TWENTY-YEAR REVIEW OF GBIF

COMMITTEE ON DATA

CODATA

INTERNATIONAL
SCIENCE COUNCIL

GBIF | Global Biodiversity
Information Facility

EXECUTIVE SUMMARY

GBIF | Global Biodiversity Information Facility

This independent report was requested from CODATA by the GBIF Governing Board and leadership, who comprise the primary intended audience. Government decision-makers and others in the biodiversity community may want to read Chapter 9 of the FULL REPORT, which contains the full conclusions and recommendations, and the body of the report for the supporting evidence and rationale. The report is also intended to provide the GBIF team leaders and data Node managers with a neutral view of GBIF by external experts.

TABLE OF CONTENTS

The Twenty-Year Review of GBIF comprises two sections. The EXECUTIVE SUMMARY (this section) and the FULL REPORT.

Ⓘ INTRODUCTION

The past two decades of the new century have brought into sharp relief many global trends, both positive and negative, that put the current review of the Global Biodiversity Information Facility (GBIF) in a larger context and inform action in the coming years. Societal challenges provide important reasons for bold action. Our planet is experiencing environmental degradation on a massive scale—whether it is climate warming (IPCC 2018), the impacts on our lands and oceans[1], or the immense loss of biodiversity. The latter, which many informed observers have referred to as the beginning of the 6th Great Extinction (Kolbert 2014), is the immediate context for GBIF and this review. Although we are aware of these huge negative trends, there is much that we do not know about biodiversity in many places on the planet and thus are not well equipped to confront the problems comprehensively, even if we mustered the requisite will to do so. And, of course, there are many other forces that work against our ability to respond positively to these crises and that need to be challenged by broadly available factual information. Good information, responsibly used, can be transforming and lead to appropriate decision-making.

These societal challenges provide a compelling rationale and urgent need for data mobilization and collection to support greater action by global information activities such as GBIF. We believe it is the positive trends, however, that increasingly provide the means to do the biodiversity data work successfully. We therefore focus here principally on the scientific and technical developments, as well as the social context in which they occurred, over the past twenty years.

From a scientific standpoint, there has been an explosion of data and information, and a concomitant paradigm shift to data-driven research and specifically biodiversity research and its myriad applications. Moreover, there are many new and "non-traditional" sources of information, such as meta-barcoding and remote sensing technologies, that are being integrated in this changing paradigm. Novel and redesigned scientific data organizations, including those supporting citizen science and integrating indigenous knowledge, are developing approaches designed specifically to take advantage of the unprecedented data opportunities in a cooperative framework. And innovative open data policies and structures are proliferating.

The inexorable progress of information technologies makes this scientific data revolution possible. As a biodiversity information facility and global infrastructure expressly set up as a focal point and institution to mediate such data, GBIF is uniquely positioned to take advantage of these opportunities and confront the challenges. This 20-year review and resulting report are intended to help GBIF and its large network make that happen.

---

[1] https://www.ipcc.ch/site/assets/uploads/sites/3/2019/09/SROCC_SPM_HeadlineStatements.pdf

❚❚ REVIEW METHODOLOGY

In 1999, the Organization for Economic Cooperation and Development's (OECD) Working Group on Biological Informatics of the Megascience Forum (now the Global Science Forum) recommended the establishment of an organization such as GBIF. The OECD report concluded that "An international mechanism is needed to make biodiversity data and information accessible worldwide" (OECD 1999).

GBIF was created to be both an intergovernmental organization governed by a nonbinding Memorandum of Understanding and an international network and research infrastructure. Funded by 40 governments as of January 2020[1], with supplementary support given by other organizations, it is aimed at providing anyone, anywhere, open access to data about all types of life on Earth. Details about the structure, organization, members, work, and the data that GBIF mediates may be found on its website at: www.gbif.org and in the body of the review report that follows.

