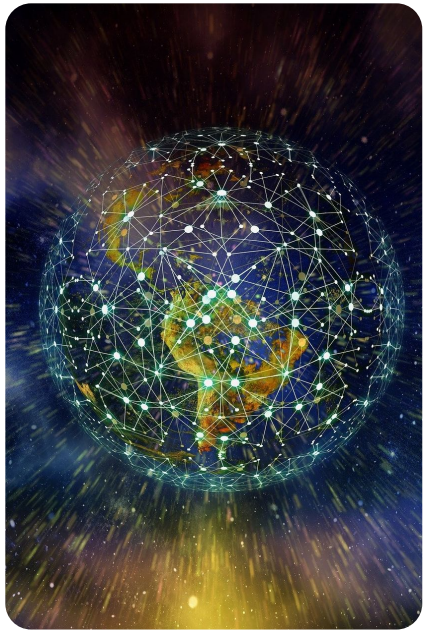
The Global Biodata Coalition: An overview

GLOBAL BIODATA COALITION

November 2025

Biodata resources: a global infrastructure



Crucial

- A healthy ecosystem of biodata resources underlies biological, life sciences & biomedical research
- Essential - ubiquitous use in academia, biotech, pharma, etc.

Distributed

- Highly distributed, scientist-led infrastructure that facilitates the efficient sharing of research data

Cost effective

- Spending to preserve research data is much lower than the cost of generating the data

With opportunity

- Life science is becoming increasingly data intensive
- A global infrastructure ready to scale for this will be a powerful enabler

Challenges for the biodata infrastructure

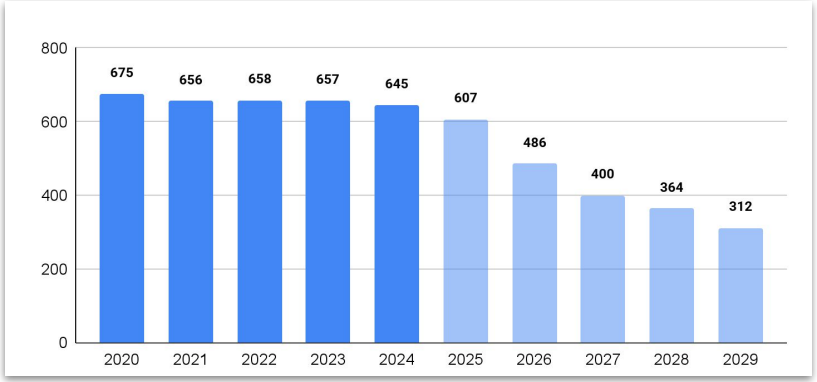
Demand increasing

- Increasing rate of data generation increases demand
- Open access policies increase demand
- Requirements for data management plans increase demand
- New technologies require new data resources

Sustainability challenges

- Funding has been haphazard, short term, and distributed unsystematically
- There is little global coordination among funders of these resources
- This global research infrastructure is not managed as an infrastructure
- The infrastructure is poorly described

GCBR Staff Funding Horizons (42 GCBRs)



GLOBAL BIODATA COALITION

Formative discussions ahead of establishing GBC

PERSPECTIVE

2015

Sustaining the big-data ecosystem

Organizing and accessing biomedical big data will require quite different business models, say **Philip E. Bourne, Jon R. Lorsch** *and* **Eric D. Green.**

recorded. All of this means that absolute numbers are hard to interpret.

These caveats notwithstanding, more details of data usage are needed to inform funding decisions. Over time, such usage patterns could tell us how best to target annotation and curation efforts, establish which data should receive the most attention and therefore incur the largest cost, and determine which data should be kept in the longer term. The cost of data regeneration can also influence decisions about keeping data.

