GNT Pretraining survey results

NB: Full text questions were summarized by AI.

Responses

41 responses from 35 Nodes
4 regional support team (not associated with a node)

By region

Africa - 6
Asia - 2
Europe and Central Asia - 17
Latin American and Caribbean - 6
North America - 3
Oceania - 1

Argentina
ASEAN Centre for Biodiversity
Belgium (4)
Benin
Brazil
Cameroon
Canadensys
Catalogue of Life / Species 2000
Colombia (2)
Denmark
DiSSCo
Finland
France
Georgia
Germany
Guatemala
iDigBio (3)
Ireland
Mauritania
Mexico
Netherlands
Norway
OBIS
SCAR
South Sudan
Spain
SPREP
Sweden
Switzerland
Symbiota Support Hub
TaiBIF
Togo
United Kingdom
Uruguay
Zimbabwe

Additional comments

👉 In short: Participants are excited to learn and network, see great value in Global Nodes Training, want **more hands-on technical guidance** (especially on data packages, cubes, and hosted portals), and request **better support for continuity and remote participation**.

1. Training and Learning Expectations

- Strong enthusiasm for **learning new skills and models** (data publishing, hosted portals, new DwC data packages, barcoding, ecological data visualization).
- Many participants emphasized the value of **Global Nodes Training** and suggested it become a **bigger part of regional meetings**.
- Interest in **hands-on sessions** for hosted portals, data cubes, and data packages, with practical guidance on workflows, automation, and visualization.
- Several mentioned it would be their **first time attending** and they look forward to **learning and networking**.

2. Hosted Portals and Technical Interests

- Some participants are **not focused on hosted portals**, instead prioritizing socio-biodiversity data, sensitive data management, and CARE principles.
- Others expressed specific goals:
    - Standing up **new portals for local projects** (e.g., modeled after the Vermont Atlas of Life).
    - Integrating hosted portals into **regional dashboards** (e.g., ASEAN Biodiversity Dashboard).
    - Exploring **SQL-based services** for occurrence data cubes.

3. Regional and Institutional Needs

- Requests for:
    - **Clear processes for onboarding new staff** into GBIF-related work.
    - **Follow-up support** after training to ensure sustained capacity.
    - **Remote/online participation options**, since not everyone can attend in person.

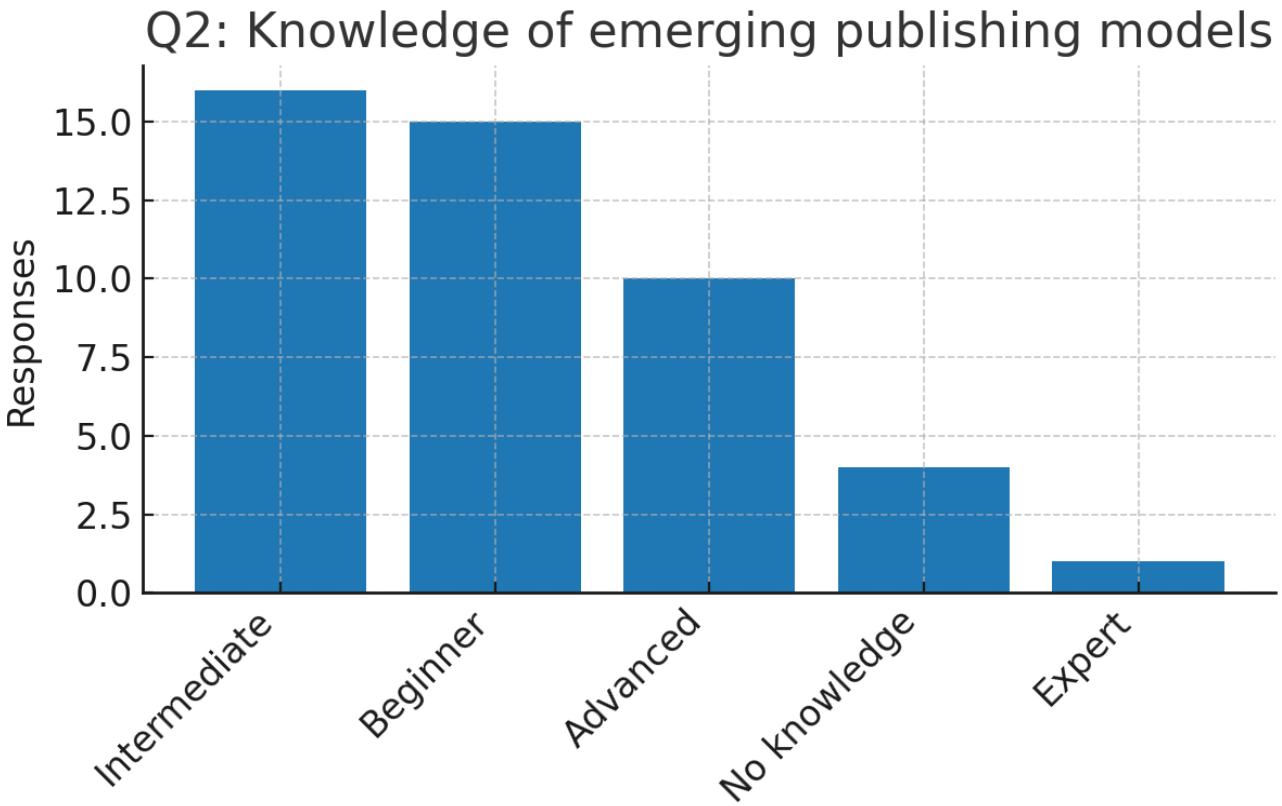
4. Tone and Personal Notes

- Generally **positive and enthusiastic**—participants expect to learn a lot, exchange experiences, and enjoy the event (with one noting: “expect to learn a lot, have fun and drink good coffee”).

Emerging publishing models

Knowledge of emerging publishing models

Intermediate: 16
Beginner: 15
Advanced: 10
No knowledge: 4
Expert: 1



What challenges or limitations have you encountered when using the Darwin Core Archive for publishing biodiversity data?

👉 In summary, **DwC-A works well for standard occurrence data**, but **breaks down with complex, multi-entity, or DNA-derived datasets**. The **star schema, metadata gaps, and reliance on extensions** were the most frequently cited pain points. However, many users still view DwC-A as “good enough” for most biodiversity publishing, with refinements needed for advanced use cases.

1. Data Model and Structural Limitations

- **Star schema constraints** (not flexible enough for complex datasets):
    - One-layer models can't capture hierarchies, subclasses, or relationships between multiple entities (e.g., samples, specimens, hosts, vectors).
    - Difficulty linking measurements or facts (MoF) to anything other than core entities.
- **Denormalization**: collapsing diverse biological classes into occurrence records results in loss of semantics.
- **Sample representation issues**: lack of fields such as `sampleID` makes tracking and publishing samples difficult.
- **Multiple specimens/items**: hard to represent cases like one herbarium sheet containing several specimens, or a single specimen stored as multiple parts (blood, skin, bone).
- **Absence/coverage data**: challenges in expressing absence records, grids, or time series data.