The OECD report also expected strong cooperation on a worldwide basis, with a nimble GBIF Secretariat in a coordinating role. However, present structures of biodiversity information systems – which we try to capture broadly in Chapter 2 on the landscape – are still strongly influenced by nationally or topically focused funding and organizational decisions of the past. A new approach to such thematic coordination, the Alliance for Biodiversity Knowledge[2] ("the Alliance"), which originated from a gathering of major players in this landscape in 2018, deserves close observation.

GBIF has now existed for almost 20 years, so it is a good time to examine the past and to look forward. Therefore, at the suggestion of the GBIF Secretariat, the governance bodies of GBIF have asked CODATA to review the organization's activities and accomplishments.

The statement of task given by the GBIF Governing Board to the CODATA review team specified:

1. Review how effective GBIF has been since 2001 in meeting the expectations from the OECD working group

2. Review the governance and sustainability of GBIF as a global network and organization (including hosting of the Secretariat[3] in Denmark)

3. Review the place of GBIF within the 2018 landscape of biodiversity and research organisations

4. Review the technical aspects of GBIF's delivery and its sustainability and trustworthiness (in particular, to researchers) as a research infrastructure

5. Consider the challenges in the next 5-10 years that GBIF needs to be prepared to meet

6. Provide recommendations on areas needing attention and improvement.

The body of the report comprises a mix of descriptive, analytical, and advisory approaches in response to this task statement. Fact finding was performed through desk research, in-depth interviews with key members of the GBIF Secretariat and Committees, and interviews with a global selection of 108 experts who focused on GBIF's strengths, weaknesses, opportunities, and threats (SWOT).

The Conclusion sections in Chapters 3 – 8 lead to the advisory portion of the report: the summary of the conclusions and the resulting recommendations are proposed in Chapter 9.

---

[1] The up-to-date Participation count can be tracked at https://www.gbif.org/the-gbif-network

[2] https://www.biodiversityinformatics.org/

[3] Some terms, such as "Secretariat", are capitalized when referring to GBIF-specific functions or agents, as further explained in section 1.6.

III SUMMARY OF CONCLUSIONS

SUCCESSES

First of all, our findings show that GBIF is the most comprehensive, openly available, application-agnostic (most unbiased), easiest-to-use, and modern access point to known digital species occurrence data. Consequently, as a global, distributed platform, GBIF is viewed widely as being a major success and a great improvement over alternative solutions; this relates equally to the data it provides and to its capacity building activities. There is now a broadening range of applications of GBIF-mediated data that matches the OECD expectations of 1999. This success would not have been possible without GBIF's foundational principles of functioning as a distributed organization and its pursuit of a dependable funding structure. Its 20-year existence alone – and the way it is funded – creates much of the trust on which most of the current, extremely high expectations depend. The organization has built and maintains relations with a large and very heterogeneous array of other actors in the fields of biodiversity informatics and conservation, and at the science-policy interface. It has established itself variously as a member, partner, and coordinator of biodiversity-related activities.

The successes also would not have been possible without the much-lauded staff at the Secretariat, which was uniformly praised as competent, dedicated, responsive, and enthusiastic. We found the Secretariat and GBIF as a whole to be an agile, learning, and evolving organization. Few of the SWOT comments will come as a surprise to the Secretariat. In particular, its leadership, which is fully aware at least of the most important issues raised, has actually addressed many of them in the current strategy and work program.

The governance of GBIF – its bodies and rules – has contributed to the essential agility and evolution of GBIF. While it is broadly inclusive of Participants and Affiliates, it has a core, consisting of the Executive Committee and the standing committees supporting it, which ensures that GBIF remains adaptable and flexible.
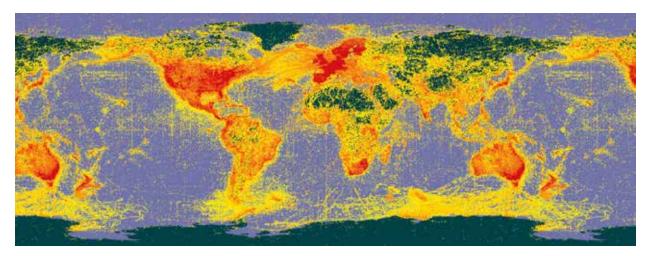
These broad measures and perception of success are built on a range of abilities and systemic features developed over the last 20 years, such as implementing its Open Access data policy, building a geographically inclusive organization, tracking the use of data, enabling Nodes and data Publishers to do common work, developing excellent technology and systems, and being trusted as a neutral broker within the community.