Funders should encourage the development of new metrics to ascertain the usage and value of data, and persuade data resources to provide such statistics for all of the data they maintain. We can learn here

Nature 2015 527: S16–S17

Towards Coordinated International Support of Core Data Resources for the Life Sciences

W. Anderson, R. Apweiler, A. Bateman, G.A. Bauer, H. Berman, J.A. Blake, N. Blomberg, S.K. Burley, G. Cochrane, V. Di Francesco, T. Donohue, C. Durinx, A. Game, E.D. Green, T. Gojobori, P. Goodhand, A. Hamosh, H. Hermjakob, M. Kanehisa, R. Kiley, J. McEntyre, R. McKibbin, S. Miyano, B. Pauly, N. Perrimon, M.A. Ragan, G. Richards, Y-Y. Teo, M. Westerfield, E. Westhof, P.F. Lasko

doi: https://doi.org/10.1101/110825

This article is a preprint and has not been certified by peer review [what does this mean?].

2017

A global coalition to sustain core data

As members of an international working group to support the rapidly growing core-data resources in the life sciences, we aim to create a sustainable and accessible data infrastructure that will benefit scientists worldwide.

Although researchers have relied on international resources such as the Protein Data Bank and Flybase for decades, the current system is unsustainable because it is largely funded by short-term grants (P. E. Bourne *et al. Nature* **527,** S16–S17; 2015). A global coalition of data resources would provide much-needed governance structure, active service manageme community-driven scient development, which toget

2017

Nature 2017 543: 179

GLOBAL BIODATA COALITION

Composition of the GBC

Members

NIH National Human Genome Research Institute

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Federal Department of Economic Affairs, Education and Research EAER
State Secretariat for Education, Research and Innovation SERI

NSF

wellcome

UKRI

Forskningsrådet
The Research Council of Norway

Chan Zuckerberg Initiative

GenomeCanada

EMBL

The African Academy of Sciences

Funds managed through the Human Frontier Science Program Organization

Observers

- Commonwealth Scientific and Industrial Research Organisation
- European Commission
  - European Research Council
- Inserm (France)
- South African Medical Research Council

GLOBAL BIODATA COALITION

A timeline of founding and establishing the GBC

2015	Nature perspective "Sustaining the big data ecosystem" published, outlining need for funder collaboration
2015-2018	Series of meetings with funders, biodata resource managers, biodata resource users
2016	Affirmation from Heads of International Research Organisations (HIROs) of need for GBC
2017	GBC Steering group established
2018-2019	First funding from NIH, NSF, Wellcome, A*Star
2019	First GBC Board meeting
2020	Launch: Secretariat, website, branding, Letter of Understanding
2020	Scientific programme initiated
2021	GBC Scientific Advisory Committee appointed
2022	Executive Director appointed
2022	Global Core Biodata Resources first selection round completed
2023	Global Core Biodata Resources second selection round completed
2023	First inventory of global biodata resources published
2025	White paper

GLOBAL BIODATA COALITION

GBC aims

- To be a forum for funders of biodata resources to better coordinate and share approaches for efficient management and growth of this infrastructure and share strategies

- To stabilize and ensure sustainable financial support for the global biodata infrastructure, with a focus on an identified and prioritized set of Global Core Biodata Resources that are crucial for sustaining the broader infrastructure

GBC Strategy and Targets 2023-2026

Aim 1. To be a forum of funders	
Goal 1	Increase number of GBC funder members
Goal 2	Promote cross-funder dialogue on sustainability
Goal 3	Advocate with funders for critical support

Aim 2. Ensure sustainable financial support for global biodata infrastructure	
Goal 1	Develop new funding models
Goal 2	Characterise components that comprise global biodata infrastructure
Goal 3	Work with GCBRs to support their sustainability

2026 Deliverables

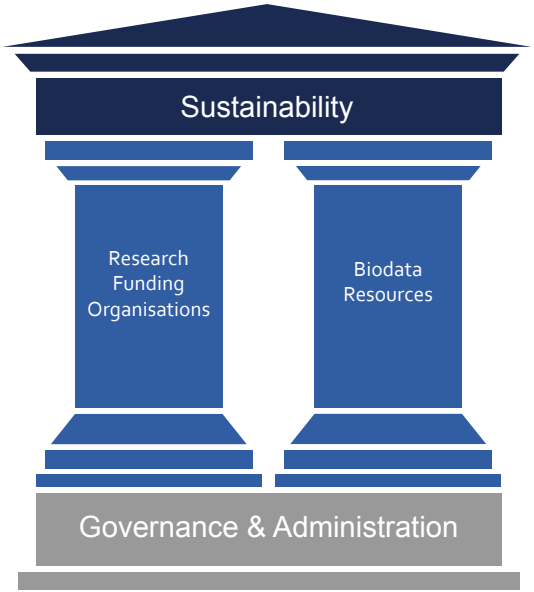
- Increase funding to GCBRs
- Increase membership to 15 funders
- Raise profile of biodata sustainability through advocacy
- Publish Working Group white papers
- Develop three case studies to support policy making
- Establish stable list of contributing GCBRs
- Pilot one new cooperative funding model

