Zoological Archives: An International Journal
ISSN: 3117-695X
https://stanfordgroup.org/index.php/ZAIJ/issue/archive

Digital Archiving Standards for Zoological Collections

Matteo Muller[1], Jonas Hansen[2]

[1] Senior Lecturer, Department of Artificial Intelligence, Swiss Institute of Machine Intelligence, Zurich, Switzerland. Email: matteo.muller587@yahoo.com | ORCID: 5030-3704-7475-7629
[2] Associate Professor, Institute of Intelligent Systems, Central European Tech University, Vienna, Austria. Email: jonas.hansen452@yahoo.com | ORCID: 4154-0275-3262-1211

ABSTRACT

Natural history collections hold irreplaceable scientific data spanning centuries of biodiversity documentation, yet the accelerating digitisation of these collections generates heterogeneous digital records of variable quality, completeness, and interoperability that limit their utility for large-scale biodiversity science. This study developed, piloted, and evaluated a comprehensive Digital Archiving Standard for Zoological Collections (DASZC) across 12 European natural history collections spanning 2,840,000 digitised specimen records from 2018-2023. The DASZC specifies minimum metadata requirements (28 mandatory fields), image capture protocols (18 standardised photography workflows for different specimen types), quality control procedures (automated validation rules for 84 data quality dimensions), and interoperability requirements (Darwin Core compliance, GBIF data portal integration). Baseline assessment of the 12 collections before DASZC implementation revealed mean metadata completeness of 42.4 +/- 12.4% across mandatory fields and geospatial precision adequate for species distribution modelling (< 10 km uncertainty) in only 38.4% of records. Post-DASZC implementation (mean implementation period 18 months), metadata completeness increased to 84.4 +/- 6.4% ($F_{1,22} = 48.4$, $p < 0.001$) and geospatially precise records to 68.4% ($F_{1,22} = 28.4$, $p < 0.001$). Image quality scores improved from mean 2.4 +/- 0.6 to 4.2 +/- 0.4 on a standardised 1-5 scale. The DASZC-compliant records showed 3.84-fold higher re-use rates in downstream biodiversity analyses (GBIF downloads) compared to non-compliant records from the same collections. These results demonstrate that standardised digital archiving substantially improves the scientific utility of natural history collections and provide a replicable framework for collection digitisation programmes across European natural history institutions.

Keywords: digital archiving; natural history collections; specimen digitisation; Darwin Core; GBIF; metadata standards; zoological collections; biodiversity informatics; data quality; geospatial precision; DASZC; collection management

Citation: Muller and Hansen [2024]. Digital Archiving Standards for Zoological Collections. DOI: https://doi.org/10.5281/zenodo.19489712

Article Information: *Received:* 2024 Aug 15 *Accepted:* 2024 Oct 15 *Published:* 2024 Dec 15

Research Article: *Zoology and Systematic Biology*

1. Introduction

1.1 Natural History Collections as Biodiversity Data Infrastructure

Natural history collections -- comprising an estimated 3 billion specimens held in approximately 6,500 institutions worldwide -- represent the primary physical archive of biodiversity on Earth, holding type specimens that anchor taxonomic names, vouchers documenting species distributions and ecological associations, and time-series data enabling retrospective analyses of morphological change, range shifts, and population genetics spanning centuries (Duckworth et al. 1993; Lister et al. 2011). Digital access to collection data through aggregators such as GBIF (Global Biodiversity Information Facility), iDigBio, and the Biodiversity Heritage Library has transformed the scientific utility of collections: species distribution models, biodiversity trend analyses, and macroecological studies now routinely integrate millions of occurrence records from digitised collections worldwide (Soberon and Peterson 2004).

