Global Core Biodata Resources Webinar for Applicants

@globalbiodata

Chuck Cook
Rachel Drysdale
David Carr
GBC Secretariat

GLOBAL BIODATA COALITION

22nd and 23rd February 2023

Housekeeping

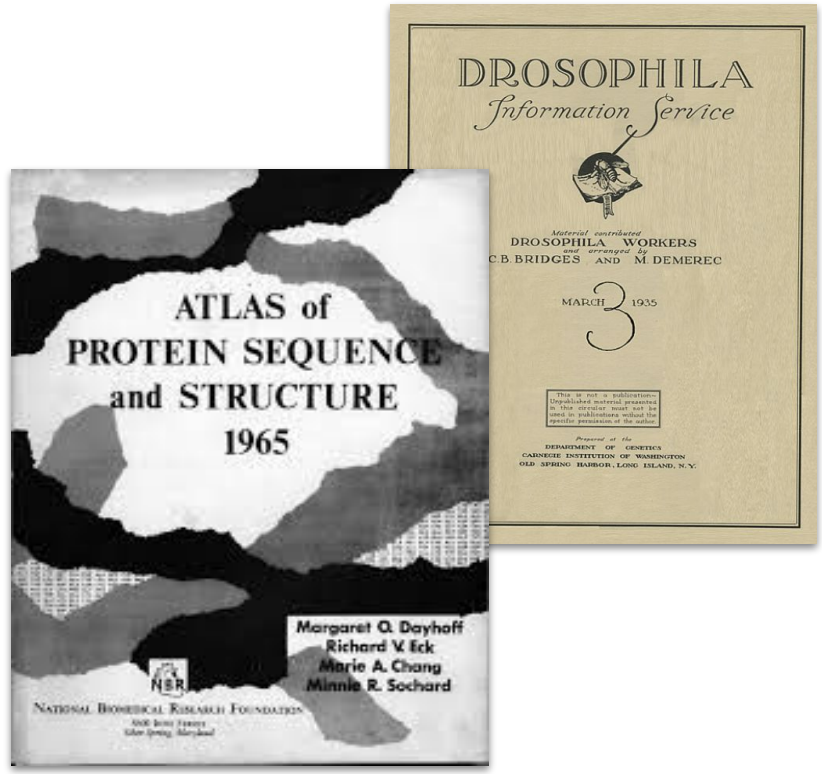
- The video is being recorded and will be made available via the GBC website after the presentation
- Attendees will be muted throughout, and video thumbnails will be turned off
- Questions entered into Q&A box will be addressed after the main presentation
- Please use the chat function to highlight any administrative issues at any time

We hope you enjoy the webinar

GLOBAL BIODATA COALITION

Biodata resources

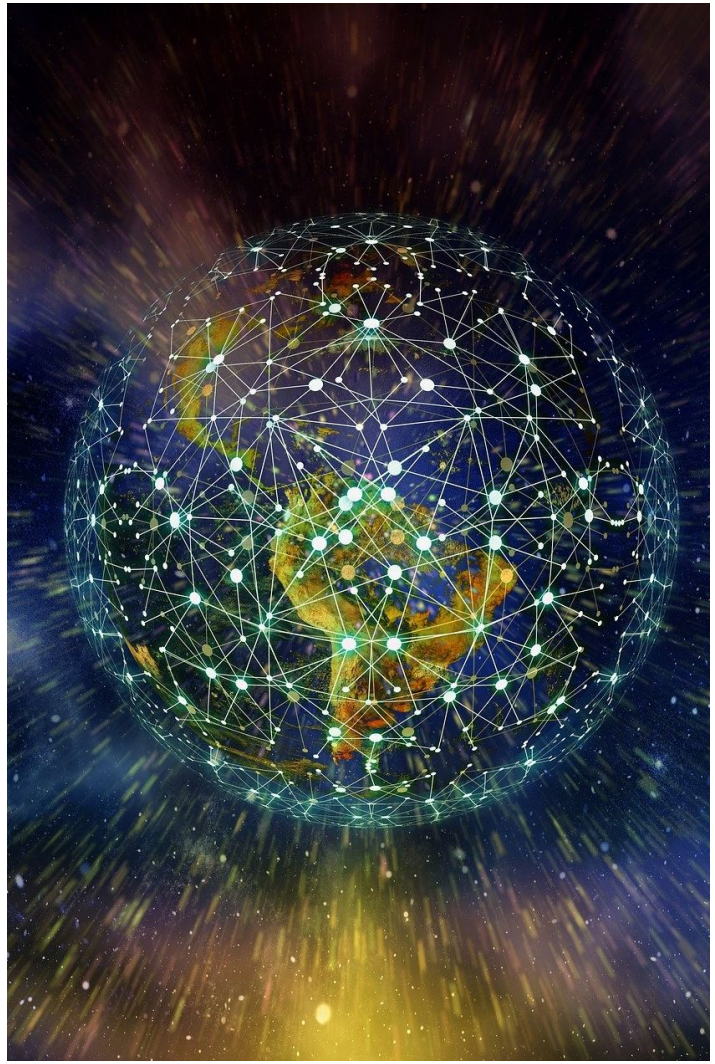
- Life sciences: long history of sharing data
- Archive primary life sciences research data
  Specialist repositories
- Added-value curation/analysis of primary data
- Extensive collaboration and data exchange
- 1,000’s of resources



