Toward the large biodiversity collider: the *alliance for biodiversity knowledge*

Kyle Copas | GBIF Secretariat

Empowering Biodiversity Research 2022 | Tervuren, Belgium—24 May 2022

alliance for biodiversity knowledge

Strawberry spiderhead (*Serruria aemula*). Photo 2015 Rolf Theodor Borlinghaus via iNaturalist (CC BY-NC 4.0). https://www.gbif.org/occurrence/1838411601

BACK TO THE FUTURE (1999)
REPORT OF OECD MEGASCIENCE FORUM BIODIVERSITY INFORMATICS SUBGROUP

The Subgroup on Biodiversity Informatics recommends that the governments of OECD countries establish and support **a distributed system of interlinked and interoperable modules (databases, software and networking tools, search engines, analytical algorithms, etc.)** that together will form a Global Biodiversity Information Facility (GBIF).

BACK TO THE FUTURE (1999)
REPORT OF OECD MEGASCIENCE FORUM BIODIVERSITY INFORMATICS SUBGROUP

The Subgroup on Biodiversity Informatics recommends that the governments of OECD countries establish and support a distributed system of interlinked and interoperable modules (databases, software and networking tools, search engines, analytical algorithms, etc.) that together will form a Global Biodiversity Information Facility (GBIF).

This facility will **enable users to navigate and put to use vast quantities of biodiversity information**, thereby

BACK TO THE FUTURE (1999)
REPORT OF OECD MEGASCIENCE FORUM BIODIVERSITY INFORMATICS SUBGROUP

The Subgroup on Biodiversity Informatics recommends that the governments of OECD countries establish and support a distributed system of interlinked and interoperable modules (databases, software and networking tools, search engines, analytical algorithms, etc.) that together will form a Global Biodiversity Information Facility (GBIF).

This facility will enable users to navigate and put to use vast quantities of biodiversity information, thereby

- advancing scientific research in areas such as **agriculture, biomedicine, biotechnology, environmental management, pest control, health, education, and conservation, among others**;

BACK TO THE FUTURE (1999)
REPORT OF OECD MEGASCIENCE FORUM BIODIVERSITY INFORMATICS SUBGROUP

The Subgroup on Biodiversity Informatics recommends that the governments of OECD countries establish and support a distributed system of interlinked and interoperable modules (databases, software and networking tools, search engines, analytical algorithms, etc.) that together will form a Global Biodiversity Information Facility (GBIF).

This facility will enable users to navigate and put to use vast quantities of biodiversity information, thereby

- advancing scientific research in areas such as agriculture, biomedicine, biotechnology, environmental management, pest control, health, education, and conservation, among others;
- serving the economic and quality-of-life **interests of society**; and

BACK TO THE FUTURE (1999)
REPORT OF OECD MEGASCIENCE FORUM BIODIVERSITY INFORMATICS SUBGROUP

The Subgroup on Biodiversity Informatics recommends that the governments of OECD countries establish and support a distributed system of interlinked and interoperable modules (databases, software and networking tools, search engines, analytical algorithms, etc.) that together will form a Global Biodiversity Information Facility (GBIF).

This facility will enable users to navigate and put to use vast quantities of biodiversity information, thereby

- advancing scientific research in areas such as agriculture, biomedicine, biotechnology, environmental management, pest control, health, education, and conservation, among others;
- serving the economic and quality-of-life interests of society; and
- providing a basis from which our knowledge of the natural world can grow rapidly and in a manner that **avoids duplication of effort and expenditure**.


It works!

BACK TO THE FUTURE: "A GLOBAL BIODIVERSITY INFORMATICS FACILITY"

"GBIF, which is intended to go online in three to four years, will contain **scores of databases** — including geospatial, chemical, molecular and genetic collections — plus a **catalogue of names** of known organisms, **digitization** of natural history data, **literature resources**, and **a bank dedicated to the discovery of new species**. It will also include **training and outreach** programmes to help scientists to use it."

Vast database offers vision of biodiversity

[PARIS] The Organization for Economic Cooperation and Development (OECD) last week approved plans aimed at creating the world's largest biodiversity databank. The Global Biodiversity Information Facility (GBIF) will be launched later this year.

Science ministers from the OECD's 29 member states also decided to redirect its Megascience Forum, a group set up seven years ago to coordinate large scientific projects among member countries. Now called the Global Science Forum, it will emphasize international cooperation to develop global science infrastructures of any dimension.

GBIF is to be launched by an interim steering committee with members from ten countries, with a permanent secretariat to be formed by mid-2000. It will knit together existing databases on biodiversity, to serve as a one-stop information resource. The multi-million dollar project will largely be funded by existing national programmes, such as the effort by the US National Science Foundation (NSF) to digitize natural history data.

Some databases that will contribute to GBIF are already running on websites, including Species 2000, a UK project aimed at indexing all species names. One of GBIF's most important tasks will be to create a comprehensive list of species names, solving problems caused by double references to organisms or misnamed species.

"Sometimes an inappropriate name is used and we find out we've been protecting the wrong organism," says James Edwards, deputy assistant director of the NSF's directorate for biological sciences and chairman

Giving science a hand, clockwise from top left: ministers from Japan, Norway, Russia, South Africa, Israel, Germany, Ireland and Iceland.

of the Megascience Forum bioinformation working group. "Or two countries use different names, so they can't communicate about an endangered species."

GBIF, which is intended to go online in three to four years, will contain scores of databases — including geospatial, chemical, molecular and genetic collections — plus a catalogue of names of known organisms, digitization of natural history data, literature resources, and a bank dedicated to the discovery of new species. It will also include training and outreach programmes to help scientists to use it. Specialists say that GBIF will change the study of biodiversity.

"Because of distance, it is humanly impossible to get a real vision of biodiversity. A few years down the road we will think 'How could we talk about biodiversity when we couldn't even see it?'," says Frank Bisby, professor of botany at the University of Reading in Britain and chairman of Species 2000.

The idea for GBIF originated three years ago in the Megascience Forum. The new Global Science Forum will tackle such issues as the rivalry between the US Superconducting Collider and Europe's proposed Large Hadron Collider. But it will also consider smaller projects, in the hope of bringing policy-makers and scientists together to resolve international issues.

"We thought, 'Why not extend the mandate so that virtually every project in science can be considered?'," says Michael Oborne, deputy director of the OECD's science, technology and industry directorate. "This will be a significant tool for the scientific community worldwide."

Particle physics and next-generation accelerators could emerge as a hot topic, says Oborne, as could nuclear waste disposal, which the Megascience Forum looked at without drawing up any solutions.

Last week's meeting also decided to form a year-long task force on radioastronomy to sort out conflicts between telecommunications companies and radioastronomers over sharing wavebands (see *Nature* 399, 513; 1999). Radioastronomers have been fighting to prevent mobile phones polluting their designated wavebands. **Heather McCabe**

McCabe H (1999) Vast database offers vision of biodiversity. Nature 400 (5). https://doi.org/10.1038/21728

“One purpose of the GBIF would be to promote the development of needed capacities so that goals of multiple organizations can be achieved without duplication of effort.”