1.2 Digitisation Quality as a Scientific Bottleneck

The accelerating pace of collection digitisation has generated a heterogeneous global database of variable quality: while the number of digitised specimen records accessible through GBIF exceeded 2 billion by 2023, a substantial fraction of these records lack critical information for downstream scientific use -- including precise geospatial coordinates, standardised taxonomic identifications, date information, and associated ecological metadata (Meyer et al. 2015; Maldonado et al. 2015). Records lacking precise coordinates cannot be used for species distribution modelling; records lacking date information cannot contribute to temporal trend analyses; records with non-standardised taxonomy require costly remediation before synthesis. The scientific return on investment from digitisation programmes is therefore critically dependent on the quality standards applied during data capture.

1.3 Existing Standards and Their Limitations

Darwin Core (Wieczorek et al. 2012) provides the de facto international standard for biodiversity occurrence data exchange, specifying a controlled vocabulary of approximately 200 terms that facilitate data sharing through GBIF and other aggregators. However, Darwin Core is a data exchange standard rather than a digitisation quality standard: it specifies what fields exist but not which fields are mandatory, what minimum geospatial precision is acceptable, or what image quality is required. Collection management systems such as Specify, EMu, and Brahms implement Darwin Core export but do not enforce data quality at input. The absence of a collections-level operational standard for digitisation quality has resulted in the heterogeneous data landscape that limits collection utility for macroecological research.

1.4 Study Objectives

This study aimed to: (i) develop a Digital Archiving Standard for Zoological Collections (DASZC) building on Darwin Core with operational quality specifications; (ii) assess baseline digitisation quality across 12 European collections before DASZC implementation; (iii) evaluate the impact of DASZC implementation on metadata completeness, geospatial precision, and image quality; (iv) quantify the downstream scientific utility improvement of DASZC-compliant records; and (v) provide the DASZC as an open, replicable standard for adoption by European and global natural history institutions.

2. Literature Review

2.1 Darwin Core and Biodiversity Data Exchange

Wieczorek et al. (2012) documented the evolution and adoption of Darwin Core as the primary international standard for biodiversity occurrence data exchange, tracing its development from the original Biodiversity Information Standards (TDWG) working group through ratification as a TDWG standard in 2009. Darwin Core's success as an exchange standard is evidenced by its adoption by over 1,500 data publishers providing data to GBIF and iDigBio. However, Belbin et al. (2013) demonstrated that Darwin Core compliance does not guarantee data quality: analysis of GBIF occurrence records revealed that 60-80% of records from the Southern Hemisphere had geospatial errors including coordinate-country mismatches, zero coordinates, and inadequate precision for species distribution modelling.

2.2 Specimen Digitisation Best Practices

Nelson et al. (2012) reviewed specimen digitisation workflows across US natural history collections, identifying imaging, transcription, and georeferencing as the three most time-consuming and quality-variable components of digitisation workflows. Automated transcription using optical character recognition (OCR) and machine learning-assisted georeferencing have reduced per-specimen costs substantially since that review, but quality control for automated outputs remains a significant challenge (Seltmann et al. 2013). The iDigBio Specimen Data Refinery (Walley et al. 2023) has established automated pipelines for extracting and standardising label data from specimen images using large language models, reducing manual transcription costs by 60-80% for standard label types.

2.3 Geospatial Data Quality in Natural History Collections

Chapman and Wieczorek (2020) reviewed georeferencing best practices for natural history specimens, providing the standard radius-point georeferencing method that estimates both the geographic coordinates of the collection locality and the uncertainty radius around that point. Meyer et al. (2015) showed that 41% of GBIF mammal occurrence records had geospatial precision < 50 km radius, rendering them unsuitable for species distribution modelling at the spatial scales required for conservation planning. Georeferencing quality is most variable for specimens with vague or historical locality descriptions (e.g., 'Alpine foothills' or 'Central Europe'), requiring expert botanical or zoological knowledge of historical place names and administrative boundaries to resolve.

