International FAIR Digital Objects Implementation Summit 2024

Extended Abstracts

https://doi.org/10.52825/ocp.v5i.1181

Published: 18 Mar. 2025

Employing "Webby" FAIR Digital Objects to Support Large-Scale Ecosystem Monitoring and Mitigation of Biodiversity Loss in the Anthropocene

Claus Weiland[1,*], Maya Beukes[1], Jonas Grieb[1], Martin Jansen[1], Karsten Wesche[2], and Alexander Wolodkin[1]

[1]Senckenberg – Leibniz Institution for Biodiversity and Earth System Research, Frankfurt, Germany

[2]Senckenberg Museum of Natural History, Goerlitz, Germany

*Correspondence: Claus Weiland, claus.weiland@senckenberg.de

Abstract: We developed the WildLIVE platform to enable curation, analysis and mobilization of data from high-throughput biodiversity monitoring for European and global cross-domain data space programs. The platform implements "webby" FAIR Digital Objects leveraging on RO-Crate and FAIR Signposting to foster self-contained operation of machines on contained digital resources (machine actionability).

Keywords: Biodiversity Monitoring, RO-Crate, FAIR Signposting, Bioschemas, Green Deal Data Space, Destination Earth, Ontology, Machine Learning, Machine Actionability

Employing "Webby" FAIR Digital Objects to Support Large-Scale Ecosystem Monitoring and Mitigation of Biodiversity Loss in the Anthropocene

Anthropogenic influences on the environment are increasing and intensifying, impacting climate, ecosystems as well as their complex interplay [1]. Comprehensive long-term environmental monitoring across multiple geographic locations is therefore more important than ever to support policy, science, and decision-making with a scientifically sound data basis for mitigation strategies against biodiversity loss, caused i.e. by increased and agricultural use, invasive alien species, and habitat destruction such as wide ranging deforestation [2].

Thus, biodiversity loss in the anthropocene highlights the importance of the development and expansion of large-scale prediction capabilities as well as the requirement of comprehensive pooling of available biodiversity data, which is envisioned in thematic European data space projects such as the Destination Earth Data Lake (DEDL [3]) and Green Deal Data Space projects such as Biodiversity Meets Data (https://bmd-project.eu).

In this talk, we present the integration of both human crowd-sourced and machine learning based curation and enrichment of biomonitoring data in Senckenberg's WildLIVE Portal (Figure 1). WildLIVE enables the mobilization of data leveraging on the FAIR Digital Object (FDO

[4]) approach providing self-contained, machine actionable data units (analogous to IP datagrams) to implement machine-to-machine communication for the exchange of structured data within the aforementioned federated data spaces.