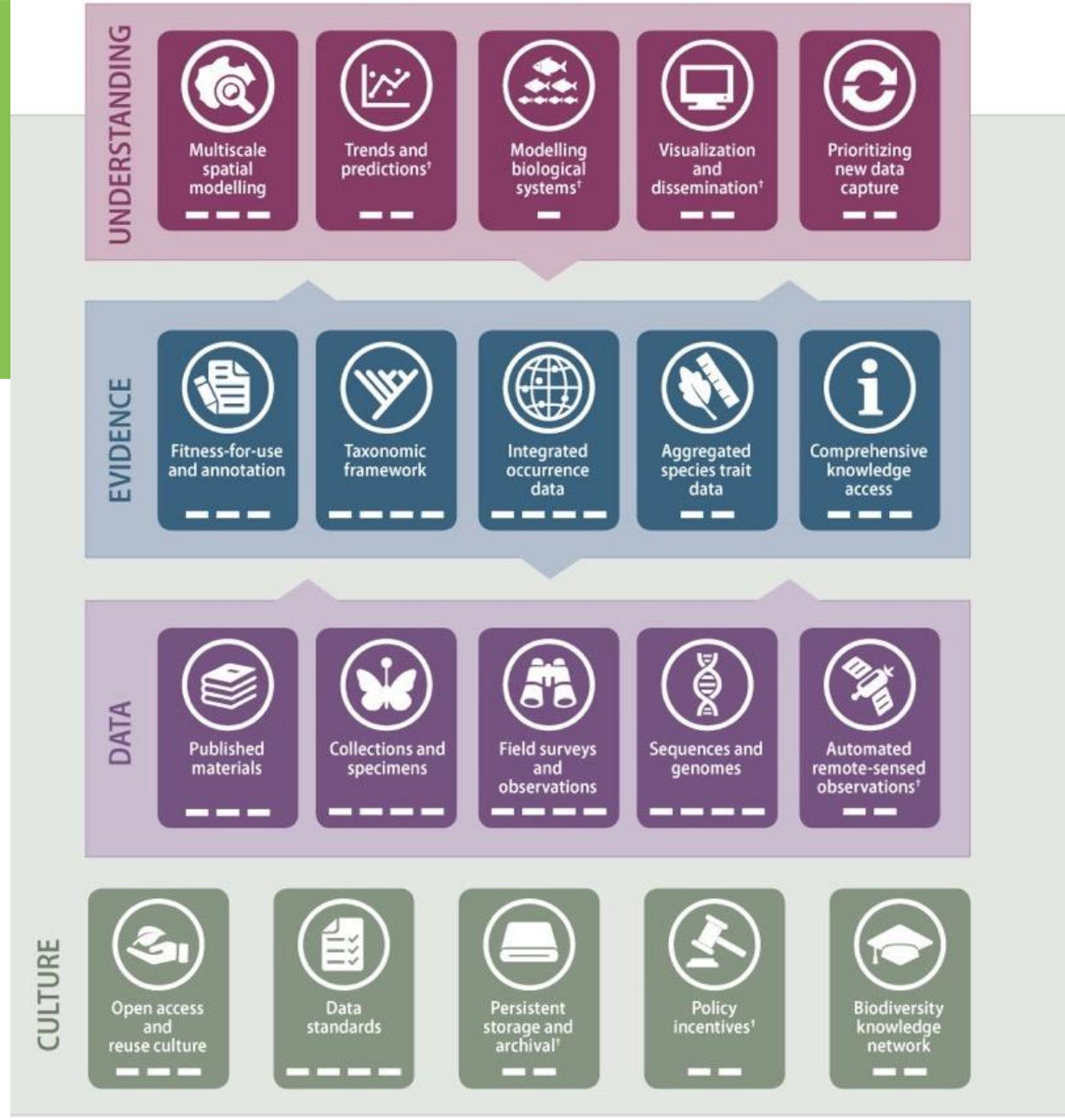
—Report of OECD Megascience Forum Biodiversity Informatics Subgroup (1999)


GBIC 2012
Global
Biodiversity
Informatics
Conference
2-4 July, 2012
University of Copenhagen

GLOBAL BIODIVERSITY INFORMATICS CONFERENCE

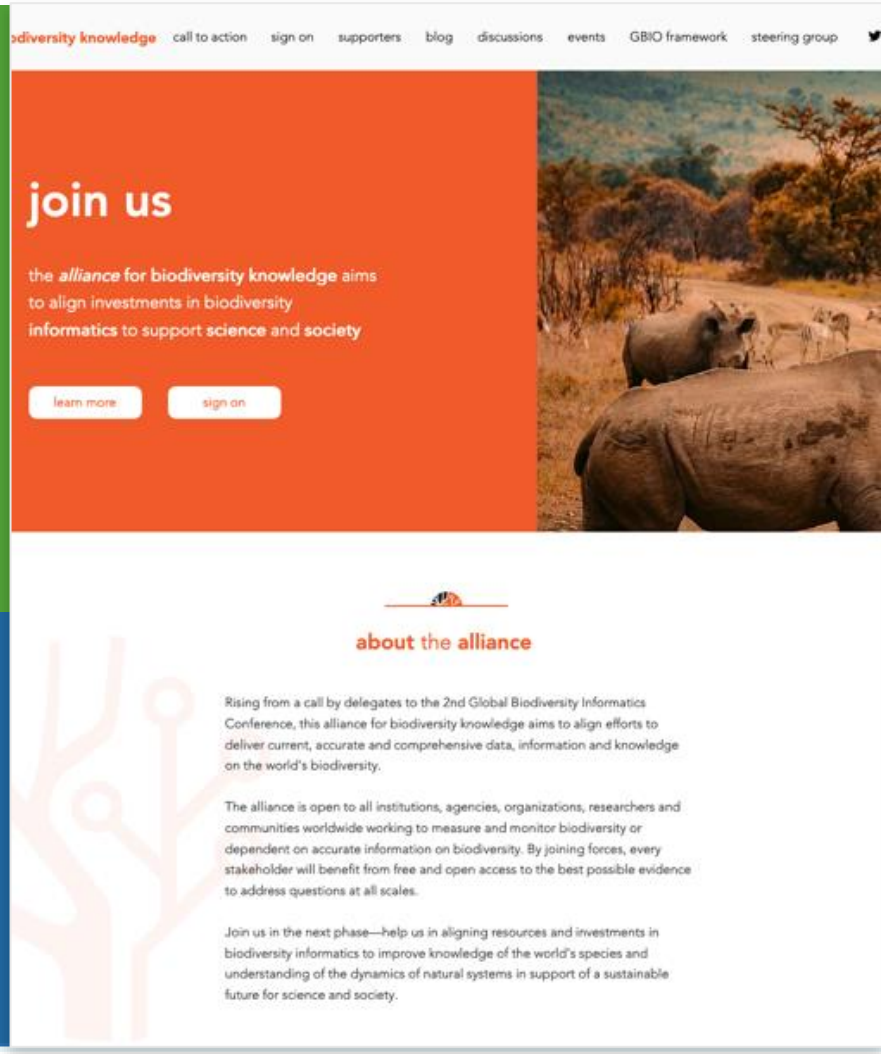
THE GBIO FRAMEWORK

- Align efforts of biodiversity informatics stakeholders to enable integrated understanding of biodiversity
- Four bottom-up focus areas
  - Understanding
  - Evidence
  - Data
  - Culture
- Each focus area includes five core components



https://www.allianceforbio.org/gbio-framework

GBIC2: 2nd GLOBAL BIODIVERSITY INFORMATICS CONFERENCE



- Follow-up event with 100+ attendees in July 2018
- Attendees reached consensus and asked GBIF to coordinate an ***alliance for biodiversity knowledge*** “to align the delivery of current, accurate and comprehensive data, information and knowledge”

https://doi.org/10.3897/BDJ.7.e33679

alliance for biodiversity knowledge

THREE PROPOSED ACTIVITY STREAMS

Complementary, self-reinforcing model for coordination that links informatics to policy goals and societal needs

Biodiversity informatics

- Big data, tools and technology

Science impact

- Big ideas and partnerships to address big problems

Science-policy interface

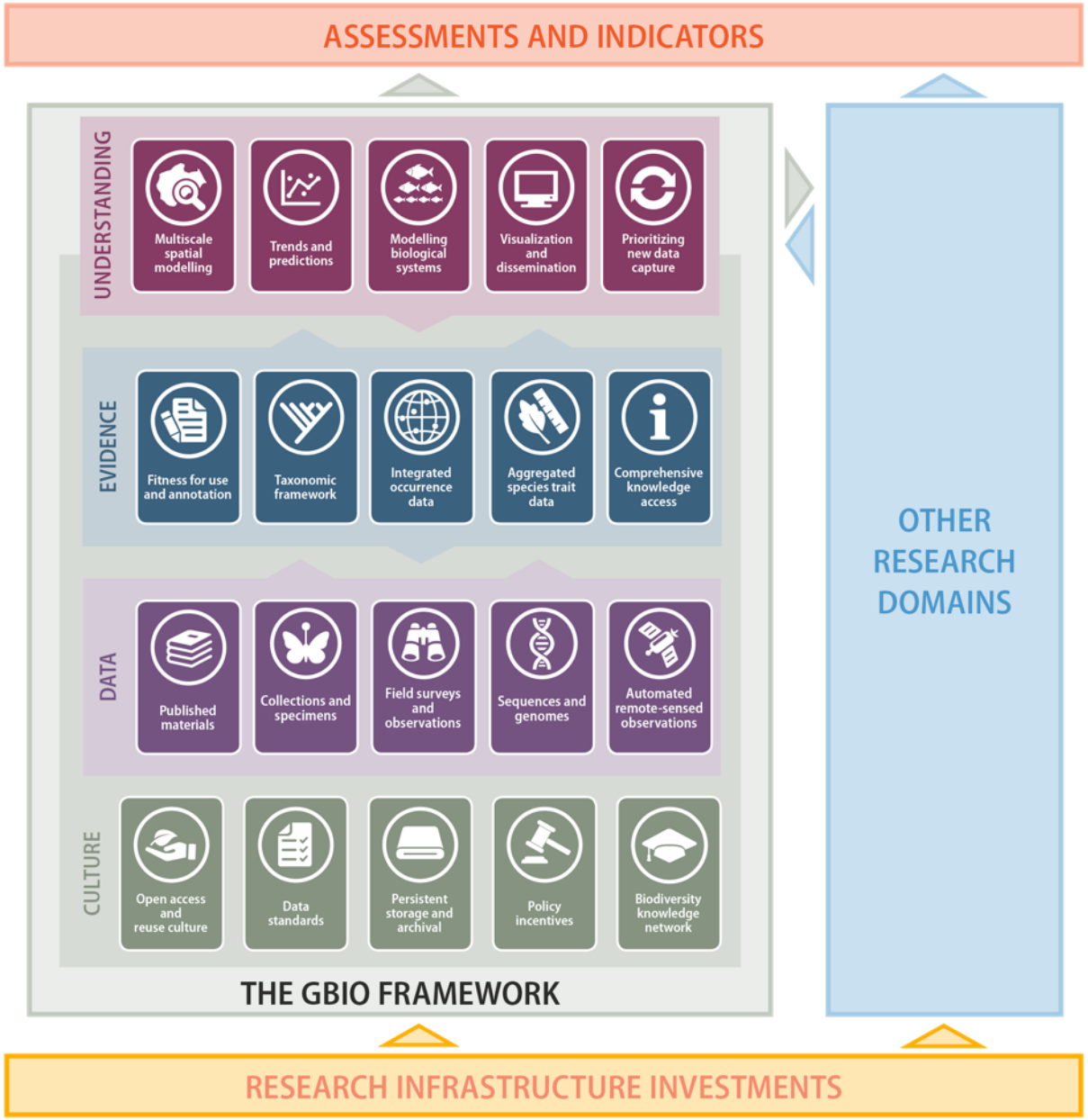
- Big advocates for providers of biodiversity evidence with national, regional & international decision-makers