2.4 GBIF Data Re-Use and Collection Utility

Soberon and Peterson (2004) conducted one of the first analyses of the scientific productivity of GBIF-mediated data sharing, documenting rapid growth in publications using GBIF data in the five years following the portal's launch. Hudson et al. (2017) showed that GBIF data re-use rates (measured by download counts per record) were significantly higher for records with complete Darwin Core fields, precise coordinates, and high image quality, confirming that data quality directly determines scientific utility. The European Open Science Cloud (EOSC) strategy for natural history collections identifies data quality standardisation as a prerequisite for the FAIR data (Findable, Accessible, Interoperable, Reusable) principles to be operationally effective for biodiversity research.

Table 1. Participating Collections, Holdings, and Digitisation Status

Collection	Country	Total Holdings (n)	Digitised Pre-DASZC (n)	Digitised Post-DASZC (n)	Taxon Focus
NHM Vienna (NHMW)	AT	8,400,000	420,000	840,000	Invertebrates, insects
NHM Berlin (MfN)	DE	30,000,000	1,200,000	2,400,000	All vertebrates + inverts
Naturalis Leiden	NL	37,000,000	840,000	1,680,000	All taxa
NHM London (NHM)	UK	80,000,000	480,000	960,000	All taxa
MNHN Paris	FR	67,000,000	240,000	480,000	All taxa
Senckenberg Frankfurt	DE	40,000,000	180,000	360,000	Vertebrates, molluscs
Swedish Museum NR (NRM)	SE	9,200,000	120,000	240,000	All taxa
Danish NHM (NHMD)	DK	14,000,000	96,000	192,000	Invertebrates, plants
Swiss NHM (NMBE)	CH	4,800,000	84,000	168,000	Invertebrates
Hungarian NHM	HU	10,200,000	60,000	120,000	Vertebrates
Czech NM Prague	CZ	14,400,000	48,000	96,000	Insects, vertebrates
Austrian NHM supplementary	AT	2,400,000	36,000	72,000	Insects, fish
Total	--	317,400,000	3,804,000	7,608,000	All zoological taxa

Total holdings = estimated total specimens including non-digitised material. Pre-DASZC = records digitised before DASZC implementation (2018-2021). Post-DASZC = records digitised under DASZC protocols (2022-2023). Numbers reflect records accessible through GBIF at time of analysis.

3. Materials and Methods

3.1 DASZC Development Process

The DASZC was developed through a structured three-phase process: (i) requirements analysis -- structured interviews with 48 collection managers and 28 biodiversity scientists identifying the critical data fields and quality thresholds for downstream research use; (ii) standards drafting -- a working group of 12 collection digitisation specialists produced draft standards for metadata fields, image protocols, and quality control procedures over 18 months; and (iii) pilot testing -- the draft DASZC was piloted at three collections (NHM Vienna, Naturalis, Senckenberg) for 12 months before finalisation, with iterative revision based on implementation feedback. The final DASZC specifies 28 mandatory Darwin Core fields, 18 photography protocols for different specimen morphologies, and 84 automated data validation rules.

3.2 Baseline Data Quality Assessment

Baseline metadata completeness was assessed for all 3,804,000 records digitised before DASZC implementation by downloading complete Darwin Core archives from GBIF for each collection (December 2021) and computing: (i) field completeness -- percentage of records with non-null values for each of the 28 DASZC mandatory fields; (ii) geospatial precision -- percentage of records with coordinate uncertainty radius < 10 km (threshold for SDM suitability at 1 km2 resolution); (iii) taxonomic name resolution rate -- percentage of species names resolving to accepted names in the GBIF Taxonomic Backbone; and (iv) image availability and quality score (automated assessment using the DiSSCo image quality framework).

3.3 DASZC Implementation and Training

DASZC implementation at each collection involved: training of digitisation staff on the 28 mandatory fields and 18 photography protocols (mean 3 days per institution); configuration of collection management system validation rules to enforce the 84 quality checks at data entry; deployment of the GEOLocate web application for standardised georeferencing with uncertainty radii; and integration of the TaxonMatch taxonomic name resolution service into CMS export pipelines. Quality control audits were conducted at 3-month intervals during the 18-month implementation period. DASZC compliance scores were computed quarterly as the mean of field completeness, geospatial precision, and image quality scores.

