact - resulting in a fair amount of creative chaos



Committed to Open Science - striving for the right balance



Figshare - research data management, empowering the research to manage and share



Altmetric - experimenting with the right balance of making parts openly available and still creating added value for a sustainable business model



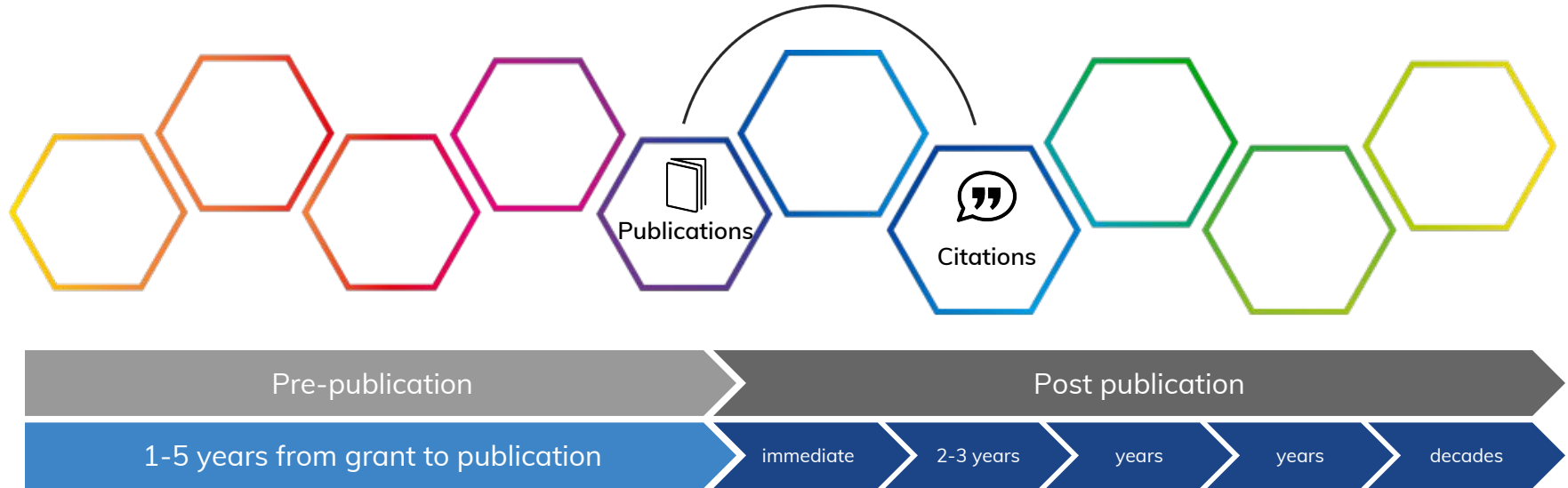
GRID: tackling the challenge of ambiguous affiliation data with a 90k organisation entries and persistent identifiers - openly available under CCO license for the research community