Bethroot (*Trillium erectum*), photo fbowrys via iNaturalist CC BY-NC 4.0. https://www.gbif.org/occurrence/2423073306

alliance for biodiversity knowledge

BIODIVERSITY INFORMATICS

- **Align investments** to deliver current, accurate and comprehensive data, information and knowledge on the world's biodiversity
- **Organize community-driven technical consultations** on data models, standards and interfaces, agree on priorities, and develop consensus road maps
- **Bridge social and cultural obstacles** to provide a collaborative framework that overcomes division and fragmentation



https://www.allianceforbio.org/

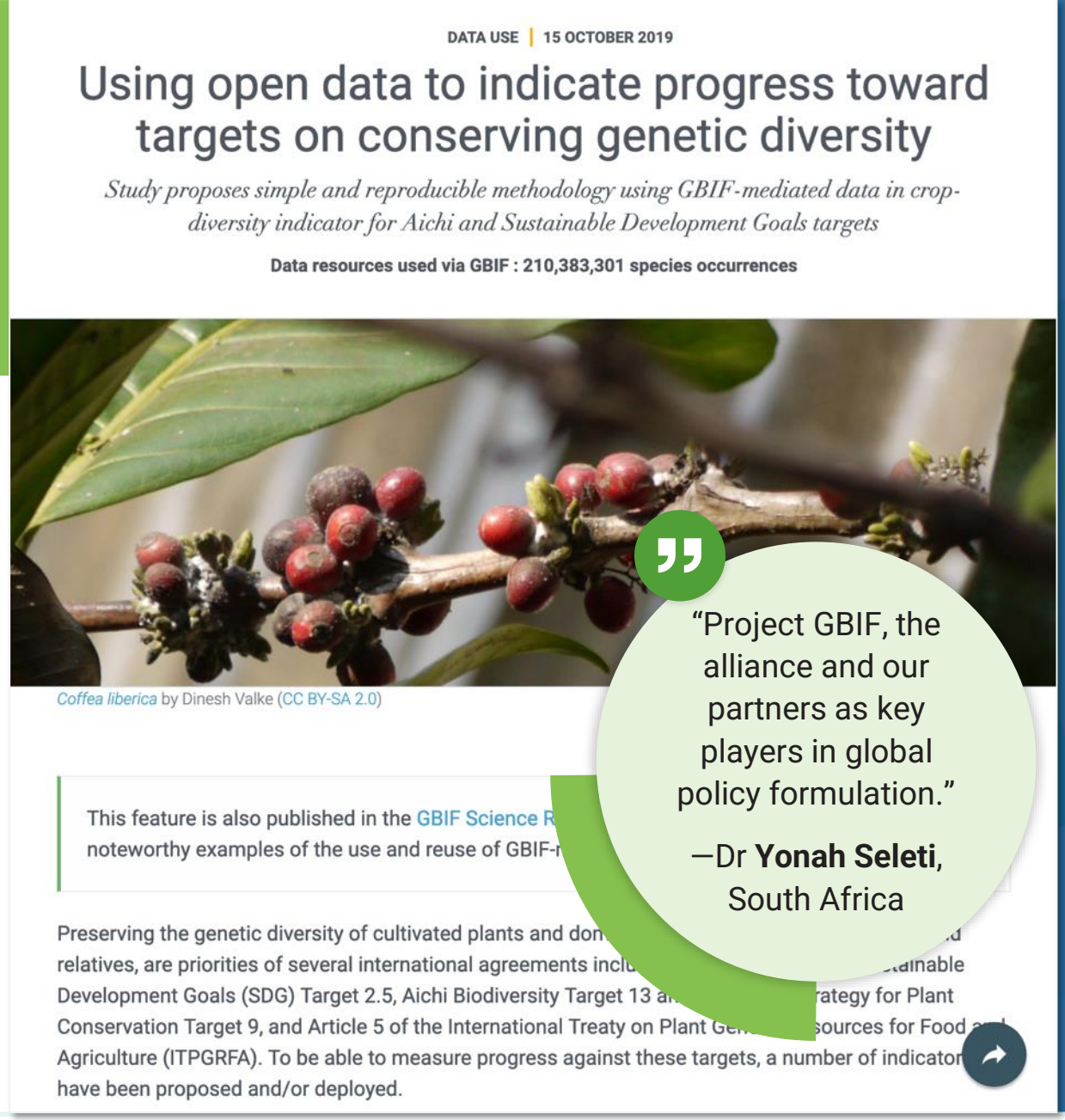
alliance for biodiversity knowledge

SCIENCE-POLICY INTERFACE

"make sure that international organizations at the science-policy interface are aware of the alliance members' strength and capabilities and have them **endorsed accordingly**"

—CODATA *Twenty-year Review of GBIF* (2020)

- Work with GBIF Participants to **build on and expand current engagement efforts** across partnership with CBD, IUCN, IPBES, CMS, FAO, GEF, IPCC, WWF, ICZN, GEO BON, OBIS, CETAF, national agencies, bilateral funders, development banks et al.



"Project GBIF, the alliance and our partners as key players in global policy formulation."

