



the hyve



# Career perspective and collaborations in open science

---

Kees van Bochove, Founder, The Hyve

@keesvanbochove

Leiden, Feb 19, 2020





We advance biology and medical research...  
... by building and serving thriving open source communities.

## Fast-growing

Started in 2012  
45+ people by now

## Core values

Share



Reuse



Specialize



## Services

Professional support for  
**open source software** in  
biomedical informatics

- Software development
- Data engineering
- Consultancy
- Hosting / SLAs

## Customer Segments

Pharma  
Life Sciences  
Healthcare

## Office Locations

Utrecht, The Netherlands  
Cambridge, MA, United States

# Teams at The Hyve: open source communities



## Health Data Networks

- Data warehouses: tranSMART, i2b2
- Cohort selection: Glowing Bear
- Request Portals: Podium



Open Targets

## Genomics

- Cancer data portal: cBioPortal
- Knowledge base: Open Targets



## Real World Data

- Real world evidence: OMOP/OHDSI
- Wearables platform: RADAR-BASE



## Research Data Management

- FAIR Data Governance consultancy
- Fairspace (meta)data management

A Year at



the hyve

**58** Projects

**13** pharma companies

**10** hospitals

**11** consortia/biobanks/health  
networks/patient organisations

**33** Projects with

patient level data

**7** Clinical Data

Models



1. Introduction to Open Science & Data Stewardship
2. Medical evidence generation is changing through open science

Statement #1

Science is broken  
but it can be fixed

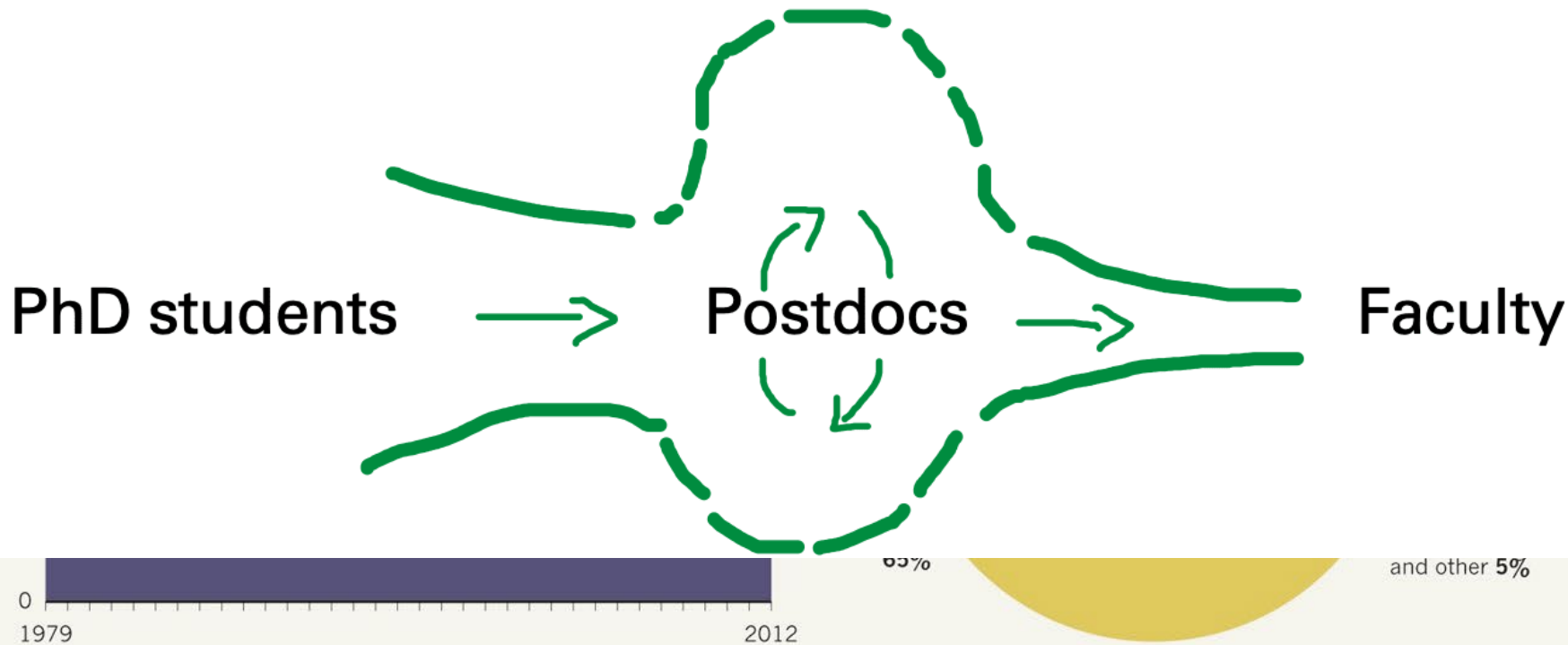
@keesvanbochove @TheHyveNL

# What's with the academic career path?



## THE POSTDOC PILE-UP

The number of researchers in US postdoctoral positions has more than tripled since 1979. The vast majority of postdocs are in the life sciences. Across fields, median salaries for postdocs are outstripped by those for non-postdoc positions, when measured up to 5 years after receiving a PhD.