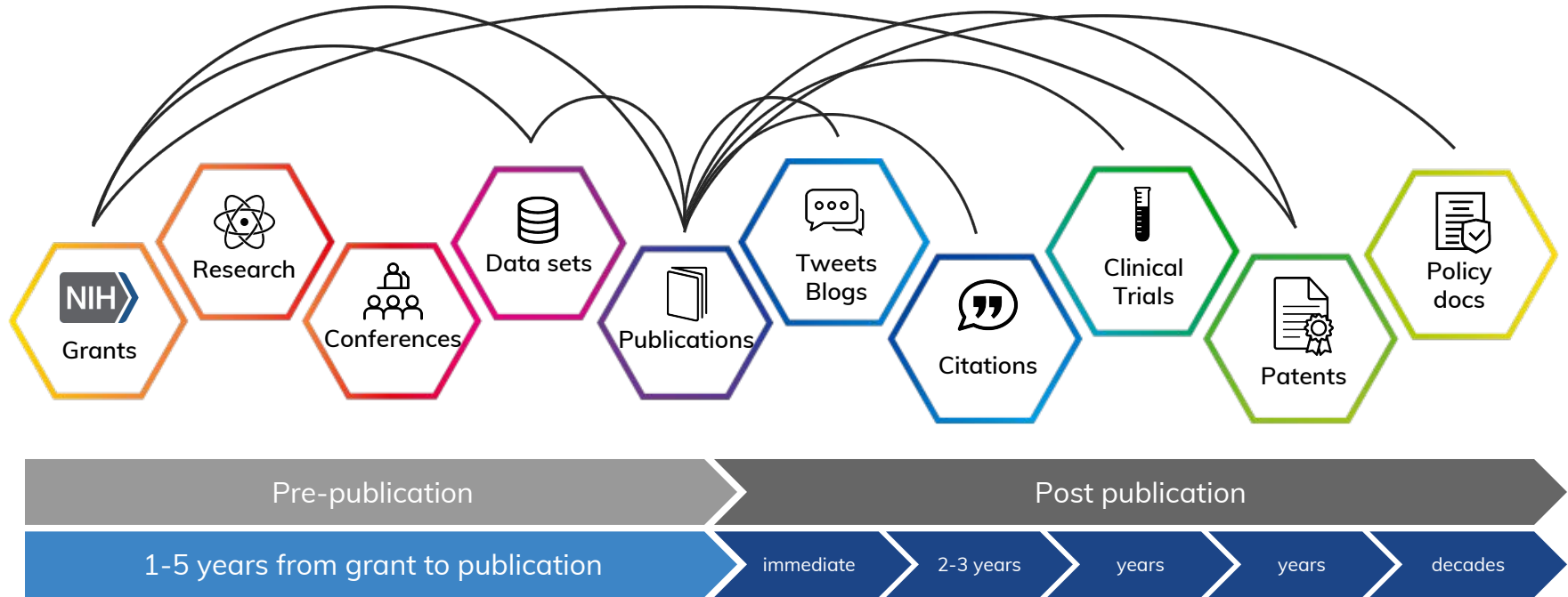
Dimensions - establishing a new, modern research information - as openly available as possible, realising a model where secondary data is not monopolised but where the business model is based on value created on top of the data

Why did we start Dimensions?

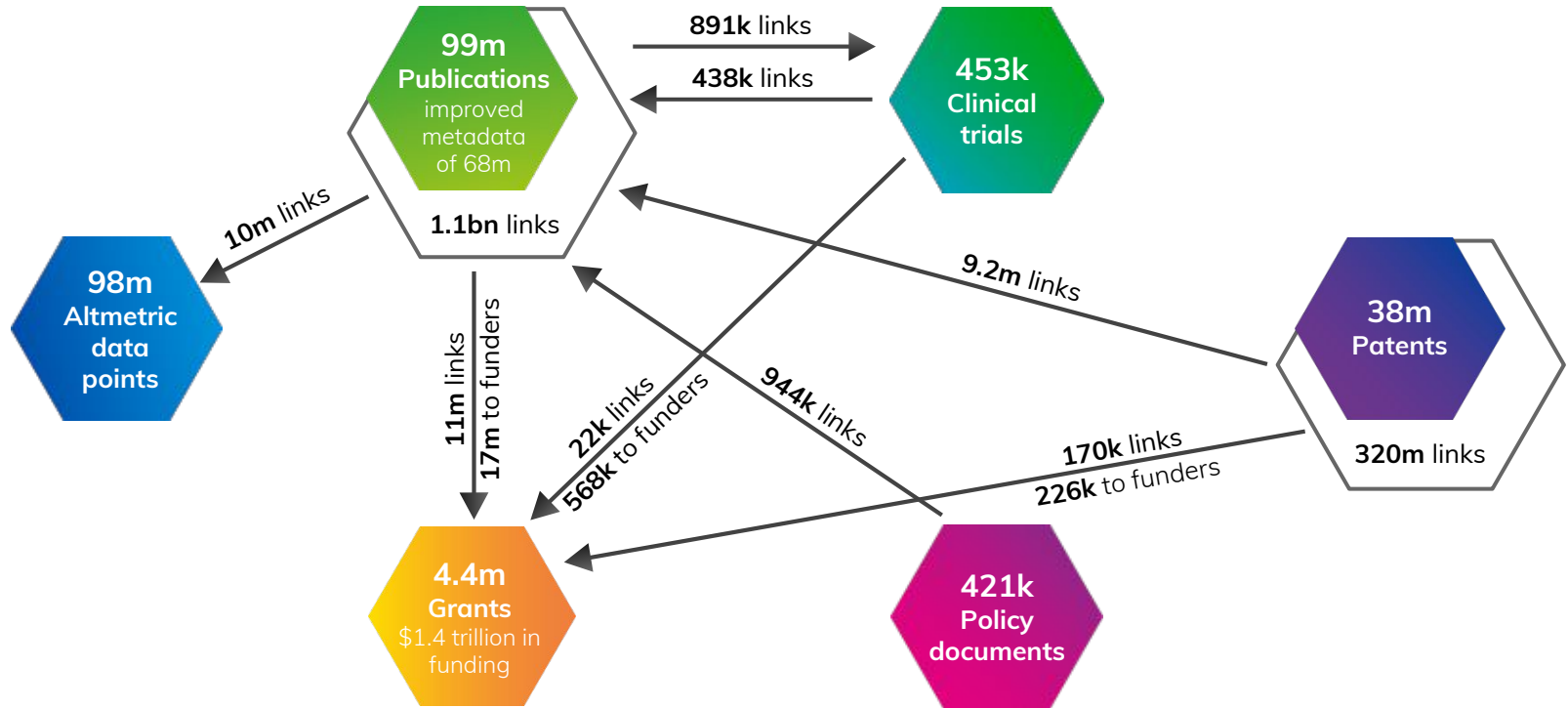
Discovery and research assessment: a narrow focus for years