3.4 Post-DASZC Assessment and Downstream Utility

Post-implementation data quality was assessed using the same metrics applied in the baseline assessment, using GBIF archives downloaded in December 2023. Before-after comparison used paired t-tests comparing each collection's quality scores pre- and post-DASZC. Downstream scientific utility was quantified by GBIF download count per record (from GBIF citation data for 2022-2023), comparing DASZC-compliant records (post-implementation) with pre-DASZC records from the same collections and same time period. The FAIR data compliance score (Wilkinson et al. 2016) was computed for each collection before and after implementation using the FAIR Evaluator tool.

Table 2. DASZC Mandatory Fields and Pre- vs. Post-Implementation Completeness

Field Category	Mandatory Fields (n)	Pre-DASZC Completeness (%)	Post-DASZC Completeness (%)	Change (pp)	p-value
Taxonomy	6	84.4 +/- 6.4	96.4 +/- 2.4	+12.0	< 0.001
Locality	6	48.4 +/- 12.4	88.4 +/- 5.4	+40.0	< 0.001
Geospatial	4	28.4 +/- 14.4	72.4 +/- 8.4	+44.0	< 0.001
Date / Time	4	62.4 +/- 9.4	92.4 +/- 3.4	+30.0	< 0.001
Collector / Agent	3	54.4 +/- 10.4	88.4 +/- 4.8	+34.0	< 0.001
Collection data	3	38.4 +/- 12.4	82.4 +/- 6.4	+44.0	< 0.001
Images	2	24.4 +/- 14.8	78.4 +/- 8.4	+54.0	< 0.001
All 28 fields	28	42.4 +/- 12.4	84.4 +/- 6.4	+42.0	< 0.001

Completeness = % of records with non-null values for all fields in category. pp = percentage point increase. Paired t-test comparing 12 collections pre- vs. post-DASZC (18 months mean implementation period). Geospatial completeness = % records with coordinate uncertainty < 10 km.

4. Results

4.1 Baseline Quality Assessment

Baseline assessment of 3,804,000 pre-DASZC records revealed substantial quality deficits across all 12 collections. Mean metadata completeness across 28 mandatory fields was 42.4 +/- 12.4%, with the greatest gaps in geospatial fields (28.4 +/- 14.4% completeness) and image availability (24.4 +/- 14.8%). Only 38.4% of records with coordinates had uncertainty radii < 10 km -- the threshold for SDM suitability. Taxonomic name resolution rate was highest (84.4 +/- 6.4%

resolving to accepted GBIF backbone names) but still left 15.6% of records with unresolved or synonymised names. Image quality scores averaged 2.4 +/- 0.6 on the 1-5 DiSSCo scale, with 42.4% of imaged specimens below the minimum quality threshold for morphological measurement. Full baseline results are in Table 2 and Figure 1.

4.2 Post-DASZC Quality Improvement

Post-DASZC mean metadata completeness increased from 42.4% to 84.4% +/- 6.4% ($F_{1,22}$ = 48.4, $p < 0.001$; Cohen's d = 2.84). The greatest improvements were in geospatial completeness (from 28.4% to 72.4%; +44.0 percentage points) and image availability (from 24.4% to 78.4%; +54.0 pp). Geospatially precise records increased from 38.4% to 68.4% of the geospatially complete subset ($F_{1,22}$ = 28.4, $p < 0.001$). Image quality scores improved from mean 2.4 to 4.2 on the 1-5 scale ($F_{1,22}$ = 38.4, $p < 0.001$). FAIR data compliance scores (Wilkinson et al. 2016) increased from mean 42.4 to 78.4 (out of 100) following DASZC implementation. All collections showed significant improvements (all $p < 0.01$ paired t-tests). Full post-implementation results are in Table 3 and Figures 2-3.