GLOBAL BIODATA COALITION

Biodata resources: a crucial global infrastructure

- Crucial
  - The healthy ecosystem of biodata resources underlies biological, life sciences & biomedical research
  - Essential - ubiquitous use in research, biotech, pharma
- Distributed
  - Scientist-led
  - Efficient sharing of research data
- Cost effective
  - Data reuse vs regeneration of data
- Expanding
  - Global science is increasingly data-intensive
  - Scalable infrastructure will promote growth
- Under threat
  - No international coordination of the infrastructure



GLOBAL BIODATA COALITION

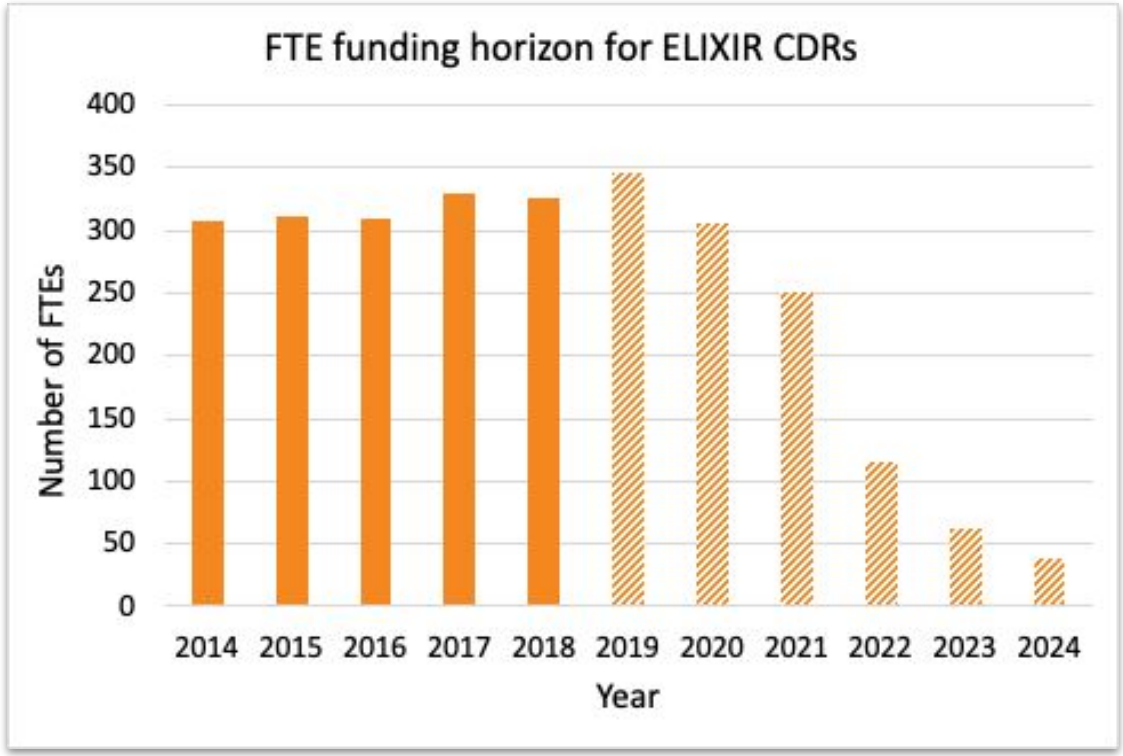
Challenges to the biodata infrastructure

Demands increasing

- Increasing rate of data generation increases demand
- Open access policies 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



https://doi.org/10.1093/bioinformatics/btz959

GLOBAL BIODATA COALITION

Establishing the Global Biodata Coalition

- **2015** *Nature* perspective “Sustaining the big data ecosystem” - need for funder collaboration
- **2016-18** Series of meetings: funders, biodata resource managers, biodata resource users
- **2016** Affirmation from Heads of International Research Organisations (HIROs)
- **2017** Steering group established, *Nature* note “A global coalition to sustain core data”
- **2018-19** Seed funding NIH, NSF, Wellcome, A*Star
- **2019** Launch: secretariat, website, logo, Letter of Understanding
- **2020** Scientific program initiated
- **2021** Scientific Advisory Committee appointed
- **2022** Executive Director appointed
- **2022** Global Core Biodata Resources Selection