2. Data Types Difficult to Express

- **DNA/metabarcoding/genomics**: current schema fits poorly with DNA-derived data, eDNA, or whole-genome sequencing.
- **Complex ecological interactions**: difficult to publish data involving parasites, vectors, and hosts simultaneously.
- **Non-traditional biodiversity data**: such as camera traps, acoustic recordings, soundscapes, imagery, and biodiversity use data.
- **Trait and functional data**: limited support for mapping traits beyond basic measurements.
- **Taxonomic complexity**: handling contentious histories, nomenclatural issues, and traits remains challenging.

3. Metadata, Semantics, and Identifier Gaps

- **Missing concepts**: lack of MaterialEntity core, absence of richer semantics.
- **Identifiers**: insufficient support for stable identifiers across linked entities.
- **Term overload**: DwC has many terms; hard for users to know which to apply, especially beginners.
- **Underuse of terms**: even where terms exist (e.g., `establishmentMeans`), they are often not applied consistently.

4. Technical and Practical Barriers

- **Loss of information in transformation**: simplifying or reshaping data to fit DwC often strips context or resolution.
- **Reliance on extensions**: while useful, they don't solve all cases and often feel like "forcing" data.
- **Integration challenges**: mapping local/national field codes or forest/marine monitoring terms into DwC requires expertise.
- **Repository limitations**: no clear way to host large associated files (imagery, acoustic, genomic).
- **Interoperability issues**: DwC was built in the biodiversity/GBIF context, limiting use across domains.

5. Usability and Learning Curve

- **Beginners**: struggle to understand core vs. extension split, star schema, and DwC term selection.
- **Mapping difficulties**: researchers often lack examples of best-practice mappings for new data types.
- **Training burden**: challenges teaching how GBIF interprets and displays DwC data vs. other networks like OBIS.

6. Positive and Neutral Feedback

- Some respondents reported **no challenges** or only **minor issues** (e.g., reshaping original data).
- For many use cases, DwC-A combined with IPT and GBIF portals provides an **effective discovery stack**.
- Several saw DwC as "good enough" for most biodiversity publishing, with limitations mainly in specialized contexts.

7. Suggested Improvements / Future Directions

- Introduce **MaterialEntity core** and richer hierarchical models.
- Improve **identifier support** and linking capabilities.
- Enhance **interoperability** by reframing DwC-A as a profile of Frictionless Data Packages.
- Provide **tools for validation and visualization**, akin to OpenRefine or Hosted Portals, to explore archives easily.
- Share more **best-practice examples** for new and complex data types.
- Ensure **engagement of Living Atlas community** with evolving data model updates.

What are your main concerns, questions or uncertainties about publishing or republishing data using the new data model or with the Darwin Core Data Package?

👉 In summary: The **biggest uncertainties** center on **clarity (guidelines, examples, training), complexity (risk of alienating publishers), and transition (migration effort, tool readiness, and backward compatibility).** While many see the **potential of DwC-DP to better accommodate complex data**, there is widespread concern that it could become **too complicated** without sufficient support, making adoption difficult for the broader community.

1. Clarity, Guidance, and Communication

- **Unclear differences** between DwC-DP and DwC-A in practice.
- **Lack of clear guidelines, documentation, and training** on how to publish with the new model.
- **Few worked examples** of data mappings to follow, making it hard to learn by doing.
- **Uncertainty about terminology and identifiers**: how to use IDs to link tables, represent agent roles, or interconnect samples, observations, and media.
- **Questions about communication**: how to explain the shift to publishers and communities who are used to DwC-A.

2. Complexity and Usability

- **Risk of overwhelming publishers**: too many classes, tables, and terms may discourage use.
- **Steep learning curve for beginners** and potential loss of engagement from non-experts.
- **Overkill for simple datasets**: concern that DwC-DP may be unnecessarily complex when DwC-A already works well.
- **Additional time and workload** required to understand, implement, and maintain the new model.
- Experiences with similar models (e.g., **CamtrapDP**) already showed challenges with back-and-forth corrections before successful publication.

3. Tools, Support, and Integration

- **Uncertainty about tool support**: whether IPT, validation services, metadata editors, and Living Atlas systems fully support DwC-DP.
- **Limited developer capacity** at GBIF may slow support, risking frustration and low adoption.
- **Compatibility concerns**:
    - With existing IPT workflows.
    - With national biodiversity databases, ASEAN Biodiversity Dashboard, and Living Atlases.
- **Backward compatibility**: can existing DwC-A datasets be enriched incrementally, or must they be fully re-published in the new format?

4. Migration and Transition Challenges

- **Effort required to convert datasets**: significant re-mapping and restructuring work.
- **Risk of information loss or errors** during migration.
- **Concerns about dataset continuity**: whether old datasets will need to be re-issued as new ones, or if they can be updated in place.
- **Capacity-building needs**: many publishers will require training and technical support to transition.

5. Functionality and Benefits

- **Concerns about whether GBIF ingestion will still reduce data** into simplified occurrence records, limiting the usefulness of richer models.
- **Questions about added value**:
    - How DwC-DP improves handling of complex data types (DNA, time series, event-based monitoring, multi-taxa surveys).
    - Whether it supports linking identifiers more effectively.
    - How richer datasets will be visualized, downloaded, and used.
- **Duplication issues**: uncertainty about overlaps with existing platforms for DNA data.

6. Community Concerns

- **Transition burden on publishers**: many are just becoming comfortable with DwC-A.
- **Knowledge transfer**: how to ensure communities are supported in adopting DwC-DP.
- **Data quality responsibility**: concerns that validation will remain largely on publishers/nodes, with limited systemic safeguards.

7. Neutral or Positive Views

- Some respondents expressed **no concerns yet**, but interest in learning more.
- Others saw DwC-DP as a **necessary and valuable step** to address limitations of DwC-A, especially for richer or more complex datasets.

Are there specific use cases or types of data (e.g., survey and monitoring data, interactions data, physical materials, checklists) you're most interested in representing with the new model?

👉 **In summary:**

- **Survey and monitoring data** is the most widely desired use case, with strong emphasis on time-series, event-based, and long-term monitoring.
- **DNA and molecular data** (eDNA, metabarcoding, barcoding) and **specimen/physical material representation** are seen as high-value areas where DwC-DP offers improvements.
- **Interactions, traits, and checklists** are important but require further development and clarity.
- There is enthusiasm for integrating **novel machine-generated data types** (camera traps, acoustics, imagery), signaling the need for flexible, extensible models.