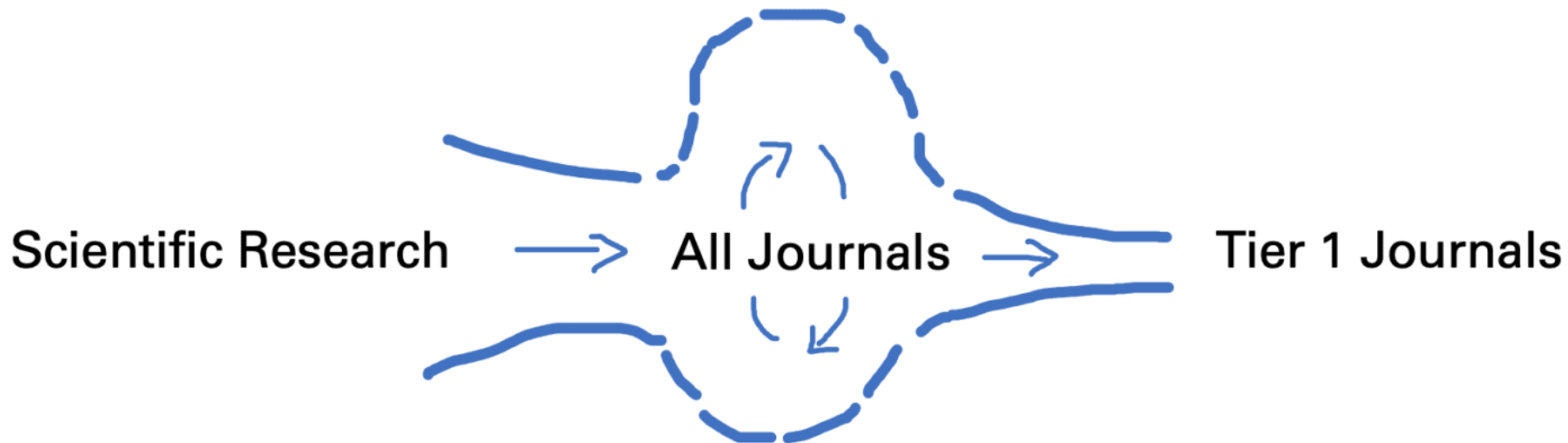
# What's with the academic career path?



## Can Twitter Save Science?

Two Truths and a Take, Season 2 Episode 4

Alex Danco Feb 16 ♡ 10 📌



<https://danco.substack.com/p/can-twitter-save-science>





## nature

---

Subscribe

COMMENT • 20 MARCH 2019

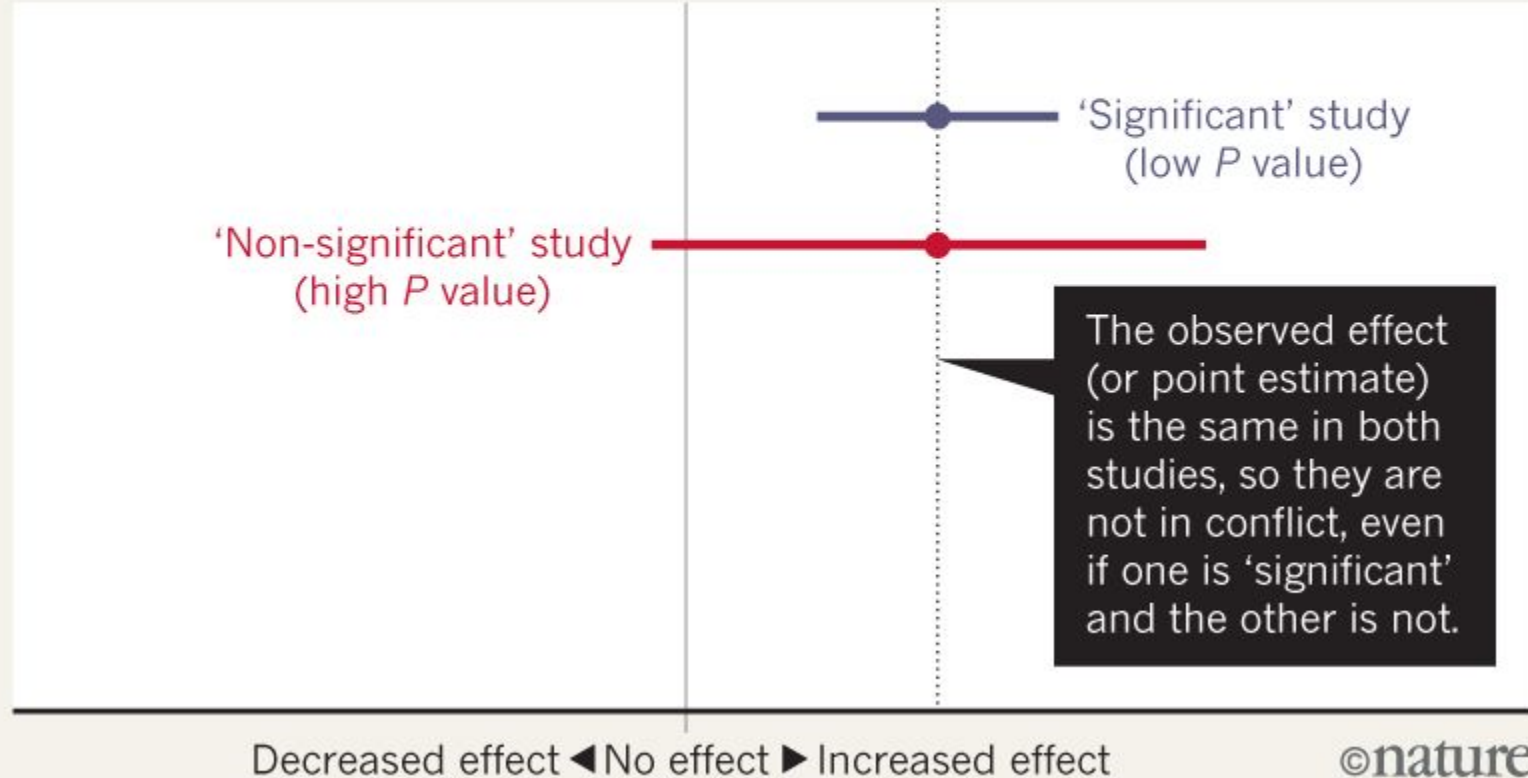
# Scientists rise up against statistical significance

Valentin Amrhein, Sander Greenland, Blake McShane and more than 800 signatories call for an end to hyped claims and the dismissal of possibly crucial effects.