GLOBAL BIODATA COALITION

The Global Biodata Coalition

Members

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

Biotechnology and Biological Sciences Research Council

Chan Zuckerberg Initiative

Forskningsrådet
The Research Council of Norway

EMBL

The African Academy of Sciences

GenomeCanada

Funds managed through the Human Frontier Science Program Organization

Observers

Agency for Science, Technology and Research (Singapore)
Canadian Institutes of Health Research
European Commission
European Research Council
Indian Council for Medical Research
Inserm (France)
Japan Agency for Science and Technology
Korean National Institutes of Health
Natural Sciences and Engineering Research Council (Canada)
Royal Society Te Apārangi (New Zealand)
South African Medical Research Council

- Letter of understanding
  - Support for Global Biodata Coalition
  - Commitment to work towards sustainable biodata resources
- Drive for more funding organisations as members
- Drive for greater diversity and global spread of funding organisation types

GLOBAL BIODATA COALITION

GLOBAL BIODATA COALITION

The Global Biodata Coalition (GBC) is a forum for research funders to better coordinate and share approaches for the efficient management and growth of biodata resources worldwide. The GBC aims to stabilize and ensure sustainable financial support for the global biodata infrastructure and has identified a set of Global Core Biodata Resources that are crucial for sustaining the broader biodata infrastructure.

Discover the benefits

https://globalbiodata.org/

GLOBAL BIODATA COALITION

GBC work program

1) Describing the biodata infrastructure
   - Inventory of global biodata resources
   - **Global Core Biodata Resources**

2) Knowledge exchange
   - Funder working groups

GLOBAL BIODATA COALITION

Global Core Biodata Resources

GLOBAL CORE BIODATA RESOURCE

"**fundamental importance** to the wider **biological and life sciences community** and the **long term preservation of biological data**"

and

"show **high levels of usage, scientific quality and service**"

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

https://doi.org/10.5281/zenodo.7468719

GLOBAL BIODATA COALITION

The Global Core Biodata Resources list - four examples

Name **Host Countries: Primary**, Additional	Overview	Funders, 2017–2022
Alliance of Genome Resources **USA**, Canada, UK	The primary mission of the Alliance of Genome Resources (the Alliance) is to develop and maintain sustainable genome information resources that facilitate the use of diverse model organisms in understanding the genetic and genomic basis of human biology, health and disease.	US National Institutes of Health (NIH)
BacDive **Germany**	BacDive is the worldwide largest database for standardized bacterial information. Its mission is to mobilize and integrate research data on strain level from diverse sources and make it freely accessible.	German Federal Ministry of Education and Research (BMBF)
BRENDA **Germany**	BRENDA is the main collection of enzyme functional data available to the scientific community.	German Federal Ministry of Education and Research (BMBF)
ChEBI **UK**	Chemical Entities of Biological Interest (ChEBI) is a freely available dictionary of molecular entities focused on 'small' chemical compounds.	European Commission, EMBL, UK Research and Innovation (UKRI)

https://globalbiodata.org/scientific-activities/global-core-biodata-resources/

GLOBAL BIODATA COALITION

Why focus on Global Core Biodata Resources?

- Global Core Biodata Resources are fundamental for the entire biological, life sciences & biomedical global infrastructure

- Global Core Biodata Resources are analogous to keystone species in an ecosystem
- Focus on Global Core Biodata Resources will help protect the entire global infrastructure

GLOBAL BIODATA COALITION

Defining GCBRs will provide benefits 1

- **For researchers** selecting repositories to archive their primary data, to comply with funders' and publishers' open data requirements, GCBR status will provide confidence in their choices.

- **For funding agencies and science publishers**, the availability of recognised lists of GCBRs will help them recommend suitable data repositories and reliable sources of information to their grantees and authors, with confidence.

- **For individual funding agencies** faced with applications for support from multiple (and sometimes competing) data resources, a list of GCBRs and the criteria that characterise them will provide insight into the established data resource ecosystem, in turn providing useful context for local decision making.

GLOBAL BIODATA COALITION

Defining GCBRs will provide benefits 2

- **For developing data resources**, GCBRs will provide examples of good practice that can inform the development of their own data resources.

- **For data resources defined as GCBRs**, the GCBR community will provide opportunities for sharing expertise, driving collaboration, and - within the context of the GBC - exploring potential solutions to the challenge posed by their precarious funding.