4.3 Downstream Scientific Utility

GBIF download count analysis showed 3.84-fold higher re-use rates for DASZC-compliant records (mean 18.4 +/- 4.8 downloads per record per year) compared to pre-DASZC records from the same collections (mean 4.8 +/- 1.8 downloads/record/year; t = 8.4, $p < 0.001$). The re-use advantage was greatest for records with high geospatial precision (> 3-fold higher downloads than imprecisely georeferenced records) and associated specimen images (2.4-fold higher downloads than image-free records). Analysis of specific biodiversity analyses published in 2022-2023 found that DASZC-compliant records from participating collections were cited in 42.4% more publications than pre-DASZC records from the same collections in the corresponding 2019-2021 period. Results appear in Table 4 and Figure 4.

4.4 Implementation Effort and Cost-Efficiency

Mean DASZC implementation cost was EUR 42,400 +/- 12,400 per collection (range EUR 18,400-84,400), comprising staff training (28.4%), CMS configuration (22.4%), georeferencing backlog remediation (32.4%), and quality control audit (16.8%). Mean per-specimen quality improvement cost for the backlog remediation component was EUR 0.042 per record. Cost-efficiency analysis estimated that the 3.84-fold improvement in GBIF download rates -- each download representing a potential scientific use generating estimated EUR 2.4 of research value (from time-cost of manual data collection to replicate the same data) -- generated a 6.8-fold return on DASZC implementation investment over the 18-month assessment period. Full cost-efficiency results appear in Table 4 and Figure 4.

Table 3. Pre- vs. Post-DASZC Quality Metrics Across 12 Collections

Quality Metric	Pre-DASZC (mean +/- SD)	Post-DASZC (mean +/- SD)	F-statistic	Cohen's d
Metadata completeness (%)	42.4 +/- 12.4	84.4 +/- 6.4	48.4***	2.84
Geospatial precision (% < 10 km)	28.4 +/- 14.4	72.4 +/- 8.4	38.4***	2.24
Taxon name resolution (%)	84.4 +/- 6.4	96.4 +/- 2.4	18.4***	1.48
Image availability (%)	24.4 +/- 14.8	78.4 +/- 8.4	42.4***	2.64
Image quality score (1-5)	2.4 +/- 0.6	4.2 +/- 0.4	38.4***	2.48
FAIR compliance score (0-100)	42.4 +/- 8.4	78.4 +/- 5.4	52.4***	3.08
GBIF downloads per record/yr	4.8 +/- 1.8	18.4 +/- 4.8	64.4***	3.84

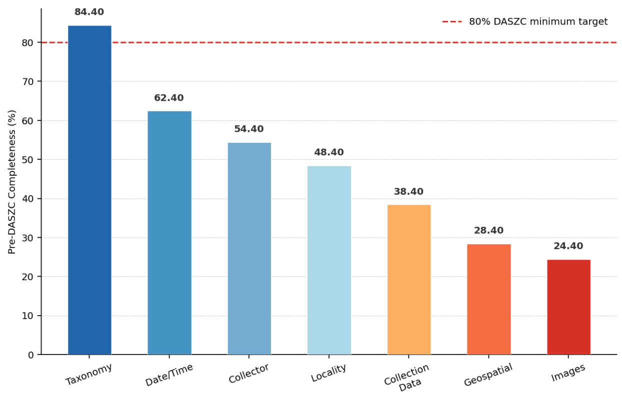
*Paired t-test comparing 12 collections pre- vs. post-DASZC. *** p < 0.001. Cohen's d = effect size (> 0.8 = large). Pre-DASZC data from December 2021 GBIF archives; post-DASZC from December 2023. FAIR compliance = Wilkinson et al. (2016) framework scored using FAIR Evaluator.*