More “Dimensions” possible and required



Broader view, links to follow the trajectory, more colorful picture



Publications / citations as openly available as possible

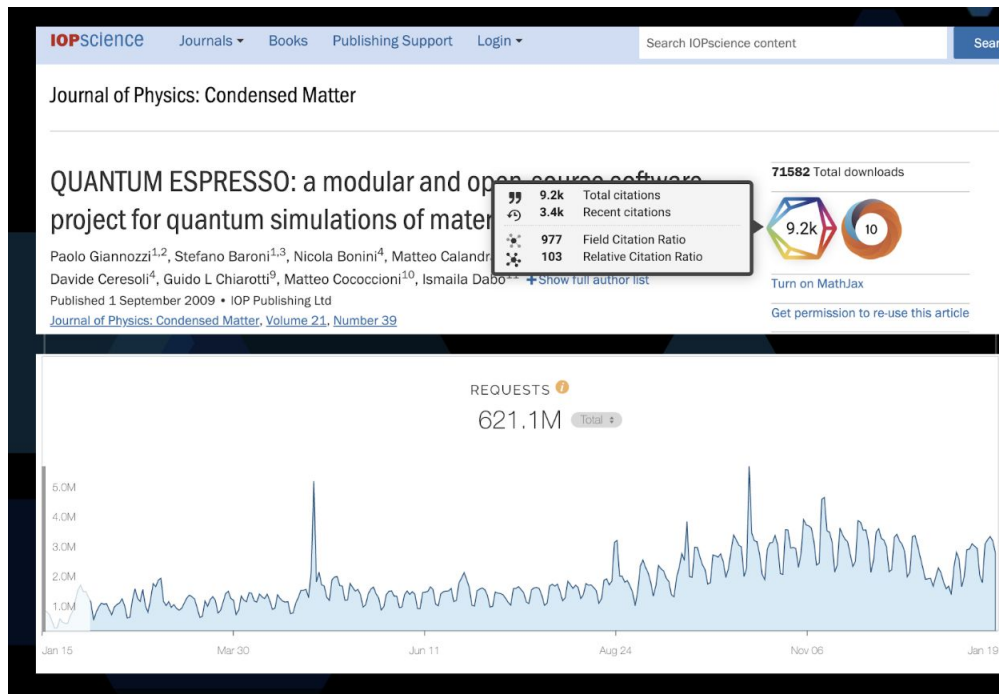
Dimensions is
freely available for
personal use -
even without a
login

99M publication
records and more
than 1bn citations

The screenshot displays the Dimensions database interface. At the top, the search bar contains the query 'e.g. plastic AND instrument'. The interface is divided into several sections:

