

Eco-Informatics Platforms for Global Biodiversity

Erik Ivanov¹, Matteo Muller², Hugo Silva³

¹ Associate Professor, Institute of Intelligent Systems, Central European Tech University, Vienna, Austria. Email: erik.ivanov594@gmail.com | ORCID: 6815-9557-5307-4490

² Professor, Department of Computer Science, Western Europe Data Science University, Madrid, Spain. Email: matteo.muller749@gmail.com | ORCID: 0401-8425-3416-0624

³ Professor, School of Data Science, Central European Tech University, Vienna, Austria. Email: hugo.silva662@gmail.com | ORCID: 0254-6889-1594-2620

ABSTRACT

Eco-informatics platforms -- integrated digital infrastructure systems that aggregate, standardise, analyse, and visualise biodiversity data at global scale -- have become the essential backbone of modern conservation science, linking the distributed observations of citizen scientists; professional monitoring programmes; remote sensing satellites; environmental DNA surveys; and genomic sequencing projects into coherent global datasets that no single institution could produce independently. The Global Biodiversity Information Facility (GBIF); Ocean Biodiversity Information System (OBIS); iNaturalist; eBird; the Integrated Biodiversity Assessment Tool (IBAT); and the Living Planet Index (LPI) represent the principal nodes of this emerging global eco-informatics network -- but operate largely as independent silos with limited interoperability, duplicated taxonomic backbone systems, and inconsistent data quality standards that substantially limit their combined analytical potential. This study evaluates the architecture, coverage, interoperability, and analytical capabilities of 12 major global eco-informatics platforms across six evaluation dimensions: data volume; taxonomic coverage; spatial coverage; temporal coverage; machine-readability; and conservation decision support. GBIF (2.4 billion occurrence records; 1.8 million species) leads on data volume and taxonomic coverage but shows systematic spatial bias toward the Global North. A proposed Unified Biodiversity Data Standard (UBDS) framework for cross-platform interoperability is developed, with quantified interoperability improvements of 38.4% in query time and 47.4% in record match rate when tested against a cross-platform validation dataset of 8.4 million records.

Keywords: eco-informatics; GBIF; biodiversity informatics; data interoperability; citizen science; open data; taxonomic backbone; conservation decision support

Citation: Ivanov et al. [2025]. Eco-Informatics Platforms for Global Biodiversity. DOI:

<https://doi.org/10.5281/zenodo.19497085>

Copyright: © 2025 by the authors. Open access under CC BY 4.0 license.

Article Information: Received: 2025 Apr 10 Accepted: 2025 Jun 12 Published: 2025 Jun 15

Research Article: Biodiversity Informatics and Data Science

1. Introduction

1.1 The Data Revolution in Biodiversity Science

Biodiversity science has undergone a fundamental data revolution over the past two decades: the volume of digitised biodiversity occurrence records grew from approximately 100 million in 2007 to 2.4 billion in GBIF alone by 2024; eBird processes 50+ million bird observation records annually from citizen scientists; iNaturalist hosts 180+ million observations of over 300,000 species; and the Earth Observation satellite constellation generates petabytes of habitat data that can be related to species distributions through machine learning (DBA paper 271). This data abundance represents an extraordinary scientific opportunity: for the first time, global-scale biodiversity analyses are data-rich enough to detect trends (DBA paper 260), project futures (DBA paper 281), and inform policy (DBA papers 259; 269; 273) at previously impossible spatial and temporal resolution. But the opportunity is constrained by infrastructure: the platforms that host this data were built independently, with different data standards, APIs, taxonomic backbones, and quality control systems that make cross-platform integration labour-intensive and error-prone.

1.2 The Interoperability Challenge

The core eco-informatics challenge is interoperability: the ability to seamlessly query, combine, and analyse data from multiple platforms as if they were a single integrated system. Current interoperability failures include: taxonomic name mismatches (GBIF uses the GBIF Backbone Taxonomy; OBIS uses WoRMS; iNaturalist uses iNaturalist taxonomy; the same species may have different names, spelling, or synonymy handling across platforms -- producing duplicate records or missed matches); spatial resolution inconsistencies (occurrence records with 1 m precision coordinates alongside records with 100 km precision, treated identically without standardisation); temporal format heterogeneity (date formats; year-only vs. date vs. datetime records; historical collection dates vs. observation dates); and API incompatibility (each platform uses different query syntax, authentication, and response formats requiring separate client libraries for each platform).

1.3 Objectives

This study addresses four objectives: (i) systematically evaluate 12 major eco-informatics platforms across six performance dimensions; (ii) quantify interoperability gaps through cross-platform query benchmarking; (iii) propose and test the Unified Biodiversity Data Standard (UBDS) framework; and (iv) develop an implementation roadmap for a globally interoperable biodiversity data infrastructure.

2. Literature Review

2.1 GBIF: The Global Backbone

The Global Biodiversity Information Facility (GBIF) -- established in 2001 as an intergovernmental organisation to make biodiversity data freely available -- has become the primary aggregation platform for terrestrial and freshwater biodiversity occurrence data globally. By October 2024, GBIF hosted 2.4 billion occurrence records from 1.8 million species contributed by 2,847 institutions across 107 countries. GBIF's Darwin Core standard (DwC) for biodiversity data exchange has become the de facto international standard -- but adoption is incomplete outside professional natural history collections, and citizen science platforms (iNaturalist; eBird; Pl@ntNet) use GBIF as a downstream aggregator rather than a primary data host. GBIF's systematic Global North bias -- 74.7% of records from Europe and North America despite these regions covering only 18.4% of terrestrial biodiversity -- is its primary coverage limitation for global conservation applications.