Most experts interviewed appeared to see GBIF as able to address or contribute to the solution to the challenges identified, even the challenges that are beyond GBIF's current remit.

CHALLENGES

Nonetheless, our analysis also indicates that major challenges lie ahead and that proceeding as before may not be an option for the continued success of GBIF. Some of the challenges are fundamental and will certainly need a decade to be resolved. Other challenges are practical in nature and can be realistically addressed in the short term, i.e., within one to three years at the current core budget levels. The overarching challenges that have been distilled from the organization's weaknesses, many opportunities, and some threats, are described below. The Conclusion sections of chapters 3 to 8 contain much more detail.

In the data and technologies domain, two major issues loom large. First, the fitness for purpose of current GBIF-mediated data needs to be made even more transparent and further improved. Transparency relates to indications of accuracy and best practices in the usage of data. Improvement depends strongly on improved and more effective and efficient quality assurance methods, in particular, employing and engaging user and community feedbacks, which may be addressable in the short term, and on closing taxonomic and geographic gaps, which is a longer-term task, mostly outside GBIF's direct control.

Species occurrences, www.gbif.org

The second major issue is the expected rise in the quantity and importance of non-traditional biodiversity data, such as from remote sensing, genomic analysis methods, other automated observations, and contributions by citizens. Some of these developments appear to accelerate dramatically through the recent application of machine learning technologies and the pervasive availability of networks of people and smart devices. Here, the question cannot be whether, but rather how and how deeply GBIF should be engaged. Many of the interviewees and the Secretariat consider any of a number of engagements necessary for GBIF to remain relevant, while other interviewees warn about "scope creep" or quality issues with specific data sources. Consequently, GBIF will need to make some well-considered choices when further increasing its data holdings, which has been a persistent and crucial goal since GBIF's inception.

Other major issues to be addressed are **in the domain of GBIF's organization and community**. The weakness of biodiversity data gaps is not specific to GBIF, but it can, in GBIF's context, partially be attributed to the first major issue: gaps in formal Participation from biodiversity-rich countries and, particularly, from Asia as a whole. But the identified strengths as well as opportunities indicate that GBIF is structurally better positioned than most to close this gap "on the ground": GBIF has been rather successful in Latin America and newly in Africa, the latter mainly based on a specific capacity-enhancement program, Biodiversity Information for Development (BID). Here in particular the principle of regional networks of Nodes has demonstrated opportunities, but also limits, which are mainly due to the shortfalls in funding (see below). In Asia, there is no disagreement in principle on the strengths of GBIF, but the expectation that a much better communication of a cost-benefit analysis would help build formal Participation.

The second major issue in this domain is that even the concept of national Nodes (or national Biodiversity Information Facilities) is limited. There are 20 national Nodes in Europe, serving it well. But there is general skepticism about the Nodes concept in the comparably sized US, with just one Node serving it. A number of well-funded national Nodes show that the GBIF concept can be implemented. However, in most countries the Nodes are not staffed to achieve full functionality. This can be compensated to some extent by a strong hosting institution, but institutions in most countries outside Europe lack such resources. In Asia, a fundamental principle of GBIF funding currently precludes a dense Nodes network to be built at all, as Nodes are tied to formal Participation. In all GBIF regions except in North America and western Europe, the language barrier appears to be a real impediment to capacity building and also in the ability of GBIF's Publishers to contribute data.