- Filters:** A sidebar on the left with 'PUBLICATION YEAR' expanded, showing a list of years from 2010 to 2019 with corresponding citation counts. Other filter categories include RESEARCHER, FIELDS OF RESEARCH, PUBLICATION TYPE, SOURCE TITLE, and JOURNAL LIST.
- Publications List:** The main area shows a list of publications. The first entry is 'Debonding mechanism of zirconia and lithium disilicate resin cemented to dentin' by Mina Aker Sagen et al. (2019). The second entry is 'The effect of antimicrobial additives on the properties of dental glass-ionomer cements: a review' by Tamer Tüzüner et al. (2019). The third entry is 'Synthesis of Ag/NiO Honeycomb Structured Nanoarrays as the Electrode Material for High Performance Asymmetric Supercapacitor Devices' by Sadayappan Nagamuthu et al. (2019). The fourth entry is 'EZ Entropy: a software application for the entropy analysis of physiological time-series' by Peng Li (2019).
- Analytical Views:** A sidebar on the right with 'FIELDS OF RESEARCH' expanded, showing a list of research fields. Below this is an 'OVERVIEW' section with a line graph showing the RCR Mean (0.76) and FCR Mean (1.32) over time, and a 'RESEARCHERS' section.

Open Metrics API: more than 621 million calls



... we receive
regular requests
from more
than 500
implementations
so far

API: not only data, also exposing the business logic layer

- Powerful API - not only for the extraction of data
- Also giving access to the business logic layer to allow users to build their own analysis and tools on top of it
- e.g. realising composite indicators across content types



Dimensions API

The Dimensions API provides access to Dimensions data directly, and makes it possible to retrieve results to precise and complex queries expressing queries using terms and structures relevant to the Dimensions data.

Query

```
search publications in full_data for "CRISPR/CAS9"
where year in [2013:2018] and type="article" and open_access="Open Access - all"
return publications [basics-issue-volume-pages+doi+times_cited] sort by times_cited
return research_orgs
return funders
```

Run

Results

```
{ 4 items
  "research_orgs": [...] 20 items
  "publications": [ 20 items
    0 : { 8 items
      "title": "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2"
      "author_affiliations": [ 7 items
        0 : { 3 items
          0 : { 6 items
            "first_name": "Michael I"
            "last_name": "Love"
            "orcid": "0000-0001-8401-0545"
            "current_organization_id": "grid.10698.36"
            "researcher_id": "ur.0612232567.50"
            "affiliations": [ 3 items
              0 : { 8 items
```

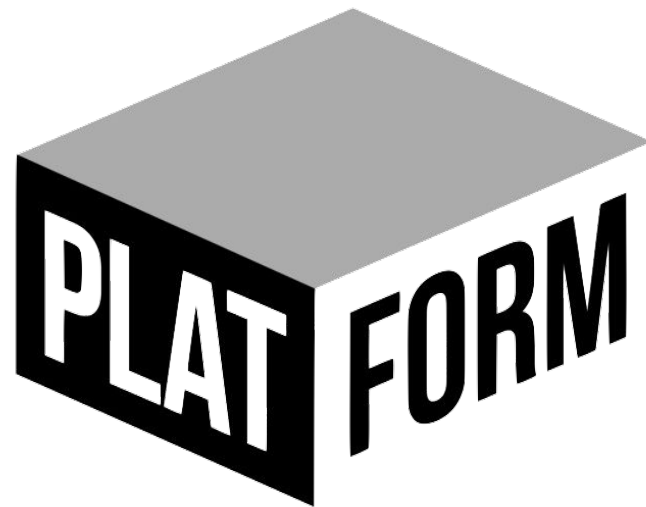
As little strings attached as possible

- The Dimensions data comes with as little strings attached as possible
- 'Possible' being define as:
 - Taking our agreement with copyright holders into account
 - Avoid large scale cannibalisation effects which endanger our sustainability
- Principle: pay for data once only - to support the efforts



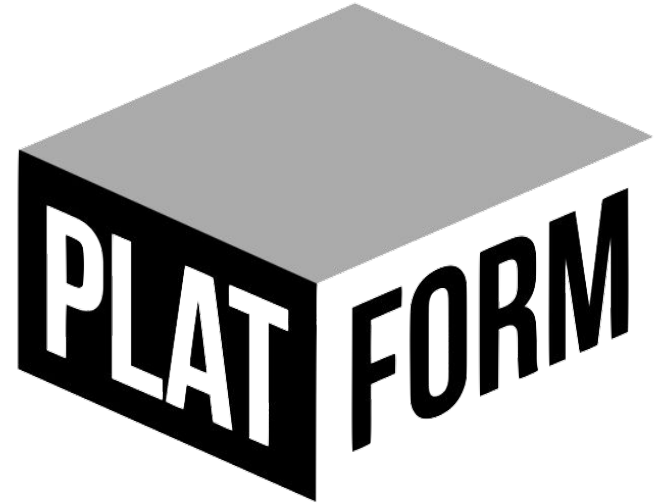
A research data platform for the research community

- Aggregating open data from various sources (citations, grants, trials, policy documents)
- Work with copyright holders to add in 'closed' content not openly available elsewhere..



