Dimensions

Performing Research Analytics at Scale using Dimensions & BigQuery

2nd June 2025
ESWC, Slovenia

Michele Pasin
VP Product, Data
m.pasin@digital-science.com

Part of DIGITAL science

Summary

1. About myself & Digital Science
2. BigQuery & Cloud based research analytics
3. Our Looker-based research analytics platform
4. Live demos

About me



Michele Pasin
VP Product, Data
Digital Science

2019 — **Digital Science**

Keys: data visualization, research analytics , bigquery, dashboards

2012 — **Springer Nature**

Keys: data architecture, digital publishing, scigraph, scientific modeling, knowledge management

2008 — **King's College, Digital Humanities**

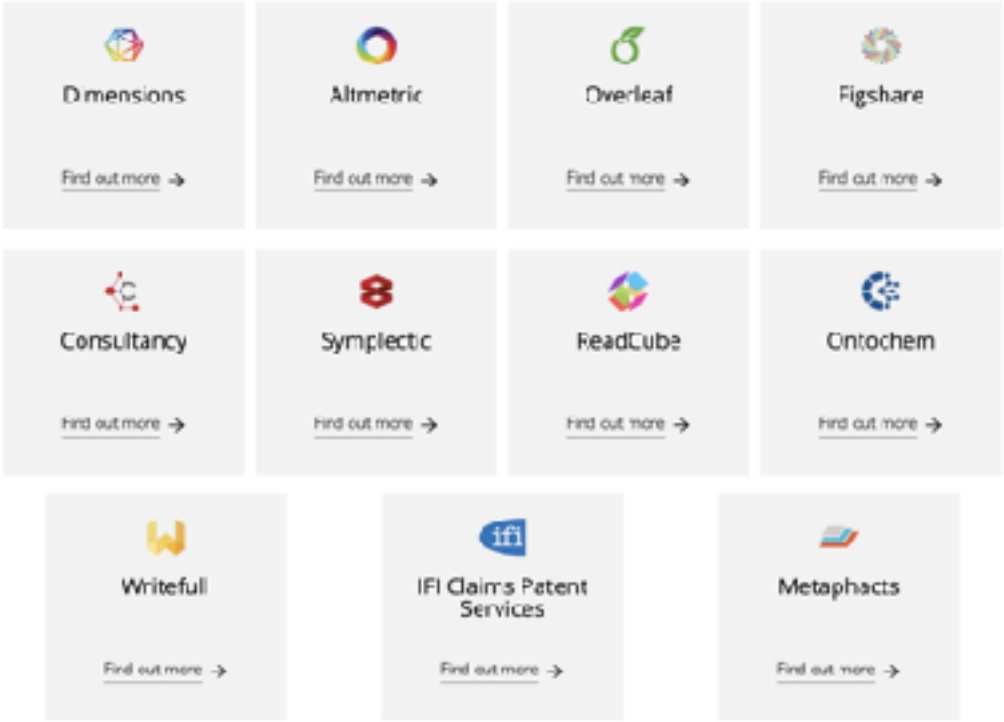
Keys: digital history, digital art, ontology, databases, Python

2003 — **PhD, Semantic Web & eLearning**

Keys: ontology, semantic web, digital narratives, RDF< OWL, LISP

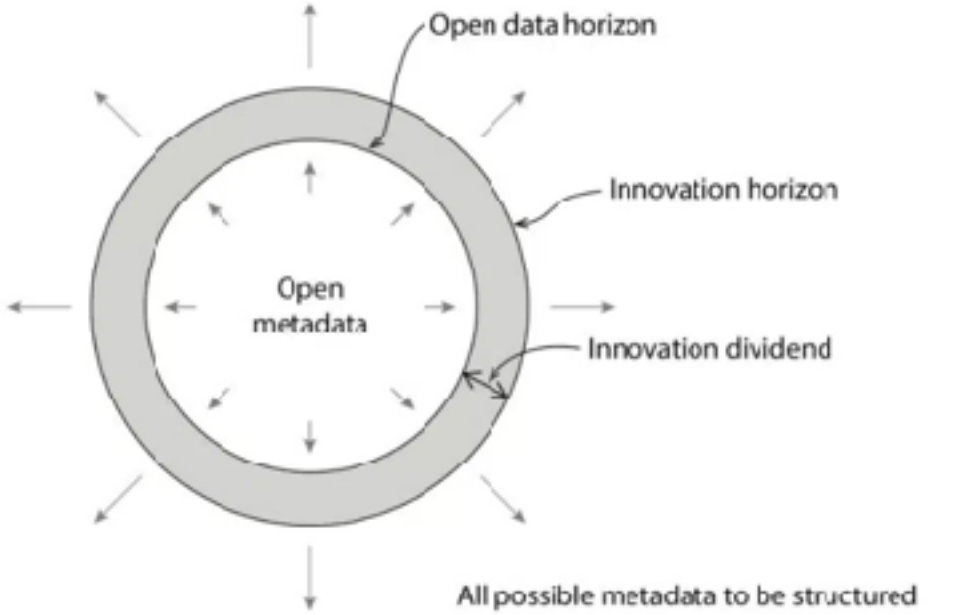
Digital Science – covering all stages of the research lifecycle

- Started in **2009**, invested in and started 12 companies, to date about 700 colleagues
- **Mission**: helping game-changing businesses that rewire the world of research for the better

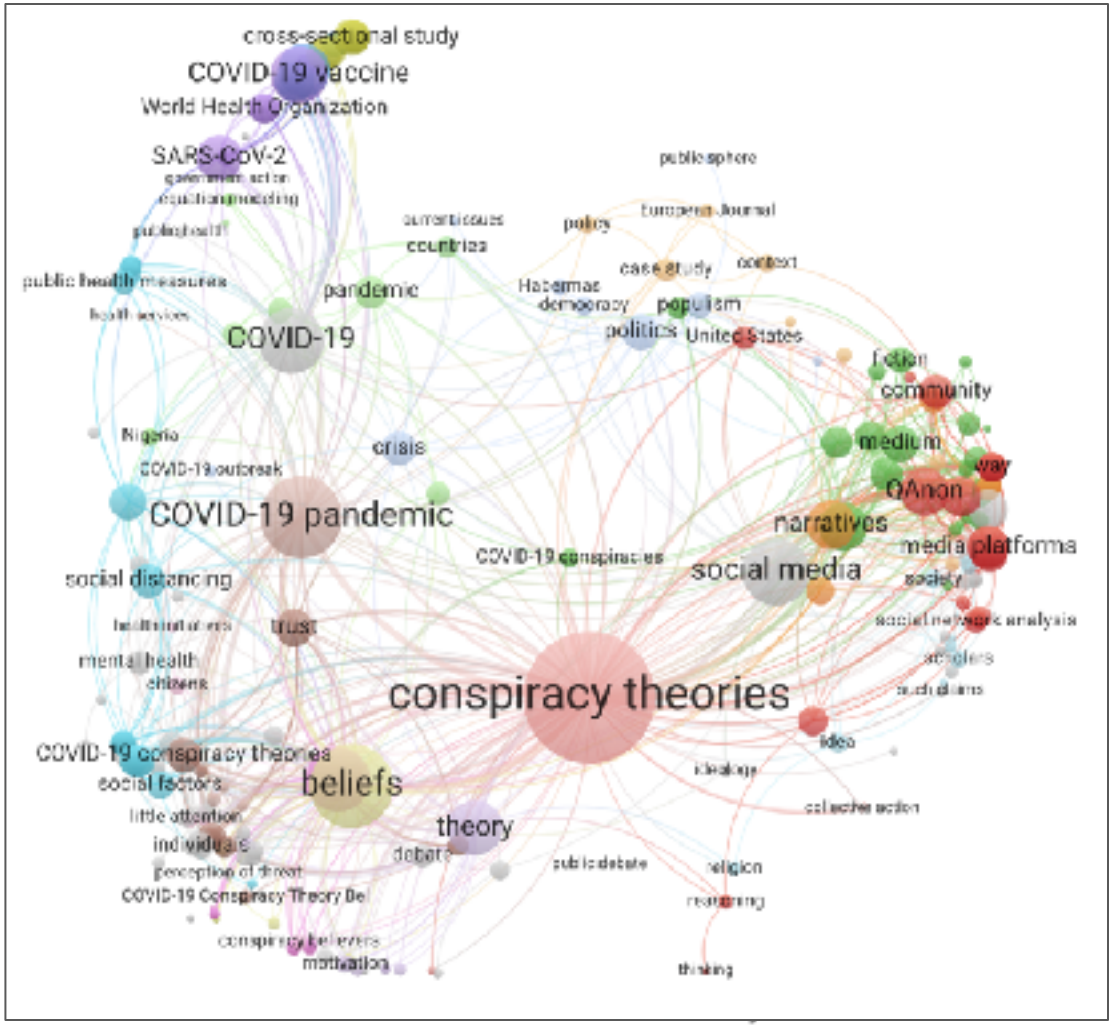


Digital Science – covering all stages of the research lifecycle

- **Rooted in research**, most founders come from an academic background, started a company around solving a self encountered problem
- **Key principle:** striking the right balance between enabling entrepreneurial spirit and activity and supporting open science and open data e.g. Barcelona Declaration