1. Survey and Monitoring Data (most frequently cited)

- Mentioned by a majority of respondents as the **top priority**.
- Includes:
    - Long-term biodiversity monitoring (forest plots, coral reef surveys, migratory bird counts).
    - Time-series data to capture trends.
    - Event-based sampling (e.g., using Humboldt Extension).
    - Pseudo-absence / absence data.
    - Coverage or grid-based data (not just point occurrences).
    - Community and environmental measurements.
    - Nature chronicles and annual monitoring datasets (e.g., Lepidoptera, Odonata, pollinators).

2. Molecular and Genetic Data

- Strong interest in using DwC-DP for **DNA-based data**, especially:
    - DNA metabarcoding.
    - eDNA and environmental samples.
    - Basic barcoding samples.
    - Linking **MaterialEntity** to molecular analyses.
    - Genetic data tied to specimens.

3. Specimens and Physical Material

- Emphasis on richer representation of **MaterialEntity**:
    - Museum specimens.
    - Environmental samples.
    - Multiple specimens on one herbarium sheet.
    - One specimen represented by multiple items (blood, skin, bones).
- Also: physical materials in general, with connections to genetic analysis.

4. Interactions and Traits

- Several responses highlighted **species interactions**:
    - Host–parasite, vector–pathogen relationships.
    - General interaction datasets.
- Related interests:
    - Traits.
    - Species use data.
    - Invasive species presence/absence over time.

5. Checklists

- Frequently mentioned alongside monitoring.
- Interest in checklists that go beyond taxonomic data, e.g.:
    - Integration with COLDP (Catalogue of Life Data Package).
    - Desire for a roadmap to better support checklist publishers.
    - Historical checklists, invasive species checklists.

6. Novel and Machine-Generated Data

- Interest in standardized ways to represent **machine-observed data**, such as:
    - Camera trap data (with CamTrapDP as inspiration).
    - Bio-logging data.
    - Acoustic data (active, passive).
    - Imagery-based data (satellite, drones, zooscan).

- Respondents expressed desire for **generalizable models** reusable across observation technologies.

7. Other Emerging Use Cases

- Indigenous knowledge.
- Human observation data linked with environmental context (soil, topography).
- Ecosystem restoration monitoring.
- Species occurrences in ASEAN Heritage Parks and protected areas.

How prepared is your node's team to adopt the new data publishing models and tools (e.g., updated IPT, Darwin Core Data Package)?

👉 **In summary:**

- A spectrum of preparedness exists, from **well-prepared early adopters** with updated IPTs and prototype datasets, to nodes with **limited experience or resources** that see themselves as "not prepared."
- The majority fall into a **moderate middle**: technically up-to-date but lacking full experience and confidence, needing clear guidance, training, and capacity-building.
- Regional readiness is uneven, and dependencies on interoperability (e.g., with Living Atlas) and training opportunities will strongly influence adoption.

1. Well Prepared / Ready

- Some nodes reported being **ready or already experimenting**:
    - Updated IPTs (e.g., v3.0.6, v3.1.5) in place.
    - Prototype DwC-DP exports created and tested.
    - Active participation in testing new models (e.g., CamTrapDP, eDNA, interactions).
    - Experienced data managers or engineers in place.
    - Confidence in managing the transition with current expertise.
- Example responses:
    - "Well prepared, we already created a prototype for DwC-DP."
    - "Our data manager is highly experienced, I think we are ready to try."

2. Moderately Prepared / Some Readiness but Gaps

- A large group indicated **medium preparedness**:
    - IPTs generally up to date.
    - Theoretical knowledge from webinars and testing, but limited hands-on cases.
    - Some nodes see readiness in principle, but acknowledge resource and training needs.
    - Preparedness uneven within regional groups (e.g., ASEAN nodes).
- Example responses:
    - “Overall medium. IPT is up to date. Limited resources for publication and training.”
    - “Moderately prepared. No active work begun, but systems allow gradual transition.”
    - “We think we could manage with webinar info, but real-world cases are needed.”

3. Not Very Prepared / Early Stage

- Several respondents said they are **not yet ready** or only partially ready:
    - Lack of hands-on experience with IPT updates.
    - Low confidence in instructing publishers.
    - Staff turnover (e.g., IT engineer leaving).
    - Haven’t yet tested DwC-DP or related tools.
- Example responses:
    - “Not so prepared.”
    - “Not well enough, but willing to improve this.”
    - “We have some theoretical knowledge but no hands-on experience—maybe 30–40% ready.”
    - “Not prepared.”

4. Constraints and Dependencies Highlighted

- **Capacity limitations**:
    - Limited staff, time, and training resources.
    - Need for mentoring and documentation.
    - Some rely on collaborations (CESP projects, Living Atlas).

- **Technical dependencies**:
    - Interoperability with Living Atlas is a blocker for some.
    - Concerns over ABCD 2.06 data integration.
    - Awaiting updates or clarity from GBIF on migration steps.
- **Regional differences**:
    - Some ASEAN member states ready, others need substantial training and infrastructure support.