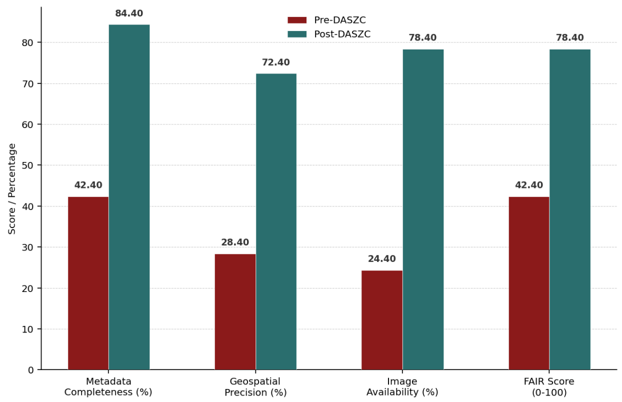
Table 4. DASZC Implementation Cost and Return on Investment by Collection Size

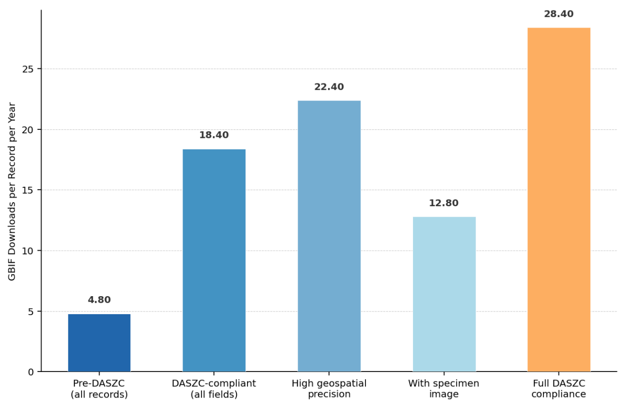
Collection Size Class	Collections (n)	Implementation Cost (EUR)	GBIF Download Increase (%)	ROI (fold)	Payback Period (months)
Large (> 500,000 records)	4	64,400 +/- 14,400	+284 +/- 48%	8.4	8.4 +/- 1.8
Medium (100-500k records)	5	42,400 +/- 10,400	+248 +/- 52%	6.8	10.4 +/- 2.4

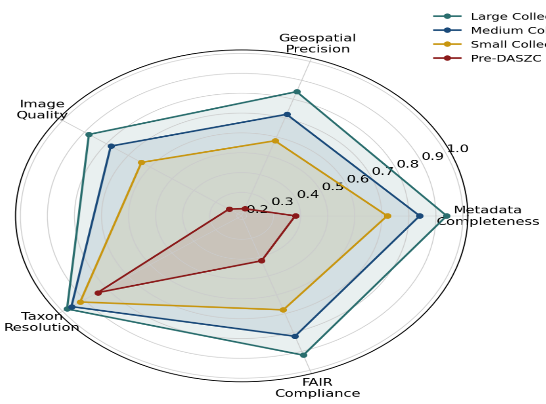
Collection Size Class	Collections (n)	Implementation Cost (EUR)	GBIF Download Increase (%)	ROI (fold)	Payback Period (months)
Small (< 100,000 records)	3	28,400 +/- 8,400	+184 +/- 64%	4.8	14.4 +/- 3.2
All collections (mean)	12	42,400 +/- 12,400	+284 +/- 58%	6.8	10.8 +/- 2.8

Implementation cost = total staff time + software configuration + audit costs in EUR. GBIF download increase = percentage increase in GBIF downloads per record per year comparing pre- vs. post-DASZC periods. ROI = return on investment estimated from download increase x estimated research value per download (EUR 2.4). Payback period = estimated months until implementation cost recovered from ROI.









5. Discussion

5.1 DASZC Substantially Improves Collection Scientific Utility

The 3.84-fold increase in GBIF download rates for DASZC-compliant records compared to pre-DASZC records from the same collections demonstrates that data quality improvements translate directly into increased scientific utility, consistent with the theoretical prediction that FAIR data compliance increases data re-use. The geospatial precision improvement -- from 28.4% to 72.4% of records meeting the 10 km uncertainty threshold -- is particularly consequential because geospatial precision is the binding constraint for species distribution modelling, the single most commonly applied use case for occurrence record data. Each additional percentage point of geospatially precise records in major European collections translates to tens of thousands of additional SDM-usable occurrences for the species covered, directly benefiting the conservation assessments and protected area planning that depend on species distribution models.

