Biodiversity Information Science and Standards 9: e181898
doi: 10.3897/biss.9.181898

Conference Abstract

MIDS in Practice: Comparing Specimen Digitisation Levels

Mathias Dillen[‡], Ben Norton[§], Cat Chapman[|], Elspeth Margaret Haston[¶], Marie Briggs[¶]

‡ Meise Botanic Garden, Meise, Belgium
§ Independent Informatics Consultant, Cary, North Carolina, United States of America
| University of Florida, Gainesville, United States of America
¶ Royal Botanic Garden Edinburgh, Edinburgh, United Kingdom

Corresponding author: Mathias Dillen (mathias.dillen@plantentuinmeise.be)

Received: 09 Dec 2025 | Published: 11 Dec 2025

Citation: Dillen M, Norton B, Chapman C, Haston E, Briggs M (2025) MIDS in Practice: Comparing Specimen Digitisation Levels. Biodiversity Information Science and Standards 9: e181898.
https://doi.org/10.3897/biss.9.181898

Abstract

The MIDS standard (Minimum Information about a Digital Specimen) aims to indicate the digitisation level of natural history specimens on a simple scale. Much progress has been made over the past year to standardise this process for optimal reusability and scalability. The Simple Standard for Sharing Ontological Mappings (SSSOM) framework (Matentzoglu et al. 2022) has been adopted by the MIDS task group for this purpose and mappings are now available for Darwin Core archives and ABCD. The use of SSSOM mapping sets should facilitate the assessment of digitisation status over time, for different kinds of specimens and for different data standards and their serialisation. In this presentation, we will show the results of calculating MIDS levels for those different use cases when applied to several datasets acquired from the Global Biodiversity Information Facility (GBIF), using the MIDSCalculator and the GBIF MIDS Checker tools. We will determine the reasons why MIDS levels were (not) met, highlighting the impact of MIDS information element definitions and, in particular, their mappings to different standards and the stability of their implementation in different kinds of software. These results should show the feasibility of enabling MIDS across different infrastructures and promote implementation of the MIDS standard.

To acquire an overview of the current digitisation status of specimen data, we downloaded an export of specimens from GBIF (GBIF.org 2025f) using its new SQL API service to minimize data storage requirements for unrequired data. To calculate the MIDS scores for this large dataset - 265 million records - we used an [adapted script]() re-using the MIDSCalculator calculation scripts in batches of 10 million (omitting the RShiny interface). We subsequently summarised the results per MIDS information element and level and performed follow-up analyses, again using the GBIF SQL API, to look into the causes of some MIDS information elements often being absent in the GBIF data. Fig. 1 shows the overall results per information element. 4% failed to meet MIDS level 0 and would be considered undigitised. 69% met MIDS level 0, 25% level 1 and 2% level 2. No specimen met level 3, but this was because Darwin Core terms mapped to MediaID are only available in multimedia extensions to Darwin Core and were not available through the SQL API, so achieving this level was impossible through this method.

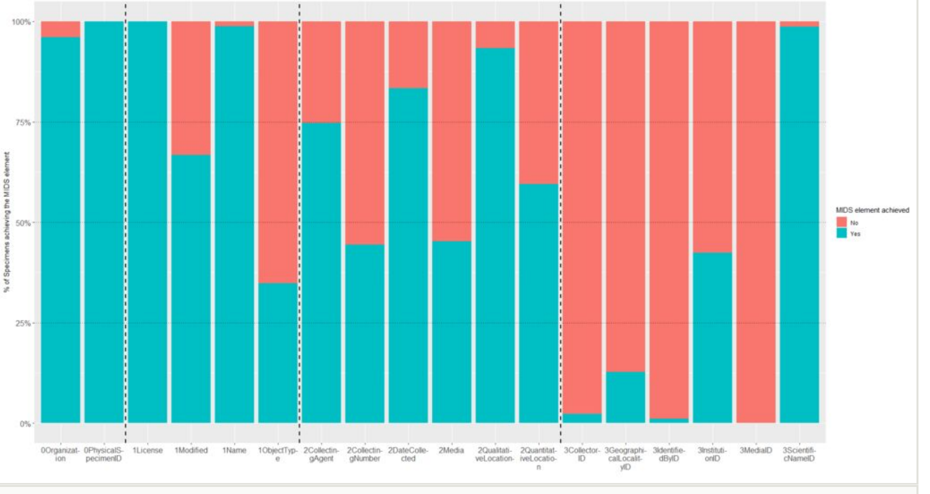

Figure 1.