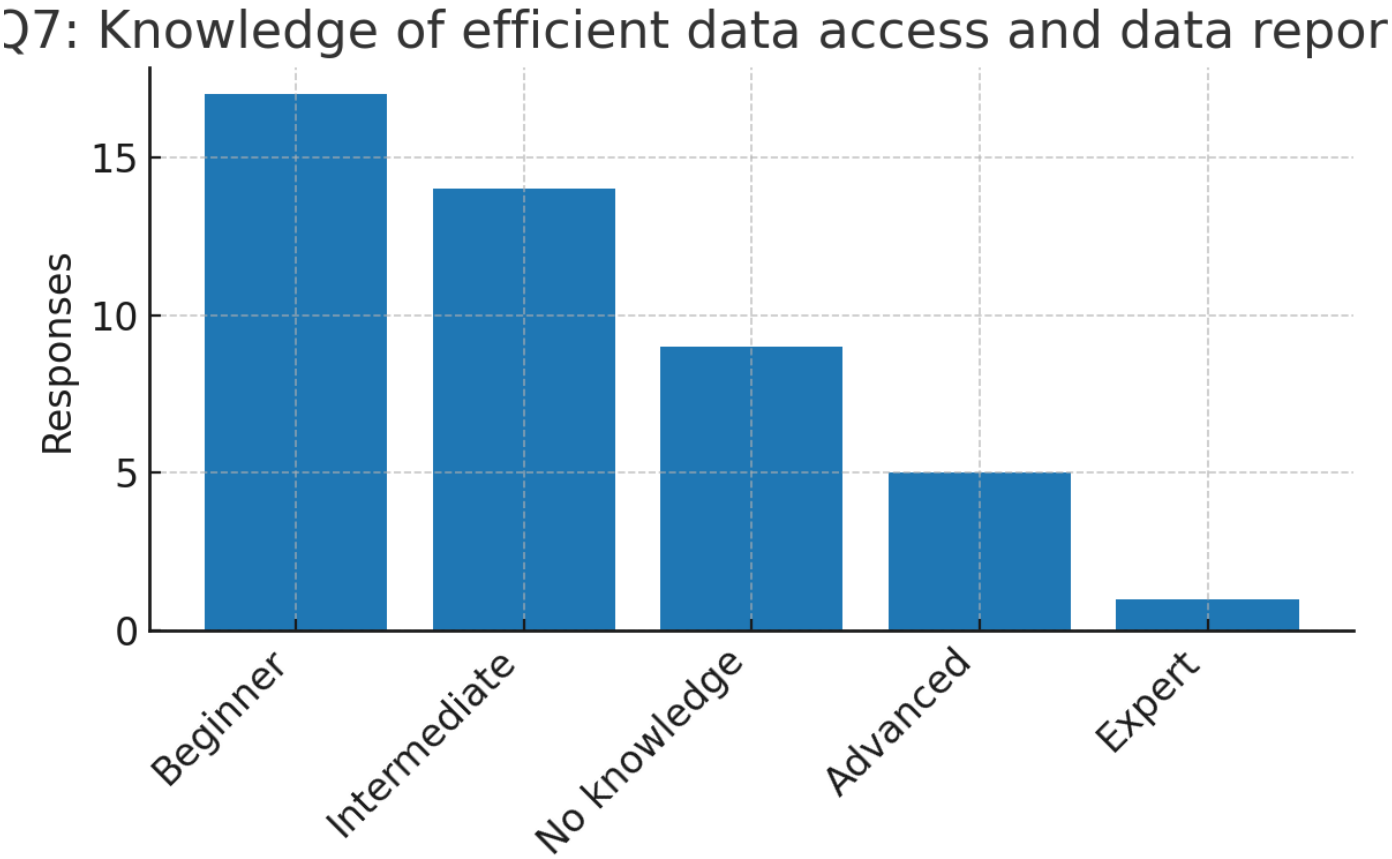
5. Attitudes and Outlook

- Even when not fully prepared, nodes are generally:
    - **Eager to adopt** the new models.
    - Motivated to enhance data publishing capabilities.
    - Looking forward to training and guidance.
- Example sentiments:
    - “We are up for the challenge, but I see a lot of work ahead.”
    - “Not very prepared but would like to embrace it.”
    - “The team is ready to commit, subject to capacity-building support.”

Efficient data access for data reporting

Knowledge of efficient data access for data reporting

Beginner: 17
Intermediate: 14
No knowledge: 9
Advanced: 5
Expert: 1



Does your node receive biodiversity data access requests from researchers?

Yes: 29

What are some examples of biodiversity data access requests that you receive from researchers?

👉 **In summary:** Most biodiversity data requests revolve around **species occurrence records and regional species lists**, often filtered by taxon or geography for use in **species distribution modeling, conservation, and policy reporting**. Increasingly, requests also include **restricted/sensitive data, thematic syntheses (threatened/endemic/invasive species), and advanced data access methods (API, large datasets, integration with other data sources)**. Many researchers—especially students—also rely on node support for **query construction, data interpretation, and quality/licensing guidance**.

1. Species Occurrence Data (Most Common)

- Requests for **occurrence records** (often plants, fish, birds, freshwater species).
- Frequent filters: species name, taxonomic group, geographic region, precise coordinates.
- Common use: **species distribution modeling**, ecological niche modeling, conservation planning.
- Some requests specify exclusions (e.g., “no citizen science data,” exclude eBird/iNaturalist).

2. Species Lists and Checklists

- National or regional checklists (e.g., COL sub-checklists, national checklists).
- Lists of species in specific areas (protected areas, reserves, ecosystems, administrative units).
- Requests tied to policy/assessment:
    - Threatened, endemic, migratory, invasive, or CITES/IUCN/NOM-059-listed species.
    - Species inventories for ASEAN Heritage Parks, Walloon reserves, etc.
    - Data for **national biodiversity strategies, red list updates, environmental reports**.

3. Regional and Thematic Syntheses

- Synthesized biodiversity data at scales beyond national (e.g., ecoregion, protected areas, southern ocean).
- Statistics by region (state, county, province).
- Requests for biodiversity indicators: threatened species, invasive alien species (IAS), endemics.
- Some nodes maintain portals to support this (e.g., *Biodiversidad en Cifras* in Colombia).

4. Specialized Data Types

- **Sensitive/restricted data**: requests for datasets with restricted use.
- **Metadata catalogs**: requests for all datasets about a given taxon.
- **Specimen data**: herbaria, museum collections, lab samples.
- **Agricultural practices** linked to biodiversity.
- **Invasive species** data (Hindu Kush Himalaya, regional IAS requests).
- **Pollinators** and climate change interaction studies.

5. Advanced / Large-Scale Data Needs

- Large downloads for meta-studies or bespoke datasets.
- Use of APIs for automated access, overcoming portal download limits.
- Requests for **high-resolution data** (precise coordinates) for modeling.
- Interest in **data integration across repositories** (e.g., with environmental datasets).
- More advanced needs: digital twins, data lakes, data spaces (beyond standard GBIF API).

6. Support and Training Requests

- Many researchers need **help navigating the GBIF portal** (filters, taxon searches, custom queries).
- Some requests arise from **misuse/misunderstanding of the search tools**, requiring staff support.
- Students (MSc, PhD) frequently request guidance for research projects (distribution, niche modeling).
- Training requests: **clarification on licensing, sensitive data handling, and data quality**.

Does your node currently produce any kind of biodiversity data report?

Yes: 23

What are some examples of biodiversity data reports that you produce?

👉 **In summary:** Nodes and institutions produce a mix of **formal government reports (national biodiversity strategies, state of environment, compliance)**, **institutional activity reports (dataset metrics, publication highlights, data gap analyses)**, and **regional/thematic syntheses (e.g., ASEAN, Colombia's portals, protected areas reports)**. Many also provide **species lists, biodiversity indicators, and outreach reports** for public or stakeholder engagement.

1. National / Government Reporting

- **State of the Environment & National Biodiversity Reports**
    - National Report on the State and Protection of the Environment (annual).
    - Reports to ministries (e.g., National Biodiversity Forum, Ministry of Environment, Ministry of Culture/Education/Research).
    - National Biodiversity Strategy and Action Plan (NBSAP) updates.
    - National Science Foundation compliance reports.
    - Reports for national or cantonal species conservation responsibilities.

- **Examples:**
    - Brazil: Quarterly reports to Biodiversity Forum + IBGE biodiversity data assessments.
    - Togo: Green Growth Index.
    - Netherlands: Annual biodiversity status report for government.

2. Institutional / Node-Level Reports

- **Annual activity & dataset reports**
    - Data publication statistics, species lists, and richness summaries.
    - Data gap analyses (e.g., comparing with neighboring countries to detect biases).
    - Highlights reports (showcasing new datasets, activities, metrics).
    - Dataset metrics and publication evolution (e.g., GBIF Germany, GBIF Spain).
- **Examples:**
    - GBIF Spain: annual report on collections (e.g., Informe_colecciones2021).
    - Germany: annual reports on datasets published via German data publishers.
    - Some nodes: reports embedded in student theses or scientific publications.