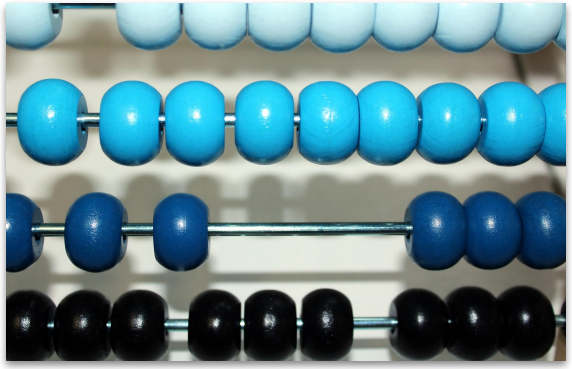
- **For all stakeholders, including private enterprise** - open data stimulates research and drives data re-use and discovery. Working toward GCBR status inspires data resources to implement more permissive open data licenses to more fully reflect the FAIR principles.

GLOBAL BIODATA COALITION

Indicators for Global Core Biodata Resources

Indicators are:

Quantitative ….



and Qualitative

“Not everything that can be counted counts,
and not everything that counts can be counted”

William Bruce Cameron*

Indicators are used to inform the peer review process

The peer review process brings expert judgement

*1963, Informal Sociology, a casual introduction to sociological thinking. Page 13, Random House, New York.

GLOBAL BIODATA COALITION

Indicators - Five Categories

Indicator Category
1: Scientific focus and quality of science
2: Community served by the resource
3: Quality of service
4: Funding, governance and legal infrastructure
5: Impact stories

https://doi.org/10.5281/zenodo.7468719

GLOBAL BIODATA COALITION

Indicators - Category 1

1. Scientific focus and quality
1a. Deposition database and/or Knowledgebase
1b. Scope statement
1c. Global dimension 1c i. Operation 1c ii. Users/contributors
1d. Staff effort

Includes the inherent scientific quality of the data and metadata, the resource's uniqueness and comprehensiveness, whether the resource is a recognised authority, and whether the resource is of fundamental importance to the broad life science community and the long-term preservation of data.

https://doi.org/10.5281/zenodo.7468719

GLOBAL BIODATA COALITION

Indicators - Category 2

2. Community
2a. Data resource usage - quantitative data
2b. Usage in research as measured through data resource citation in the scientific literature
2c. Citation of key publications describing the data resource
2d. Connections to other data resources

Reflects the size and the measured demand of the communities that are served by the resource in several dimensions, and includes web statistics, user reach, citation data and global usage data.

https://doi.org/10.5281/zenodo.7468719

GLOBAL BIODATA COALITION

Indicators - Category 3

3. Quality of service
3a. Identifier use
3b. Data volume
3c. Technical performance: 3c i. Uptime 3c ii. Response times of key web pages 3c iii. Back-up and disaster recovery
3d. Use of standards
3e. Documentation 3e i. Data Curation 3e ii. Provenance and Evidence 3e iii. Quality Assurance

3. Quality (continued)
3f. Data availability 3f i. Data sharing services 3f ii. Data sharing formats
3g. User support 3g i. Helpdesk 3g ii. User feedback 3g iii. Training 3g iv. Communications 3g v. Language

Service levels, reliability, and technical performance, use of data and metadata standards, data availability, provenance compliance and user support.

https://doi.org/10.5281/zenodo.7468719

GLOBAL BIODATA COALITION

Indicators - Category 4

4. Funding, governance and legal infrastructure
4a. Funding
4b. Scientific Advisory Board
4c. Data preservation
4d. Open Science
4e. Privacy policy
4f. Ethics policy

Covers the funding provision, governance and legal footing, including open science licensing, and privacy and ethics policy considerations.

https://doi.org/10.5281/zenodo.7468719

GLOBAL BIODATA COALITION

Indicators - Category 5

5. Impact stories
5a. Accelerating science
5b. Counterfactual

Includes evaluation of how effectively the data resource is meeting the needs of the scientific community via counterfactual and accelerating science stories.

Descriptions of all 23 Indicators by which the Global Core Biodata Resources will be selected can be found in the materials listed at the end of this presentation.

https://doi.org/10.5281/zenodo.7468719

GLOBAL BIODATA COALITION

The GCBR 2023 selection process - announcement

THE GLOBAL BIODATA COALITION TO HOLD A 2023 ROUND OF SELECTION OF GLOBAL CORE BIODATA RESOURCES

Feb 13, 2023

The Global Biodata Coalition (GBC) is pleased to announce a second round of selection of Global Core Biodata Resources, that will run throughout 2023.