The major challenges in the previous domains inevitably lead to a consideration of **the domain of funding**. Here the first major issue is that the current level of GBIF funding – including the national funding of Nodes and supplementary funds – is barely sufficient to perform its activities at the current scope of its strategy and work programs. An increase in formal Participation may help, but it definitely would not be sufficient to undertake big strategic moves, if they are determined to be necessary in the long term. Still, the mostly hidden cost of adequately staffing Nodes may need the most attention. National funding for additional staffing is, and may remain, unavailable in most countries.

Due to developments in the political realm, some observers expect a second major funding issue that GBIF may face over the coming years will be increased instability of core funding[1]. However, GBIF is prepared to absorb major funding fluctuations. Some major successes have been achieved using supplementary funding, but GBIF's governance functions are largely adapted to managing these as moderate contributions to its existing work programs. Consequently, substantial third-party funding, for example by philanthropic organizations or industry, cannot bolster the core of GBIF's work program without major changes in governance and possibly strategy.

The final major challenge lies in the **management of expectations**[2]. The weaknesses, opportunities, and threats brought forward in interviews reflect the expectations of GBIF's stakeholders. The fact that its community assumes that GBIF could meet great expectations, is, above all, a very good sign about its reputation, but it reveals in many cases fundamental misconceptions about GBIF's resources or the effort required. Consequently, while addressing those challenges that are in fact within its remit and its limits of resources, GBIF will urgently need to manage expectations about what it can achieve and when. If the organization allows overoptimistic expectations to stand, it might experience an unnecessary decrease in its reputation.

---

[1]   Throughout the report, "core funding" is the GBIF term for Voting Participants' (long-term) contributions.

[2]   "management of expectations" is a technical term used in software development with an aim to reduce unrealistic expectations of costumer or users about what can be achieved and when it can be achieved.

Ⅳ SUMMARY OF RECOMMENDATIONS

The high expectations on GBIF as well as the technical and scientific challenges ahead strongly suggest that, while GBIF is well positioned to meet them, a scenario of "business as usual" may fail to preserve GBIF's relevance. Strong growth should be considered seriously, with implications for GBIF's technology, services, organization, and funding.

Our specific recommendations are generally ordered in a sequence that reflects a plausible phasing of related activities. Thus, we present our recommendations in three sub-sections. The first one sets forth generic guidelines mainly formulated to maintain a focus on what we have identified as the main factors in GBIF's success and reputation. Specific short-term recommendations in the second section aim to encourage progress on the most urgent challenges, which could be achieved at the current funding level and structure, and without revising the organization's strategy. The third section provides recommendations related to all of the challenges, but which probably can only be approached by a longer-term activity or by a substantial expansion of funding in the next two funding cycles (of five years, each). That third set of recommendations thus might require an explicit expression in strategy, organization, or even governance.

The recommendations that follow are cross-referenced to pages in the FULL REPORT.

GENERAL RECOMMENDATIONS

In order to maintain and strengthen GBIF's relevance and standing, we recommend the consideration of growth paths in a number of dimensions. Such growth will enable GBIF to support biodiversity research broadly in the future. To do so, the organization will probably also need to participate in and contribute to the rapid development of its scientific data methods and standards.

R1a, Data Quality and Quantity (p.121): GBIF needs to build on and maintain its reputation as the most comprehensive source of openly available global occurrence data. This means that it should continue the trajectory of growth of the data that it mediates, in quantity as in quality, by extending and deepening relations within its network, and by supporting non-traditional types of biodiversity data.

R1b, Technology and standards (p.121): GBIF needs to maintain or attain leadership in essential technological and standardization areas related to biodiversity informatics. In order to do so, it should continue to work actively with other stakeholders in the "landscape", such as the researchers at the forefront of such fields as metagenomics, remote sensing, and observation and cloud technologies, to keep abreast of developing data sources, standards, and technologies. GBIF should add a long-term focus on IT security.

R2, Networking (p.121): GBIF should maintain and even strengthen its capabilities to network its stakeholders and to lead them to cooperate and build consensus. GBIF, and particularly the Secretariat, should continue to be seen as a neutral broker.

