



**Karolinska  
Institutet**

# **Global landscape of open science biomolecular knowledgebases**

**David Herron PhD**  
**Biomolecular research liaison**  
**Karolinska Institutet University Library**  
**[david.herron@ki.se](mailto:david.herron@ki.se)**  
**+08 524 84113/070 399 4135**

# Aim

- To paint a picture of the global landscape of open science biomolecular knowledgebases and show some specific examples

# Contents

- A database on databases
- A long list
- Three big clusters
- The European sustainability effort
- The “Breakthrough” databases
- A few demos

# A database on databases

- **Database Commons – a catalog of biological databases**
- @ the National Genomics Data Center (NGDC), Beijing, China
- <https://ngdc.cncb.ac.cn/databasecommons/>
- About 5500 databases (May 2022)
- Ranked citation data on databases
  - e.g.demo
  - [browse]
  - **David**

# A long list

- **Journal of Nucleic Acid Research (NAR) online molecular biology database collection**
- @ Oxford University Press, Oxford, England
- <https://www.oxfordjournals.org/nar/database/c/>
- About 1700 databases (Jan 2022)
- About 80 new and 80 leave per year
- All linked to a reference article in NAR
  - e.g. demo
  - Category/paper list
  - Search Summary Papers
  - [biomarkers]
  - **MarkerDB (TP53)**

# The big three clusters (1)

- European Bioinformatics Institute (EMBL-EBI)
- @ Hinxton, Cambridge, UK
- <https://www.ebi.ac.uk/>
- EBI search integrated database retrieval system
- Allows searches in about 40 databases
  - e.g. demo
  - [TP53] in All
  - UniprotKB P04637 (P53\_HUMAN)

## The big three clusters (2)

- National Center for Biotechnology information (NCBI)
- @ NIH, NLM, Bethesda, USA
- <https://www.ncbi.nlm.nih.gov/>
- Entrez integrated database retrieval system
- Allows searches across about 35 databases
  - e.g. demo
  - [TP53] in All databases
  - **Gene (TP53)**

## The big three clusters (3)

- National Genomics Data Center (NGDC)
- @ Beijing, China
- <https://ngdc.cncb.ac.cn/>
- BIG search integrated database retrieval system
- Allows searches across about 50 databases
  - e.g. demo
  - TP53 in All Databases
  - NGCD and Partners Databases
  - Database Commons (TP53 databases)



# The European sustainability effort (ELIXIR)

## ELIXIR Core Data Resources

<https://elixir-europe.org/platforms/data/core-data-resources>

e.g. Europe PMC, Human Protein Atlas, UniProtKB

## ELIXIR Deposition Databases for Biomolecular Data

<https://elixir-europe.org/platforms/data/elixir-deposition-databases>

e.g. BioStudies

# The “Breakthrough” databases (NAR)

- “NAR assigns “Breakthrough” status to papers that solve longstanding problems, or which are otherwise considered exceptionally important” (NAR db edition 2022).
- ScoV2-MD, AlphaFold Protein Structure Database (2022)
- ABCDDb, Scop, Alliance of Genome Resources (2020)
- Virual Metabolic Human, Vibrism (2019)
- Denovo-db, Monarch Initiative, Open Targets (2017)

## A few of demos

- Europe PMC  
→ <https://europepmc.org/>
- Open targets platform  
→ <https://platform.opentargets.org/>
- Human Protein Atlas  
→ <https://www.proteinatlas.org/>