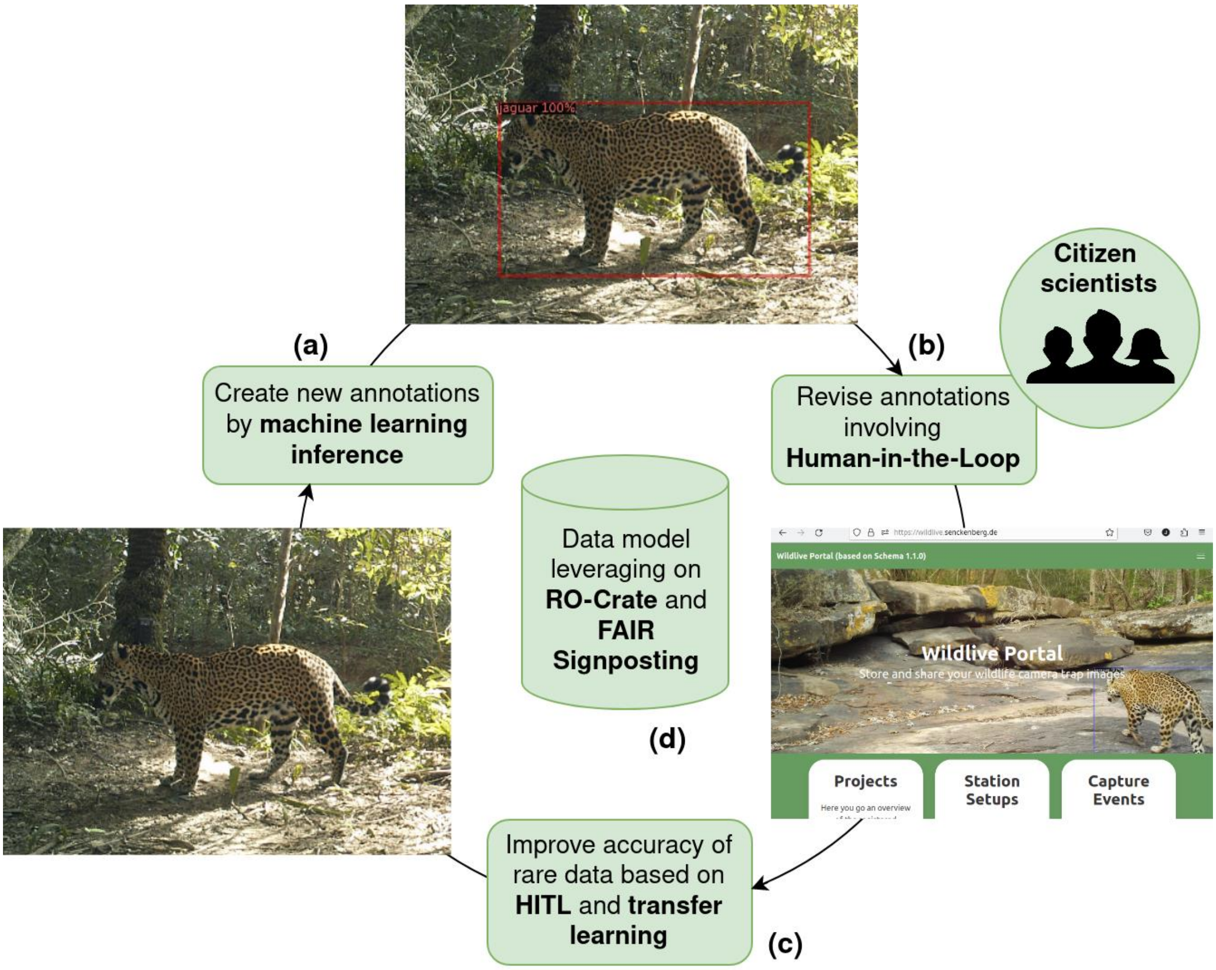

Figure 1. *Data flow in the WildLIVE platform (https://wildlive.senckenberg.de). (a.) A machine learning service provides automatic high-throughput annotations as baseline. (b) These annotations are subject to review and refinement by citizen scientists (Human-in-the-Loop, HILT). (c.) Revised annotated data are subsequently compiled into new training data. (d.)The data model based on RO-Crate and FAIR Signposting provides the operational semantics of how to process data and capture contextual information (provenance) from both human and machine-based operations.*

The service builds upon a technology stack based extensively on common web technologies involving RO-Crate (lightweight packaging of research outputs along with metadata [5]), Bioschemas (structured metadata [6]) and FAIR Signposting (web linking of digital objects [7]) combined with the digital object middleware Cordra (https://www.cordra.org). Combining these technologies facilitates the implementation of “webby” FDOs [8] that bind relevant information (e.g. metadata of sensors, geolocation and links to content stream) from an observing process together with operational semantics to provide clients with compatible information needed to operate on these data, in particular to perform machine learning-based object location and adjacent species classification [9].

To describe the semantics of fundamental concepts and relations used in WildLIVE in a FAIR-compliant way (*use[ing] a formal, accessible, shared, and broadly applicable language for knowledge representation,* https://www.go-fair.org/fair-principles/i1-metadata-use-formal-accessible-shared-broadly-applicable-language-knowledge-representation), we designed the WildLife Monitoring Ontology (WLMO, https://wildlive.senckenberg.de/wlmo), an ontology modeled using the W3C Web Ontology Language (OWL, https://www.w3.org/TR/owl-ref). Besides introducing new terms for example to represent a series of observations that were triggered by the same event (*wlmo:CaptureEvent*) or the setup of one or several joint monitoring sensors (*wlmo:StationSetup*), WLMO imports and reuses terms from existing domain ontologies like the Web Annotation Ontology (OA, https://www.w3.org/ns/oa) and the Semantic Sensor Network Ontology (SSN [10], https://www.w3.org/TR/vocab-ssn).

The talk will be closing with a preview on WildLIVE's upcoming project phase AI4WildLIVE where we will improve and upscale the service's predictive and AI capabilities within the German Federal Ministry of Education and Research's framework "Artificial Intelligence Methods as a Tool for Biodiversity Research“ (https://www.feda.bio/en/projects/biodivki).