Daniel Hook, 2024, Barcelona: A beautiful horizon., https://www.digital-science.com/tldr/article/barcelona-a-beautiful-horizon/


cross-sectional study
COVID-19 vaccine
World Health Organization
SARS-CoV-2
public health
public health measures
health services
pandemic
COVID-19
current issues
countries
policy
public sphere
European Journal
case study
context
Habermas
democracy
populism
politics
United States
fiction
community
medium
way
QAnon
narratives
media platforms
social media
society
social network analysis
scholars
conspiracy theories
idea
ideology
collective action
religion
reasoning
thinking
public debate
Nigeria
COVID-19 outbreak
crisis
COVID-19 pandemic
COVID-19 conspiracies
social distancing
trust
mental health
citizens
COVID-19 conspiracy theories
social factors
beliefs
little attention
individuals
theory
debate
perception of threat
COVID-19 Conspiracy Theory Bel
conspiracy believers
motivation

Dimensions

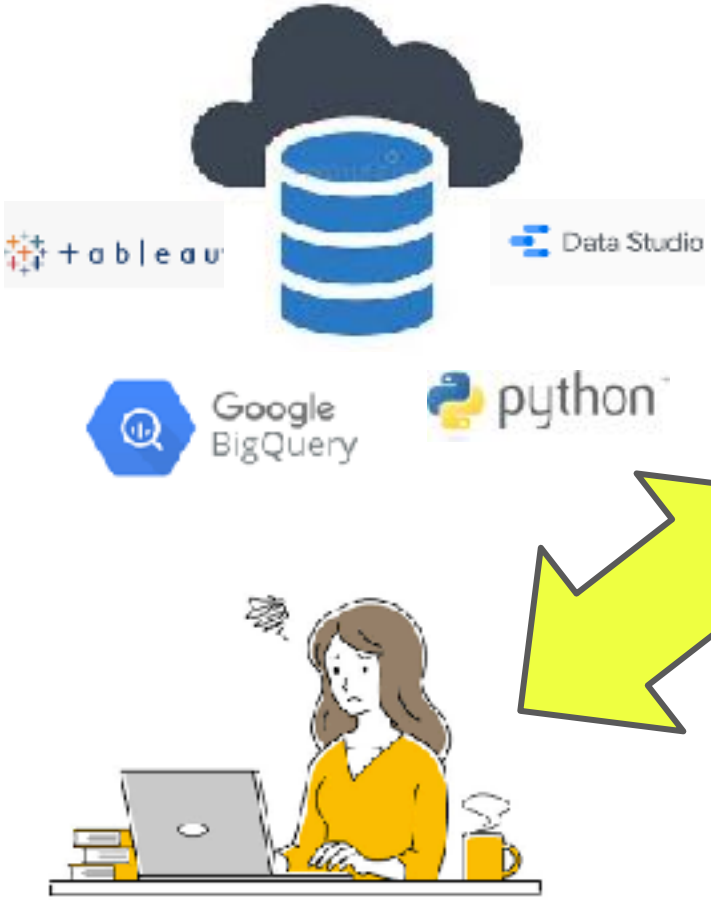


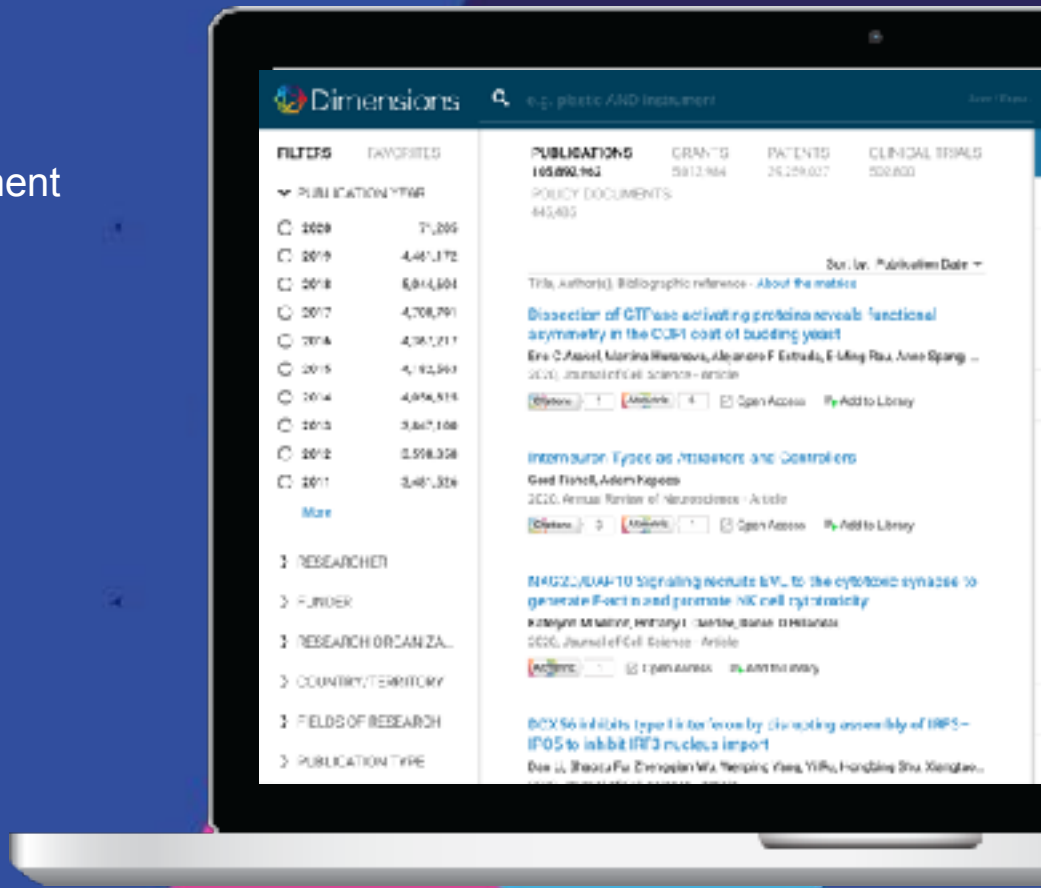
tableau
Data Studio
Google BigQuery
python
Faster
Scalable
Reproducible
cross-sectional study
COVID-19 vaccine
World Health Organization
SARS-CoV-2
public health measures
pandemic
COVID-19
Nigeria
COVID-19 outbreak
crisis
COVID-19 pandemic
social distancing
trust
mental health
COVID-19 conspiracy theories
beliefs
theory
conspiracy theories
public sphere
policy
countries
case study
context
Habermas
democracy
populism
politics
United States
community
medium
war
QAnon
narratives
media platforms
COVID-19 conspiracies
social media
society
social network analysis
scholars
idea
ideology
collective action
public debate
religion
reasoning
thinking

What is Dimensions?

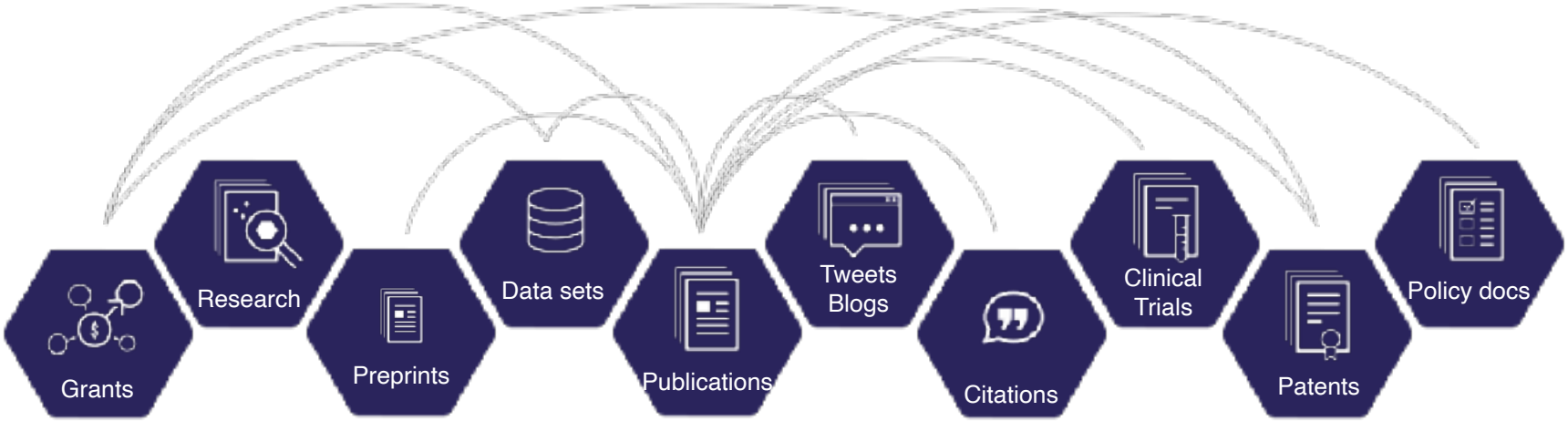
Dimensions delivers an array of search and discovery, analytical, and research management tools, all in a single platform.

- **Moving beyond citations** to better capture the research lifecycle: the **links among** publications, grants, policy documents, clinical trials, datasets, patents, researchers are the key to understanding this landscape.