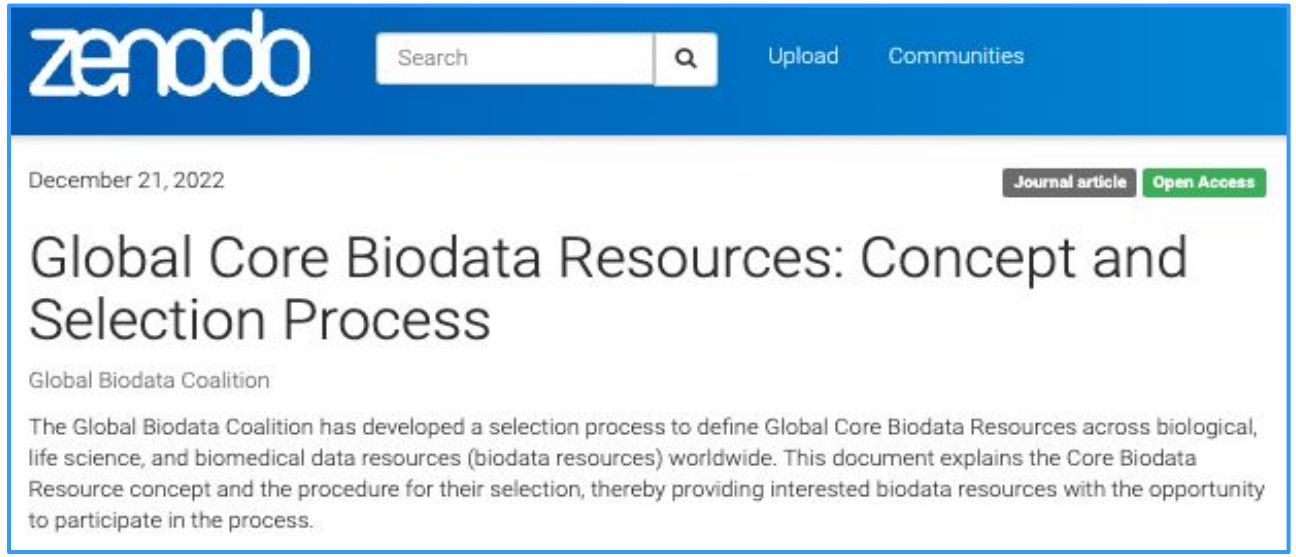
"Background information and instructions for applicants can be found on the Global Core Biodata Resource Selection 2023 web page and in the paper "Global Core Biodata Resources: Concept and Selection Process"

https://globalbiodata.org/gcbr-2023-process-launch/

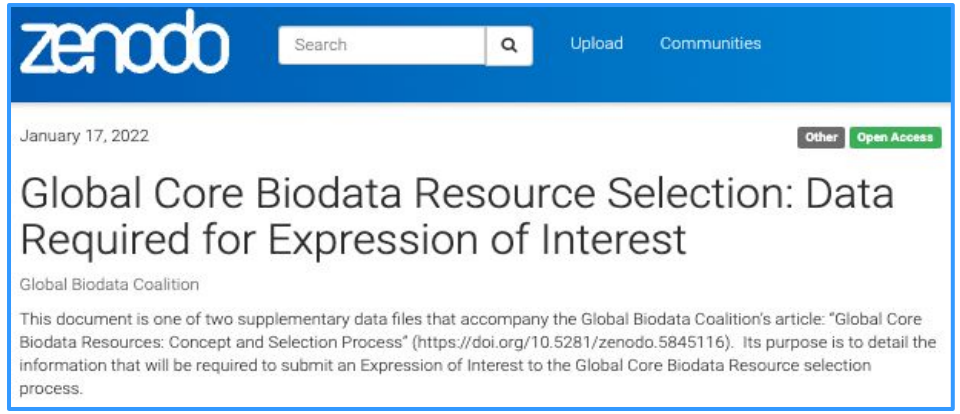
GLOBAL BIODATA COALITION

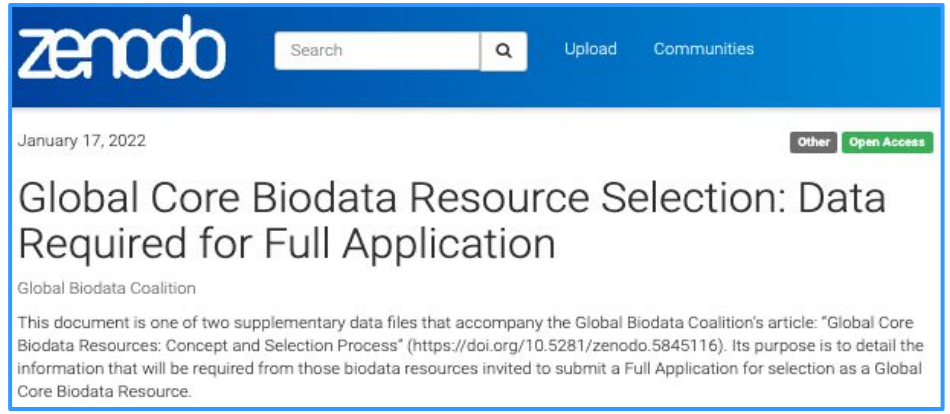
The GCBR 2023 selection process - articles

The primary paper - https://doi.org/10.5281/zenodo.7468719



The two supplementary data documents: https://doi.org/10.5281/zenodo.5846742 and /zenodo.5846758





GLOBAL BIODATA COALITION

Selection is a two-stage process

1. Expressions of Interest:
   - Description
   - Eligibility Criteria
   - Short Answer questions

Review Committee will decide the applicants to be invited to the Full Application stage

2. Full Application: Detailed questions about the complete set of 23 indicators used for Global Core Biodata Resource selection.