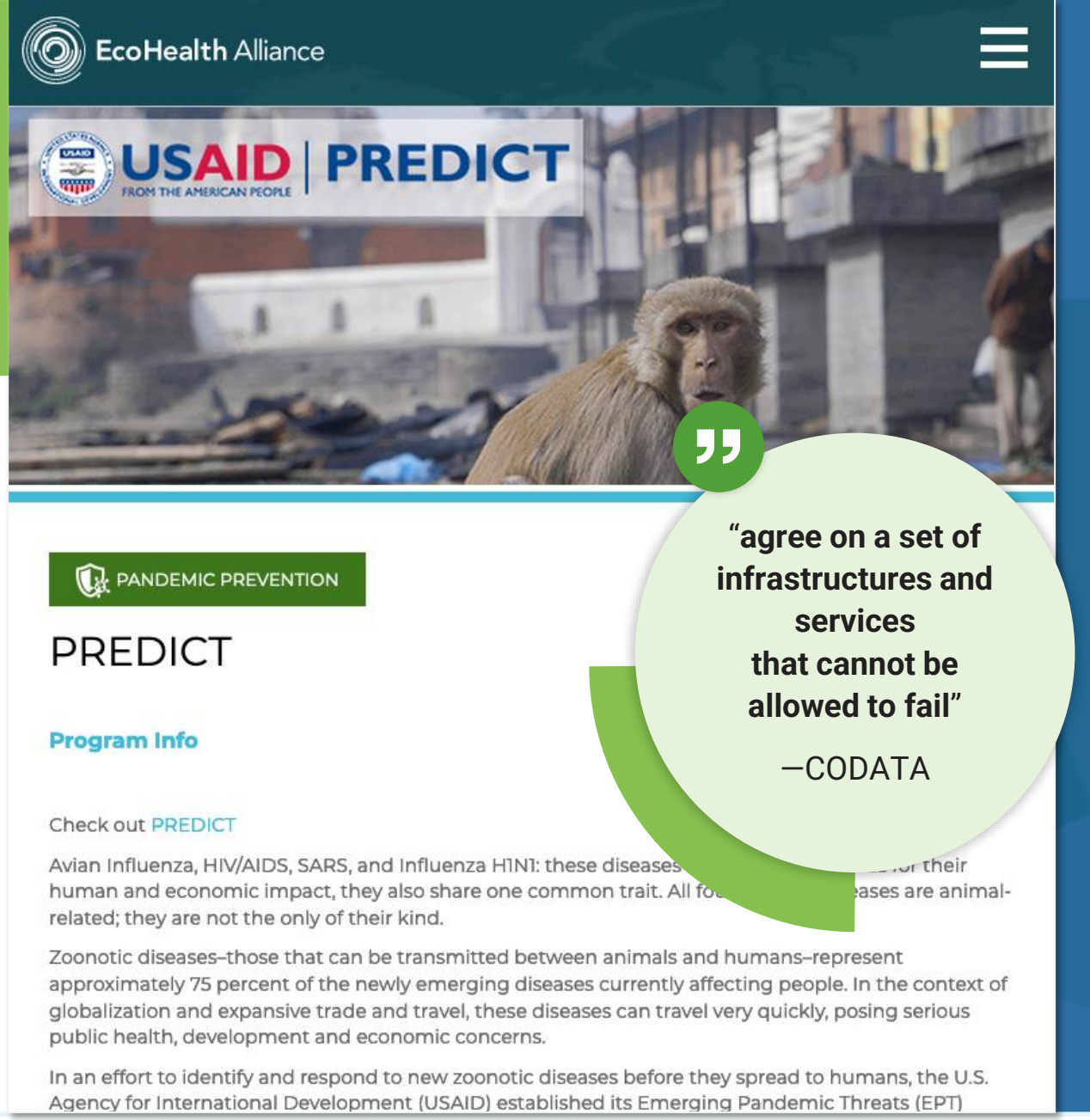
—Dr **Yonah Seleti**, South Africa

https://www.gbif.org/data-use/1g2NQ8TlMFI3nNKLlqJJYa/

alliance for biodiversity knowledge

SCIENCE IMPACT

- Precarious financial and political situations demand broader and more proactive scientific consensus about **what constitutes critical knowledge infrastructure**
- Interdisciplinary engagement on top-priority thematic and sectoral needs with **maximum societal value and impact**
- Coordinate links that share interoperable data and tools across **adjacent domains and infrastructures**



"agree on a set of infrastructures and services that cannot be allowed to fail"

—CODATA

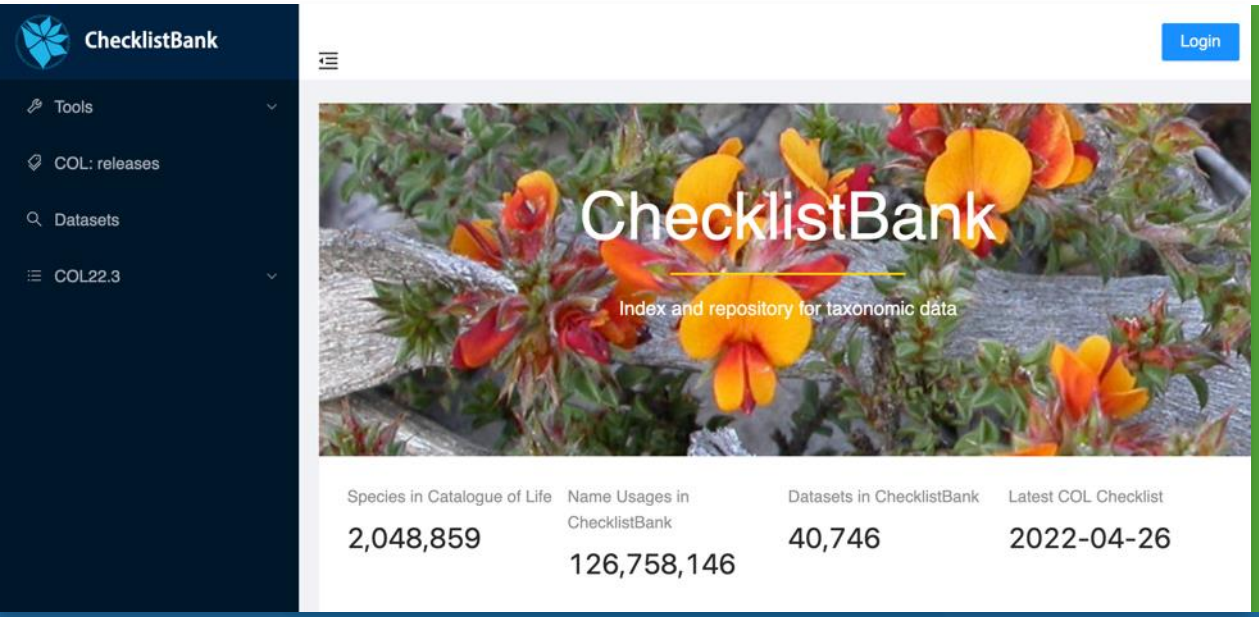
alliance for biodiversity knowledge

RECENT ACTIVITIES

- “Pipelines” collaborations to develop **shared infrastructure and code base** (with ALA and iDigBio)
- **Collections catalogue**
- **ChecklistBank**
- **Digital Extended Specimen**
- **BiCIKL**

Specimens at the Trinidad National Herbarium. Photo 2019 Maheva Bagard Laursen | GBIF (CC BY 2.0)

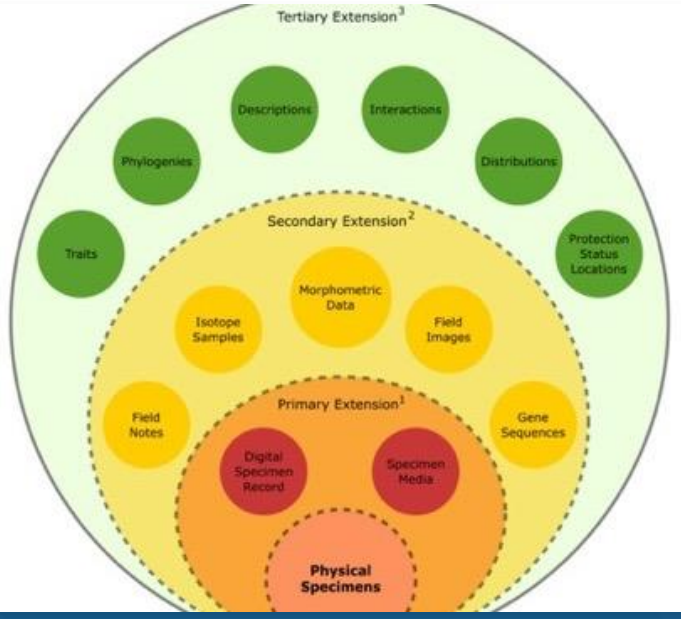
ChecklistBank



- Joint development by COL and GBIF has produced a shared and sustainable infrastructure that can support wider community efforts
- Establishes repository for all COL datasets and any other publicly published species checklists (including those registered with GBIF)
- Provides standardized and web-accessible interpretation of multiple data formats

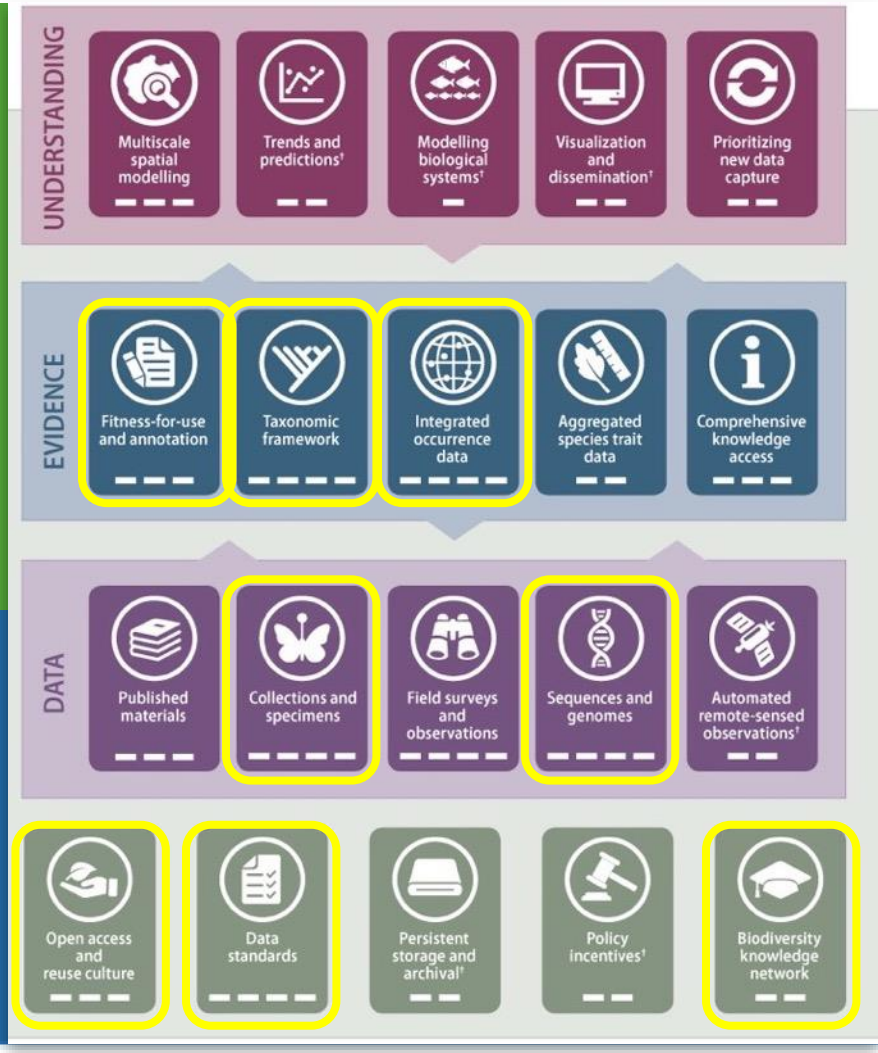
https://www.checklistbank.org

DIGITAL EXTENDED SPECIMEN



Two international community consultations aimed at steering evolving European and U.S. data models toward collaboration