Breakdown of MIDS level scores through an indicator bar for each information element what fraction of specimens (n = 265 million) achieved this element or not. The underlying data comes from a snapshot of GBIF occurrences with dwc:basisOfRecord = PreservedSpecimen created on 2025-06-26. This graph was made by Mathias Dillen and is available under the CC-BY 4.0 license.

Significant impacts were observed for several information elements: An unmapped Organization caused ca. 10 million specimens to remain at the level of "undigitized" (technically MIDS level -1) (GBIF.org 2025a). More than half of the datasets with specimens seemed to omit a last Modified date, preventing the specimens in them from reaching MIDS level 1 (GBIF.org 2025b). Object Type was even more impactful, missing from 173 million specimens or almost 2 out of every 3 specimens (GBIF.org 2025e). Collecting Number was missing for more than half of the specimens, but this one is difficult to assess properly in the absence of a reliable vocabulary for unknown, blank and non-applicable values (GBIF.org 2025g, GBIF.org 2025d, GBIF.org 2025c). This number is very important for a significant amount of specimens, as it is often used to refer to them in literature, but many specimens do not use these numbers at all.

These results show that calculating MIDS even at larger scales is possible and insightful. The technical workflow used could still be optimized in several ways, including parallel computing, more efficient storage methods and more powerful software frameworks than R, and potentially be leveraged to show the evolution of MIDS values, and hence digitization, through time using past GBIF snapshots.

Keywords

SSSOM, darwin core, GBIF, mapping

Presenting author

Mathias Dillen

Presented at

Living Data 2025

Acknowledgements

MD received funding from the Fonds Wetenschappelijk Onderzoek (FWO) of the Flemish government as part of the DiSSCo Flanders 2 project (I001625N).

Conflicts of interest

The authors have declared that no competing interests exist.

References

- GBIF.org (2025a) GBIF Occurrence Download. Release date: 2025-7-17. URL: https://doi.org/10.15468/dl.va72ra
- GBIF.org (2025b) GBIF Occurrence Download. Release date: 2025-9-08. URL: https://doi.org/10.15468/dl.cakga6
- GBIF.org (2025c) GBIF Occurrence Download. Release date: 2025-10-02. URL: https://doi.org/10.15468/dl.u8ajtf
- GBIF.org (2025d) GBIF Occurrence Download. Release date: 2025-10-02. URL: https://doi.org/10.15468/dl.w2f8y3
- GBIF.org (2025e) GBIF Occurrence Download. Release date: 2025-8-01. URL: https://doi.org/10.15468/dl.28wapg

- GBIF.org (2025f) GBIF Occurrence Download. Release date: 2025-6-26. URL: https://doi.org/10.15468/dl.taw6xr
- GBIF.org (2025g) GBIF Occurrence Download. Release date: 2025-9-26. URL: https://doi.org/10.15468/dl.3e5qpd
- Matentzoglu N, Balhoff JP, Bello SM, Bizon C, Brush M, Callahan TJ, Chute CG, Duncan WD, Evelo CT, Gabriel D, Graybeal J, Gray A, Gyori BM, Haendel M, Harmse H, Harris NL, Harrow I, Hegde HB, Hoyt AL, Hoyt CT, Jiao D, Jiménez-Ruiz E, Jupp S, Kim H, Koehler S, Liener T, Long Q, Malone J, McLaughlin JA, McMurry JA, Moxon S, Munoz-Torres MC, Osumi-Sutherland D, Overton JA, Peters B, Putman T, Queralt-Rosinach N, Shefchek K, Solbrig H, Thessen A, Tudorache T, Vasilevsky N, Wagner AH, Mungall CJ (2022) A Simple Standard for Sharing Ontological Mappings (SSSOM). Database 2022 https://doi.org/10.1093/database/baac035