SHORT-TERM RECOMMENDATIONS

Most major challenges are not resolvable by GBIF on its own or by simple measures. Therefore, many of those can only be solved in the long term. Still, it is necessary to address them even in the short-term, if only to reassure stakeholders of GBIF's awareness and of progress that the organization is making. GBIF should demonstrate that it is working on them or even provide what could be seen as short-term workarounds.

Many of our short-term recommendations are related to improved communication and GBIF's visibility and to managing stakeholders' expectations. Most of this could and should be done in the next two to three years:

R3, Visibility (p.122): Establish greater visibility among decision-makers and the policy community in the biodiversity arena, as well as throughout the scientific community.

R4a, Value Proposition (p.122): Develop compelling value propositions, especially as a tool for GBIF champions and the Nodes.

R4b, Grand Challenge (p.122): Choose a "grand" but doable real-world challenge (i.e., low-hanging fruit). In this regard, GBIF should identify associated data gaps and focus community effort on filling those gaps, work with scientists to perform the analysis, and actively disseminate how the collections, countries, and GBIF are coordinated to solve or contribute to the solution of such a significant real-world problem.

R5, Eastern Participation (p.122): Build Participation in the East. This is a long-term activity, but flexibility and creativity should be employed through short-term measures, which would show progress and build good will for full Participation in GBIF, particularly in Asia.

R6, Strengthen Nodes (p.123): Make the full Nodes concept work in all regions. Each GBIF region has different needs and issues – therefore, each needs its own engagement strategy.

R7, Communicate Successes (p.123): Discuss with stakeholders and show prominently and appealingly what has and can be done with GBIF-mediated data in a scientifically sound way. Focus on high-value products (see R4).

R8, Improve Data Fitness (p.123): Develop criteria and metrics to measure data quality and demonstrate improvements in the short term.

R9, Publish Methods and Applications (p.124): Continue exploring and publish information on how each new type of biodiversity data is relevant in the context of "traditional" data and vice versa, and explain the problems and compromises involved.

R10a, Manage Expectations (p.124): In managing expectations about what GBIF can or cannot deliver, the organization's leaders should do more to communicate its mission, the magnitude of its funding, as well as scalability issues in general.

R10b, Leverage the Network (p.124): GBIF should identify in a strategic planning process where it could leverage its network and could succeed in brokering support for urgently requested activities.

To be able to execute some or all of the medium-term to long-term recommendations below, we recommend the following action in the short-term:

R11, Urgent Deliberations and Decisions of the Governing Board (p.124): The Governing Board needs to discuss and establish guidelines regarding the role of national Nodes within the GBIF structure, the staffing of Nodes, the (co-)funding of GBIF work programs by intergovernmental or governmental third parties and by philanthropic organizations or industry, and explore how core funding from Participants could be increased.

LONGER-TERM RECOMMENDATIONS

In the following, we use "medium term" to mean no longer than the next 5-year GBIF funding interval and "long term" to mean up to 10 or more years. Note that some of the recommendations, particularly R15 to R20, may need significant additional funding.

Because we see major challenges and consequential decisions ahead for the organization's leadership we recommend a strengthening of observation, advice, and support with views provided both from the outside and the inside on a regular basis:

R12, Create Advisory Board (p.125): The Executive Committee should create a small, but inclusive, high-level advisory body consisting of external experts to advise the Governing Board on strategies, particularly on matters of fundamental importance to the mission, organizational principles, governance, and funding of GBIF.

R13, Foresight and Monitoring (p.125): The Executive Committee should, with the support of the Secretariat and its standing committees, develop an explicit, strategic foresight and monitoring process, including landscape analysis updates and estimates of the penetration of the potential user base, with the aim to adjust priorities within the work programs and suggest the timely adaptation of strategic approaches and directions to the Governing Board.