A research data platform for the research community

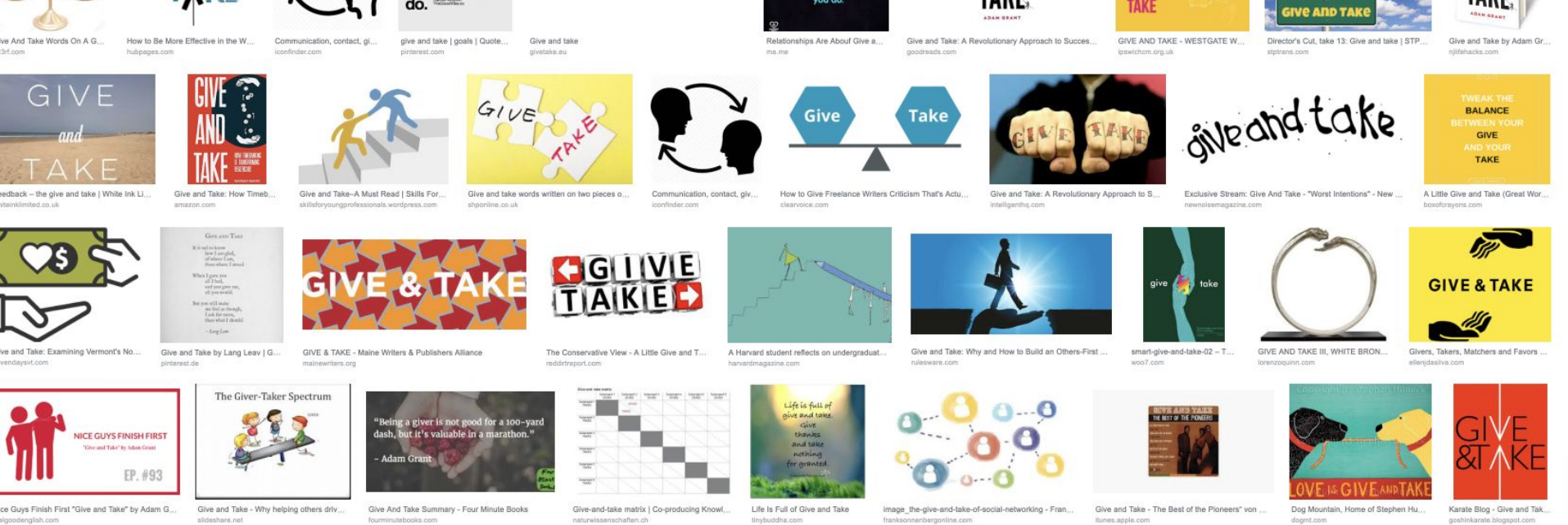
- Aggregating open data from various sources (citations, grants, trials, policy documents)
- Work with copyright holders to add in 'closed' content not openly available elsewhere (Elsevier...)
- Investing heavily in harmonisation and curation, attach persistent identifiers, create linked dataset across content types
- The 'small task' of operating a redundant platform with more than 143M documents and billions of connections, appr. 1.8M unique users and 24/7 on call dev-ops team,
- ... not to mention support, engagement and product development etc.



Dimensions as an infrastructure for the research community

- The business model around Dimensions has been determined with development partners
- Costs engineered to be roughly what these partners would be willing to spend for a reliable data infrastructure which keeps evolving
- We think we as Digital Science should not be in the business of developing proprietary indicators
- Instead: the scientometric community is invited to make use of the platform and the API
- ... to do research on broader, refined and new indicators, as diverse as possible





And how are we trying to live up to these statements?