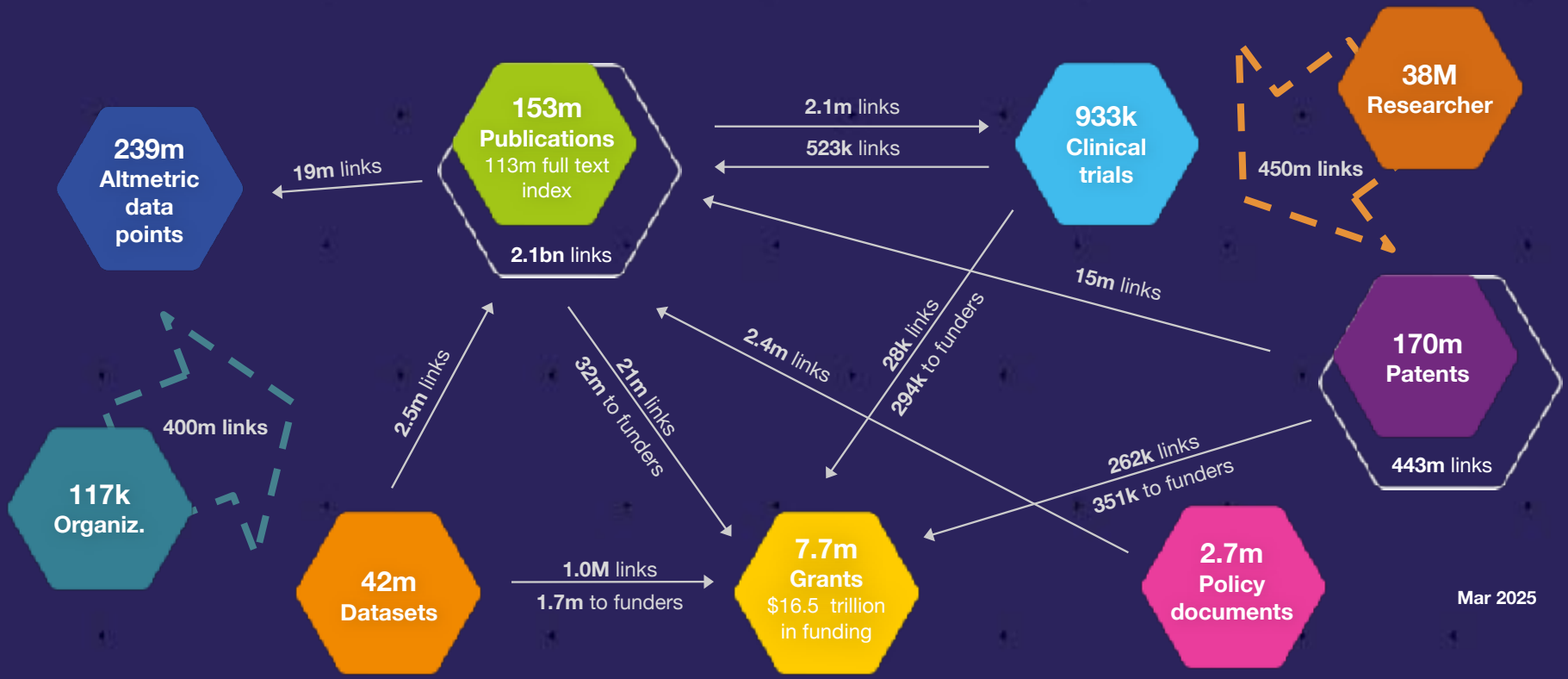
https://www.dimensions.ai



We capture a much broader picture of the research landscape



The science data jungle



Dimensions

Altmetric

ifi CLAIMS

Dimensions ‘flavors’

Web App



- Search & discovery; top analytical use cases
- Dedicated UI, inbuilt visualizations
- In the browser, no specialized knowledge required

For everyone

API



- Ad hoc analysis & single data-type analyses
- Full-text search & special functions e.g. affiliation extraction
- Product integrations e.g. CRIS

For API users + data & analytics teams

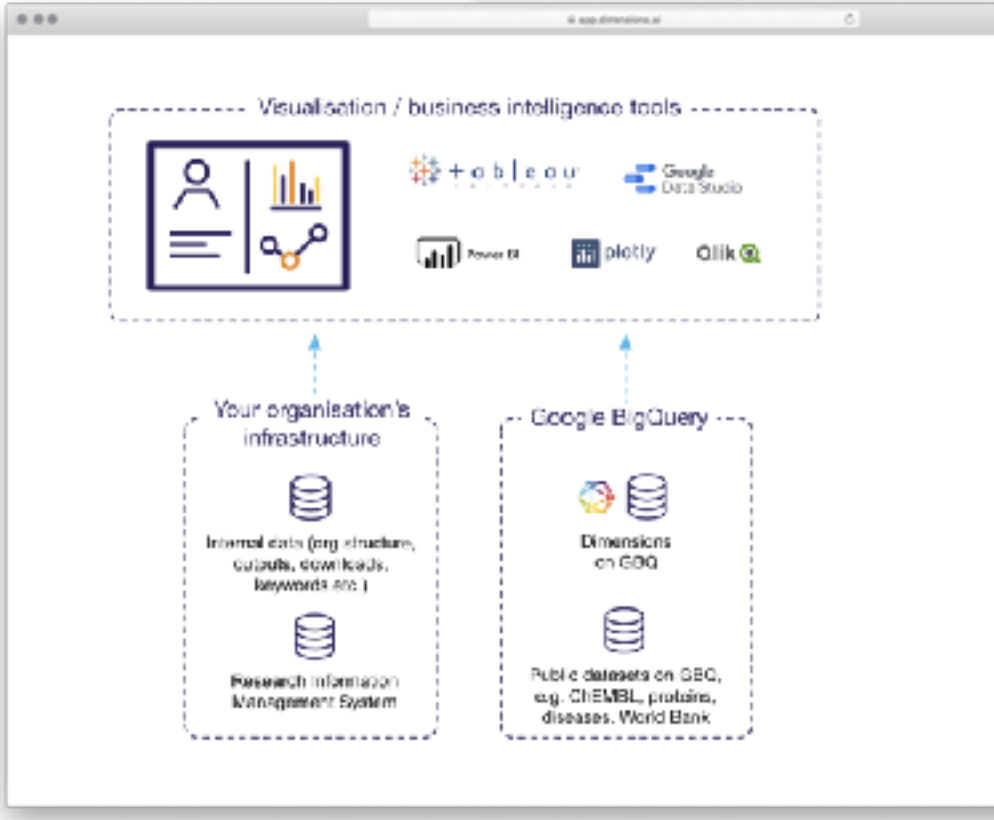
Google BigQuery



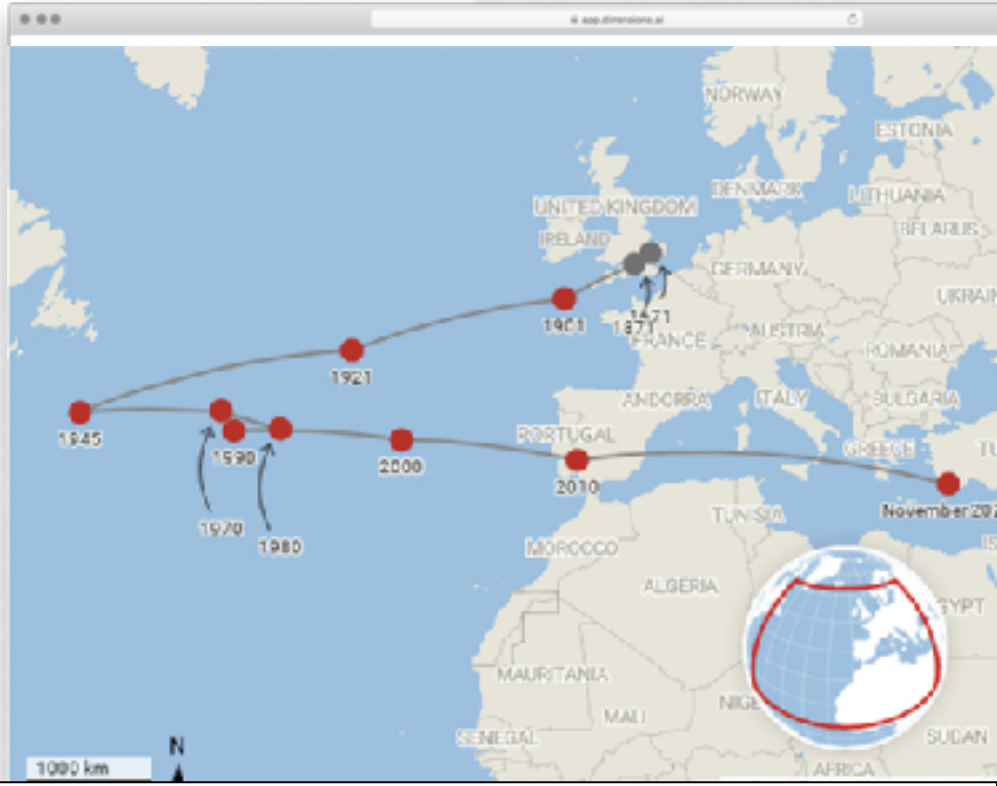
- Fast, large scale analyses; dynamic dashboards
- Direct integration with BI tools e.g. Tableau, Qlik, PowerBI
- Join with private & public data

For data & analytics teams + dashboards for everyone

- **How they work**
  - Cost-efficient storage
  - Distribute access control
  - Perform complex calculations with an on-demand infrastructure