Valentin Amrhein , Sander Greenland & Blake McShane

## BEWARE FALSE CONCLUSIONS

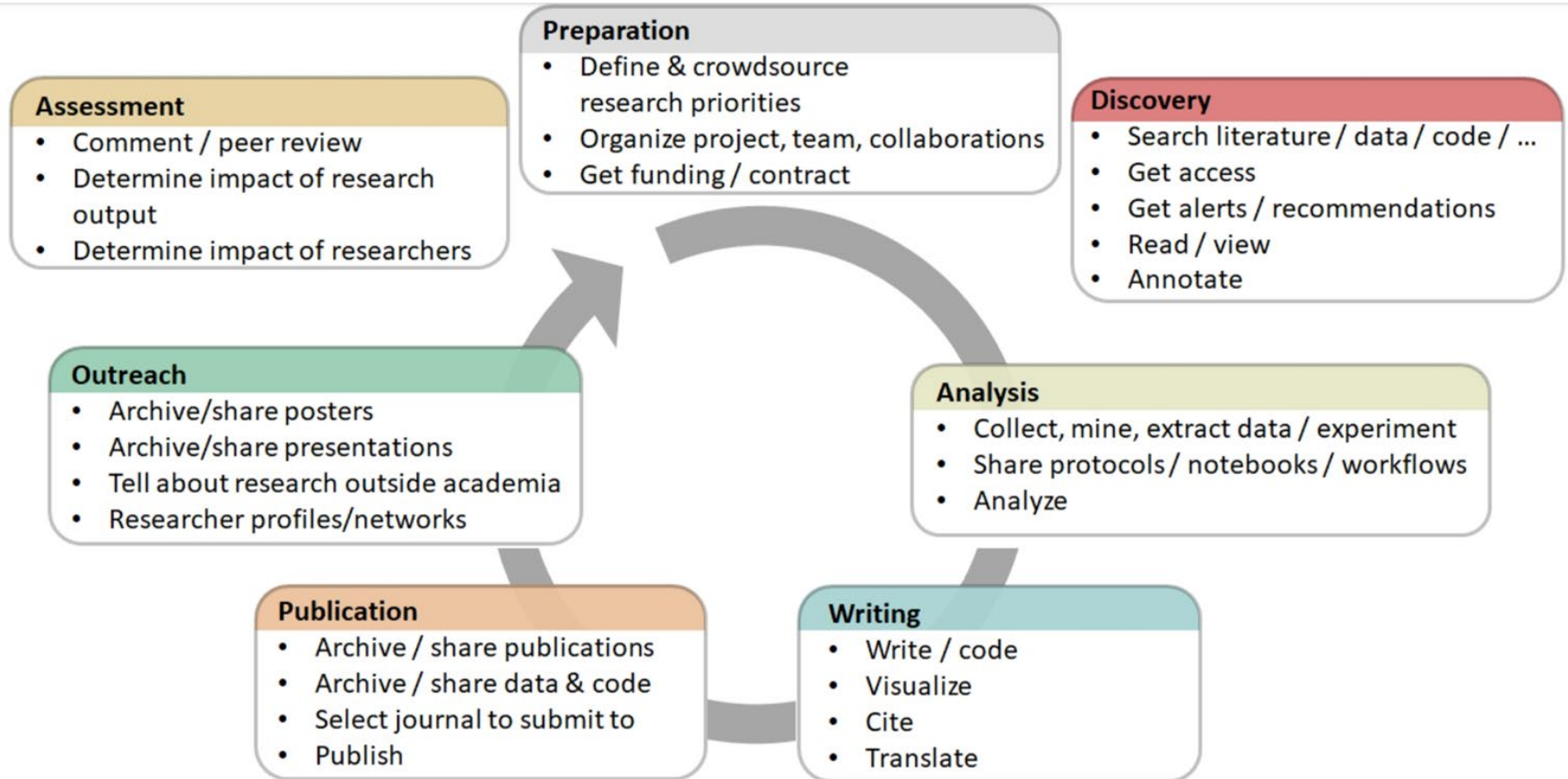
Studies currently dubbed 'statistically significant' and 'statistically non-significant' need not be contradictory, and such designations might cause genuine effects to be dismissed.



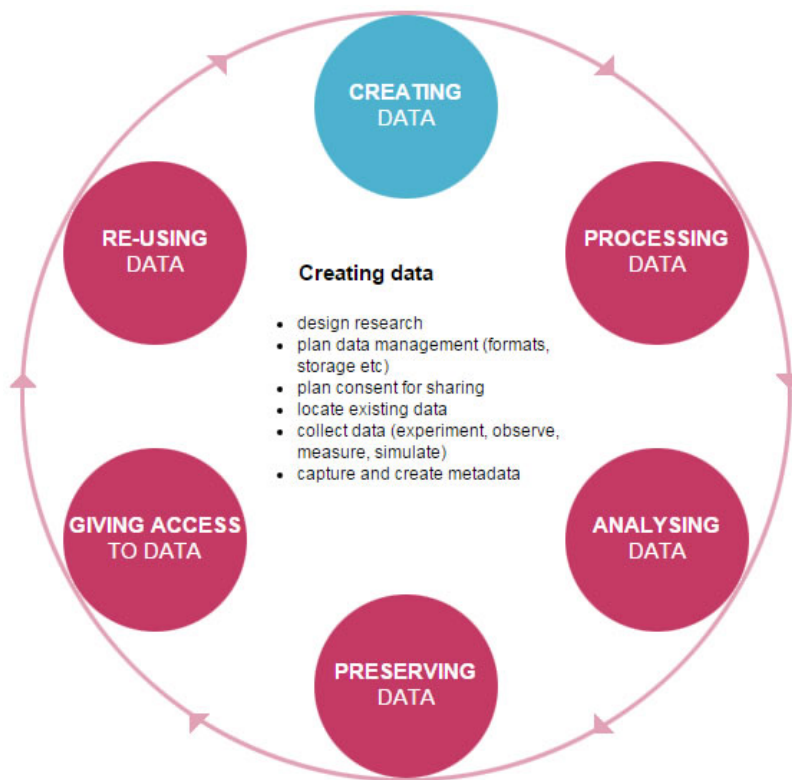
13155

- 42 News stories
- 40 Blog posts
- 21,654 Tweets
- 127 Facebook posts
- 2 Wikipedia mentions
- 1 Reddit post
- 9 F1000 reviews
- 4 Videos
- 736 Readers on Mendeley
- 198 Citations on Dimensions