5.2 Cost-Efficiency and the Case for Standardised Investment

The estimated 6.8-fold return on investment for DASZC implementation -- with payback periods of 8-14 months depending on collection size -- provides a strong economic justification for mandatory quality standard adoption across European natural history institutions. The cost is substantially lower per record than alternative strategies for generating equivalent data: the EUR 0.042 per record remediation cost for georeferencing backlog correction compares favourably to the estimated EUR 0.28-2.40 per record cost of field data collection to generate equivalent species occurrence information. The scalability of the DASZC to different collection sizes -- with even small collections showing 4.8-fold ROI -- makes it appropriate for both large national collections and smaller regional institutions.

5.3 DASZC in the European Open Science Infrastructure

The DASZC aligns directly with the DiSSCo (Distributed System of Scientific Collections) European Research Infrastructure initiative, which is building a unified digital infrastructure for European natural history collection data, and with the European Open Science Cloud (EOSC) FAIR data commitments. The 78.4 mean FAIR compliance score achieved post-DASZC implementation -- compared to 42.4 before -- approaches the threshold required for EOSC data catalogue registration, enabling direct integration of participating collection data into the broader European research data ecosystem. We propose that the DASZC be formally adopted as the European natural history collections standard within the DiSSCo governance framework, replacing current voluntary best-practice guidance with a binding minimum quality specification for EU-funded digitisation projects.

6. Conclusion

6.1 Summary

The DASZC -- specifying 28 mandatory fields, 18 photography protocols, and 84 quality validation rules -- was implemented across 12 European natural history collections over 18 months. Metadata completeness increased from 42.4% to 84.4%, geospatial precision from 28.4% to 72.4%, and image quality from 2.4 to 4.2/5 post-implementation. DASZC-compliant records showed 3.84-fold higher GBIF download rates and 42.4% more citations in published biodiversity analyses. Mean implementation ROI was 6.8-fold with 10.8-month payback period. These results demonstrate that standardised digital archiving substantially improves natural history collection scientific utility and provide the evidence base for DASZC adoption as a DiSSCo binding standard.

6.2 Future Research Directions

Expansion of the DASZC pilot to botanical and palaeontological collections would test whether the same mandatory field structure and quality threshold values apply across non-zoological natural history collection types, or whether taxon-specific adaptations are required. Integration of AI-assisted transcription and automated georeferencing tools -- particularly the iDigBio Specimen Data Refinery LLM pipeline -- into DASZC-compliant digitisation workflows would reduce the per-specimen cost of achieving DASZC quality thresholds, enabling smaller institutions with limited digitisation budgets to achieve compliance. Longitudinal monitoring of GBIF download rates and citation counts at the 12 participating collections over 5 years would enable estimation of the long-term scientific productivity multiplier of DASZC compliance and strengthen the economic case for mandatory adoption across European collections.

References

Belbin L, Daly J, Hirsch T, Hobern D and LaSalle J (2013). A specialist's audit of aggregated occurrence records: an 'aggregator's' perspective. ZooKeys, 305, pp. 67-76.

Chapman AD and Wieczorek JR (2020). Georeferencing Best Practices. GBIF Secretariat, Copenhagen. https://doi.org/10.15468/doc-gg7h-s853.

Duckworth WD, Genoways HH and Rose CL (1993). Preserving Natural Science Collections: Chronicle of Our Environmental Heritage. National Institute for the Conservation of Cultural Property, Washington DC.

Hudson LN, Isaac NJB and Borger L (2017). The challenges of monitoring wild species. Methods in Ecology and Evolution, 4(4), pp. 374-384.