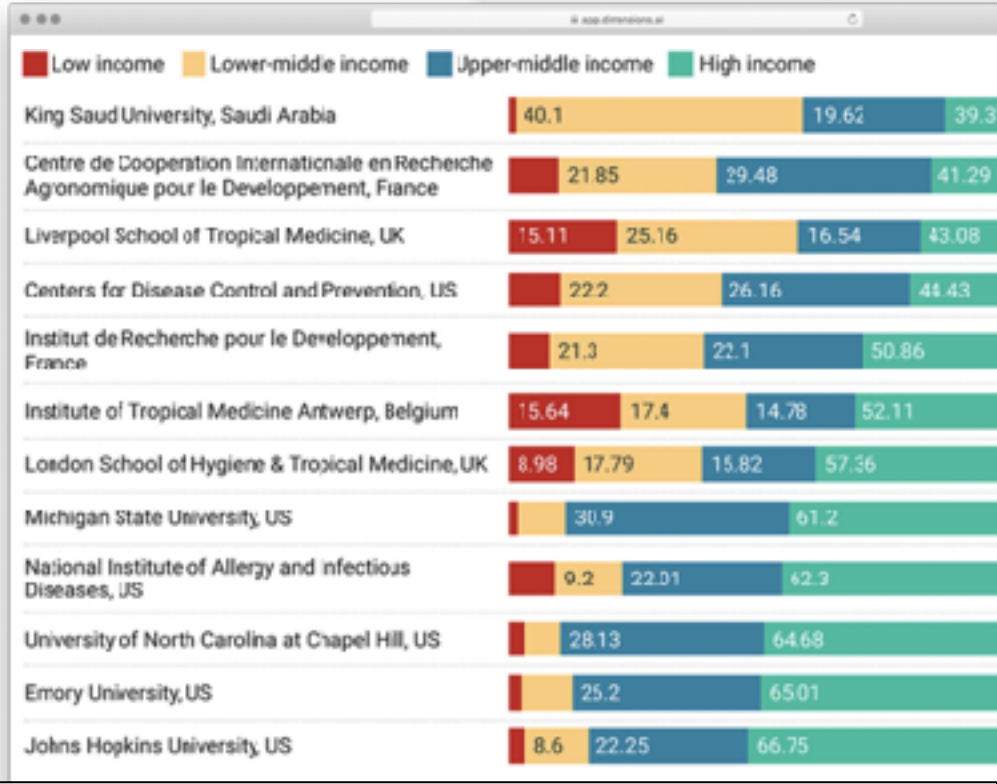
- **How they work**
  - Cost-efficient storage
  - Distribute access control
  - Perform complex calculations with an on-demand infrastructure
- **Case study: “centre of mass of research”**
  - authors affiliations locations weighted by citations to outputs



Hook, D. W., Porter, S. J., (2021). Scaling Scientometrics: Dimensions on Google BigQuery as an Infrastructure for Large-Scale Analysis. *Front. Res. Metr. Anal.* 6.

Cloud-based scientometrics databases

- How they work
  - Cost-efficient storage
  - Distribute access control
  - Perform complex calculations with an on-demand infrastructure
- Case study: “centre of mass of research”
  - authors affiliations locations weighted by citations to outputs
- **Case study: “connecting scientometrics”**
  - GRID + World Bank Data
  - High/Low-income-country research organizations collaborations



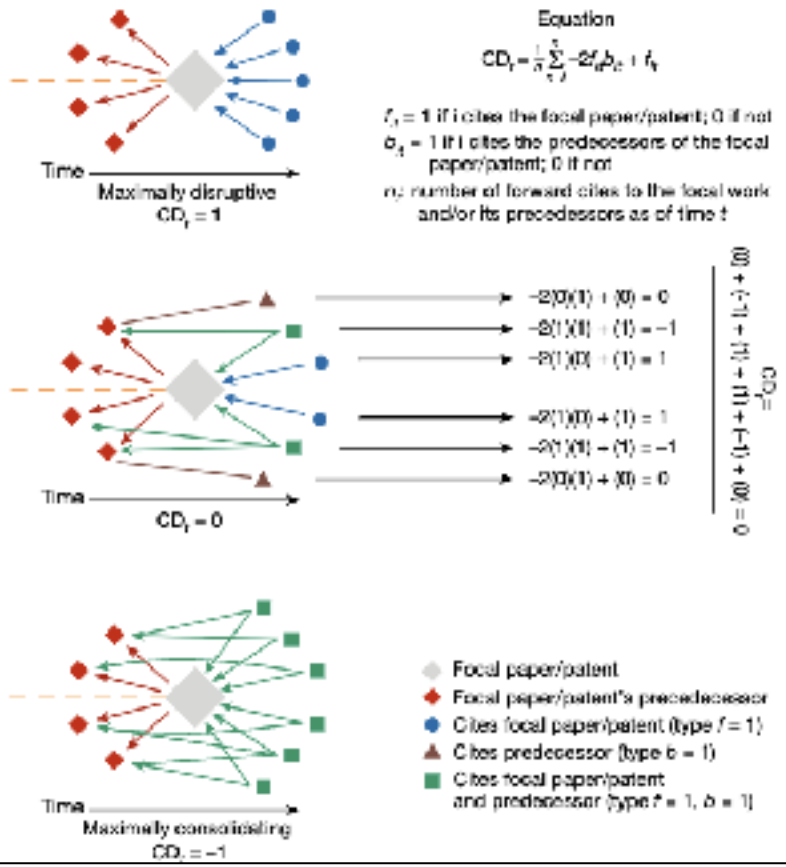
Porter, Hook, (2022). Connecting Scientometrics: Dimensions as a Route to Broadening Context for Analyses. *Front. Res. Metr. Anal.*

Dimensions

Google BigQuery

What is disruptive research?

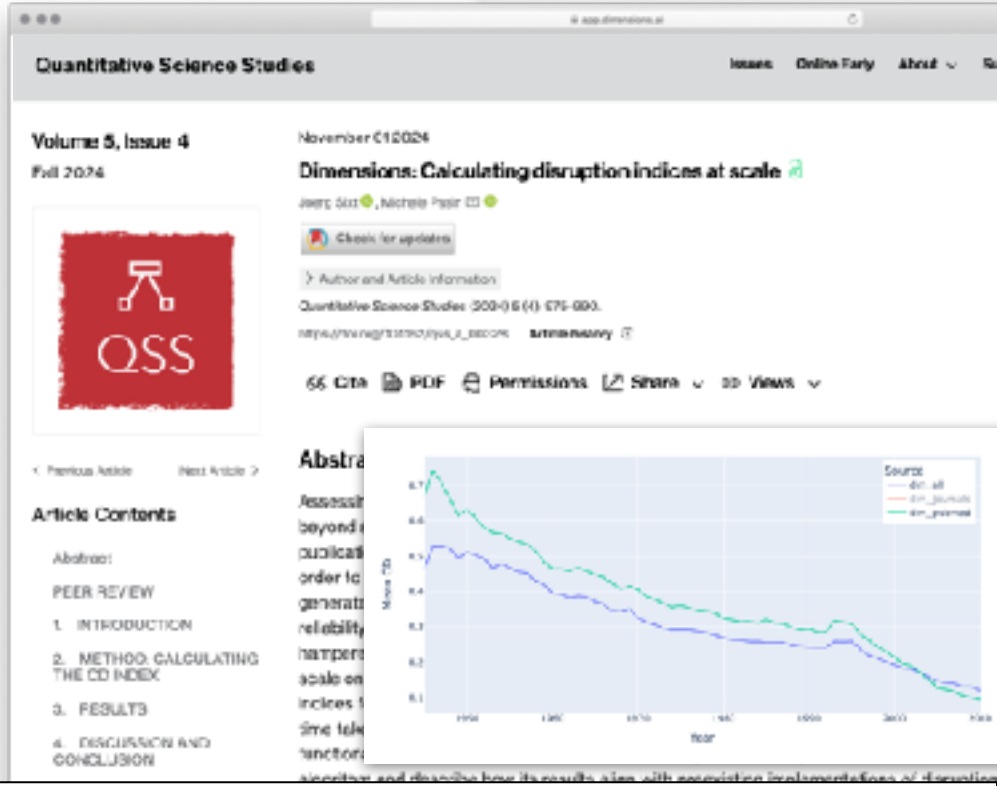
- Russell Funk (2023) created metrics (“CD” and “mCD” indices) that try to classify papers into “disruptive”, “neutral” and “consolidating papers”
- The calculation is based on comparing citations to a paper and citations to its references
    - Nature article & widely echoed in mainstream media e.g. New York Times: *Papers and patents are becoming less disruptive over time*
- **Challenge**:
    - What is this metric really telling us?
    - Hard to reproduce, many variables and costly to run



J. Sixt, M. Pasin (2024) Dimensions: Calculating Disruption Indices at Scale - *Quantitative Science Studies*

Dimensions

- **Dimensions**: implementation using SQL and BigQuery
  - < 5h (laptop) Vs days
  - Multiple runs (all pubs, pubmed, journal articles etc..)
- **Results**: it is possible to experiment with multiple versions of index formula
  - ie years, factors
- Is the CD index **useful**?
  - Not sure still..