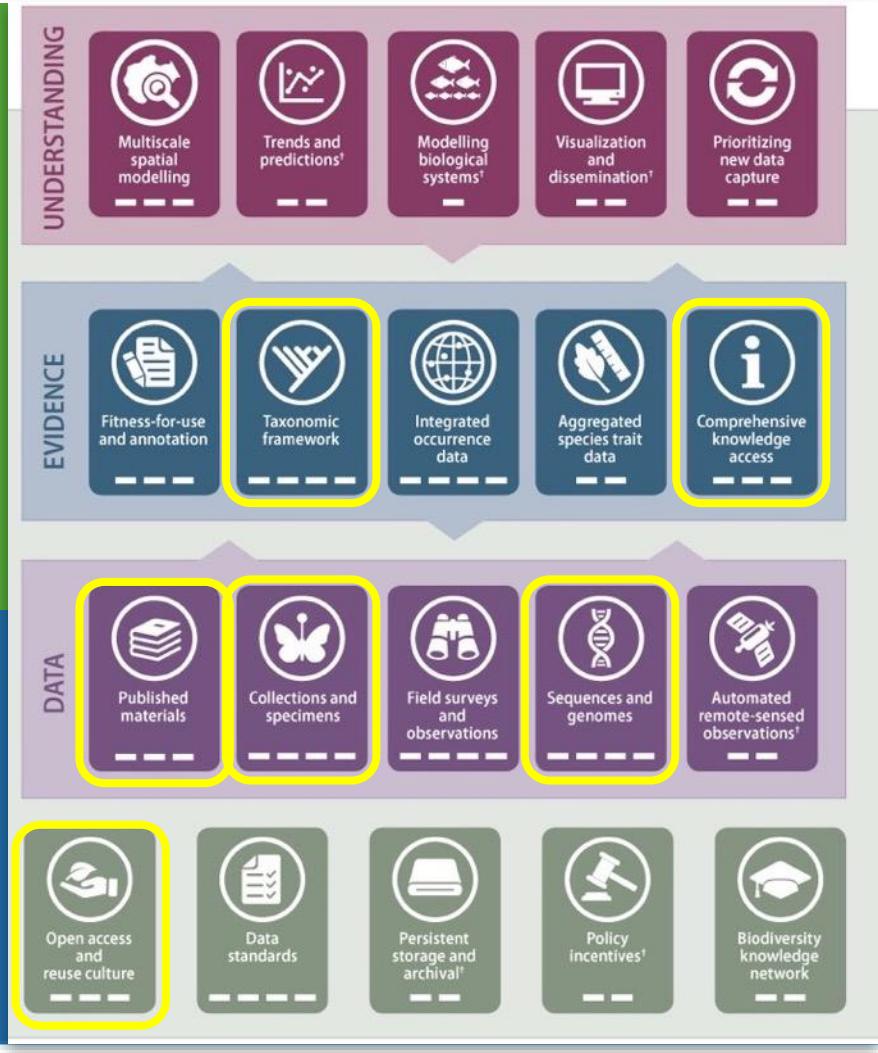
- Ongoing discussions about persistent identifiers
- Partners have submitted an application to the U.S. National Science Foundation to support Research Coordination Network



Thiers B, Monfils A, Zaspel J et al. (2019) Extending U.S. Biodiversity collections to promote research and education.
https://bcon.aibs.org/wp-content/uploads/2019/04/ExtendingBiodiversity-Collections-Full-Report.pdf

BICIKL
BIODIVERSITY COMMUNITY INTEGRATED KNOWLEDGE LIBRARY

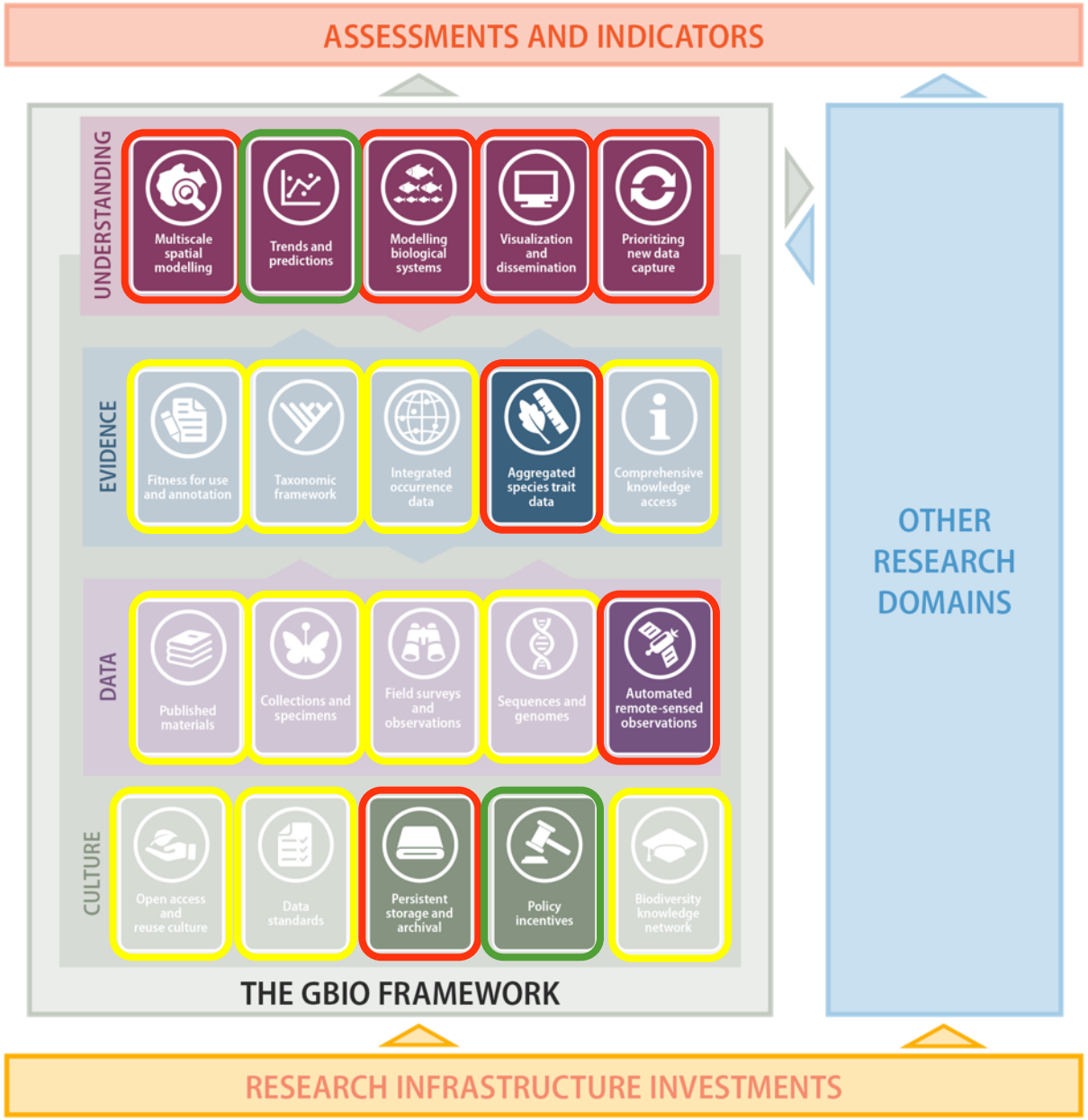
- Horizon 2020-funded project providing external funding allocated to the *alliance*
- WP2 activities aligned with *alliance* governance goals
- Focus on “specimens, genomics, taxonomy and data in publications”
- Enhances and expands on work of **ChecklistBank** and **Digital Extended Specimen (DES)**



alliance for biodiversity knowledge

OPPORTUNITIES FOR WIDER ENGAGEMENT

- Current activities cover only 11/20 focus areas of the GBIO Framework
- Some of the most challenging topics remain: looking at you, traits data and spectral ecology
- Work on engaged focus areas remains iterative and ongoing



https://www.allianceforbio.org/

alliance for biodiversity knowledge GOALS FOR 2022

- Expand range of workshops and consultations to topics beyond scientific collections
- Work with a contractor to develop:
  - Analysis of progress to date and possible models for *alliance* governance
  - Recommendations for establishing a project office
  - *Recruitment profile for programme officer?*
- Plan toward GBIC3 in 2023 with focus on governance

HOW CAN YOU GET INVOLVED?