3. Regional / Multinational Reports

- **ASEAN Centre for Biodiversity (ACB)**
    - *ASEAN Biodiversity Outlook* (periodic, policy + biodiversity data trends).
    - Reports on ASEAN Heritage Parks, thematic topics (e.g., mangroves, coral reefs, migratory species).
- **Other regional syntheses**
    - Biodiversidad en Cifras (Colombia, annual subnational syntheses).
    - Campaign for National Parks “Health Check” reports (UK).
    - Parques Nacionales Cómo Vamos (Colombia, biodiversity in national parks).
    - Reporte BIO (Colombia, national and regional biodiversity state/trends).

4. Protected Area / Thematic Reports

- Species lists and biodiversity inventories for:
    - Protected areas, parks, reserves.
    - Strategic ecosystems (mangroves, coral reefs, migratory species).
- Indicators: species richness, endemics, threatened species, invasive alien species.

5. Communication & Outreach Reports

- Public-facing data portals and updates:
    - Biodiversidad en Cifras (annual updates).
    - Humboldt Institute portals (*Reporte BIO*, *Biodiversidad.co*).
    - Social media highlights of new datasets.
- Campaign reports for awareness (e.g., *Campaign for National Parks*).

6. Project-Based Reports

- Reports generated within externally funded projects or monitoring programs.
- Often shared with project partners or ministries, not always public.

Do you receive requests from third parties (e.g., government agencies, environmental consultants, private sector) to help access GBIF-mediated data for use in formal reporting?

Yes: 22

What types of reporting frameworks or obligations are involved?

👉 **In summary:** Nodes and institutions contribute biodiversity data to a **layered set of reporting frameworks**:

- **Global:** CBD, GBF, IUCN Red List, CITES, Ramsar, CMS.
- **Regional:** EU directives, African Biodiversity Strategy/COMIFAC, Latin American socio-biodiversity frameworks.
- **National:** State of the Environment reports, national biodiversity indicators, NBSAPs, red lists, and ministry-led obligations.
- **Project-based & thematic:** BID projects, Living Planet Index, CARE principles, environmental impact studies.

1. Global Frameworks & Conventions

- **Convention on Biological Diversity (CBD)**
    - National Reports to the CBD.
    - Contribution to **CBD / Kunming–Montreal Global Biodiversity Framework (GBF)** indicators.
    - Support for **National Biodiversity Strategies and Action Plans (NBSAPs)**.
    - Monitoring and reporting on threatened and exotic species.
- **Other Multilateral Environmental Agreements (MEAs):**
    - **CITES** (Convention on International Trade in Endangered Species).
    - **Ramsar Convention** (wetlands).
    - **CMS** (Convention on Migratory Species).
- **IUCN Red List / National Red Lists** – species assessments and updates.
- **Living Planet Index** (e.g., Belgium, via WWF).

2. Regional Frameworks & Directives

- **Europe:**
    - European biodiversity networks (e.g., **EASIN**, invasive species).
    - Regional directives and development plans (species/taxa-specific).
- **Africa:**
    - African Biodiversity Strategy, **COMIFAC**, African Union reporting.
- **Latin America & Caribbean:**
    - Regional guidelines, socio-biodiversity and species-use reporting.
- **Other regional frameworks** supporting ecological data and indicators.

3. National Reporting Obligations

- **National State of the Environment Reports.**
- **National biodiversity indicators** (species, ecosystems, socio-biodiversity).
- **Support to Ministries of Environment** (e.g., SEMARNAT in Mexico, Norway's Artskart system).
- National data contributions to conservation conventions.
- National technical opinions (e.g., environmental impact, risk analysis).
- Private sector reports (annual biodiversity/environmental performance).

4. Thematic / Project-Based Frameworks

- **Biodiversity Information for Development (BID) projects**
    - Mobilization of data to support CBD/GBF indicators and national decision-making.
- **Ecological & Socio-biodiversity Initiatives**
    - CARE Principles for Indigenous Data Governance.
    - Data streams for sustainable use and conservation planning.

What types of data are typically required to support these reporting needs?

👉 **In summary:** Reporting frameworks rely on a **combination of species lists, occurrence data, conservation/risk status, ecological and environmental variables, and derived biodiversity indicators**. These are often integrated with metadata, usage statistics, and technical tools (GBIF API, national portals) to generate standardized, policy-relevant outputs.

1. Species & Taxonomic Data

- **Species lists** (national, regional, local).
- Taxonomic details: validated scientific names, synonymy, classification.
- Species' **regional/national status** (native, endemic, invasive, threatened, regulated).
- **Checklists** organized by themes such as threat status, invasiveness, conservation priority.
- Metadata catalogs for taxonomic datasets.

2. Occurrence & Observation Data

- **Occurrence records**: species, location (coordinates), date, observer/collector.
- **Population data**: abundance, presence/absence, monitoring effort.
- **Trends in distribution** and species range shifts.
- Number of observations and records in specific areas.
- Data disaggregated by state, ecoregion, or protected areas.

3. Conservation & Risk Data

- **IUCN Red List / extinction risk assessments.**
- Threatened/endangered species status.
- Invasive Alien Species (IAS) occurrence and monitoring.
- Conservation measures and protected area species inventories.
- Indicators such as species richness, biodiversity statistics, and pressure/threat data.

4. Environmental & Ecological Data

- Habitat data: ecosystems, land cover, forest types, habitat characteristics.
- Climate and environmental parameters.
- Ecological data on species use, associated traditional knowledge, derived products.
- Disease vectors and other ecosystem health indicators.
- Integration with social and economic variables for decision-making.

5. Derived Metrics & Indicators

- Species richness (by region or taxonomic group).
- Number of downloads, citations, dataset providers (usage statistics).
- Biodiversity indicators for national and regional reports.
- Time-series data for monitoring population trends.

6. Data Access & Technical Support

- Support with **GBIF API** and **web portal** for data extraction.
- Use of **national portals** (e.g., Norway's Artskart) for enhanced spatial and red list services.
- Mobilization of specialized datasets (e.g., IAS, freshwater, regulated/protected species).
- Coordination of biodiversity data with environmental/social datasets.
- Metadata on sampling protocols, data quality, reliability.

What is the typical level of biodiversity or technical expertise of these requesters?

👉 **In summary:** Requesters include a broad spectrum from students and technicians to expert researchers and government agencies. While biodiversity knowledge can be high, technical skills (data management, FAIR principles, advanced analysis) are often limited. As a result, many depend on GBIF nodes or intermediaries to **interpret, process, and package biodiversity data** into accessible formats for reporting and decision-making.

1. Overall Range

- Expertise **varies widely**: from **non-experts** and policymakers to **advanced researchers** and biodiversity specialists.
- In general, many users have **low to moderate technical expertise**, even if their biodiversity knowledge is stronger.