# #OpenScience



# Research Data Lifecycle



UK Data Service / JISC

# Data Stewardship

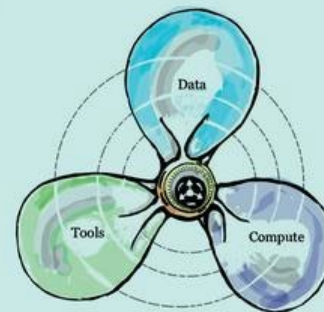
Rob Hooft

The screenshot shows the 'Knowledge model editor' interface of the Data Stewardship Wizard. On the left is a sidebar with navigation links: 'Data Stewardship Wizard', 'Organization', 'User Management', '404 Editor', '404 Packages', and 'DS Planner'. The main area is titled 'Knowledge model editor' and contains a 'Question' form. The form has fields for 'Title' (containing 'Is there any pre-existing data?'), 'Short UUID', and 'Text' (containing 'Are there any data sets available in the world that are relevant to your planned research?'). Below these are 'Question type' (set to 'Options'), 'Answers' (with 'No' and 'Yes' options), and 'References' (with '5.1' listed). On the right, a 'Current changes' panel shows a tree of changes under 'Design of experiment', including questions about pre-existing data, reference data, and data sets, with associated version numbers like 1.1, 1.2, 1.3, and 1.7.

<https://ds-wizard.org>

## DATA STEWARDSHIP FOR OPEN SCIENCE

Implementing FAIR Principles



BAREND MONS



## Statement #1

Science is  
broken, but it can  
be fixed

- The classical way of doing science has too many perverse incentives
- Digital collaboration can change that for the better
- From data management to data stewardship

# Intermezzo

Your research project

- Go to [ds-wizard.org](https://ds-wizard.org), create an account
- Browse some of the knowledge models, if you are not sure use **Life Sciences DSW Knowledge Model, 2.0.0**
- Create a questionnaire for your project, and choose 2 or 3 sections to fill in
- Bonus / alternative: Look at what a computer knows about you as scientist: find or create your ORCID

## Statement #2

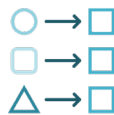
Medical evidence  
generation is  
changing through  
open science

- Scale
- Speed
- Impact





# OBJECTIVES OF THE EH DEN PROJECT



## Harmonisation

**Harmonise** in excess of **100 million** anonymised **health records** to the OMOP common data model, supported by an ecosystem of certified SMEs, and technical architecture for a federated network



## Evidence

Impact our understanding of, and improvement of, clinical **outcomes for patients** within diverse healthcare systems in the EU

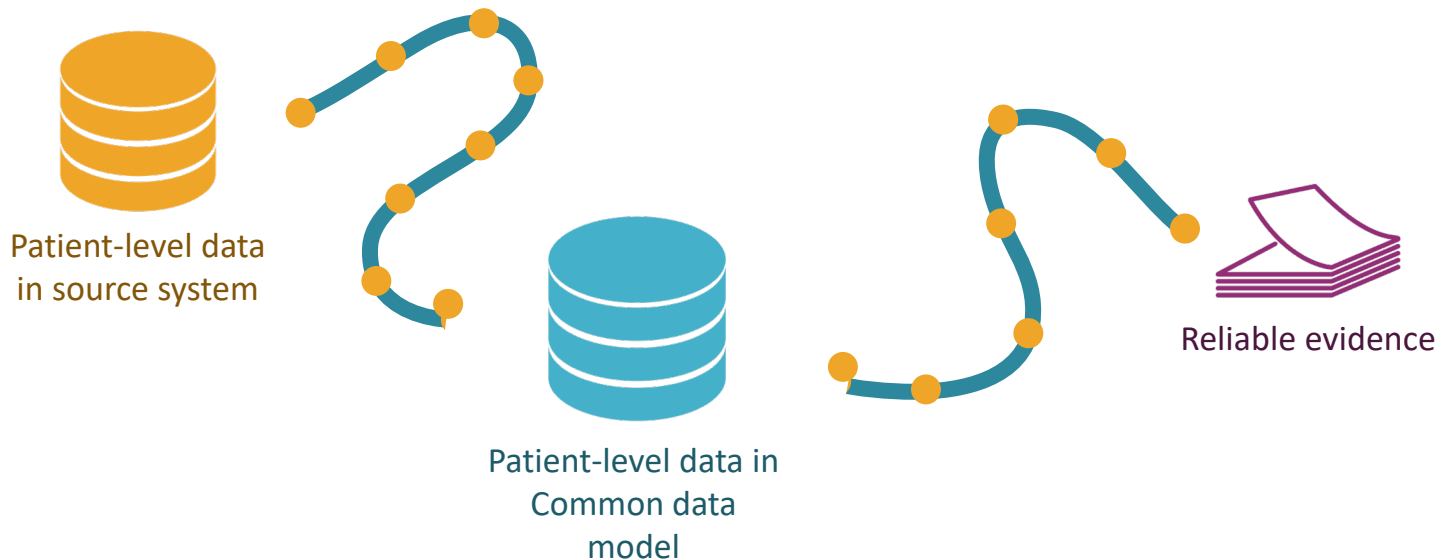


## Community

Establish a self-sustaining **open science collaboration** in Europe, supporting academia, industry, regulators, payers, government, NGOs and others



# THE JOURNEY TO REAL WORLD EVIDENCE



## Reproducible data flow

Documented manipulations and procedures.  
Automated, end-to-end analysis code.