- Join community events and consultations
- Subscribe/follow for alerts on opportunities
- Keep doing exemplary work to test scalable solutions to address larger problems



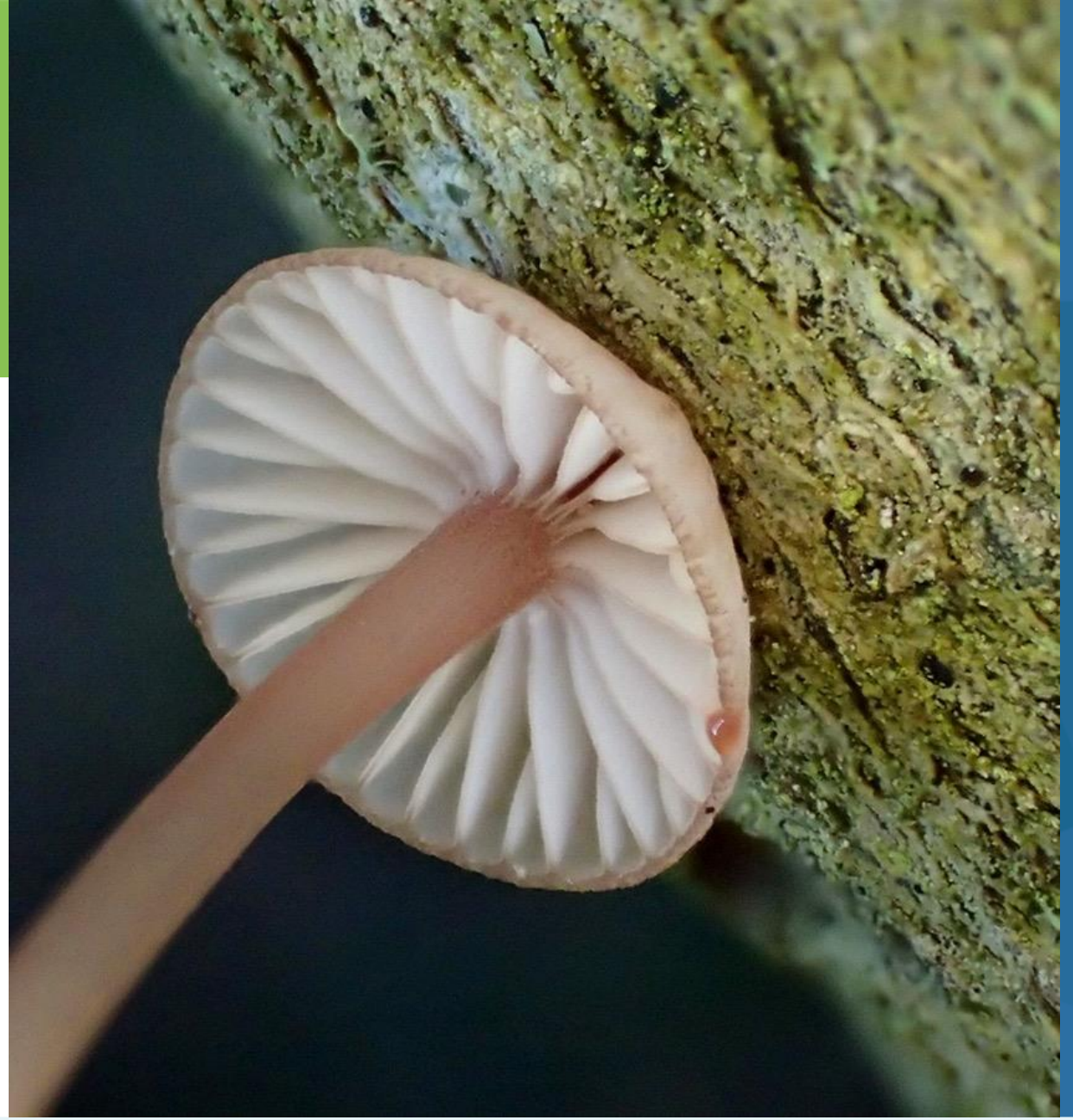



ANALYSIS OF BIODIVERSITY DATA NEEDS IN THE POST-2020 FRAMEWORK

Community consultation to focus on three topics:

- Improving data models and standards for better biodiversity monitoring
- Addressing spatial, temporal and taxonomic biases in primary biodiversity data
- Increasing transparency in processes for indicator development

White paper: **bit.ly/post2020whitepaper**



Bleeding bonnet (*Mycena sanguinolenta*), observed in Norway. Photo 2020 © Kirsti Anne Mandal via Norwegian Species Observation Service. https://www.gbif.org/occurrence/2848866491

ANALYSIS OF BIODIVERSITY DATA NEEDS IN THE POST-2020 FRAMEWORK

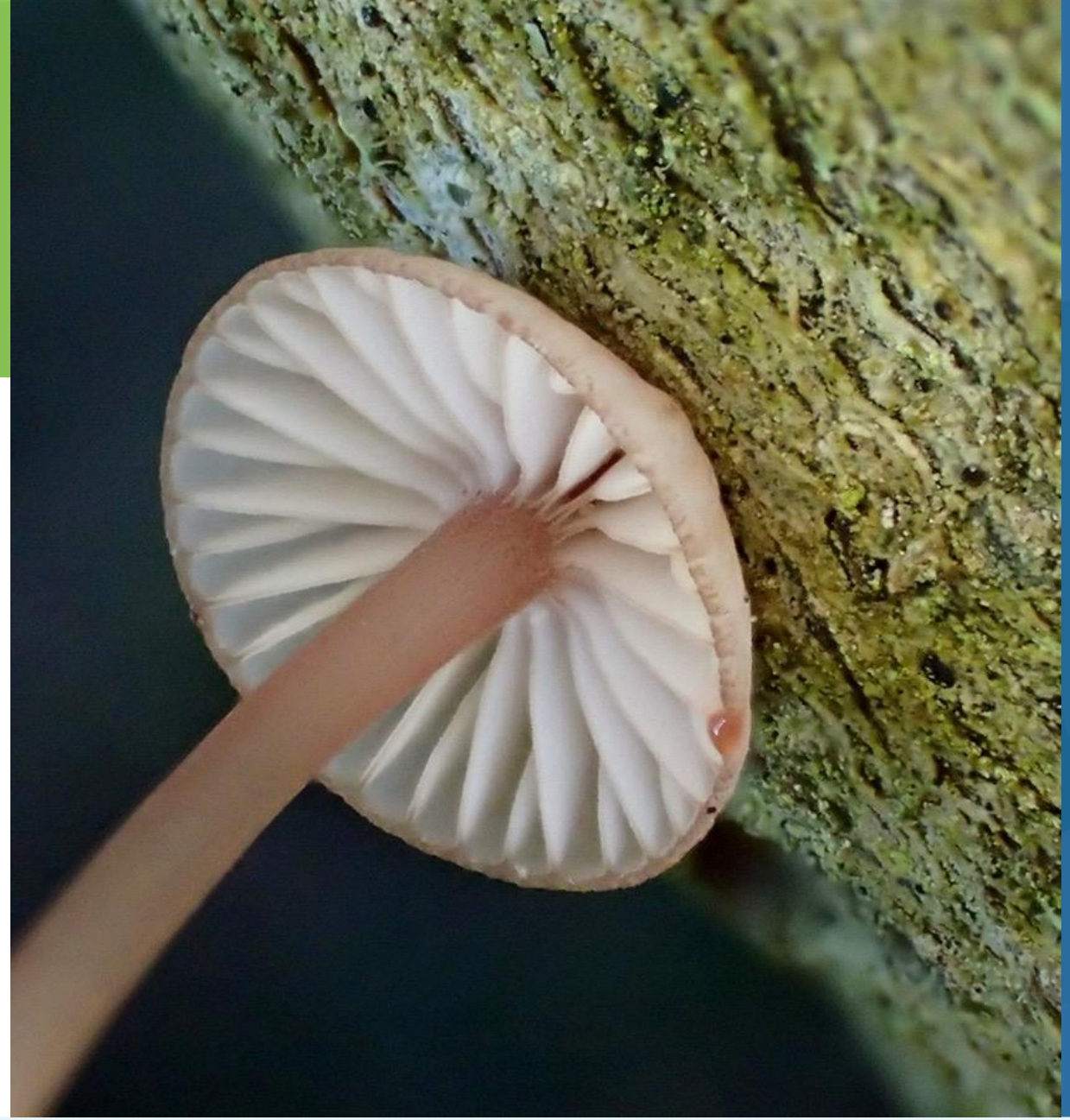
Webinar on 21 July opens two-week consultation