Review Committee will decide which data resources they recommend to be included in the initial list of GCBRs. The review committee's recommendations will be forwarded to the GBC Board of Funders for the formal final decision.

https://globalbiodata.org/scientific-activities/global-core-biodata-resources/gcbr-selection-2023/

GLOBAL BIODATA COALITION

Expression of Interest - 3 sections

Section 1 - Description of Biodata Resource

D1: Concise description of your biodata resource, addressing how it falls within scope as a Global Core Biodata Resource (200 words)

D2: Is this biodata resource……
A deposition database serving as a repository of record for a specified data type(s)?
or
A knowledgebase where, for example, expert manual curation adds value?

D3: Consortium context (100 words)

Global Core Biodata Resource Selection: Data Required for Expression of Interest
https://doi.org/10.5281/zenodo.5846742

GLOBAL BIODATA COALITION

Expression of Interest - 3 sections

Section 2 - Eligibility Criteria

10 Yes/No questions regarding

- data resource maturity and stability
- ability to provide quantitative usage data
- metadata and standards
- data availability and format
- governance
- open data policy

The purpose of the Eligibility Criteria is to focus the applicant's attention on the factors required to qualify for GCBR status, and enable them to judge whether continuing with their application is worth their investment

For the complete list of EC questions see:
Global Core Biodata Resource Selection: Data Required for Expression of Interest
https://doi.org/10.5281/zenodo.5846742

GLOBAL BIODATA COALITION

Eligibility Criteria - examples

EC2: Is the resource able to provide usage statistics covering the scale of users and a statement of geographic distribution of users?

To qualify as a Core Biodata Resource it needs to be able to demonstrate global usage beyond the country in which the data resource is housed, at scale.

EC5: Does the resource have a Scientific Advisory Board or equivalent formal advisory body?

A dedicated advisory body will be required for eligibility.

EC6: Does the biodata resource have an open data licence with no requirement for permission from a data access committee or other authority for access to the resource or data within?

An open data license will be required for eligibility, e.g. Creative Commons licenses CC0, CC-BY or CC-BY-SA are all conformant with the Open Definition (http://opendefinition.org/licenses/).

For the complete list of EC questions see:

Global Core Biodata Resource Selection: Data Required for Expression of Interest
https://doi.org/10.5281/zenodo.5846742

GLOBAL BIODATA COALITION

Expression of Interest - 3 sections

Section 3 - Short Answer Questions

Five questions relating to the alignment of the data resource with the goals of the GBC and the characteristics of the resource - in terms of the Indicators introduced above - that qualify it to be considered a GCBR.

Answers given in Section 3 will be assessed in terms of how well they indicate that the biodata resource is suitable to progress to the Full Application stage. i.e. does the answer to each question suggests that this biodata resource has a realistic chance of attaining GCBR status, for the corresponding indicators.