2. Groups of Requesters & Expertise Levels

- **Researchers & Academics**
    - Typically **high biodiversity/taxonomic expertise** (ecology, taxonomy, environmental sciences).
    - Can use **raw data directly** (occurrence data, time series, GIS).
    - Graduate students or early-career researchers often need **methodological support**.
- **Government Institutions / Policy-makers / Agencies**
    - Varying expertise: some employ biodiversity specialists, others rely on **synthesized data** (indicators, maps, tables).
    - Often expect **species lists** with breakdowns (native/exotic, IUCN status).
    - Data usually requested in **Excel or simple formats**.

- **NGOs & Conservation Stakeholders**
    - Intermediate expertise, applied focus (protected areas, threats, monitoring of priority species).
    - Require **interpreted outputs** for practical conservation work.
- **Decision-makers & Planners**
    - **Limited technical expertise**.
    - Typically request **aggregated/processed data** presented as reports, indicators, or visualizations.
- **Technicians & Supervisors**
    - Moderate capacity, operational focus on data handling but limited advanced analysis.
- **Natural Resource Managers & Public Health Specialists**
    - Practical applied expertise, but often require support in accessing and interpreting biodiversity datasets.

3. Common Patterns

- **High biodiversity expertise, low technical expertise** is a recurring theme.
- Most users lack advanced **data management** skills or training in FAIR principles, even if aware of them.
- Many need **guidance on accessing and interpreting GBIF data** to ensure taxonomic and spatial accuracy.

What could help your node to better meet data access or reporting requests?

👉 **In summary:** Nodes highlight the need for **more diverse and higher-quality data**, stronger **technical infrastructure (APIs, automation, dashboards)**, and sustained **capacity building** to support reporting. Clearer recognition of GBIF's role by governments and funders, along with **stable financing**, will ensure that nodes can better deliver on biodiversity data access and reporting obligations.

1. Expand and Improve Data Coverage

- Increase **quantity and diversity of datasets**, especially underrepresented ecosystems (wetlands, drylands) and taxa (invertebrates, fungi, microbes).
- Mobilize **historical data and literature records**.
- Encourage **institutions and researchers** to publish datasets by clarifying **benefits of open data** and addressing **intellectual property concerns**.
- Promote GBIF recognition and visibility among **governments, academia, and regional institutions**.

2. Enhance Data Quality and Integration

- Improve **data cleaning and synthesis capacity** (requires significant computing resources).
- Strengthen links between **national species lists** and international references (e.g., Catalogue of Life) to reduce biases in the GBIF Taxonomic Backbone.
- Improve **taxonomic, spatial, and metadata quality** across datasets.
- Enable synchronization between **national portals and GBIF datasets** without full re-ingestion (incremental updates).

3. Develop Technical Tools and Infrastructure

- Automated **reporting routines** and **API interfaces** for common reporting formats.
- Tools to provide **statistics to publishers** (downloads, views, citations).
- **Centralized data portals** (e.g., unified hosted portal instead of multiple fragmented ones).
- Dashboards to track **digitisation progress** and **data quality improvements**.
- Step-by-step tools and **multilingual training materials** (including videos with transcripts).
- Tools for **assisting report preparation** and **automated extraction of GBIF-mediated data**.

4. Capacity Building & Training

- Training in **data management, analysis, visualisation** (GIS, statistics, Darwin Core, IPT).
- **Data analytics training** for biodiversity metrics, species assemblages, and trends over time.
- Development of **communication and dissemination skills** to support diverse audiences.
- Sharing of **use cases and lessons learned from other nodes**.
- Continuous engagement in **biodiversity informatics activities** to strengthen expertise.

5. Institutional, Regional & Community Support

- Increase **awareness and promotion** of GBIF in underrepresented regions (Pacific, ASEAN, Africa).
- Encourage **regional collaboration and integration** of biodiversity platforms.
- Engage the **scientific community** to publish research data.
- Build **closer communication with GBIF Secretariat** and other nodes.
- Encourage **feedback loops from data users** to improve services.
- Support from **governments, ministries, and international institutions** to use GBIF as a core data pipeline.