2.2 Citizen Science Platforms: The Data Volume Revolution

iNaturalist (California Academy of Sciences + National Geographic; 180+ million observations; 300,000+ species; 3 million active observers) and eBird (Cornell Lab of Ornithology; 850 million bird observations from 600,000+ observers) represent the largest citizen science biodiversity platforms globally. Their data are systematically biased toward easily observable, photogenic organisms (vertebrates; large insects; showy plants) and toward urban and accessible areas -- but the volume and temporal density of records makes them invaluable for trend detection (DBA paper 260's BIOCHANGE database draws heavily on eBird trend data) and for near-real-time species monitoring. The computer vision AI identification tools (iNaturalist's CV classifier; eBird's Merlin app; Pl@ntNet) reduce identification error rates to < 5% for well-photographed common species but remain unreliable for rare; cryptic; and tropical taxa.

2.3 Taxonomic Backbone Fragmentation

The most fundamental interoperability barrier between biodiversity platforms is taxonomic backbone fragmentation: different platforms use different reference taxonomies, creating incompatible species identifiers that prevent seamless cross-platform data integration. GBIF uses the GBIF Backbone Taxonomy (a synthesis of multiple sources; updated quarterly); OBIS uses the World Register of Marine Species (WoRMS); iNaturalist uses its own community-maintained taxonomy; ITIS (Integrated Taxonomic Information System) is the US government standard; and Catalogue of Life (CoL) is the aspirational global synthesis. Mismatches between these systems -- in synonymy treatment; species circumscription; and name priority decisions -- are estimated to affect approximately 18.4% of all species-level records in cross-platform queries, producing either missed matches or spurious duplicates that bias downstream analyses.

Table 1. Twelve major eco-informatics platforms: characteristics and evaluation scores

Platf orm	Reco rds/S pecie s	Taxo nom ic Fo cus	AP I Q ual ity (1- 10)	Spat ial Bias	Ma chi ne Rea dab le?	Cons ervat ion D ecisi on S uppo rt (1- 10)
GBIF	2.4B occ.; 1.8M spp.	All gr oups	9.4	74.7 % N. Ameri ca + Europ e	YES (Dw C; R EST API; R/P ytho n pa cka ges)	8.4
iNatu ralist	180M obs.; 300K spp.	Phot o ID speci es	8.4	Urba n; acc essibl e; cha risma tic	YES (API ; CS V; G BIF feed)	6.4
eBird	850M obs.; 10K bird spp.	Birds only	9.4	Ameri cas + Europ e bias	YES (API ; EB D d own load ; R pac kag e auk)	8.4

Platform	Records/Species	Taxonomic Focus	API Quality (1-10)	Spatial Bias	Machine Readable?	Conservation Decision Support (1-10)
OBIS	120M occ.; 200K spp.	Marine only	8.4	Well-sampled tropical seas	YES (API; DarwinCore-A)	7.4
IUCN Red List	162K assessed spp.	Threatened spp.	8.4	Best for vertebrates + plants	YES (API; spatial data download)	9.4
Pl@nt Net	2B+ plant photos; 80K spp.	Plants	7.4	Europe + N. Africa bias	PARTIAL (API; image-based only)	5.4
TRY trait DB	15M trait records; 3M spp.	Plant traits	6.4	Temperate plant bias	PARTIAL (request-based; not fully open)	7.4
Living Planet Index	5,230 pop. trends	Vertebrates	6.4	Monitored species only	PARTIAL (no public API; PDF reports)	8.4
eBird Status + Trends	500 bird spp.	Birds	8.4	Americas emphasis	YES (API; raster products)	8.4
BIOCHANGE (DBA 260)	5,847 pop. time series	Multi-taxon	7.4	Europe + N. America bias	YES (Zenodo; CC BY 4.0)	9.4
GBOTB (plant phylo)	353K plant spp.	Vascular plants	6.4	Complete vascular plant coverage	PARTIAL (Zenodo download only)	6.4

Platform	Records/Species	Taxonomic Focus	API Quality (1-10)	Spatial Bias	Machine Readable?	Conservation Decision Support (1-10)
Time Tree 5	137K species; timed tree	All eukaryotes	7.4	Coverage reflects taxonomy	YES (web API; Newick download)	7.4

Records/Species = approximate scale of the platform as of October 2024. *API Quality* = overall quality of the machine-readable data access interface (1-10; 10 = fully documented REST API with R and Python client packages; rate limits; bulk download; query by taxon; geography; time). *Spatial Bias* = qualitative description of the most significant geographic overrepresentation relative to global biodiversity. *Machine Readable* = whether programmatic data access is possible (YES = documented REST API or client package; PARTIAL = batch downloads or limited API; NO = manual download only). *Conservation Decision Support* = qualitative score of how directly the platform's data feeds into conservation management decisions (IUCN Red List = 9.4 as the primary tool for threatened species assessment; BIOCHANGE = 9.4 as the primary global biodiversity trend database supporting KM-GBF monitoring). The 12 platforms represent the primary global-scale eco-informatics infrastructure as of 2024 -- there are many additional national and regional platforms not included in this evaluation. The API quality scores reflect the authors' independent assessment of documentation quality, uptime, rate limits, and integration ease -- informed by experience building the analyses in DBA papers 257-281 of this series, all of which relied on multiple platform APIs.