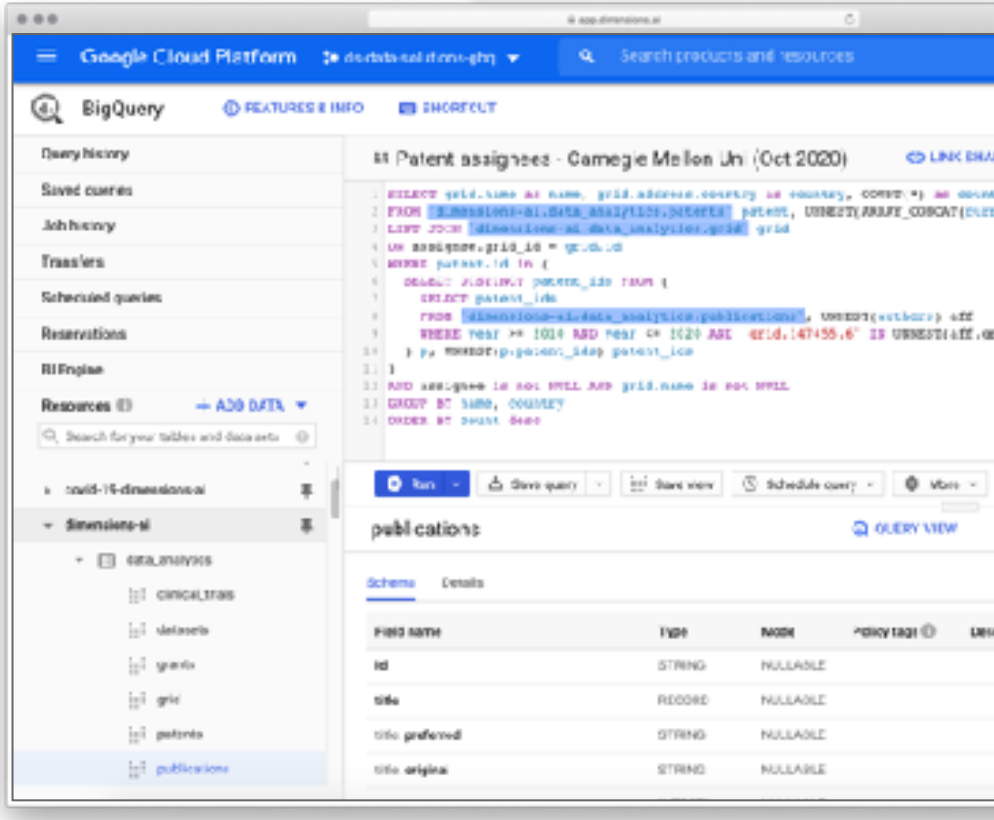


J. Sixt, M. Pasin (2024) Dimensions: Calculating Disruption Indices at Scale - *Quantitative Science Studies*

Dimensions

End user challenges

1. DB Structure can be complex
2. SQL Query can be complex
3. Cloud systems have their own peculiarities EG nested fields



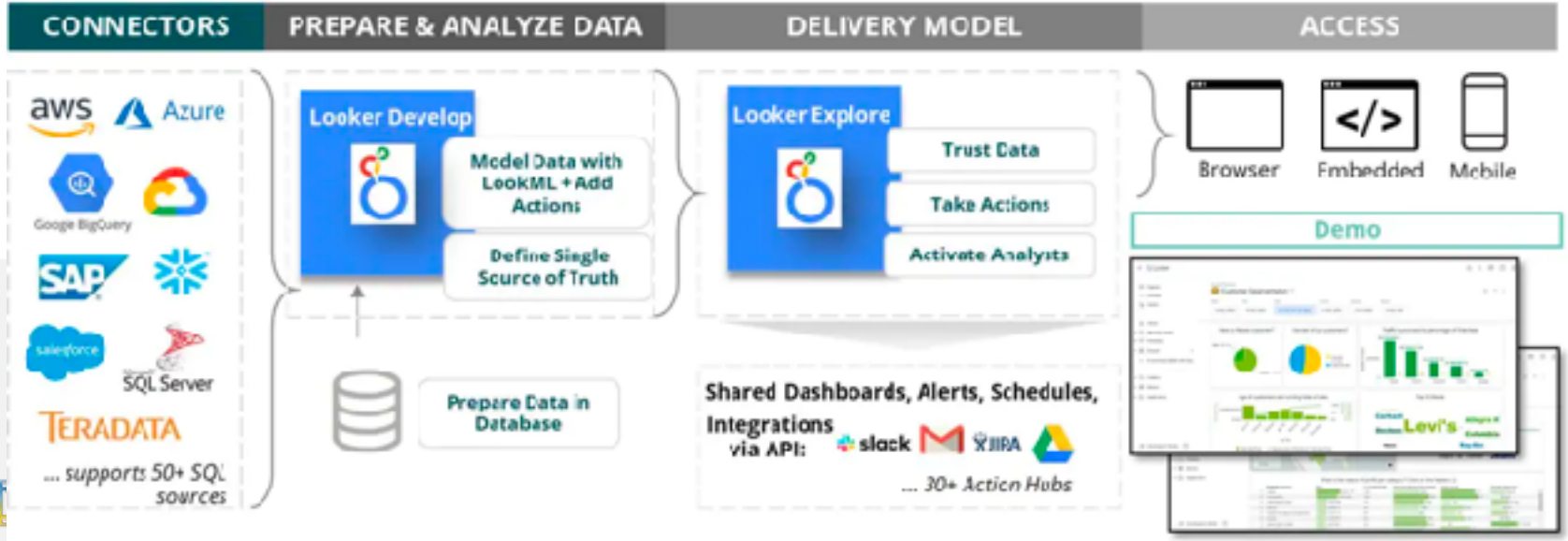
Dimensions

Part of DIGITAL science

PART 2

What is Looker? (It's not ~~Google~~ Looker Studio!)

- **Data analytics platform**
  - Google SaaS
  - BI focus
  - Enterprise level
- **Full pipeline support**
  - Prepare / Model data
  - Query / Explore / Visualize
  - Dashboards / Applications delivery
- **Developers friendly**
  - LookML
  - Git
  - JS plugins



LookML language

- Semantic layer for modeling your data
  - DB agnostic
  - GBQ-friendly
- Single source of truth
  - Maximise reuse
  - Transparency
- Dashboards as LookML
  - Programmatic generation

Example: [modelling ISBN]



Dimensions

Looker pros and cons

Pros

- Fast development and insights
- We can focus on data and visualisation and **not** on front end development, etc.
- Works seamlessly with Google Big Query which is Dimensions' preferred tool i.e. easy to compare analysis with colleagues who use GBQ

Cons

- Can be inflexible if you want other features (e.g. complex filtering, alerts, etc.)
- Embedding into Dimensions Analytics took some time

Example 1: publications and patents

Link

1. Number of publications 2010-2020
2. Map of publications
3. Development of US publications
4. How many US publications are funded by Chinese organizations over time?
5. Research to Application: country of publication to assignee country of patent by citation

Example 2: Landscape & Discovery app

- Landscape & Discovery is…
- Shows how it all comes together
- [Demo]

Thanks

PUBLICATION DATA

PUBLICATIONS

- Journal articles, pre-prints, conference proceedings and books/book chapters
- Over 100 million records based on metadata; 70+ million with full-text indexing
- Metadata and citations derived from multiple available databases
- OA tagging
- Over 1 billion citations

JOURNALS / BOOKS	PRE-PRINT / OA	Pipeline 2019
Crossref	bioRxiv	arXiv.org
PubMed	ChemRxiv™	RePEc
Europe PMC	SSRN	

Find experts, collaborators, speakers and more

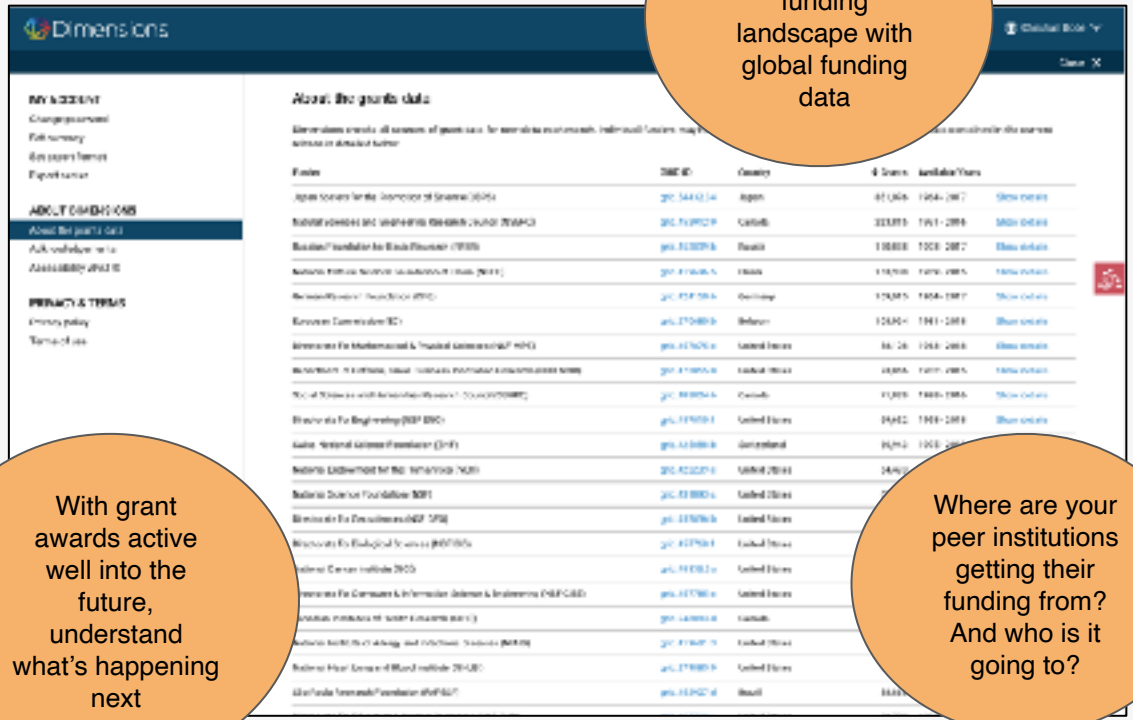
Where should I submit?