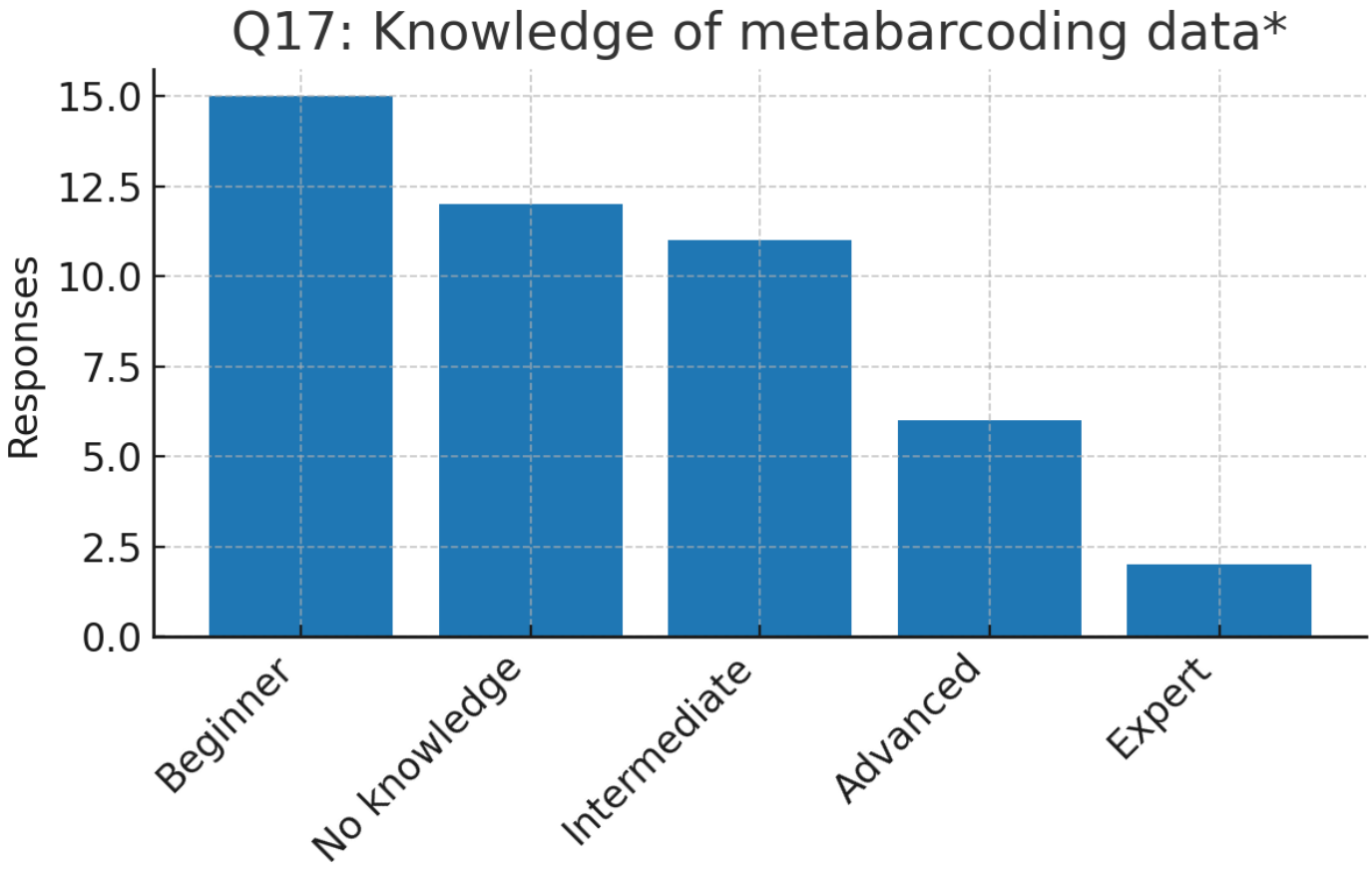
6. Sustainable Resources

- Secure **long-term funding** for node operations, data maintenance, and new digitisation projects.
- Ensure availability of **technical staff** to meet increasing reporting demands.
- Institutional support for **sustainability of node activities**.

Metabarcoding data

Knowledge of metabarcoding data

Beginner: 15
No knowledge: 12
Intermediate: 11
Advanced: Expert: 2



Is your node currently participating or planning to participate in the pilot phase of the Metabarcoding Data Programme?

Yes: 18
Planning on it: 17
No: 11

What best describes your experience with publishing metabarcoding data to GBIF?

I'm interested in publishing metabarcoding data to GBIF but haven't had the chance to learn much about how to do it yet.: 19
I have published metabarcoding data to GBIF.: 10
I have explored how to publish metabarcoding data to GBIF (e.g. through training or reference materials), but haven't published yet.: 10
I am not interested in publishing metabarcoding data to GBIF.: 4
I'm familiar with metabarcoding data and feel confident I could publish to GBIF, but haven't had the opportunity.: 3

Are you aware of any research groups, institutions, or monitoring programs in your node's network currently working with metabarcoding or eDNA data?

Yes: 36

Have you been in contact with these research groups, or do you see opportunities to support them in publishing or sharing data?

👉 **In summary:**
Most nodes **have engaged or are engaging** with research groups generating metabarcoding/eDNA data, often through **pilot projects, workshops, and publication support** (via IPT, MDT, or national portals). While some groups are already publishing datasets, others are in the **early stages of contact and exploration**. A minority report no contact yet, though they see **future opportunities** for collaboration.

1. Active Contact and Ongoing Support

- Many nodes reported **direct contact with research groups** and active support for publishing metabarcoding/eDNA data.
- Several are already **supporting publication** through GBIF's IPT or specialized portals (e.g., Swedish ASV portal, SIBUy, Academia Sinica, Colombian OBIS, INBO, Naturalis).
- Some groups are **already publishers**, while others are in the process of becoming publishers.

2. Testing and Use of Tools

- Nodes are experimenting with or using the **Metabarcoding Data Toolkit (MDT)** to support researchers.
- Some report that **IPT currently fulfills most needs**, with MDT being tested for potential future use.
- Workshops and demonstrations (e.g., UK DNA working group, Miwa FAIRe workshop) have been held to encourage adoption

3. Emerging Collaborations

- Contacts established with groups such as:
    - **Uruguayan Barcode of Life Initiative** (published 1 dataset, 2 more planned).
    - **Colombian OBIS node and academic institutions**.
    - **Africa Herbaria Consortium**.
    - **Universities working on marine biodiversity**.
    - **Ilia State University (Georgia, barcoding project)**.
- Several institutions have expressed **strong interest in publishing**, but concrete steps are still developing.

4. Positive but Limited Engagement

- Some nodes report **limited or early-stage contact** — interest has been expressed but **few datasets published yet**.
- Engagement often depends on **availability of resources** and **staff capacity**.
- In some cases, conversations are ongoing but have **not yet led to concrete collaboration**.

5. No or Uncertain Contact

- A small number of responses indicated **no contact yet** or uncertainty about whether national agencies have been in touch with relevant groups.

What challenges or questions do you anticipate in supporting the integration of metabarcoding data into GBIF?

👉 **In summary:** Nodes anticipate challenges in **standards and representation (MaterialEntity, metadata, data duplication)**, **workflow integration (MDT, IPT, automatic updates)**, and **ensuring data quality** for large, complex datasets. They highlight gaps in **training and capacity**, concerns about **governance and sensitive data**, and the need for **better infrastructure and institutional coordination**.

1. Standards, Models, and Data Representation

- Concerns about how well the **GBIF data model** can accommodate:
    - Sequence metadata, replicates, bioinformatic pipeline documentation.
    - The **MaterialEntity class** to represent environmental samples themselves (not just species occurrences).
    - Alignment with Darwin Core and emerging sequence data standards.
- Questions around **what exactly should be published where** (GBIF vs. GenBank vs. BOLD) to avoid duplication and confusion.

2. Tools, Workflows, and Infrastructure

- Need for **simplified, well-documented workflows** and examples to guide non-specialists.
- Support for **automatic workflows** (e.g., annotations, reporting, updating datasets).
- Integration of the **Metabarcoding Data Toolkit (MDT)** into national pipelines, ensuring smooth validation from MDT → IPT → GBIF.
- Requests for features such as:
    - Earlier DOI minting in IPT (for embargo periods).
    - Multi-institution dataset ownership.
    - Improved support for very large dataset downloads.

3. Data Quality, Cleaning, and Validation

- Existing validation tools don't cover metabarcoding/eDNA data; need **new approaches for quality assurance**.
- **Data cleaning practices differ by country/project**, requiring harmonization.
- Incomplete or inconsistent **metadata** and **reference libraries** (especially regionally) can limit accurate species identification.
- Risk of **reduced data integrity** when converting MDT outputs into occurrence-core-only systems (e.g., NBN Atlas).

4. Capacity, Training, and Support

- Strong need for:
    - **Training programs, mentorship, and regionally relevant resources**.
    - Building institutional and technical capacity to handle large sequence datasets.
    - Motivating and guiding researchers through publication workflows.
- Limited **staff expertise** in nodes on eDNA/metabarcoding data remains a barrier.