There are and certainly will continue to be many issues to resolve. Currently, the most promising approach appears to be through the emerging Alliance of Biodiversity Knowledge (the Alliance). We therefore recommend that:

R14, Leadership (p.126): GBIF should assume, as requested by major voices in the community, leadership of the Alliance in order to create consensus and collaboration at a much broader scale.

We offer the following three specific recommendations to be implemented by GBIF in the medium term, informed by previous explorations and, as far as possible, by outputs of the work of the Alliance, but not waiting for its conclusion:

R15, Feedback and Quality Assurance (p.126): GBIF should speed up the work on feedback and quality assurance, and their implementation, so that major results are shown in the medium term.

R16, Full-Service Portal (p.127): Prototype and demonstrate a concept for a "full service" portal, which would provide additional types of data and services.

R17, Long-Term Strategy (p.127): Formulate strategic plans for the long term, underpinned by cost estimates. These need to resolve whether to host data other than occurrences, how to link them, whether and how to perform standard analytics or to offer a platform to execute user analytics, and how to host GBIF services in the future.

In order to address the full range of challenges, we recommend that:

R18, Asian Participation (p.127): To establish a permanent foothold in Asia, GBIF should consider a branch office in an Asian country which is, like Denmark, small but wealthy enough to support hosting, politically (relatively) neutral, multilingual, and safe.

R19, Staffing Increase (p.128): Staffing at the Secretariat in support of the various biodiversity communities' engagement should be doubled in the medium term.

R20, Organization and Funding of Nodes (p.128): If success in implementing fully functional Nodes in most, if not all, of the Participant countries or developing their sustainable funding cannot be achieved in the medium term, a major adaptation of either the funding of Nodes or the organizational concept of GBIF should be considered by the Governing Board, supported by its newly established high-level advisory board.

R21a, Funding Guidelines and Ability to Execute (p.128): GBIF should be prepared with guidelines from the Governing Board to fund new strategic initiatives and be organizationally able to absorb the workload or distribute it to reliable partners.

R21b, Increased Core Funding (p.128): The Governing Board should work to increase core funding, perhaps doubling it in the medium term (the next funding period of 5 years). For the longer term – the following 5 years – we confirm the recommendation of a previous review (CODATA 2005) to arrive at 12 million EUR per year, in 2019 Euros (which would constitute another doubling).

If this increased level of funding cannot be achieved by the collection of reasonable and fair dues from national Participants, our last two recommendations are about seeking other funding sources:

R21c, Core Funding Sources (p.128): The Governing Board should consider whether to invite third parties (non-Participants) to contribute to the funding of the core program. Such funding should involve predominantly longer-term commitments, i.e., for a minimum of 5 years.

R21d, Node Funding Sources (p.128): The long-term funding of many Nodes will very likely need to be supplemented by third parties. Both this and eventual third-party contributions to the core funding should be strictly monitored, however, to exclude undue influence on Nodes or on GBIF as a whole.

In summary, we see the organization as being able to rise to the challenges and today's expectations. To succeed, it must be enabled – financially, structurally, and with good leadership – to continue to perform the requisite exploratory work and implement the necessary steps in a timely manner.

We hope to have provided some recommendations that will enable GBIF to sustain its high standing in the biodiversity informatics landscape and to further continue to improve its support of biodiversity research and societal welfare. These recommendations should help GBIF face the challenges of as yet unforeseeable demands from the science-policy interface, or applications in the health, agricultural, or commercial domains. Such developing expectations require the ability to adapt strategies and technologies flexibly and quickly, particularly as methods of systematic biodiversity data acquisition and scientific analysis may undergo some radical changes in the near future, and the pace of technological development for informatics and biotechnology accelerates.

CODATA

The Committee on Data of the
International Science Council

Connect with us at:

www.codata.org
simon@codata.org

International Science Council
5 rue Auguste Vacquerie
75016 Paris, France

🐦 www.twitter.com/CODATANews

f www.facebook.com/codata.org

Cover image: Species occurrences,
www.gbif.org