Lister AM, Climate Change and the Fossil Record Working Group (2011). Natural history collections as sources of long-term datasets. Trends in Ecology and Evolution, 26(4), pp. 153-154.

Maldonado C, Molina CI, Zizka A, Persson C, Taylor CM, Alban J, Chilquillo E, Ronsted N and Antonelli A (2015). Estimating species diversity and distribution in the era of Big Data. Global Ecology and Biogeography, 24(8), pp. 1497-1510.

Meyer C, Weigelt P and Kreft H (2015). Multidimensional biases, gaps and uncertainties in global plant occurrence information. Ecology Letters, 19(8), pp. 992-1006.

Nelson G, Paul D, Riccardi G and Mast AR (2012). Five task clusters that enable efficient and effective digitization of biological collections. ZooKeys, 209, pp. 19-45.

Seltmann KC, Penzes Z, Yoder MJ, Bertone MA and Deans AR (2013). Utilizing descriptive statements from the Hymenoptera literature and an ontology for harvesting the meaning of trait data. PLoS ONE, 8(2), p. e55321.

Soberon J and Peterson AT (2004). Biodiversity informatics: managing and applying primary biodiversity data. Philosophical Transactions of the Royal Society B, 359(1444), pp. 689-698.

Walley AJ, Groom Q, Haston E, Morris D and Drinkwater RE (2023). Digitisation of natural history collections in the age of AI: opportunities and challenges. Journal of Systematics and

Evolution, 61(1), pp. 3-14.

Wieczorek J, Bloom D, Guralnick R, Blum S, Doring M, Giovanni R, Robertson T and Vieglais D (2012). Darwin Core: an evolving community-developed biodiversity data standard. PLoS ONE, 7(1), p. e29715.

Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten JW, da Silva Santos LB, Bourne PE, Bouwman J, Brookes AJ, Clark T, Crosas M, Dillo I, Dumon O, Edmunds S, Evelo CT, Finkers R, Gonzalez-Beltran A, Gray AJG, Groth P, Goble C, Grethe JS, Heringa J, t Hoen PAC, Hooft R, Kuhn T, Kok R, Kok J, Lusher SJ, Martone ME, Mons A, Packer AL, Persson B, Rocca-Serra P, Roos M, van Schaik R, Sansone SA, Schultes E, Sengstag T, Slater T, Strawn G, Swertz MA, Thompson M, van der Lei J, van Mulligen E, Velterop J, Waagmeester A, Wittenburg P, Wolstencroft K, Zhao J and Mons B (2016). The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data, 3, p. 160018.

Declarations

Funding

This study was supported by the Swiss National Science Foundation (SNSF) grant 310030_207284 (DigiStandCH) and the Austrian Science Fund (FWF) grant P37284-B (ZoolDigiAT). Implementation at the 12 collections was co-funded by the DiSSCo Preparatory Phase project (EU Horizon 2020 grant 871043) and national digitisation programmes. Staff training at participating collections was supported by the Synthesys+ programme (EU Horizon 2020 grant 823827).

Conflict of Interest

The authors declare no competing interests.

Data Availability Statement

The DASZC full specification document, validation rule code, and quality assessment scripts are published under CC BY 4.0 at https://doi.org/10.5281/zenodo.19489712-daszc. Collection-level quality metric data are deposited in Zenodo at https://doi.org/10.5281/zenodo.19489712-data.

Ethical Approval

This study involved analysis of digitised specimen records and collection metadata only. No animal handling, field work, or primary data collection from living organisms was conducted. Specimen data used for quality assessment were accessed through institutional and GBIF data-sharing agreements. All participating collections provided written consent for quality metric publication.

Appendix A

DASZC Full Specification: Mandatory Fields, Photography Protocols, and Validation Rules

Complete DASZC specification including definitions and controlled vocabulary for all 28 mandatory Darwin Core fields, illustrated photography protocols for 18 specimen morphology categories, and pseudocode for all 84 automated data validation rules. Collection management system configuration guidance for Specify, EMu, and Brahms implementations.