Data availability statement

The experimental data used in this study (WildLIVE! - Annotated Images from Camera Traps in Eastern Bolivia) are available under Creative Commons Attribution 4.0 International from the "Senckenberg (meta)data portal" with the identifier https://doi.org/10.12761/34zr-fh25.

Author contributions

CW, JG: Writing - Original draft preparation, Conceptualization, Methodology, Software. AW: Software, Visualization, Writing – Review and Editing. KW: Writing – Review and Editing. MB, MJ: Data Curation, Writing – Review and Editing. CW, MJ, KW: Funding acquisition.

Competing interests

The authors declare that they have no competing interests.

Funding

This work has been supported by the Deutsche Forschungsgemeinschaft (DFG; project CAMTRAPPER – CAMera TRap Archive and Public Portal for Exploration & Research - GA no. 437771903) and the Bundesministerium für Bildung und Forschung (BMBF; project AI4WILDLIVE, 16LW0445).

Acknowledgement

Stian Soiland-Reyes, Leyla Jael Castro, Rohita Ravinder, Alexander Rogers (all BioHackEU23).

References

[1] Hurtt, G.C., Chini, L., Sahajpal, R., Frolking, S., Bodirsky, B.L., Calvin, K., Doelman, J.C., Fisk, J., Fujimori, S., Klein Goldewijk, K. and Hasegawa, T., (2020) Harmonization of global land use change and management for the period 850–2100 (LUH2) for CMIP6. Geoscientific Model Development, 13(11), pp.5425-5464. https://doi.org/10.5194/gmd-13-5425-2020

[2] Newbold, T., Hudson, L.N., Hill, S.L., Contu, S., Lysenko, I., Senior, R.A., Börger, L., Bennett, D.J., Choimes, A., Collen, B. and Day, J., (2015) Global effects of land use on local terrestrial biodiversity. *Nature*, *520*(7545), pp.45-50. https://doi.org/10.1038/nature14324

[3] Duatis Juarez, J., Schick, M., Puechmaille, D., Stoicescu, M., and Saulyak, B.: Destination Earth Data Lake, EGU General Assembly (2023) Vienna, Austria, 24–28 Apr 2023, EGU23-7177, https://doi.org/10.5194/egusphere-egu23-7177
[4] Wittenburg, P., Schwardmann, U., Blanchi, C., Weiland, C. (2023) FDOs to Enable Cross-Silo Work. Vol. 1 (2023): 1st Conference on Research Data Infrastructure (CoRDI) - Connecting Communities. https://doi.org/10.52825/cordi.v1i.263
[5] Soiland-Reyes S., Goble C., Groth P (2024): Evaluating FAIR Digital Object and Linked Data as distributed object systems. PeerJ Computer Science (accepted) arXiv 2306.07436 [cs.DC] https://doi.org/10.48550/arXiv.2306.07436
[6] Gray, A. J. G., Goble, C., & Jimenez, R. C. (2017). Bioschemas: From Potato Salad to Protein Annotation. In N. Nikitina, D. Song, A. Fokoue, & P. Haase (Eds.), ISWC 2017 Posters & Demonstrations and Industry Tracks: Proceedings of the ISWC 2017 Posters & Demonstrations and Industry Tracks co-located with 16th International Semantic Web Conference (ISWC 2017) (urn:nbn:de:0074-1963-7 ed.). (CEUR workshop proceedings; Vol. 1963). RWTH Aachen University. http://ceur-ws.org/Vol-1963/paper579.pdf
[7] Van de Sompel, H. (2023) FAIR Digital Objects and FAIR Signposting. Presentation slides. https://doi.org/10.5281/zenodo.7977333
[8] Soiland-Reyes S, Castro L, Ravinder R, Weiland C, Grieb J, Rogers A, Blanchi C, Van de Sompel H. BioHackEU23 report: Enabling FAIR Digital Objects with RO-Crate, Signposting and Bioschemas. 2024; https://doi.org/10.37044/osf.io/gmk2h
[9] Younis S, Schmidt M, Weiland C, Dressler S, Seeger B, Hickler T (2020) Detection and annotation of plant organs from digitised herbarium scans using deep learning. Biodiversity Data Journal 8: e57090. https://doi.org/10.3897/BDJ.8.e57090
[10] Janowicz K, Haller A, Cox S, Phuoc DL, Lefrancois M, (2018) SOSA: A Lightweight Ontology for Sensors, Observations, Samples, and Actuators. Available at SSRN: http://dx.doi.org/10.2139/ssrn.3248499