https://doi.org/10.5281/zenodo.5846742

GLOBAL BIODATA COALITION

Expression of Interest - 3 sections

Section 3 - Short Answer Questions - 200 words

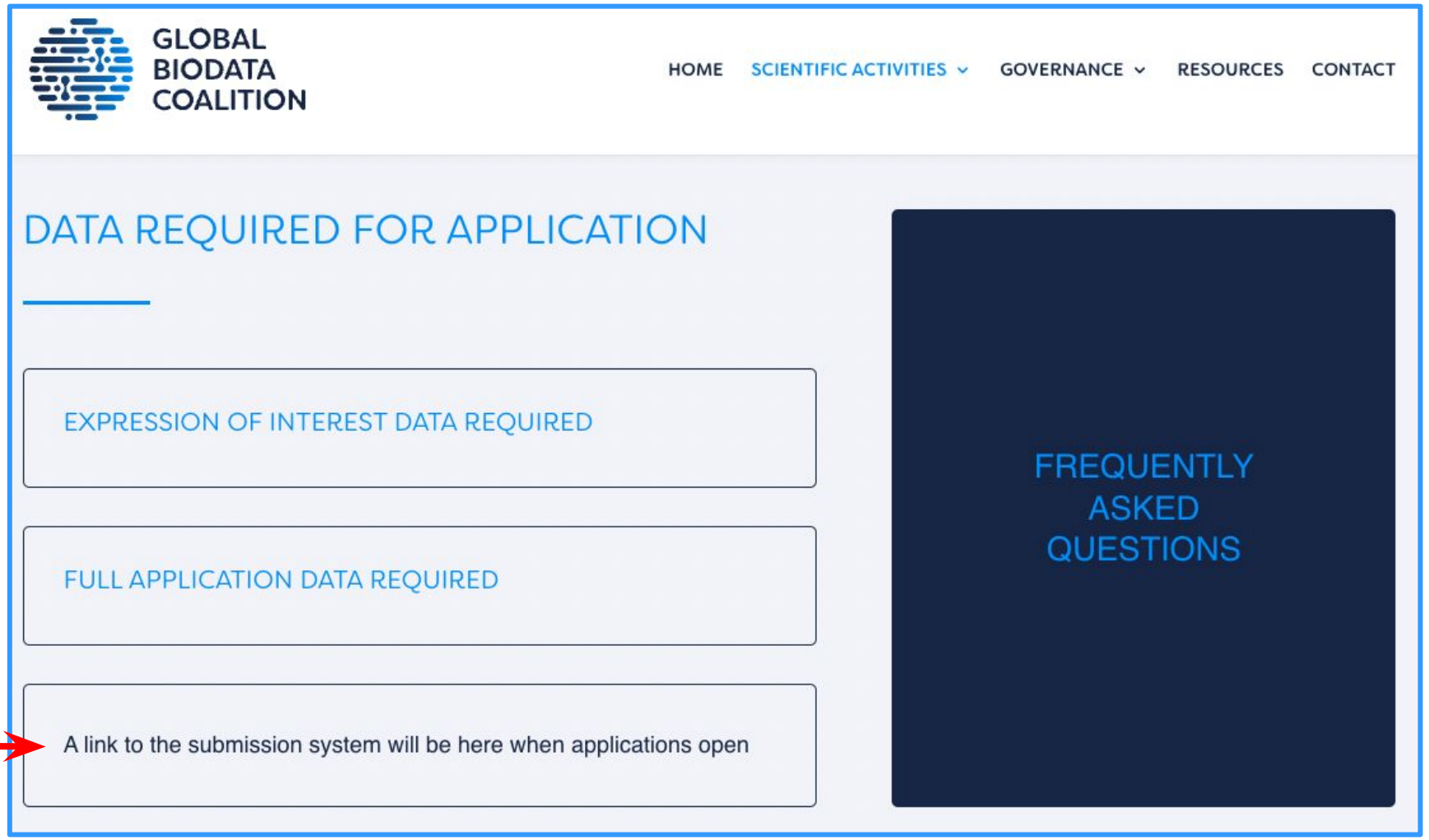
SA1: How does the resource serve as a fundamental resource across scientific approaches and disciplines, rather than for example serving a specific project or field of research?
SA2: Approximately how many users access the data resource per year? Please explain the basis for the count given, and include a statement addressing the geographical distribution of the users.
SA3: What would be the impact on other data resources and/or on the biodata resources ecosystem were this data resource to be withdrawn from service?
SA4: How does the data resource support users? Examples might include a help desk, user training or workshops or social media Q&A. What opportunities for gathering user feedback do you provide?
SA5: What management structure and governance measures are in place for the data resource? Is there a dedicated resource manager? How frequently does the oversight committee meet? What is the relationship between the data resource and the host institution?

GLOBAL BIODATA COALITION

How do I submit an Expression of Interest?

Interval for Applications is April 11th until May 8th

Submission system entry point will be **HERE**



https://globalbiodata.org/scientific-activities/global-core-biodata-resources/gcbr-selection-2023/

GLOBAL BIODATA COALITION

Review process

Operationally similar to that used by funding agencies in selecting proposals for research grant funding.

Reviewers:

- The GBC Scientific Advisory Committee
- Experts in biodata resource management
- Reviewers do not include the GBC Secretariat

Reviewers will sign:

- Confidentiality statement
- Conflict of Interest declaration

GLOBAL BIODATA COALITION

Expression of Interest Timeline

Date	Action
13th February 2023	News announcement for the 2023 round of Global Core Biodata Resource selection published here. FAQ for potential applicants is available here.
22nd-23rd February 2023	Outreach - Webinars with Q&A session for potential applicants
11th April 2023	Submission of expressions of interest opens
8th May 2023 23:59 UTC	Deadline for submission of expressions of interest
Week of 19th-23rd June 2023	Expression of interest review committee meets to determine which data resources will be invited to proceed to full application
3rd July 2023	Notification of outcome of expression of interest to submitters

Timeline: https://globalbiodata.org/scientific-activities/gcbr-selection/

GLOBAL BIODATA COALITION

Full Application Timeline

Date	Action
28th Aug-1st Sep 2023 To be confirmed*	Webinar for those data resources invited to submit a formal full application
4th Sep-1st Oct 2023 To be confirmed*	Submission window for full applications
30th Oct-13th Nov 2023 To be confirmed*	Reporting of reviewer scores and comments to submitters: followed by a 2 week opportunity to respond
20th-24th Nov 2023 To be confirmed*	Full application review committee meets to generate recommendations for selection
Q4 Board meeting: 4th–13th Dec 2023 To be confirmed*	GBC Board approves decision for inclusion in list of GCBRs ● Applicants formally notified of outcome ● Announcement of the amended GCBR list - GBC web site, news release…