What are the trends in open access publications?

GRANT DATA

$$$ GRANTS

- Project funding
- Over 4.1M grants, from ~360 funders globally
- $1.3 trillion of funding
- Sourcing
  - Direct relationships with funders
  - Data available via APIs
  - Data available via websites which we crawl



Investigate an expanded funding landscape with global funding data

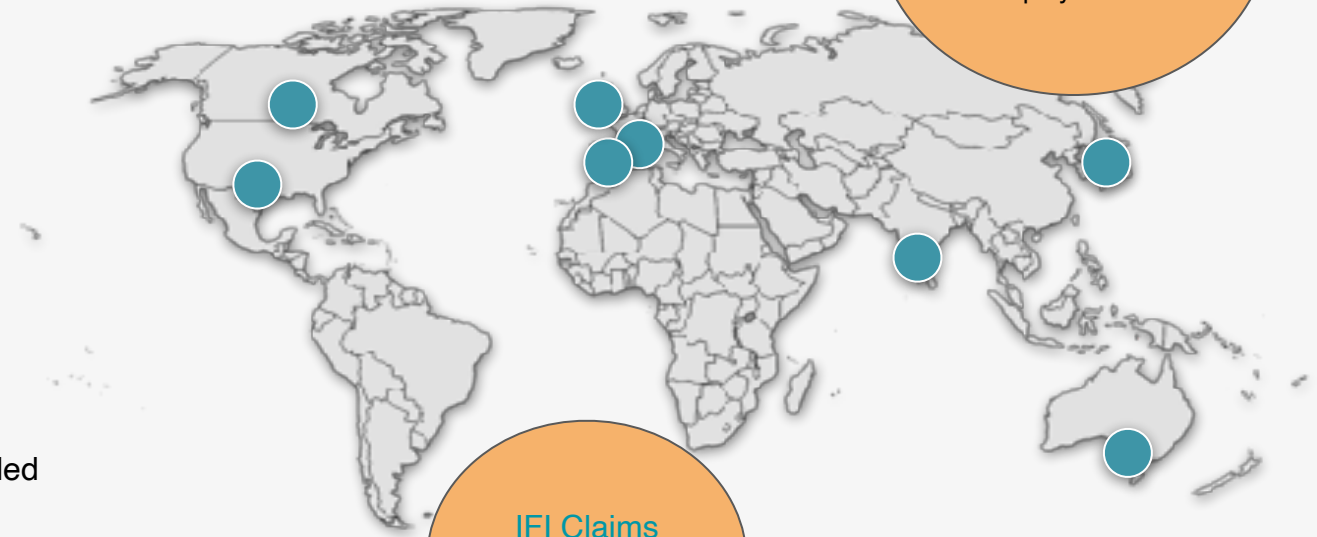
With grant awards active well into the future, understand what's happening next

Where are your peer institutions getting their funding from? And who is it going to?

PATENT DATA

PATENTS

- US
- EP
- WIPO
- DE
- CA
- IN
- AU
- GB
- FR
- Hong Kong
- Russia
- China will be the next to be added



What research is being commercialized? Who are the major players?

IFI Claims provides our patent data

Dimensions

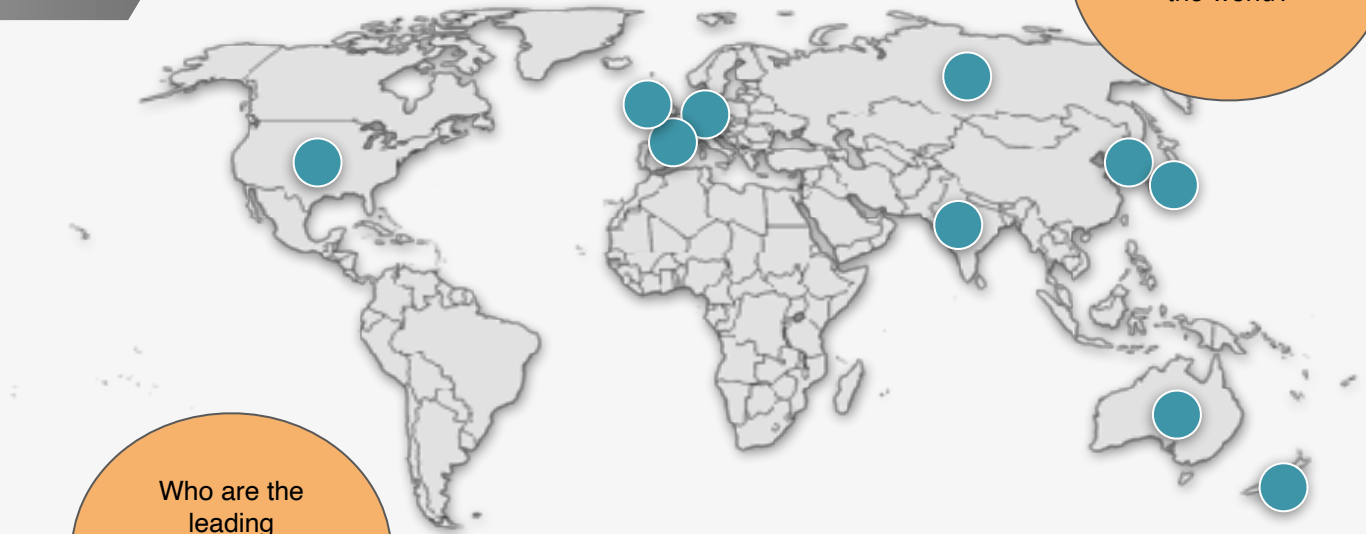
DIGITAL science

CLINICAL TRIAL DATA

CLINICAL TRIALS

ClinicalTrials.gov
EU-CTR
UMIN-CTR
ISRCTN
ANZCTR
CHICTR
NTR
GCTR
CTRI
CRIS

… and more are coming



What's moving ahead around the world?

Who are the leading organizations in a given area?

Dimensions

DIGITAL science

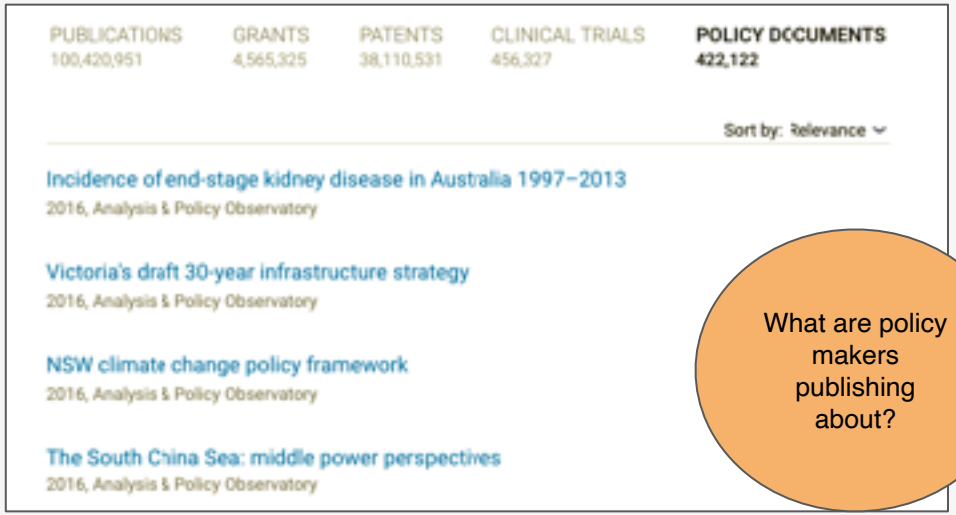
POLICY DOCUMENTS

POLICY DOCUMENTS

Over 400,000 policy document records, linked to publication records

Including but not limited to:

- World Health Organization
- World Bank
- Food & Agriculture Organization of the UN
- National Academies Press
- Centers for Disease Control & Prevention
- Publications Office of the European Union
- Government of the United Kingdom
- National Bureau of Economic Research
- Economic Policy Institute



What are policy makers publishing about?

What research are they citing?

Dimensions

DIGITAL science

Templates

Subheading Title

Plain Title Slide

White background **DS red highlight**

Subheading Title

INTRODUCTION SLIDE

Introduction to using this slide deck template for Digital Science. Add in your text here, go on, we know you want to.

Sub heading 18 bold.

The font used throughout is Open Sans. Caps are used for the slide title only. You can use bold for subheadings and to highlight words.

Slide titles are font size 28, slide introduction is 22, subheadings are bold and 18. Bullet points are 18 and body copy is 18 too. It can be 16 where there is more text such as this slide but no less than size 14 for legibility.

COLOUR OPTIONS

For images or logos go to the DS Drive here. Full DS Brand Guidelines can be found here.

Colour Codes

The palette works with the wide range of portfolio colours.

Bullets, headings

Yellow: #ffcc20

Black: #00000
Can also be used as background colour

Boxes and large areas

Wild Sand: #f2f9f6

Haze #f2f9f6

White #fffff

Highlights

Red #cc1423

BULLET POINT SLIDE

Introduction to a slide in regular and set to black as per the style guide.