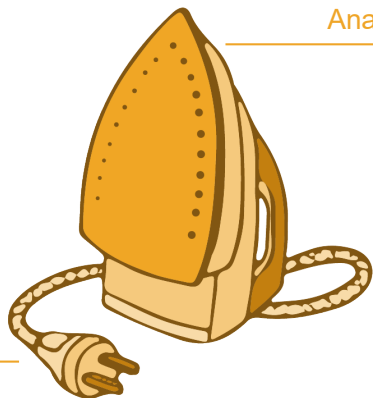
## Common data model & standardised vocabulary

Two-step process: standardisation before analysis.  
ETL & source code separated from analysis.  
Re-use of data & analysis.



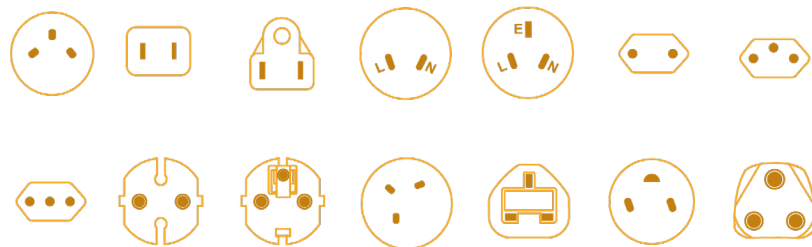
# WHY IS THIS NOT CURRENT PRACTICE?

Analytical method

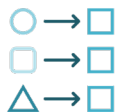


Link to data

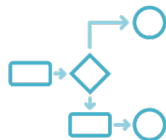
The data...



## What will it require?



Data interoperability



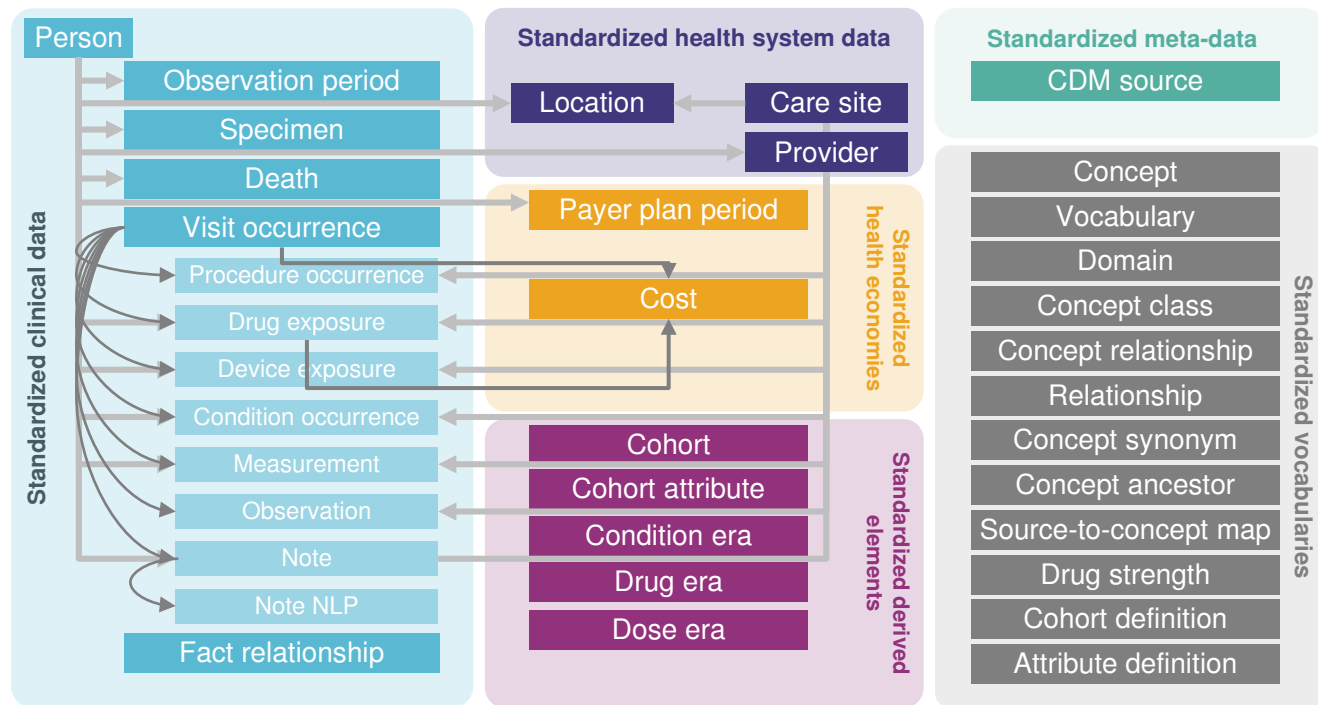
Standardised analytics



Data network



Strong community



Patient-centric  
Tabular  
Extendable  
Built for analytics  
Relational design





Estimation methods

## Cohort Method

New-user cohort studies using large-scale regression for propensity and outcome models

## Self-Controlled Case Series

Self-Controlled Case Series analysis using few or many predictors, includes splines for age and seasonality.

## Self-Controlled Cohort

A self-controlled cohort design, where time preceding exposure is used as control.

## IC Temporal Pattern Disc.

A self-controlled design, but using temporal patterns around other exposures and outcomes to correct for time-varying confounding.

## Case-control

Case-control studies, matching controls on age, gender, provider, and visit date. Allows nesting of the study in another cohort.

## Case-crossover

Case-crossover design including the option to adjust for time-trends in exposures (so-called case-time-control).

Prediction methods

## Patient Level Prediction

Build and evaluate predictive models for user-specified outcomes, using a wide array of machine learning algorithms.