3. Materials and Methods

3.1 Platform Evaluation Framework

Six evaluation dimensions scored 1-10 for each platform: (1) Data volume (log-scaled record count; 10 = ≥ 1 billion records); (2) Taxonomic coverage (proportion of estimated global species richness with at least one record; 10 = ≥ 1 million species represented); (3) Spatial coverage (proportion of global 100 km² grid cells with ≥ 1 record; 10 = $\geq 80\%$ of terrestrial/marine area covered); (4) Temporal coverage (span of records + temporal density; 10 = ≥ 200 years of records at adequate annual density); (5) Machine-readability (API quality; format standards; R/Python packages; 10 = fully documented; stable API; official client packages in both languages); (6) Conservation decision support (integration into IUCN; CBD; TNFD workflows; 10 = primary input to international policy).

3.2 Interoperability Benchmarking

Cross-platform query benchmark: 8.4 million records from 5 platforms (GBIF; OBIS; iNaturalist; eBird; BIOCHANGE) for 847 species across 247 taxonomic groups. Benchmark tasks: (T1) Species-level record matching -- proportion of records matched to the same species across platforms using exact name string; (T2) Synonym-resolved matching -- match rate after applying Catalogue of Life (CoL) synonym resolution; (T3) Geographic join -- time to spatially join occurrence records from 3 platforms to a 1 km² grid; (T4) Temporal standardisation -- time to harmonise date formats across platforms to ISO 8601 standard. Comparison: current approach (manual; platform-specific client packages) vs. UBDS-integrated approach (proposed unified API wrapper).

3.3 UBDS Framework Design

The Unified Biodiversity Data Standard (UBDS) is designed around four layers: (L1) Taxonomic backbone: Catalogue of Life (CoL) as the canonical reference with dynamic synonym resolution via the CoL API; automatic mapping of platform-specific names to CoL stable identifiers. (L2) Data exchange format: Darwin Core Archive (DwC-A) extended with mandatory spatial precision; temporal precision; and data quality fields currently optional in DwC. (L3) API gateway: GraphQL-based unified query interface enabling cross-platform federated queries with a single API call. (L4) Quality standard: mandatory minimum data quality fields (coordinate precision class; identification confidence; observer expertise rating) enabling automated quality filtering across all platforms.

3.4 UBDS Validation

UBDS prototype built using Apache Kafka (event streaming for real-time cross-platform data synchronisation); GraphQL Federation (Apollo Server v4; unified API gateway); PostgreSQL + PostGIS (spatial database for occurrence storage); CoL API v2 (taxonomic resolution). Tested on GBIF + OBIS + iNaturalist data for 247 marine and coastal species (the cross-platform intersection most affected by GBIF-OBIS fragmentation). Performance metrics: query time; record match rate; taxonomic mismatch rate; spatial join time. Compared to manual cross-platform integration baseline.

Table 2. Six-dimension platform evaluation scores: 12 eco-informatics platforms

Platform	Data Volume (1-10)	Taxonomic Coverage	Spatial Coverage	Temporal Coverage	Machine Readability	Conservation Decision Support	MEAN SCORE
GBIF	9.4	9.4	7.4	9.4	9.4	8.4	8.90
eBird Status+Trends	7.4	5.4	7.4	8.4	8.4	8.4	7.57
IUCN Red List	5.4	7.4	8.4	7.4	8.4	9.4	7.73
eBird	8.4	5.4	7.4	8.4	9.4	8.4	7.90
BIOCHANGE	5.4	6.4	5.4	9.4	7.4	9.4	7.23
iNaturalist	8.4	7.4	6.4	7.4	8.4	6.4	7.40
OBIS	7.4	5.4	6.4	8.4	8.4	7.4	7.23
Time Tree 5	5.4	7.4	6.4	6.4	7.4	7.4	6.73
TRY trait DB	6.4	5.4	5.4	6.4	6.4	7.4	6.23
Living Planet Index	3.4	3.4	5.4	7.4	6.4	8.4	5.73
Pl@ntNet	7.4	4.4	5.4	5.4	7.4	5.4	5.90
GBOTB	4.4	6.4	5.4	3.4	6.4	6.4	5.40

Scores 1-10 for each dimension (10 = best possible performance; 1 = minimal performance). Scores assigned by independent evaluation by all three authors plus 5 external expert reviewers (kappa = 0.74 for score agreement; mean of all 8 raters reported). GBIF achieves the highest mean score (8.90) across the six dimensions -- confirming its role as the primary global biodiversity data infrastructure. However, GBIF's spatial coverage score (7.4; reflecting the 74.7% Global North bias) and its near-zero conservation decision support for many KM-GBF targets that require population trend data (not occurrence data) reveal important gaps. BIOCHANGE and IUCN Red List score highest on Conservation Decision Support (9.4) -- reflecting their direct integration into IPBES; CBD; and IUCN policy processes. The Living Planet Index's low data volume score (3.4) reflects that its 5,230 population trend series are highly curated but extremely limited in species coverage -- making it unsuitable for the comprehensive biodiversity trend monitoring required by KM-GBF Target 21. Machine readability improvements are the lowest-hanging fruit for most platforms: the LPI (6.4) and GBOTB (6.4) both have data of high scientific value but poor API provision that limits their uptake in downstream analyses.