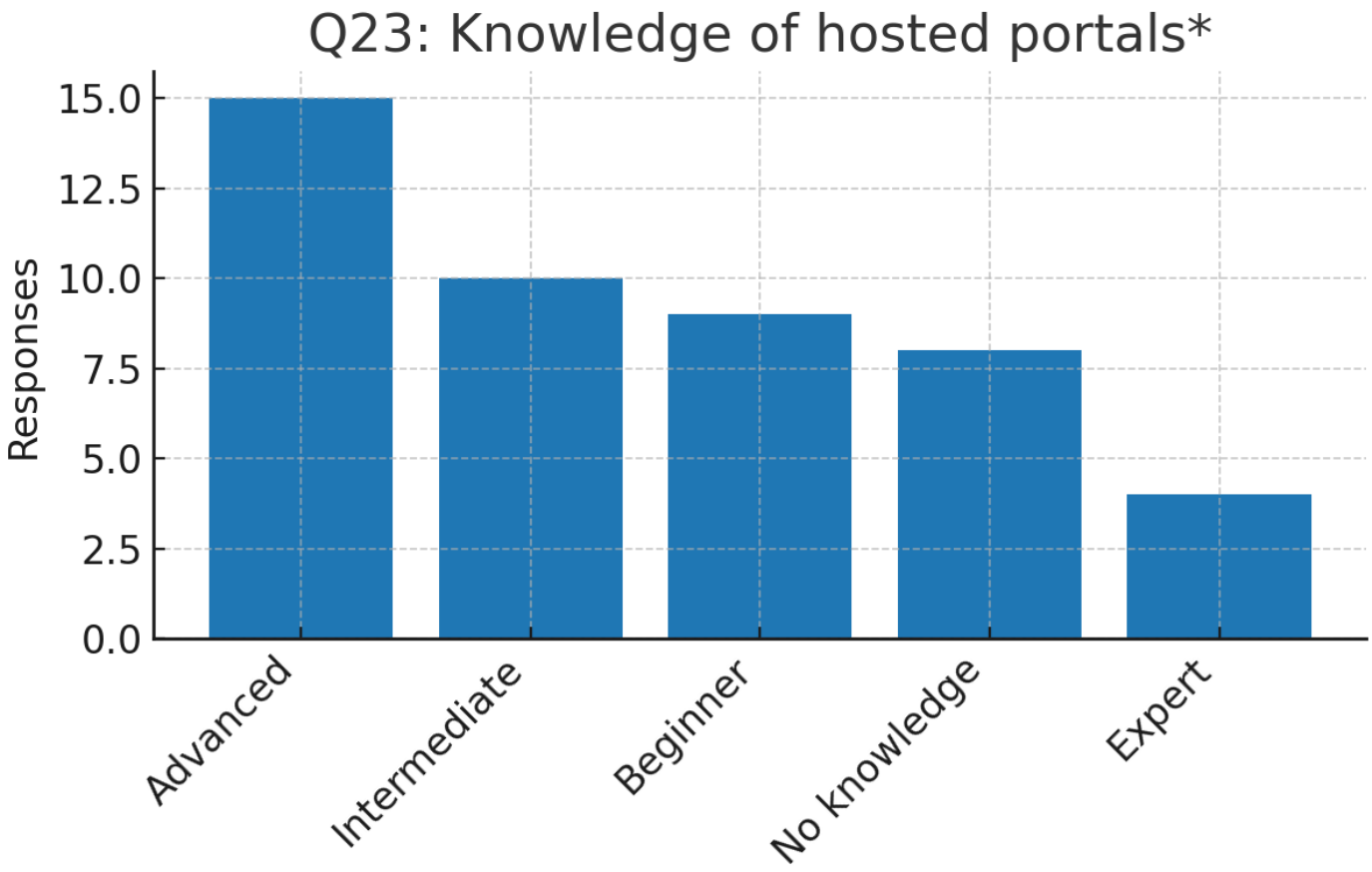
5. Governance, Sensitivity, and Coordination

- Concerns about **governance of genetic data**, intellectual property, and sensitivity (e.g., threatened or commercially valuable species).
- Questions on how to handle **restricted-access datasets**.
- Requirement for **delegated curation** of datasets from institutions publishing very large volumes.
- Challenges in **coordinating across diverse institutions** with different policies, infrastructures, and levels of readiness.

GBIF Hosted Portals

Knowledge of hosted portals

Advanced: 15
Intermediate: 10
Beginner: 9
No knowledge: 8
Expert: 4



Hosted portal group

Group 1 (Offering hosted portals): 15
Group 2 (Setting up first portal): 13
Group 3 (Advancing existing portal): 14
Not attending in Bogotá: 4

Have you completed the hosted portal application form? Not yet

Cameroon
Georgia
iDigBio
Mauritania
OBIS
Regional Support, Africa
Regional Support, ECA
Symbiota Support Hub
Zimbabwe

What feature do you want to add or improve on your hosted portal?

Canadensys, Legume Data Portal, North America regional portal

- We would like to add features like Institution search, dataset search, and collections search (We are currently working on that). Also, we would like to implement something similar to what Legume HP has: https://www.legumedata.org/taxonomy/browse, with the national taxonomic Backbone.
- Automatically replicate some GBIF news section in the news feed of the HP.

biodiversity.aq - https://sibuy.ambiente.gub.uy/

- We would like to filter in the map with other layers of information, for example, our protected areas.
- Maybe a filter by common name.
- A section connected with another platform for education.

CoL

- We do not have a hosted portal but from COL I would like to understand how we could implement Checklistbank capabilities on the hosted portals to display checklists using COL's api such as the legum portal https://www.legumedata.org/taxonomy/search, and what are the additional needs nodes have to better represent their national checklist on their own hosted portals

biodiversidad.co

- IUCN Threat Charts match with IUCN color coding.
- Others are in Github repo issues.

gbif france

- statistics of use

What is something you've learned about hosted portals that you want to make sure others know?

👉 In short: Hosted portals are **powerful, flexible, and relatively easy to manage** tools that provide access to diverse data and real-time integration with GBIF, but they require some ongoing effort to keep content updated.

1. Features and Capabilities

- Hosted portals allow access to **multiple data types** (occurrence, collections, literature).
- Useful features include:
    - **Dashboards** for visualization.
    - **Optional filters** (e.g., JSON code to customize default map display).
    - **Change of taxonomic backbone** to compare classifications.
    - **Integration of GBIF news feeds**.
    - **Translations** for broader accessibility.
- Strong **interoperability** with GBIF infrastructure, enabling **real-time updates and visualizations**.

2. Ease of Use and Management

- **Easier to set up than expected**, manageable even with a very small team.
- However, the **more content you add, the more maintenance is required** to keep it updated—important to plan team capacity accordingly.

3. Support

- **GBIF technical support is very helpful**, providing guidance and troubleshooting when needed.

Do you have any interest in sharing your hosted portal experience for participants in group 1?

- ASEAN Centre for Biodiversity (network)
- Belgian GBIF Node (GBIF participant - country)
- Canadensys (GBIF participant - associate)
- GBIF Uruguay (GBIF participant - country)
- SCAR (network)
- SiB Colombia (GBIF participant - country)
- LPWG (network - thematic)