https://globalbiodata.org/whathttps://globalbiodata.org/what-we-do/gbc-strategy-2023-2026/-we-do/gbc-strategy-2023-2026/

GLOBAL BIODATA COALITION

Constructing sustainability

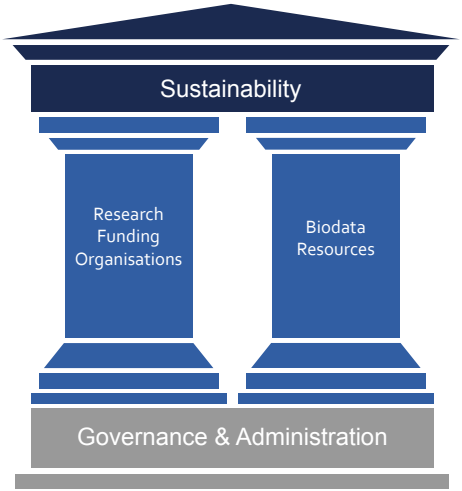
- Pillar communities
- Research funding organisations
- Biodata resources
- Models and mechanisms
- Governance and administration



GLOBAL BIODATA COALITION

Working with research funders

- Exploration of existing funding
- Quest for opportunities
- Internationally aligned approaches
- Dynamic and rewarding membership programme
- Consideration of comparative models of biodata sustay



Example of ongoing work

- Funder knowledge exchange
- Board working group on sustainability
  - National practices
  - Biodata resource life cycle management
  - Towards global sustainability models

GLOBAL BIODATA COALITION

Global Core Biodata Resources

GLOBAL CORE BIODATA RESOURCE

GCBRs are biodata resources that are of fundamental importance to the wider biological and life sciences community and the long term preservation of biological data:

- Provide free and open access to their data
- Are used extensively both in terms of the number and distribution of users
- Are mature and comprehensive, and considered authoritative in their field
- Show high scientific quality, and provide a professional standard of service delivery

52 Global Core Biodata Resources

GBC defined core selection criteria for GCBRs

Two selection rounds for Global Core Biodata Resources undertaken in 2022 and 2023

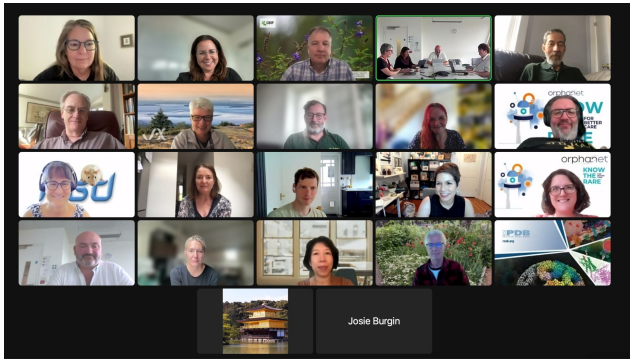
List of 52 GCBRs includes resources hosted in 13 countries, spanning a wide range of data types and domains and serving hundreds of thousands of users worldwide

GLOBAL BIODATA COALITION

GCBR Forum

Comprises the Managers of the current set of

- 52 GCBRs
- Over 75 representatives including alternates
- Meetings chaired by Forum co-chairs
  - Guy Cochrane (GBC)
  - Teri Klein (PharmGKB and ClinGen)
- Meets every 2-3 months
- Terms of Reference:



"The GCBR Forum brings together representatives of GCBRs to share experience and good practice, and engage with GBC to enhance the sustainability of the global biodata infrastructure."