## Feature Extraction


Automatically extract large sets of features for user-specified cohorts using data in the CDM.

Method characterization

## Empirical Calibration

Use negative control exposure-outcome pairs to profile and calibrate a particular analysis design.

## Method Evaluation

Use real data and established reference sets as well as simulations injected in real data to evaluate the performance of methods. 

Supporting packages

## Database Connector

Connect directly to a wide range of database platforms, including SQL Server, Oracle, and PostgreSQL.

## Sql Render

Generate SQL on the fly for the various SQL dialects.

## Cyclops

Highly efficient implementation of regularized logistic, Poisson and Cox regression.

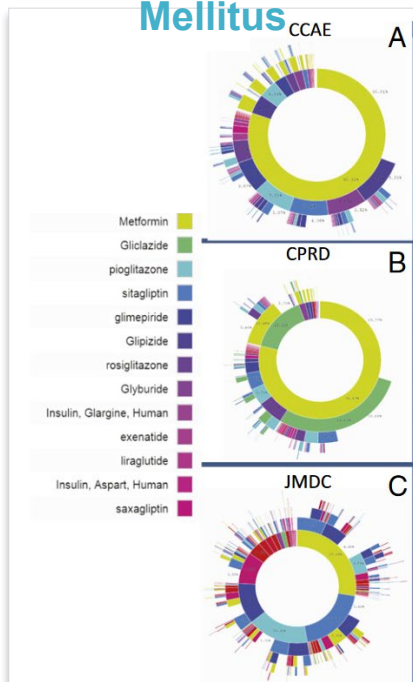
## Ohdsi R Tools

Support tools that didn't fit other categories, including tools for maintaining R libraries.

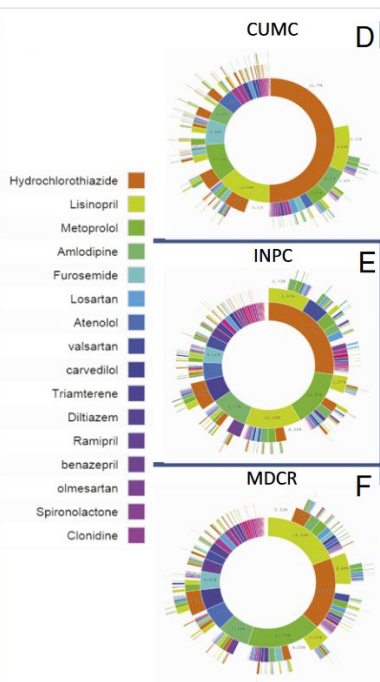


# LARGE-SCALE OBSERVATIONAL RESEARCH IS FEASIBLE.

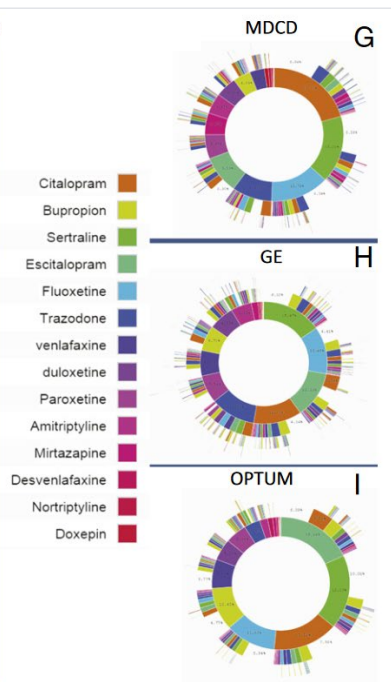
## T2 Diabetes Mellitus



## Hypertension



## Depression



11 Data sources



4 Countries



> 250 million patients

*“Characterizing treatment pathways at scale using the OHDSI network.”*

George Hripcsak et al. - PNAS (2016)27:7329–7336





**OHDSI**  
OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS



## THE LANCET

ARTICLES | [VOLUME 394, ISSUE 10211, P1816-1826, NOVEMBER 16, 2019](#)

### Comprehensive comparative effectiveness and safety of first-line antihypertensive drug classes: a systematic, multinational, large-scale analysis

[Prof Marc A Suchard, MD](#)   • [Martijn J Schuemie, PhD](#) • [Prof Harlan M Krumholz, MD](#) • [Seng Chan You, MD](#) • [RuiJun Chen, MD](#) • [Nicole Pratt, PhD](#) • et al. [Show all authors](#)

Published: October 24, 2019 • DOI: [https://doi.org/10.1016/S0140-6736\(19\)32317-7](https://doi.org/10.1016/S0140-6736(19)32317-7) •



Check for updates





# Lancet Paper Shows Most Popular Hypertension Drug Isn't Most Effective, Per OHDSI's LEGEND Study

VIDEO FEATURES: [Collaborators Discuss The Hypertension Study/Findings](#) | [Introducing The LEGEND Project](#)

## Lancet Study Link

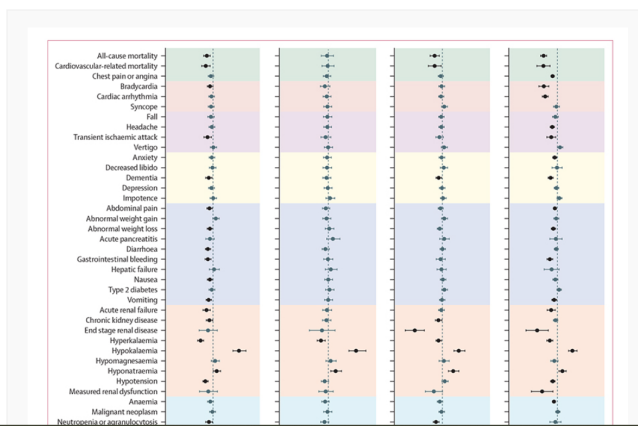
Thiazide diuretics demonstrate better effectiveness and cause fewer side effects than ACE inhibitors as first-line antihypertensive drugs, [according to a report published Oct. 24 in \*The Lancet\*](#). The study factors insurance claim data and electronic health records from 4.9 million patients across nine observational databases, making it the most comprehensive one ever on first-line antihypertensives, and it provides additional context to the 2017 guidelines for high blood pressure treatment developed by the American College of Cardiology (ACC) and American Heart Association (AHA).