- Timing dictated by shifting CBD schedule

Steering Committee members

- VertNet, CBD, UNEP WCMC, CSIRO, Senckenberg/IPBES, GEO BON, Map Of Life

Register: **bit.ly/alliance-post2020**



Bleeding bonnet (*Mycena sanguinolenta*), observed in Norway. Photo 2020 © Kirsti Anne Mandal via Norwegian Species Observation Service. https://www.gbif.org/occurrence/2848866491

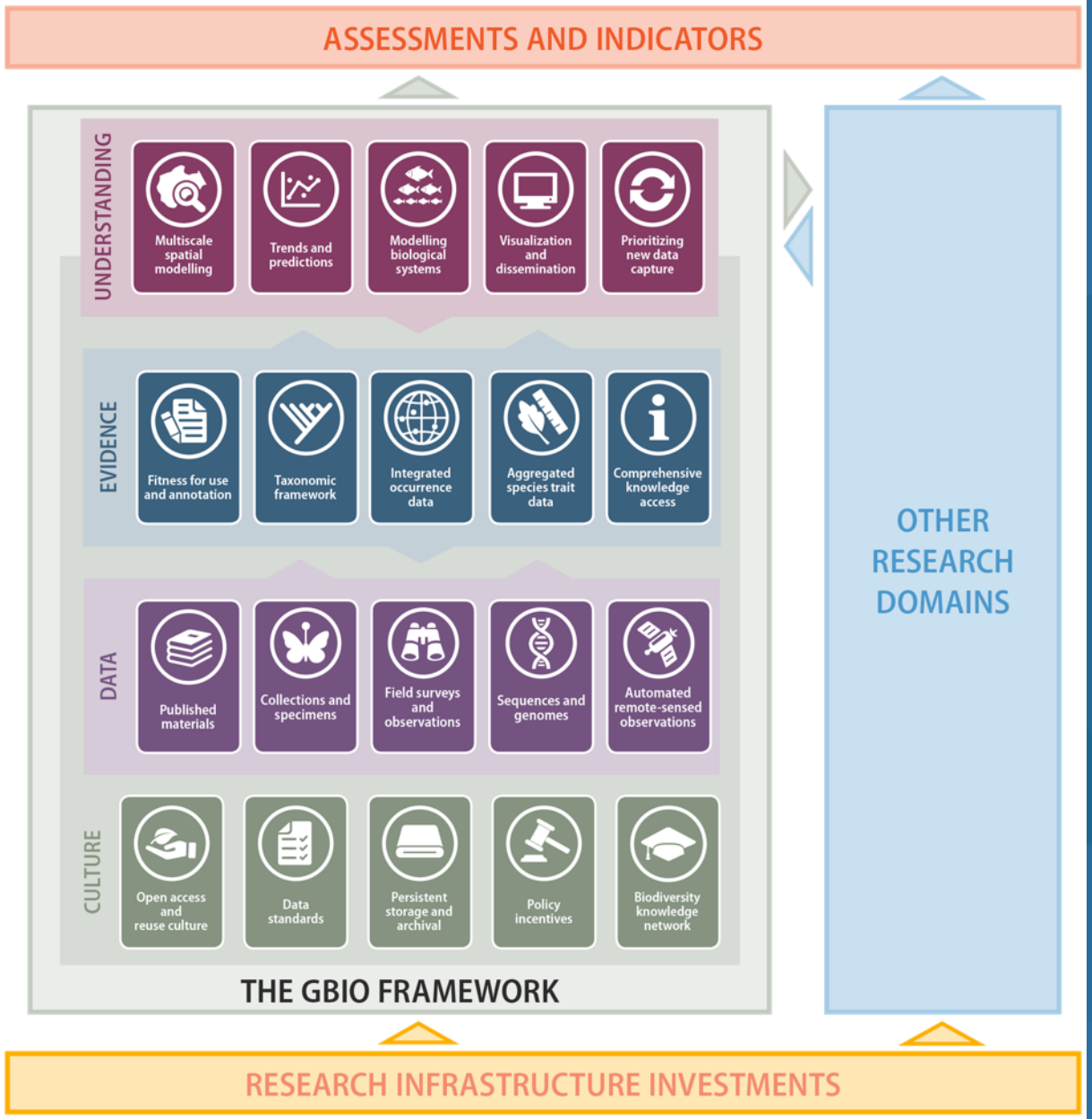
PHYLOGENETIC DIVERSITY IN THE CLOUD

- Supported by **Microsoft AI for Earth | GEO BON** grant programme to explore growing interest how policy options could account for phylogenetic diversity
- Collaboration between GBIF, UNITE, the IUCN SSC Phylogenetic Diversity Task Force, Open Tree of Life, and the University of New South Wales
- Developing cloud computing-based workflows to use GBIF-mediated data to explore how evolutionary history can contribute to indicators

Cookeina tricholoma observed in French Guiana by Julien Renoult. Photo via iNaturalist (CC BY-NC 4.0) https://www.gbif.org/occurrence/1453219544

GBIC3
3rd Global Biodiversity Informatics Conference

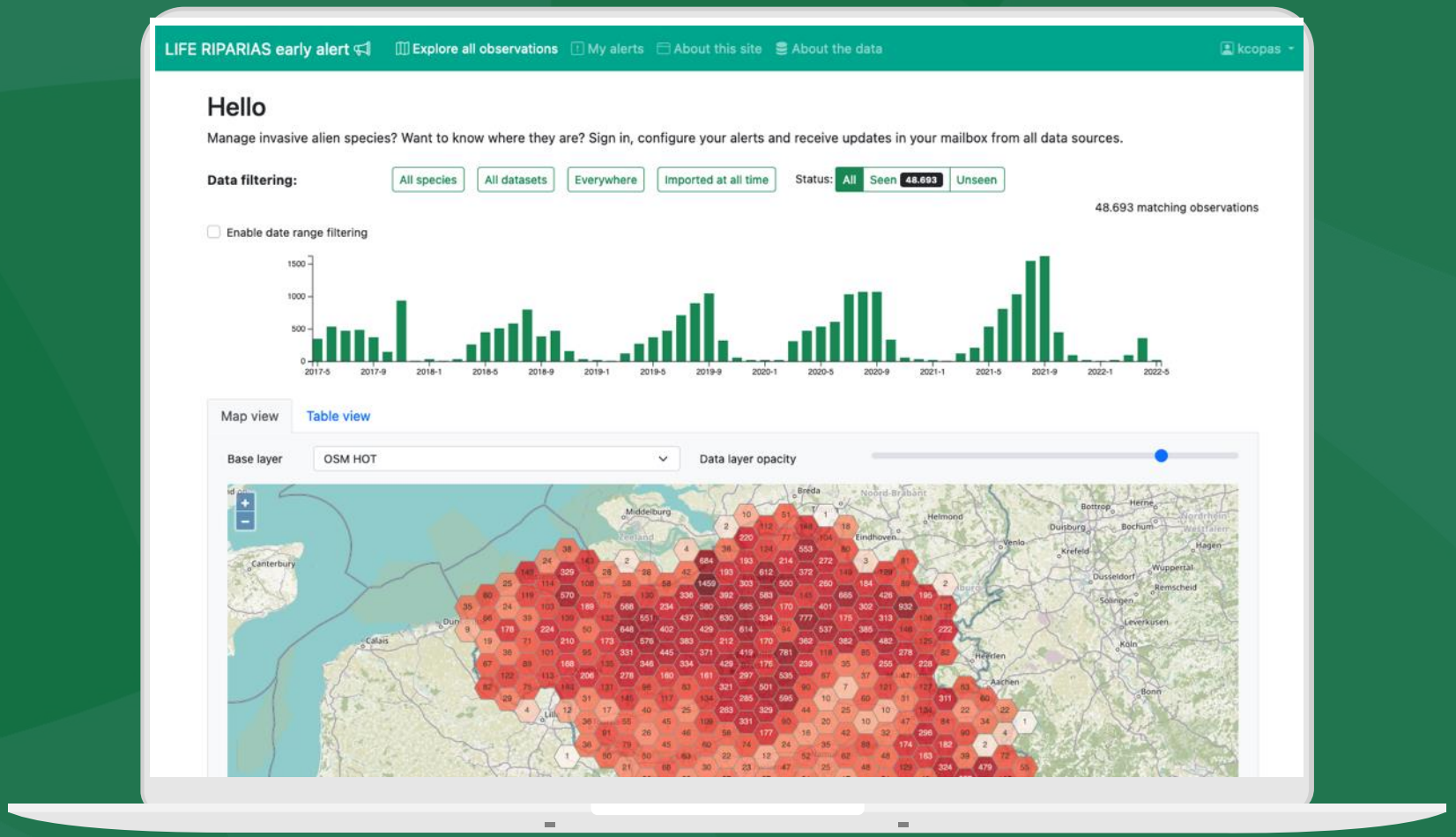
- Virtual format: date TBD (early 2023?)
- Review findings of contractor's report
- Organize discussions around governance
- Guide implementation of *alliance* programme office (along with work of newly hired programme officer)