Concretely: access to data and tools for scientometric research purposes

Simple process:

1. Indicate interest (via short form, directly)
2. Review by Dimensions team, led by Stacy Konkiel
3. Short no-cost license agreement
4. Access

Currently more than 50 research cooperations with focus on

- Funding - output - impact analysis
- co-authorship networks
- citation analyses to
- patent-grant relationships

Increase of interest in the last months, more and more requests are coming in

-> https://dimensions.ai/data_access



Application for non-commercial use of Dimensions

Please complete this form to apply for a non-commercial agreement to use Dimensions tools and data. Non-commercial use includes research projects, open source software development, and teaching/learning. We will review your application and contact you via email (usually within 10 business days) to discuss next steps.

*** Required**

Full name *
First name and surname
Your answer

Email address *
Your work email
Your answer

Employer *
List the organization where your project will be carried out.
Your answer

Job title *
Your answer

If you are a student, please list the name and email address of your faculty supervisor
Your answer

Mailing address *
List your work mailing address, including city, state/province, country, and postal code. We will not use this information for any other purpose than to complete a no-cost agreement if your application is accepted.
Your answer

Briefly describe your work or research experience relevant to your proposed project *
Your answer

In as much detail as possible, describe the project you plan to use Dimensions tools and data for *
As you include your project's goals, list expected results for the project (e.g., from data gathering or analysis to publications) the kinds of data you plan to ingest or use as part of the project, any information on related research projects, and any related grants or publications. If you would like to use Dimensions data for more than one project, please list and describe them separately using an ordered list (e.g., "1. Research project..."; "2. Webapp project..."; etc.).

Strike the right, responsible balance!

Supporting Open Science and pushing for open in general:

- Making data available where it requires little efforts and can scale - do it at scale, as efficiently as possible
- Push the boundaries as long as the long term sustainability is not endangered
- Create understanding and trust that sustainability will be achieved only by working together - and that includes commercial arrangement
- We hope that what can be done with the data is more important than whether it is

Organisation

- Keep small, empowered groups, driven by a strong sense of value
- Move from data availability to feature/function/value competition



... and listening the past two days a little bit, these are the projects I can see coming out of the OPERA workshop

- Continued cooperation with the OPERA project, to use Dimensions data API for a large VIVO pilot - agreed, agreement in preparation
- Joint research project to use citation context / sentiment analysis for better search results - potentially processing 65M full text record - Birger Larsen, Aalborg University Copenhagen
- Open Science Monitor - our data is as openly available as we can afford for indicators and metrics - and transparent and we have no legacy to defend
- Data and metrics need to be disentangled - take us up on it!
- DTU RAP - automation of the collaboration platform with the Dimensions API.... pulling it in real time...
- Meeting to explore beyond the OPERA project how we could fit in into an open indicator architecture - only 2:40mins train ride from Cologne to Hannover!



So?





?



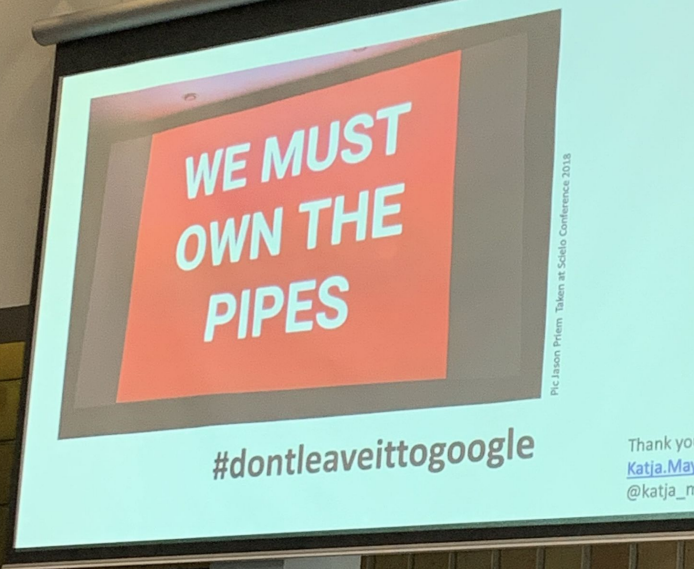
?

Perhaps there is the possibility of including

- transparent,
- mid size commercial partners
- which are aligned with the research community,
- predictable and still innovative

as accepted partners and included
in the 'we' in 'We must own the pipes'?

Picture: [link](#)



Credit to J.Priem via
K.Mayer

Thank you! / Questions?

Christian Herzog, MD,
Chief Portfolio Officer / CEO Dimensions
+49-171-5678917
c.herzog@digital-science.com