Collaborators in the Observational Health Data Sciences and Informatics (OHDSI) network produced the paper “*Comprehensive comparative effectiveness and safety of first-line antihypertensive drug classes: a systematic, multinational, large-scale analysis*” as part of the collaborative's ongoing Large-Scale Evidence Generation and Evaluation across a Network of Databases (LEGEND) project, which applies high-level analytics to perform observational research on hundreds of millions of patient records within OHDSI's international database network.

OHDSI researchers believe LEGEND will continue to significantly enhance how real-world evidence is used to study important healthcare questions that impact millions of patients worldwide.

## First-Line Thiazide Diuretic Users Experience 15% Fewer Adverse Cardiovascular Outcomes Than ACE Inhibitor Users

The 2017 ACC/AHA guidelines on antihypertensives recommend initiating hypertension (high blood pressure) treatment with prescription medications from any of five drug classes, including both thiazides and ACE inhibitors. Within the LEGEND project, ACE inhibitors produced both worse cardiovascular outcomes and worse side effects than thiazides.



First-line thiazide new-users experienced three major medical

[Home](#)[About IMI](#)[Get  
involved](#)[Apply  
for funding](#)[Projects  
& results](#)

[Home](#) | [News & Events](#) | [Newsroom](#) | Can real world data replicate a clinical trial? EHDEN study suggests yes

## Can real world data replicate a clinical trial? EHDEN study suggests yes

IMI's EHDEN project dramatically demonstrated the power of using clinical data in research by replicating, during a five-day 'study-a-thon', the results of a systematic review covering 20 years of research, and a multi-year clinical trial.





# FAST OBSERVATIONAL RESEARCH IS FEASIBLE (STUDY-A-THON)

“To compare the **risk** of post-operative **complications** and **mortality** between unicompartmental vs total knee replacement.”

## Monday

Group consensus on the **problem**  
Draft cohort definitions

## Tuesday

Review clinical characterisation  
Draft patient-level prediction design

## Wednesday

Review patient-level prediction results  
Externally validate prediction model

## Thursday

Draft population-level effect estimation design  
Review population-level effect estimation diagnostics

## Friday

Review of results  
Plan for completing **publications**



# COHORT DEFINITION

ATLAS

Home

Data Sources

Search

Concept Sets

Cohort Definitions

Characterizations

Cohort Pathways

Incidence Rates

Profiles

Estimation

Prediction

Jobs

Configuration

Feedback

Apache 2.0  
open source software

provided by  
**OHDSI**  
join the journey

www.ohdsi.org/web/atlas/#/cohortdefinition/1769729

Zoeken

Cohort #1769729

[Oxford studyathon\_Day2] Patients with total knee replacement

Definition ? Concept Sets Generation Reporting Export

enter a cohort definition description here

Cohort Entry Events

Events having any of the following criteria:

+ Add Initial Event

a procedure occurrence of [Oxford studyathon] Total knee ...

+ Add attribute...

Delete Criteria

for the first time in the person's history

with continuous observation of at least 365 days before and 0 days after event index date

Limit initial events to: earliest event per person.

Restrict initial events

Inclusion Criteria

New inclusion criteria

has no prior knee replacement surgery or history of knee replacement any time prior

Copy Delete

enter an inclusion rule description

having all of the following criteria:

+ Add criteria to group...

with exactly 0 using all occurrences of:

a procedure occurrence of [Oxford studyathon] Total knee ...

+ Add attribute...

where event starts between All days Before and 1 days Before index start date add additional\_constraint

☐ restrict to the same visit occurrence

☐ allow events from outside observation period

Delete Criteria

and with exactly 0 using all occurrences of:

a procedure occurrence of [Oxford studyathon] Unicomp...

+ Add attribute...

where event starts between All days Before and 0 days After index start date add additional\_constraint

☐ restrict to the same visit occurrence

☐ allow events from outside observation period

Delete Criteria

1. Age >= 40

2. has no prior knee replacement surgery or history of knee replacement any time prior

3. No prior other knee surgery any time prior

4. no knee fracture on or any time prior

5. has no rheumatoid arthritis or other inflammatory arthropathies on or any time prior

6. No septic arthritis on or any time prior

7. Does not have symptomatic hip pathology

8. Does not have symptomatic spinal pathology

9. Does not have symptomatic foot pathology



# FAST OBSERVATIONAL RESEARCH IS FEASIBLE (STUDY-A-THON)

“To compare the **risk** of post-operative **complications** and **mortality** between unicompartmental vs total knee replacement.”

## Monday

Group consensus on the **problem**  
Draft cohort definitions

## Tuesday

Review clinical characterisation  
Draft patient-level prediction design

## Wednesday

Review patient-level prediction results  
Externally validate prediction model

## Thursday

Draft population-level effect estimation design  
Review population-level effect estimation diagnostics

## Friday

Review of results  
Plan for completing **publications**



# PUBLICATION OF THE PROTOCOL AND RESULTS

## Prospective validation of a randomised trial of unicompartmental and total knee replacement: real-world evidence from the OHDSI network



<http://data.ohdsi.org/UkaTkaSafetyEffectiveness/>



# FAST OBSERVATIONAL RESEARCH IS FEASIBLE (STUDY-A-THON)

“To compare the **risk** of post-operative **complications** and **mortality** between unicompartmental vs total knee replacement.”