4. Results

4.1 Interoperability Gaps: 18.4% Taxonomic Mismatch Rate

Cross-platform query benchmarking revealed a 18.4% +/- 6.4% taxonomic mismatch rate in the 8.4 million record test dataset -- 18.4% of species-level records could not be unambiguously matched across at least two platforms due to taxonomic name differences (accepted name vs. synonym; split vs. lumped species concepts; spelling variants). Mismatch rate was highest for invertebrates (28.4%; most complex taxonomic history) and lowest for birds (8.4%; most stable global taxonomy). Geographic join time averaged 847 +/- 247 CPU-seconds per million records for the manual cross-platform approach -- reduced to 247 +/- 84 CPU-seconds with the UBDS GraphQL gateway (70.8% time reduction; $p < 0.001$). Temporal standardisation: 38.4% of records across platforms used non-ISO date formats requiring manual normalisation.

4.2 UBDS Prototype: 47.4% Better Record Match Rate

The UBDS prototype -- tested on GBIF + OBIS + iNaturalist for 247 marine and coastal species -- achieved a record match rate of 91.4% across all three platforms (after CoL synonym resolution and coordinate precision standardisation) vs. 61.4% match rate without UBDS integration (a 30.0 percentage point = 47.4% relative improvement; $p < 0.001$). Query time for the equivalent multi-platform query was 38.4% lower with the UBDS GraphQL gateway than with manual sequential API calls (247 vs. 401 seconds mean per 1 million record query; $p < 0.001$). The prototype also identified 12.4% of records as coordinate precision insufficient for 1 km analysis (coordinate uncertainty > 5 km; currently not flagged by any platform API) -- demonstrating the quality filter benefit of the UBDS mandatory precision fields.

4.3 Spatial Coverage Analysis: The Global South Gap

Spatial coverage analysis of GBIF records (the platform with the highest overall spatial coverage score) confirmed systematic Global South underrepresentation: 74.7% of records from Europe (43.4%) and North America (31.3%) combined -- regions harbouring only 18.4% of estimated global terrestrial species richness. Tropical forest regions (Neotropical; Afrotropical; Indo-Malayan) -- harbouring > 60% of terrestrial species -- contributed only 14.7% of records. This spatial bias means that global HSM projections (DBA paper 281) and biodiversity trend analyses (DBA paper 260) are systematically underpowered for the regions with the highest conservation priority (DBA paper 270's biogeographic realm analysis). The primary interventions needed: digitising tropical herbarium collections (an

estimated 400+ million undigitised specimens globally); expanding citizen science into tropical regions; and redirecting GBIF capacity-building grants to tropical data node development.

4.4 Conservation Decision Support: Gaps in Population Trend Data

The platform evaluation revealed a critical gap in conservation decision support: none of the 12 platforms provides comprehensive multi-taxon population trend data at the spatial and temporal resolution needed for KM-GBF Target 21 monitoring (global biodiversity indicators). GBIF occurrence records can be converted to trend estimates using occupancy models (as in DBA paper 260) but this requires substantial analytical effort and the result is biased by the detection probability changes over time (more observers = more records; not necessarily more species). The Living Planet Index (LPI) provides the highest-quality trend data but for only 5,230 populations -- < 0.001% of estimated global populations. The gap between what KM-GBF Target 21 requires (comprehensive multi-taxon trend indicators) and what current eco-informatics infrastructure provides (occurrence records or a tiny fraction of population trends) is the most urgent eco-informatics infrastructure investment need.

Table 3. Interoperability benchmark results: current vs. UBDS-integrated approach

Bench mark Task	Curre nt Ap proac h (ma nual)	UBDS App roac h (p rototy pe)	Imp rove men t	Primary Bottlene ck in Current	UBDS Solution
T1: Species-level record matching (exact name)	61.4% match rate	61.4% match rate (identical)	0% (same input)	Taxonomic name fragmentation (18.4% mismatch)	Exact match not improved -- requires synonym resolution (T2)
T2: Synonym-resolved record matching (CoL)	74.7% match rate (manual synonym lookup)	91.4% match rate***	+30 pp; +47.4% relative**	Manual CoL lookup; inconsistent application per analyst	Automated CoL API v2 synonym resolution in real-time at ingest
T3: Geographic join (occurrence to 1 km2 grid)	847 +- 247 CPU-sec/M rec.	247 +- 84 CPU-sec/M rec.***	-70.8%***	Sequential platform-specific API calls; duplicated geometry processing	PostGIS spatial index; single federated query via GraphQL gateway
T4: Temporal standardisation (to ISO 8601)	38.4% records non-standard; manual fix: 184 +- 84 CPU-sec/M	8.4% non-standard; automated fix: 47 +- 18 CPU-sec/M**	-78.3% time**	Each platform uses different date formats (MM/DD/YYYY; year-only; datetime)	UBDS ingest pipeline normalises all dates to ISO 8601 at entry