DO (AND SHARE) EXEMPLARY SCALABLE WORK

RIPARIAS

- Early warning and detection tool from **INBO Open science lab for biodiversity**
- Leverages data from GBIF network to enable rapid response to new invasive species records
- Provides alerts on target species in specific locations and originating from specific datasets

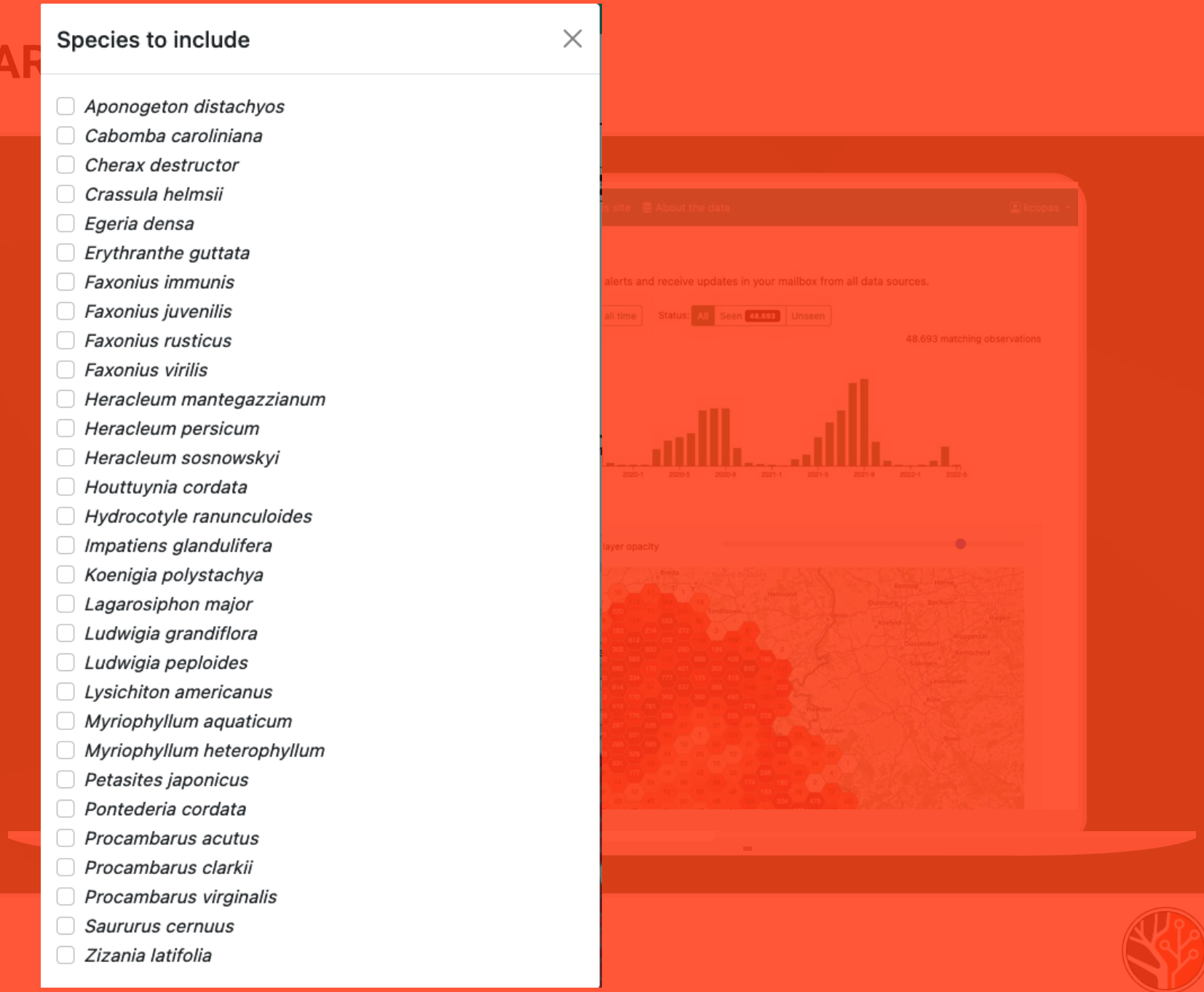


https://alerts.riparias.be

DO (AND SHARE) EXEMPLAR

RIPARIAS

- Early warning and detection tool from **INBO Open science lab for biodiversity**
- Leverages data from GBIF network to enable rapid response to new invasive species records
- Provides alerts on target species in specific locations and originating from specific datasets

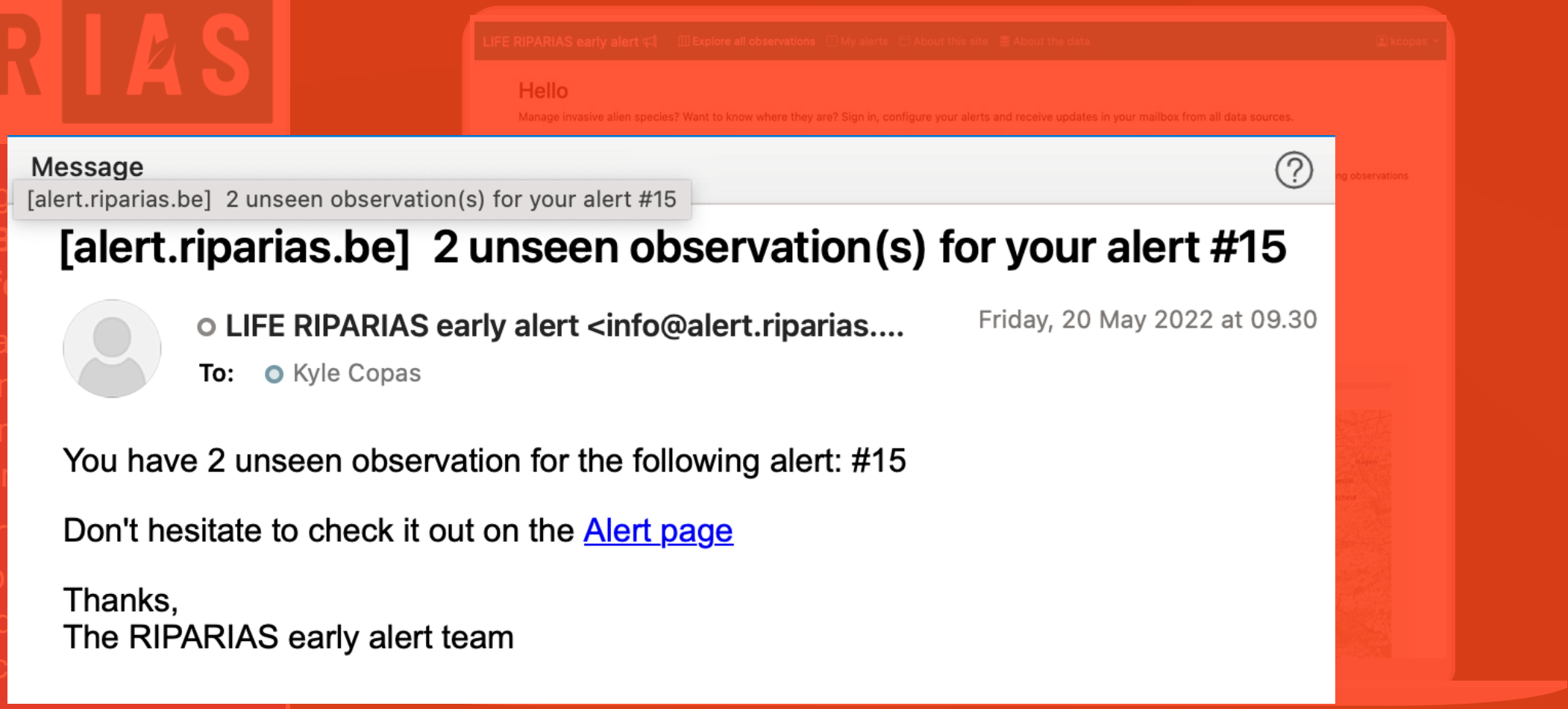


https://alerts.riparias.be

DO (AND SHARE) EXEMPLARY SCALABLE WORK

RIPARIAS

- Early warning
  tool from **INB**
  science lab f
- Leverages da
  network to er
  response to r
  species recor
- Provides aler
  species in sp
  locations and
  from specific



THANK YOU!

Kyle Copas
kcopas@gbif.org

www.gbif.org
@gbif

www.allianceforbio.org
@allianceforbio

alliance for biodiversity knowledge