Maybe this is the contents slide

- For approved icons - find these here
- For Images - find these here
- For logos - find these here
- Anything else get in touch with DSCM
- We are more than happy to help

SPEAKER SLIDE



FIRST NAME LAST NAME

Job Title

Company Name

d.hook@digitalscience.com

0000-0001-9746-1193

@twitterhandlename

MULTIPLE SPEAKERS



Your Name

And their details can go here depending what their details are

@twitterhandle



Your Name

And their details can go here depending what their details are

@twitterhandle



Your Name

And their details can go here depending what their details are

@twitterhandle

TEAM SLIDE OPTION 1



Your Name
Job Title
@twitterhandle



Your Name
Job Title
@twitterhandle



Your Name
Job Title
@twitterhandle



Your Name
Job Title
@twitterhandle



Your Name
Job Title
@twitterhandle



Your Name
Job Title
@twitterhandle

IMAGE SLIDE OPTION 1

Introduction to a slide with space to insert a graphic / graph / diagram



Title

- Bullet
- Bullet
- Bullet

DATA SLIDE OPTION 2

Introduction to a slide with space to insert a graphic / graph / diagram

- Bullet
- Bullet
- Bullet

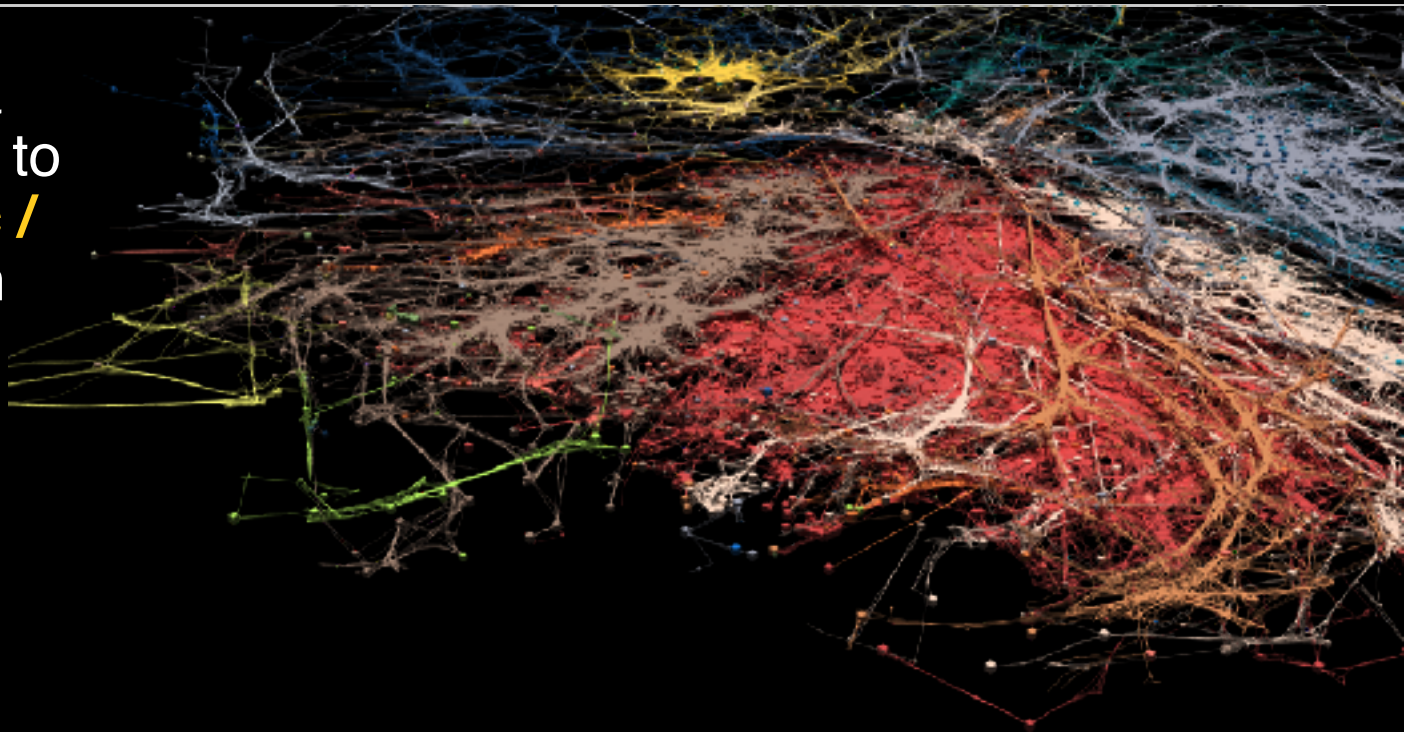


Source: xxxxxx

DATA SLIDE OPTION 2

Introduction to a slide with space to insert **a graphic / graph** / diagram

- Bullet
- Bullet
- Bullet



Source: xxxxxx

IMAGE SLIDE OPTION 3



Blocks for an image with a black caption block and some text

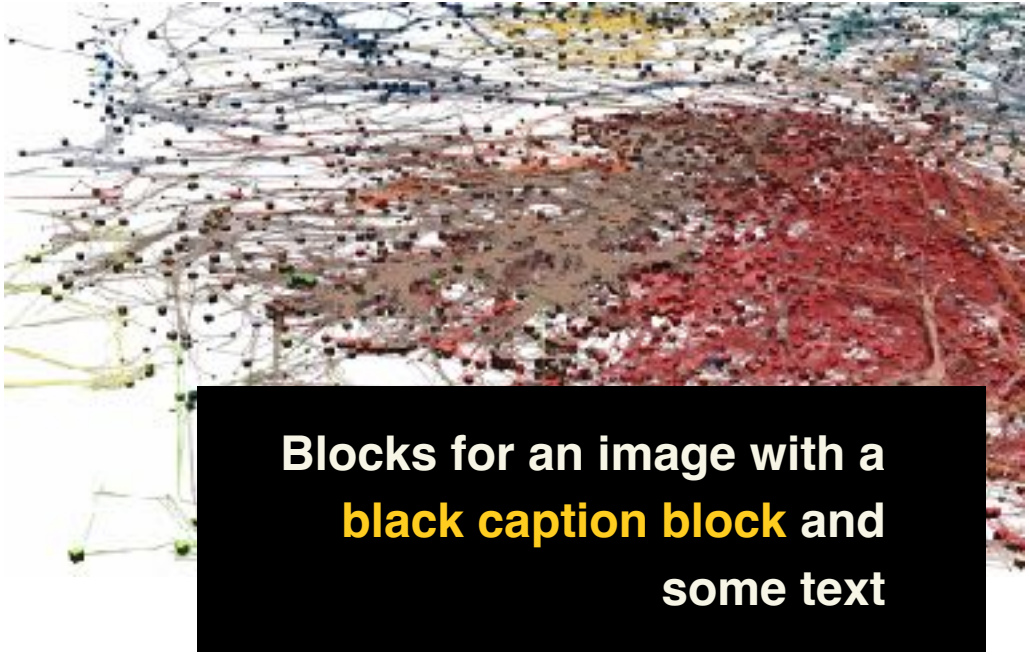
There are blocks in the style guide for text over a block of colour

There are blocks in the style guide for text over a block of colour

IMAGE SLIDE OPTION 4

There are blocks in the style guide for text over a block of colour

There are blocks in the style guide for text over a block of colour



Blocks for an image with a black caption block and some text

TABLE SLIDE OPTION 1

A heading about a table with lots of information

Column title	Column	Feature	Details of the column line item
ITEM	Item	Research	Lorem ipsum dolor sit amet, consectetur adipiscing elit
ITEM	Item	Research	Lorem ipsum dolor sit amet, consectetur adipiscing elit
ITEM	Item	Research	Lorem ipsum dolor sit amet, consectetur adipiscing
ITEM	Item	Research	Lorem ipsum dolor sit amet, consectetur adipiscing elit
ITEM	Item	Research	Lorem ipsum dolor sit amet, consectetur adipiscing elit

TABLE SLIDE OPTION 2

A heading about a table with lots of information

Column title	Column title	Column title	Column title
ITEM	Lorem ipsum dolor sit amet, consectetur adipiscing elit	Lorem ipsum dolor sit amet, consectetur adipiscing elit	Lorem ipsum dolor sit amet, consectetur adipiscing elit
ITEM	Lorem ipsum dolor sit amet, consectetur adipiscing elit	Lorem ipsum dolor sit amet, consectetur adipiscing elit	Lorem ipsum dolor sit amet, consectetur adipiscing elit
ITEM	Lorem ipsum dolor sit amet, consectetur adipiscing elit	Lorem ipsum dolor sit amet, consectetur adipiscing elit	Lorem ipsum dolor sit amet, consectetur adipiscing elit

TABLE SLIDE OPTION 3

A heading can go here or be removed

You can pick a colour	You can pick a colour	You can pick a colour	You can pick a colour
Lorem ipsum dolor sit amet, consectetur adipiscing elit	Lorem ipsum dolor sit amet, consectetur adipiscing elit	Lorem ipsum dolor sit amet, consectetur adipiscing elit	Lorem ipsum dolor sit amet, consectetur adipiscing elit

Add text here or remove

COLUMN SLIDE 1

Introduction to a slide that’s got three columns of content with icons.

Here’s a slide with three columns

Icon can go here

Title text

These icons are placeholders for now.

Icon can go here

Title text

These icons are placeholders for now.