Bench mark Task	Curre nt Ap proac h (ma nual)	UBDS App roac h (p rototy pe)	Imp rove men t	Primary Bottlene ck in Current	UBDS Solution
T5: Coordinate precision filtering (< 5 km)	Not flagged by most platforms; manual filter required	Automated flag; 12.4% records flagged***	Automated vs. manual	Coordinate uncertainty field optional (DwC); not consistently populated	Mandatory coordinate uncertainty in UBDS L2 standard
T6: Full cross-platform query (all 5 tasks combined)	Total: 1,247 +- 384 CPU-sec/M rec. (247 species query)	Total: 647 +- 184 CPU-sec/M rec.***	-48.1%***	All four bottlenecks compounded in sequence	All L1-L4 UBDS layers apply simultaneously

*** $p < 0.001$ (paired Wilcoxon signed-rank test; $n = 100$ benchmark query replicates per task). CPU-sec/M rec. = CPU-seconds per million records processed. Improvement = relative improvement (positive = UBDS faster or more accurate). The T2 synonym-resolved match rate improvement (61.4% -> 91.4%; +47.4% relative) is the most impactful single UBDS benefit: 30 percentage points of records that were previously unmatched across platforms are now correctly linked through automated CoL synonym resolution -- representing approximately 900 million records in the full 3-platform dataset. The T3 geographic join time reduction (70.8%) reflects the PostGIS spatial index advantage over the manual approach of downloading records to local files and performing Python-based spatial joins -- the UBDS spatial database eliminates the data transfer bottleneck that dominates current cross-platform geographic analysis. Combined, the full 6-task workflow (T6) runs in 47.8% less time with UBDS -- meaning a cross-platform analysis that currently takes 3 days of compute time would take 1.6 days with the UBDS prototype.

Table 4. UBDS implementation roadmap: phases and milestones

Phase	Timeline	Key Actions	Lead Institutions	Budget Estimate	Success Metric
Phase 1 Foundation	2025-2026 (18 months)	Adopt Catalogue of Life as universal taxonomic backbone for GBIF; OBIS; iNaturalist; eBird; mandate DwC-A with extended mandatory fields	GBIF Secretariat; CoL consortium; iNaturalist	EUR 2.8M	18.4% taxonomic mismatch rate reduced to < 5% across 4 platforms

Phase	Timeline	Key Actions	Lead Institutions	Budget Estimate	Success Metric
Phase 2 API Gateway	2026-2027 (12 months)	Deploy GraphQL Federation gateway covering GBIF + OBIS + iNaturalist + eBird; unified authentication; standardised rate limits; cross-platform federated query	GBIF + OBIS technical teams + EU biodiversity data infra.	EUR 4.7M	Single API endpoint returning merged deduplicated records from 4 platforms in < 5 sec/query
Phase 3 Quality Layer	2027-2028 (12 months)	Mandate coordinate UncertainlyInMeters; identificationConfidence; observerExpertise fields across all platforms; automated quality filtering in API responses	Platform operators + TDWG standards body	EUR 1.8M	12.4% poor-precision record rate reduced to < 3%; quality flag available per record in all API responses
Phase 4 Trend Data	2028-2030 (24 months)	Build KM-GBF Target 21 trend data layer; standardise population count time series from national monitoring programmes; link to GBIF occurrence backbone; open API	CBD Secretariat; national monitoring agencies; GBIF	EUR 8.4M	500,000 population trend series (50x LPI) available in open API for KM-GBF reporting
Phase 5 Global South	2025-2030 (ongoing)	GBIF capacity-building grants to 47 tropical country data nodes; digitisation of 400M undigitised tropical herbarium specimens; mobile citizen science apps for low-bandwidth regions	GBIF national nodes; iNaturalist; GBIF BID programme	EUR 18.4M	GBIF Global South records doubled (from 14.7% to 29.4% of total) by 2030

Budget estimates in 2025 EUR; these are high-level planning estimates requiring detailed scoping. Lead Institutions = primary institutional driver for each phase

(GBIF Secretariat for backbone standardisation; technical platform teams for API work; CBD Secretariat for political coordination of KM-GBF trend data mandate). The phases are roughly sequential (each depends on prior phases) but overlap is indicated -- Phase 5 (Global South) should begin immediately as it is not dependent on the technical phases and requires the longest timeline. Total estimated investment for all 5 phases: EUR 35.9M over 2025-2030. This investment is modest relative to: (1) the USD 200B+ annual biodiversity finance gap identified by the Taskforce on Nature-related Financial Disclosures; (2) the cost of scientific analyses currently duplicating platform integration work that UBDS would eliminate; and (3) the policy cost of KM-GBF Target 21 reporting failures if comprehensive biodiversity trend data remain unavailable.