## Monday

Group consensus on the **problem**  
Draft cohort definitions

## Wednesday

Review patient-level prediction results  
Externally validate prediction model

## Friday

Review of results  
Plan for completing **publications**

## Tuesday

Review clinical characterisation  
Draft patient-level prediction design

## Thursday

Draft population-level effect estimation design  
Review population-level effect estimation diagnostics



## THE LANCET Rheumatology

ARTICLES | [VOLUME 1, ISSUE 4, PE229-E236, DECEMBER 01, 2019](#)

### Opioid use, postoperative complications, and implant survival after unicompartmental versus total knee replacement: a population-based network study

[Edward Burn, MSc \\*](#) • [James Weaver, MSc \\*](#) • [Daniel Morales, PhD](#) • [Albert Prats-Urbe, MPH](#) •  
[Antonella Delmestri, PhD](#) • [Victoria Y Strauss, PhD](#) • et al. [Show all authors](#) • [Show footnotes](#)

Published: November 07, 2019 • DOI: [https://doi.org/10.1016/S2665-9913\(19\)30075-X](https://doi.org/10.1016/S2665-9913(19)30075-X) •



Check for updates





## METHODS ELABORATION

**Background :** There is uncertainty around whether to use unicompartmental knee replacement (UKR) or total knee replacement (TKR) for individuals with osteoarthritis confined to a single compartment of the knee. We aimed to **emulate the design** of the Total or Partial Knee Arthroplasty Trial (TOPKAT) **using routinely collected data** to assess whether the efficacy results reported in the trial **translate into effectiveness in routine practice** , and to assess comparative safety.

**Methods:** We did a population-based network study using data from four US and one UK health-care database, part of the Observational Health Data Sciences and Informatics network. The inclusion criteria were the same as those for TOPKAT; briefly, we identified patients aged at least 40 years with osteoarthritis who had undergone UKR or TKR and who had available data for at least one year prior to surgery. Patients were excluded if they had evidence of previous knee arthroplasty, knee fracture, knee surgery (except diagnostic), rheumatoid arthritis, inflammatory arthropathies, or septic arthritis. Opioid use from 91–365 days after surgery, as a proxy for persistent pain, was assessed for all participants in all databases. Postoperative complications (ie, venous thromboembolism, infection, readmission, and mortality) were assessed over the 60 days after surgery and implant survival (as measured by revision procedures) was assessed over the 5 years after surgery. Outcomes were assessed in all databases, except for readmission, which was assessed in three of the databases, and mortality, which was assessed in two of the databases. Propensity score matched Cox proportional hazards models were fitted for each outcome. Calibrated hazard ratios (cHRs) were generated for each database to account for observed differences in control outcomes, and cHRs were then combined using meta-analysis.



## FINDINGS AND INTERPRETATION

**Findings:** 33,867 individuals who received UKR and 557,831 individuals who received TKR between Jan 1, 2005, and April 30, 2018, were eligible for matching. 32,379 with UKR and 250,377 with TKR were propensity score matched and informed the analyses. UKR was associated with a reduced risk of postoperative opioid use (cHR from meta-analysis 0.81, 95% CI 0.73–0.90) and a reduced risk of venous thromboembolism (0.62, 0.36–0.95), whereas no difference was seen for infection (0.85, 0.51–1.37) and readmission (0.79, 0.47–1.25). Evidence was insufficient to conclude whether there was a reduction in risk of mortality. UKR was also associated with an increased risk of revision (1.64, 1.40–1.94).

**Interpretation:** UKR was associated with a reduced risk of postoperative opioid use compared with TKR, which might indicate a reduced risk of persistent pain after surgery. UKR was associated with a lower risk of venous thromboembolism but an **increased risk of revision** compared with TKR. These findings can help to inform shared decision making for individuals eligible for knee replacement surgery.



From question to  
publication in 5 days

How EHDEN and OHDSI  
change medical evidence  
generation through open  
science



This project has received funding from the Innovative Medicines Initiative 2 Joint Undertaking (JU) under grant agreement No 806868. The JU receives support from the European Union's Horizon 2020 research and innovation programme and EPPIA.



**KEES VAN BOCHOVE**

CEO TheHyve (NL)



**DANIEL PRIETO-ALHAMBRA**

Professor at University of Oxford (UK)  
NIHR Clinician Scientist

## Webinar

March 28, 2019  
16h00-17h00 CET

Register at  
[bit.ly/EHDENWeb1](https://bit.ly/EHDENWeb1)

Review the webinar for a run-through of the study-a-thon with the organizer, Dani Prieto from University of Oxford.



<https://youtu.be/X5yuoJoL6xs>

<https://thehyve.nl/resources>

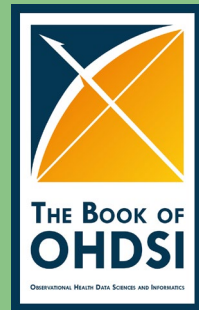
## Statement #2

Medical evidence  
generation is  
changing through  
open science

- The scale is changing: “a paradigm shift from **single study** and single estimate medical research to **large-scale systematic evidence** generation”

<https://ohdsi.github.io/TheBookOfOhdsi/OpenScience.html>

- The speed is changing:  
study-a-thon example





*We advance biology and medical  
sciences by building and serving  
thriving open source communities*

A banner for the OHDSI Europe Symposium. The background is a high-angle photograph of a historic university building with a large dome and many windows. The text is overlaid on the image.

Home Info Symposium 2020 Forum Github Past Events Photos Contact

EHDEN

 **OHDSI**  
OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

**From data to impact: the journey towards  
improving clinical practice**

March 27th Symposium, March 28-29th Tutorials  
Join us at Oxford University, UK

[More Info](#)

[Login](#)





## Working at The Hyve

**Our team** consists of software developers, solutions architects, data engineers and other support staff who share the same vision: to enhance the quality and impact of research by enabling scientists in life sciences and healthcare R&D with open source software, open data and open standards. The Hyve has a diverse and multinational team which also helps to foster our values of collaboration and learning.

Want to know how it is like to work at The Hyve? Follow our **Hyver stories!**