Timeline: https://globalbiodata.org/scientific-activities/gcbr-selection/

GLOBAL BIODATA COALITION

References

Global Core Biodata Resources: Concept and Selection Process
https://doi.org/10.5281/zenodo.7468719

Global Core Biodata Resource Selection: Data Required for Expression of Interest
https://doi.org/10.5281/zenodo.5846742

Global Core Biodata Resource Selection: Data Required for Full Application
https://doi.org/10.5281/zenodo.5846758

GCBR 2023 selection web page
https://globalbiodata.org/scientific-activities/global-core-biodata-resources/gcbr-selection-2023/

GCBR 2023 selection process FAQ web page
https://globalbiodata.org/scientific-activities/global-core-biodata-resources/gcbr-selection-2023/faq-2023/

These slides will be deposited in the GBC Zenodo channel when webinars are concluded
https://zenodo.org/communities/gbc/

GLOBAL BIODATA COALITION

Further Reading:
the 2022 round of GCBR Selection

GLOBAL BIODATA COALITION ANNOUNCES THE FIRST SET OF GLOBAL CORE BIODATA RESOURCES

https://globalbiodata.org/global-biodata-coalition-announces-the-first-set-of-global-core-biodata-resources/

EXECUTIVE DIRECTOR'S BLOG: ON THE GCBR ANNOUNCEMENT

https://globalbiodata.org/executive-directors-blog-on-the-gcbr-announcement/

Summary of 2022 GCBR selection process and outcomes - slides

https://globalbiodata.org/wp-content/uploads/The-GCBRs-a-summary-selection-process-2022.pdf

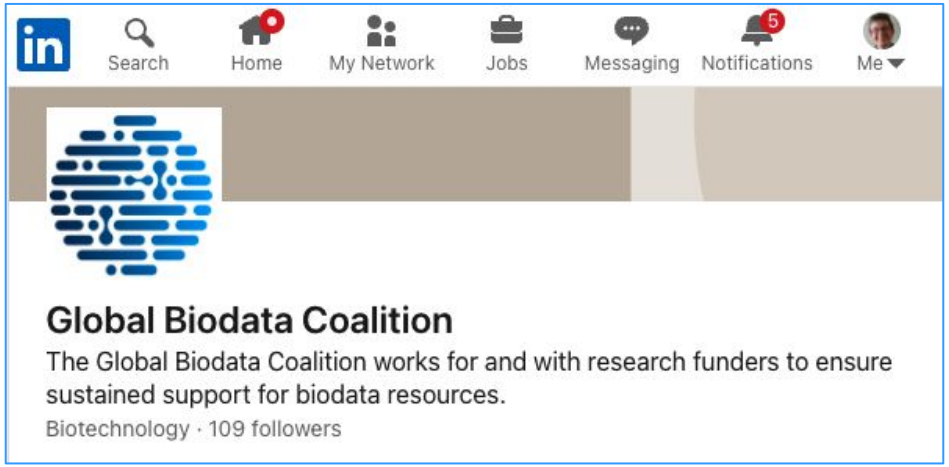
GLOBAL BIODATA COALITION

Contact

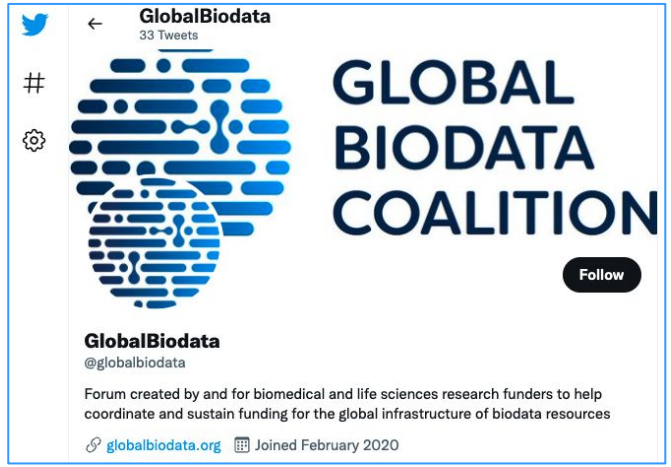
For Questions and Clarifications about GCBR selection process:
Email: gcbr-selection@globalbiodata.org

For general communications with the GBC:
Email: info@globalbiodata.org

https://www.linkedin.com/company/global-biodata-coalition/



https://twitter.com/globalbiodata



GLOBAL
BIODATA
COALITION