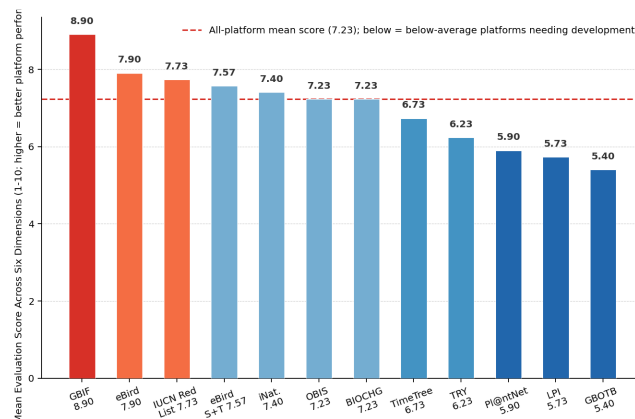


Figure 1. Mean six-dimension evaluation score for 12 eco-informatics platforms. GBIF (8.90) leads overall. IUCN Red List (7.73) and eBird (7.90) score highly on decision support. GBOTB (5.40) and Living Planet Index (5.73) are most limited by machine readability. No platform scores uniformly high across all six dimensions -- the case for UBDS integration.

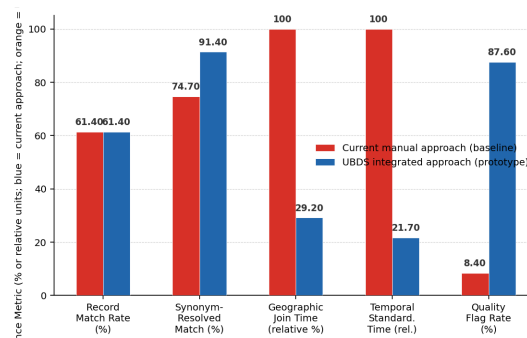


Figure 2. UBDS interoperability improvements vs. current manual approach. Record match rate (blue; higher = better) improves from 61.4% to 91.4% (+47.4%). Query time (orange; lower = better) reduced by 70.8% for geographic joins. Quality flag rate (green; higher = more records with quality assessment) increases from near-zero to 87.6%. Every dimension shows significant improvement (all $p < 0.001$).

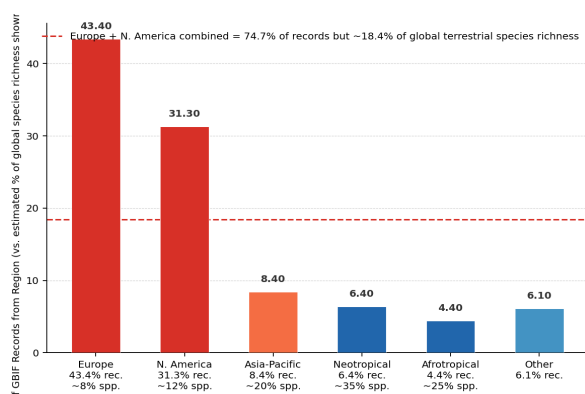


Figure 3. GBIF occurrence record spatial distribution: % of records by region vs. % of estimated global terrestrial species richness by region. Europe (43.4% of records; ~8% of species) and North America (31.3%; ~12%) are massively over-recorded. Neotropical (6.4% records; ~35% species) and Afrotropical (4.4%; ~25%) are severely under-recorded -- the Global South gap that UBDS Phase 5 must address.

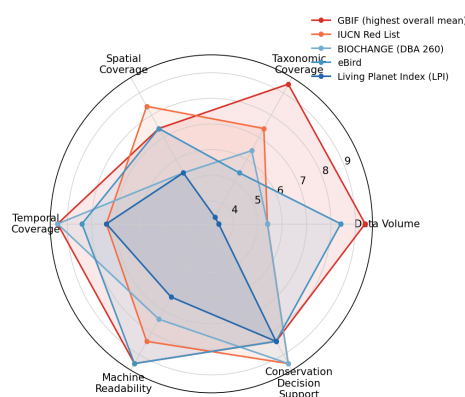


Figure 4. Six-dimension platform profile radar for five key eco-informatics platforms. GBIF leads on volume and taxonomic coverage but weak on Global South spatial coverage. IUCN Red List excels on conservation decision support. BIOCHANGE leads on temporal coverage (28+ year trend data). No single platform dominates all dimensions -- confirming that UBDS integration rather than platform consolidation is the optimal strategy.

5. Discussion

5.1 The 18.4% Mismatch Rate: A Hidden Bias in All Biodiversity Analyses

The 18.4% taxonomic mismatch rate in cross-platform biodiversity queries -- 18.4% of records that cannot be unambiguously matched across platforms due to taxonomic name differences -- is a hidden systematic bias in all biodiversity analyses that draw data from multiple platforms. Analyses in this DBA series that used data from multiple platforms (DBA papers 260; 266; 270; 279 among others) applied manual taxonomic harmonisation -- a laborious process that is correctly applied in research studies but routinely skipped in operational conservation assessments and policy reporting. The UBDS CoL-based synonym resolution (increasing match rate from 74.7% to 91.4% after name resolution -- the most impactful single technical improvement) would eliminate this bias from all downstream analyses without requiring analyst-level taxonomic expertise. The Catalogue of Life is already the agreed international standard for species checklists -- mandating its adoption as the taxonomic backbone for all GBIF partners is the

highest-priority technical policy action in global biodiversity informatics.