Icon can go here

Title text

These icons are placeholders for now.

COLUMN SLIDE 2

For list of approved icons please see the DS collateral drive and brand guidelines.

Column 1



Icon for Industry



Icon for Publisher



Icon for Institution



Icon for Funder



Icon for Government



Icon for Pharma

Column 2

Some text goes here.

This is some text and some more text to see what this look like when there's text to go in here.

COLUMN SLIDE 3

Introduction to a slide with two columns of text and each has a subheading.

Heading

Bold to highlight something in body copy.

This is a slide with lots of text so the font size for body copy is a minimum of 12.

Lorem ipsum dolor sit amet, consectetur.

Heading

Lorem ipsum dolor sit amet, cons ectetur adipiscing elit, sed do eiu smod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad min im veniam, quis nostrud exer citation ullamco laboris nisi ut aliquip ex ea commodo consequat.

COLUMN SLIDE 4

Introduction to a slide with two columns and lots of text. The icon optional.

Heading one

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad

Heading three

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore.

Heading two

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua.

Heading four

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore.

GENERAL TEXT SLIDE OPTION 1

If you really, really need something with a bit more text (remember less is more), here is a slide for you.

Here's a subheading

Lorem ipsum dolor elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Lorem ipsum dolor elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Lorem ipsum dolor elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua.

Lorem ipsum dolor elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Lorem ipsum dolor elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Lorem ipsum dolor elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua.

GENERAL SLIDE OPTION 2

If you really, really need something with a bit more text (remember less is more), here is a slide for you

Here's a subheading

100% Statistics | **100%** Statistics | **100%** Statistics

Lorem ipsum dolor elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Lorem ipsum dolor elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Lorem ipsum dolor elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua.

Lorem ipsum dolor elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Lorem ipsum dolor elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua.

BULLET SLIDE OPTION 1

If you really, really need something with a bit more text (remember less is more), here is a slide for you.

- Lorem ipsum dolor elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua.
- Lorem ipsum dolor elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua.
- Lorem ipsum dolor elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua.
- Lorem ipsum dolor elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua.

BULLET SLIDE OPTION 2

Another option if you require a slide with bullet points

- **A start of a bullet** - Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua.
- **A start of a bullet** - Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua.
- **A start of a bullet** - Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua.

BULLET SLIDE OPTION 3

Another option if you require a slide with bullet points

1. **A start of a bullet** - Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua.
2. **A start of a bullet** - Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua.
3. **A start of a bullet** - Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua.

SLIDE WITH LOGOS OPTION 1

Introduction to a slide with some text and an option to have some logos

Slide with logos

These could be Digital Science portfolio logos or customer logos. You could also use this to highlight product functionality and use icons instead of logos.











Investing in today for the research of tomorrow

Every single aspect of Digital Science is focused on helping game-changing businesses that rewire the world of research for the better. Our portfolio is home to ….



























Logo slide grey scale

Every single aspect of Digital Science is focused on helping game-changing businesses that rewire the world of research for the better. Our portfolio is home to ….



DEVICE SLIDE OPTION 1

Introduction to a slide with some text and an option to have an image

Slide with an image

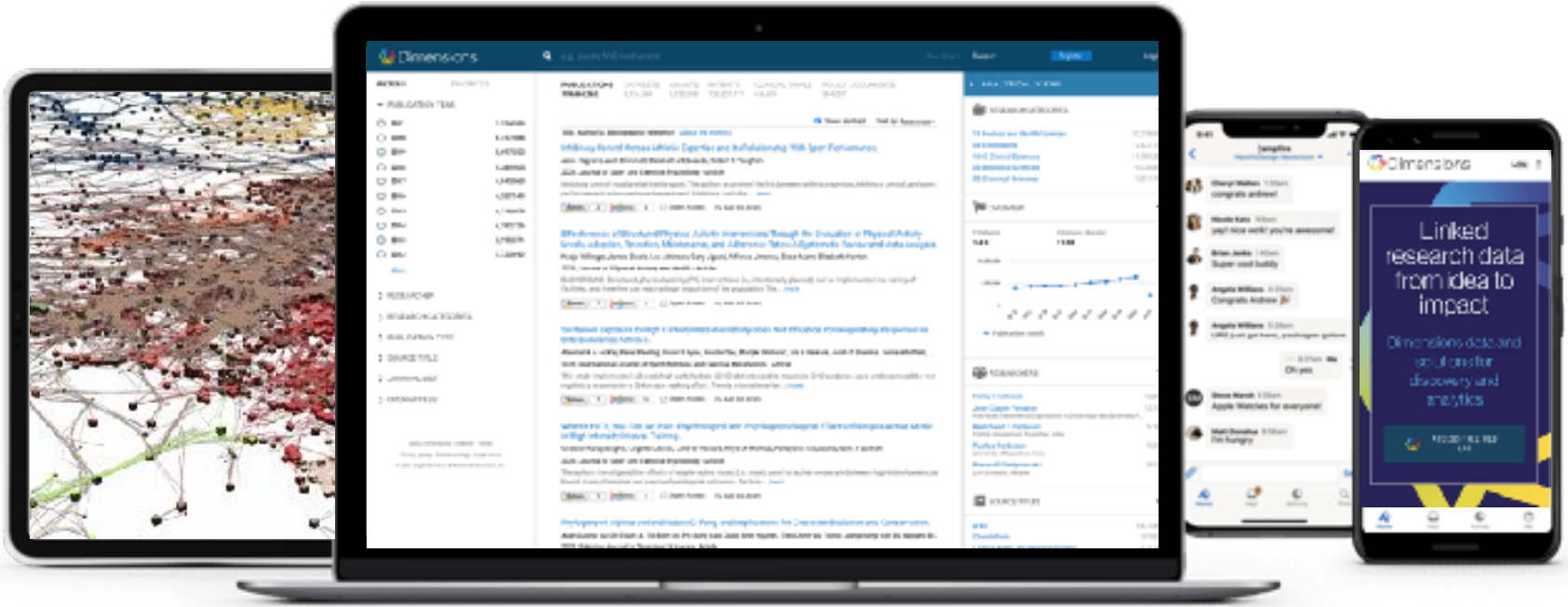
A full range of device option images can be found here.

Dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation.



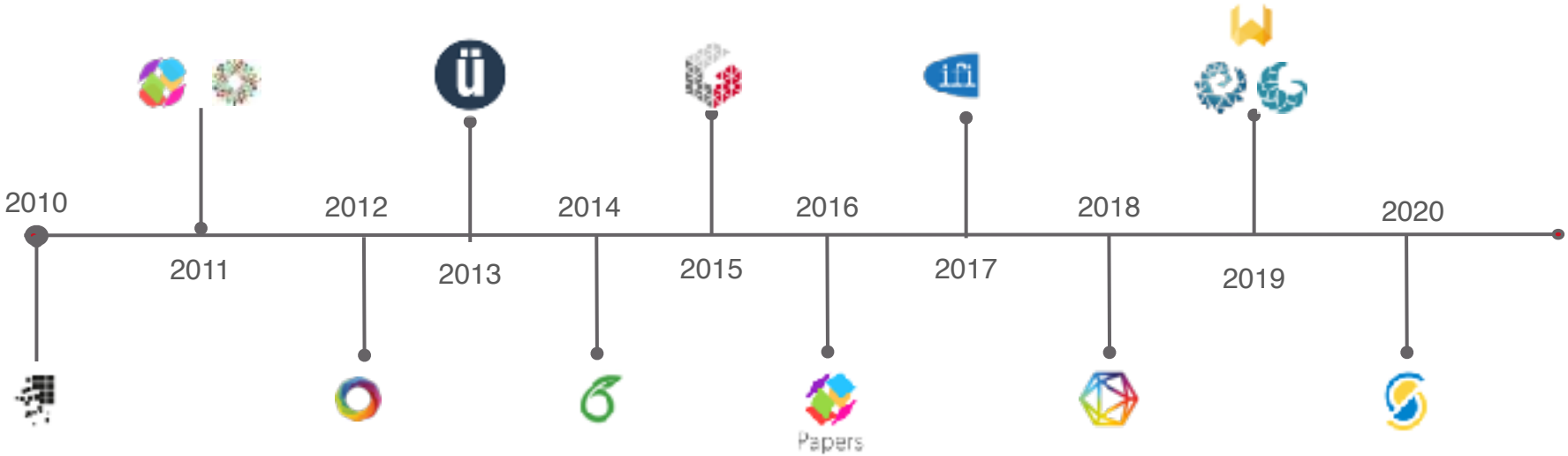
DEVICE SLIDE OPTION 2

A full range of device option images can be found here.



The Digital Science Family

Example of how the timeline slide can be used



QUOTE SLIDE OPTION 1

DIGITAL science

> **Continued waves of innovation in policy and technology are needed to take us from the current state of Open Research to some future equilibrium**

DANIEL HOOK

DIGITAL science

“Quote option 2 - etc policy and technology are needed to take us from the current state of Open Research to some future equilibrium

DANIEL HOOK


DIGITAL
science
Quote slide option 3 this is a SOOD cover image. Other images available.
DANIEL HOOK
Source of quote here

This is a yellow slide

Have fun

This is a yellow slide

Have fun

This is a yellow slide

have fun

Useful Information

How to change the ratio of my presentation.

- These slides are set up to widescreen size (ratio of 16:9). To change format please go to >> Page Setup >> change the ratio. Screenshot below