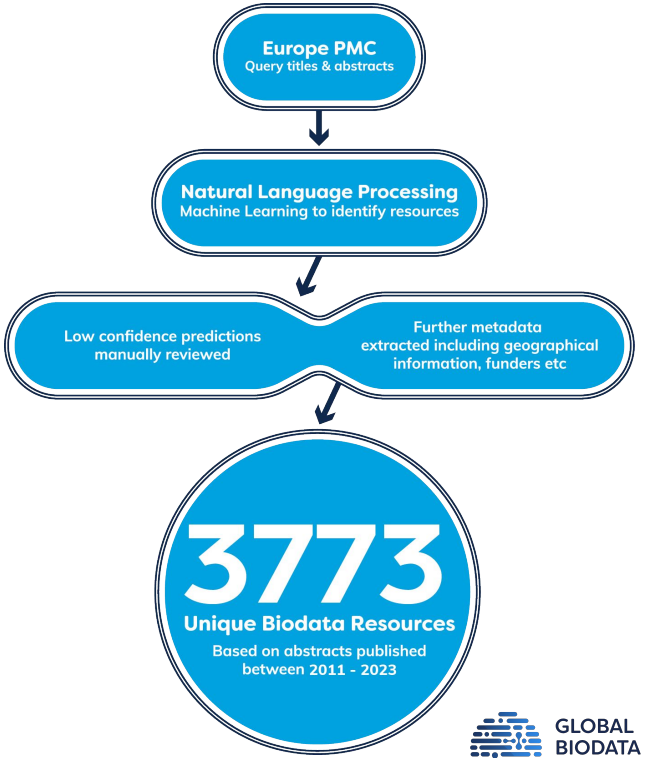
GLOBAL BIODATA COALITION

GCBR Forum: Activities and outputs

Activity	Output
Exploration of sustainability planning in place at the GCBRs	Per-GCBR sustainability plan with some common structure to allow inter-comparability
Exploration of biodata resources sustainability	Sustainability indicators
Evaluation of proposed sustainability mechanisms from Board Working Group Consultation Paper	Feedback on proposed models and iteration towards viable solutions
Raising awareness of GBC and sustainability issues with GCBR funders	Outreach to, and communications with, as-yet unreached funders
Community building and discussion and mutual awareness	Publication of GCBR set and reflecting common views

Inventory of global biodata resources

- Discovery of biodata resources from the literature
- 3,773 biodata resources
- 58.5% of biodata resources name at least one of 2,265 funders
- Code configured into an executable pipeline with documentation in GitHub to facilitate reuse and future updates
- Publications
  - https://doi.org/10.1371/journal.pone.0294812
  - https://zenodo.org/record/7767794#.ZGafrk_MLD5



GLOBAL BIODATA COALITION

GBC/Uni. Illinois Urbana-Champaign: Heidi J. Imker; GBC/Uni, Arizona: Kenneth E Schackart; CZI: Dario Taraborelli, Ana-Maria Istrate, Donghui Li, Gully Burns; EMBL-EBI: Melissa Harrison, Aravind Venkatesan, Mohamed Selim.

Supporting open data, global inclusion and working across disciplines

- Open data is a value multiplier and essential for sustainability
- Recognising the processes of data science and contributions to the onward value to biodata resource sustainability
- Science is global, as should be involvement in the global biodata infrastructure
- Life sciences connecting to other disciplines

Examples of ongoing work

Engagement with stakeholders

- Standards organisations
- Infrastructures beyond the life sciences (e.g. health, environmental)
- Policy makers (e.g. Digital Sequence Information discussion at COP15)

Board working group on open data strategies

- Global coordination and equity
- Implementing and monitoring data management plans
- Learning from other disciplines
- Signposting repositories
- Training and incentives

GBC 2024 Highlights Brochure

- Digital publication
- View and download here:
- A record of GBC's activities and achievements to date, as well as future plans

GLOBAL BIODATA COALITION

Follow us:

Subscribe to our newsletter:

Subscribe

https://globalbiodata.org/

info@globalbiodata.org

GLOBAL BIODATA COALITION