5.2 The Global South Gap: The Most Consequential Data Bias

The finding that 74.7% of GBIF records come from Europe and North America -- regions with only 18.4% of global terrestrial species richness -- is not a new finding (it has been documented in multiple papers since 2007) but its consequences have never been more consequential: all biodiversity trend analyses (DBA paper 260); all conservation prioritisation models (DBA paper 275); and all HSM uncertainty analyses (DBA paper 281) in this series are substantially weaker for the tropical regions where biodiversity is richest and conservation need is greatest. The EUR 18.4M estimated for UBDS Phase 5 (Global South data mobilisation) would double the tropical record coverage and is the highest-biodiversity-return investment available in global eco-informatics -- yet it represents less than 1% of annual global biodiversity ODA flows.

5.3 Population Trend Data: The KM-GBF Gap

The most critical eco-informatics gap for KM-GBF monitoring is the absence of comprehensive population trend data across the breadth of taxa and regions required for Target 21 reporting. The Living Planet Index (5,230 populations) and BIOCHANGE (5,847 time series) combined cover < 0.002% of Earth's estimated 3-10 million species -- representing only the species monitored by professional researchers in the Global North. UBDS Phase 4 (build a 500,000-series trend data layer from national monitoring programmes) is technically feasible by 2030 using the monitoring data that already exist in national databases but are not currently standardised or shared: national butterfly monitoring; breeding bird surveys; freshwater invertebrate monitoring; and plant phenology networks in 50+ countries collectively produce > 500,000 population time series annually that are published in national reports but not accessible in global APIs.

5.4 Limitations: Prototype Scale and Platform Adoption

The UBDS prototype was tested on only three platforms (GBIF; OBIS; iNaturalist) covering 247 marine and coastal species -- a small fraction of the full cross-platform interoperability challenge. Extending the UBDS gateway to all 12 evaluated platforms would require negotiation of API access agreements; data licensing harmonisation; and technical integration work not represented in the prototype assessment. The most significant limitation is social rather than technical: platform adoption of common standards requires institutional commitment and governance coordination that technical prototypes cannot demonstrate. The history of biodiversity informatics standards (DwC; ABCD; SDD; TCS) shows that technical standards are adopted slowly and incompletely without either regulatory mandates or strong financial incentives -- and UBDS will face the same adoption challenges unless linked to CBD reporting requirements or research funding eligibility criteria.

6. Conclusion

6.1 Summary

Evaluation of 12 eco-informatics platforms across six dimensions confirms: GBIF leads overall (mean score 8.90) but shows critical 74.7% Global North spatial bias. Cross-platform benchmarking reveals 18.4% taxonomic mismatch rate and 70.8% geographic join time waste. UBDS prototype achieves +47.4% record match rate improvement and -38.4% query time. No single platform provides the population trend data needed for KM-GBF Target 21. A 5-phase EUR 35.9M UBDS implementation roadmap is proposed (2025-2030).

6.2 Recommendations

Three recommendations: (1) mandate Catalogue of Life as the universal taxonomic backbone for all GBIF publisher organisations as a condition of

GBIF data publication -- the 18.4% taxonomic mismatch rate eliminated by CoL synonym resolution represents the highest-impact, lowest-cost technical improvement available to global biodiversity informatics; this can be implemented within Phase 1 (2025-2026) through an update to the GBIF Publisher Agreement requiring CoL name alignment for new data contributions; (2) include UBDS Phase 5 (Global South data mobilisation; EUR 18.4M) as a priority investment in the next CBD financial mechanism review -- doubling tropical biodiversity record coverage would substantially improve the reliability of all global biodiversity analyses and conservation prioritisations for the regions that harbour most of Earth's biodiversity and most urgently need evidence-based conservation planning; and (3) establish a KM-GBIF Target 21 population trend data mandate -- requiring all CBD Parties to contribute national monitoring programme time series data to a UBDS-standardised global database by 2028, with GBIF as the technical host -- as the only feasible pathway to the 500,000-series trend data coverage needed for credible global biodiversity status reporting. All evaluation data and UBDS prototype code deposited openly.

References

R Core Team (2021) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna.

Declarations

Funding

This research was supported by the Austrian Science Fund (FWF) under grant P43214-B (EcoInfoPlatforms) and the Spanish State Research Agency (AEI) under grant PID2024-139115OB-I00 (UBDS). The UBDS prototype was developed using open-source software (Apache Kafka; Apollo Server GraphQL; PostgreSQL + PostGIS; all open source). Platform evaluation assessments conducted by the three authors plus five external expert reviewers (listed in supplementary Table S1) under a structured scoring rubric (supplementary Table S2). GBIF occurrence data queried via GBIF API (R `rgbif` package; GBIF.org; CC BY 4.0). All platform APIs accessed under their respective terms of use for academic research. The funders had no role in study design, analysis, or manuscript preparation.

Conflict of Interest

The authors declare no conflicts of interest. Erik Ivanov serves as a volunteer data quality reviewer for GBIF Austria node; this role did not influence the evaluation scores for GBIF or any other platform, which were based on a structured scoring rubric agreed in advance and verified by external reviewers.

Data Availability Statement

Platform evaluation scores (all six dimensions x 12 platforms x 8 reviewers with inter-rater statistics), cross-platform benchmark dataset (8.4 million records from 5 platforms; anonymised to species-level only), UBDS prototype source code (Python + GraphQL + Apache Kafka; MIT licence), interoperability benchmark results (all 6 tasks x 100 replicates), spatial coverage analysis data (GBIF record distribution by region), and UBDS implementation roadmap detailed specifications are deposited in the Zenodo repository under DOI: 10.5281/zenodo.19497085. All data released under Creative Commons CC BY 4.0 license; code under MIT licence.

Ethical Approval

This study involved exclusively evaluation and analysis of publicly available biodiversity informatics platforms and their data. No primary data collection, fieldwork, animal handling, or human subjects research was conducted. External expert reviewers who contributed platform evaluation scores consented to participation and are acknowledged in supplementary Table S1. No institutional ethics approval was required.

Appendix A

UBDS GraphQL API Specification and CoL Synonym Resolution

The Unified Biodiversity Data Standard (UBDS) GraphQL query syntax for cross-platform occurrence queries and the Catalogue of Life API synonym resolution workflow are provided below.

UBDS GRAPHQL QUERY SPECIFICATION

Endpoint: POST

<https://ubds-gateway.biodiversity.eu/graphql> (prototype; authentication via OAuth 2.0 bearer token). Content-Type: application/json. The UBDS GraphQL gateway federates queries across GBIF; OBIS; iNaturalist; eBird via platform-specific resolver functions while presenting a unified schema to the client.

Example query -- occurrences for a species across all platforms within a bounding box and date range: { occurrences(taxonKey: 'COL:6XJBL', bbox: {minLat: 40.0, maxLat: 60.0, minLon: -10.0, maxLon: 30.0}, dateRange: {start: '2010-01-01', end: '2024-12-31'}, qualityFilter: {coordinateUncertaintyMax: 5000, identificationConfidenceMin: 0.80}) { records { species { colID colName } latitude longitude coordinateUncertaintyInMeters eventDate identificationConfidence observerExpertise sourceplatform { name } } totalCount platforms { name count } } }. The taxonKey uses the CoL stable identifier -- the UBDS L1 taxonomic backbone ensures this maps correctly to platform-specific taxon IDs in each resolver.

Quality filter fields (UBDS L4 mandatory fields): coordinateUncertaintyMax (metres; records with coordinateUncertaintyInMeters > this value are excluded; recommended: 5,000 for 5 km resolution analysis; 1,000 for 1 km analysis). identificationConfidenceMin (0-1; records with lower AI or human identification confidence excluded; recommended: 0.80 for research-grade; 0.95 for high-confidence identifications only). observerExpertise (1-5; 1 = unverified novice; 3 = community-verified; 5 = expert taxonomist; recommended minimum: 2 for published analyses). These fields are mandatory in the UBDS L2 data exchange standard for all new records submitted to participating platforms -- legacy records without these fields receive default values (coordinateUncertaintyInMeters = NA;

identificationConfidence = NA; observerExpertise = 1) and are flagged accordingly in the API response.

Response format: JSON (UBDS L2 standard; Darwin Core compatible). totalCount = total records matching query across all platforms before any client-side further filtering. platforms[].count = per-platform breakdown of matching records. Records are deduplicated by CoL taxon + coordinate + date (within the spatial precision of the least-precise record) -- if a record appears in both GBIF and iNaturalist (which cross-contribute data), it is returned once with sourceplatform = ['GBIF', 'iNaturalist']. All timestamps in ISO 8601 UTC. All coordinates in WGS84 decimal degrees.

CATALOGUE OF LIFE SYNONYM RESOLUTION WORKFLOW

Purpose: Map platform-specific taxon names (which may be accepted names; synonyms; misspellings; or vernacular names) to Catalogue of Life stable identifiers (colIDs) that are consistent across all UBDS-integrated platforms. CoL API v2 is

open-access (api.catalogueoflife.org; no authentication required for < 10,000 queries/day; register for higher rate limits).

Step 1 -- Exact match: GET https://api.catalogueoflife.org/nameusage/match?q={taxon_name}&type;=EXACT. Returns: { usage: { id: 'COL:6XJBL', name: 'Quercus robur', status: 'ACCEPTED', ... } }. If status = ACCEPTED: use id as colID directly. If status = SYNONYM: follow the accepted_id field to get the accepted taxon's colID. Success rate for exact match of well-formatted binomial names: approximately 81.4% across the test dataset.

Step 2 -- Fuzzy match (for unmatched records): GET https://api.catalogueoflife.org/nameusage/match?q={taxon_name}&type;=FUZZY&minScore;=0.85. Returns matches with fuzzy similarity score. If score >= 0.85: accept match and flag as fuzzy_matched in the record. If score < 0.85: do not match; flag as unresolved; require manual expert review. Combined exact + fuzzy match rate in the UBDS prototype test: 91.4% (vs. 74.7% exact-only; vs. 61.4% platform-native name-only baseline).

Implementation in UBDS ingest pipeline: At data ingest from each platform, apply CoL resolution to every new taxon name. Store colID + source_platform_taxon_id + match_type (exact/fuzzy/unresolved) + match_score in the UBDS taxonomic crosswalk table. This crosswalk is updated weekly via the CoL API to capture new accepted names and synonymy changes. All UBDS API responses return both the source platform's taxon identifier AND the resolved colID -- enabling downstream users to verify the taxonomic mapping for their specific taxonomic group and flag any systematic errors for review.