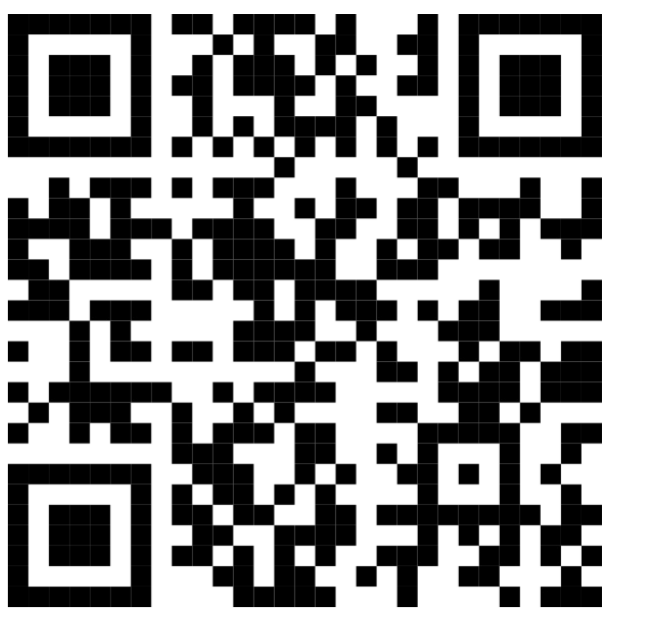# FAIR DATASPACE INTEGRATION FOR AGROBIODIVERSITY DIGITAL TWINS

Daniel Bauer[‡], Erik Kusch[§], Claus Weiland[‡], Jonas Grieb[‡], Desalegn Chala[§], and Dag Endresen[§]

‡ Senckenberg Society for Nature Research, Frankfurt am Main, Germany; § University of Oslo, Natural History Museum, Oslo, Norway

Correspondence: daniel.bauer@senckenberg.de

Platform Link
https://217.71.193.143

Introduction

With the rapid increase of available data, much of it scattered across various repositories, utilizing this data efficiently is difficult. Scientific workflows that adhere to the FAIR (**F**indable, **A**ccessible, **I**nteroperable, **R**eusable) data principles can address key challenges like reproducibility, data provenance, and transparency. Our workflow platform allows the execution of FAIR-compliant workflows and publishes results as FAIR Digital Objects for easy discoverability and reuse.

Towards a FAIR Workflow Platform

- Workflows can be submitted as Workflow Run RO-Crates[1,2]
- Workflow execution is handled by Argo (DEDL Hooks)
- Access to DestinE data via the Harmonized Data Access (HDA) API; external sources can also be used
- Workflow results stored as Digital Objects in CNRI's Cordra
- Automatic tracking of workflow provenance
- Results published as "webby" FAIR Digital Objects, discoverable via FAIR Signposting[3]

Crop Wild Relatives Digital Twin

Crop wild relatives (CWR) are wild plants closely related to domesticated crops, having highly diverse gene pools. They are valuable targets in plant breeding to enhance crop resilience against emerging challenges imposed by climate change such as diseases, drought, and other stresses. The **CWR Prototype Digital Twin**[4] utilizes species distribution modeling based on species occurrences from GBIF combined with climate and earth science data to predict habitat suitability maps. These maps can assist plant breeders in finding CWR populations with relevant genetic pools.
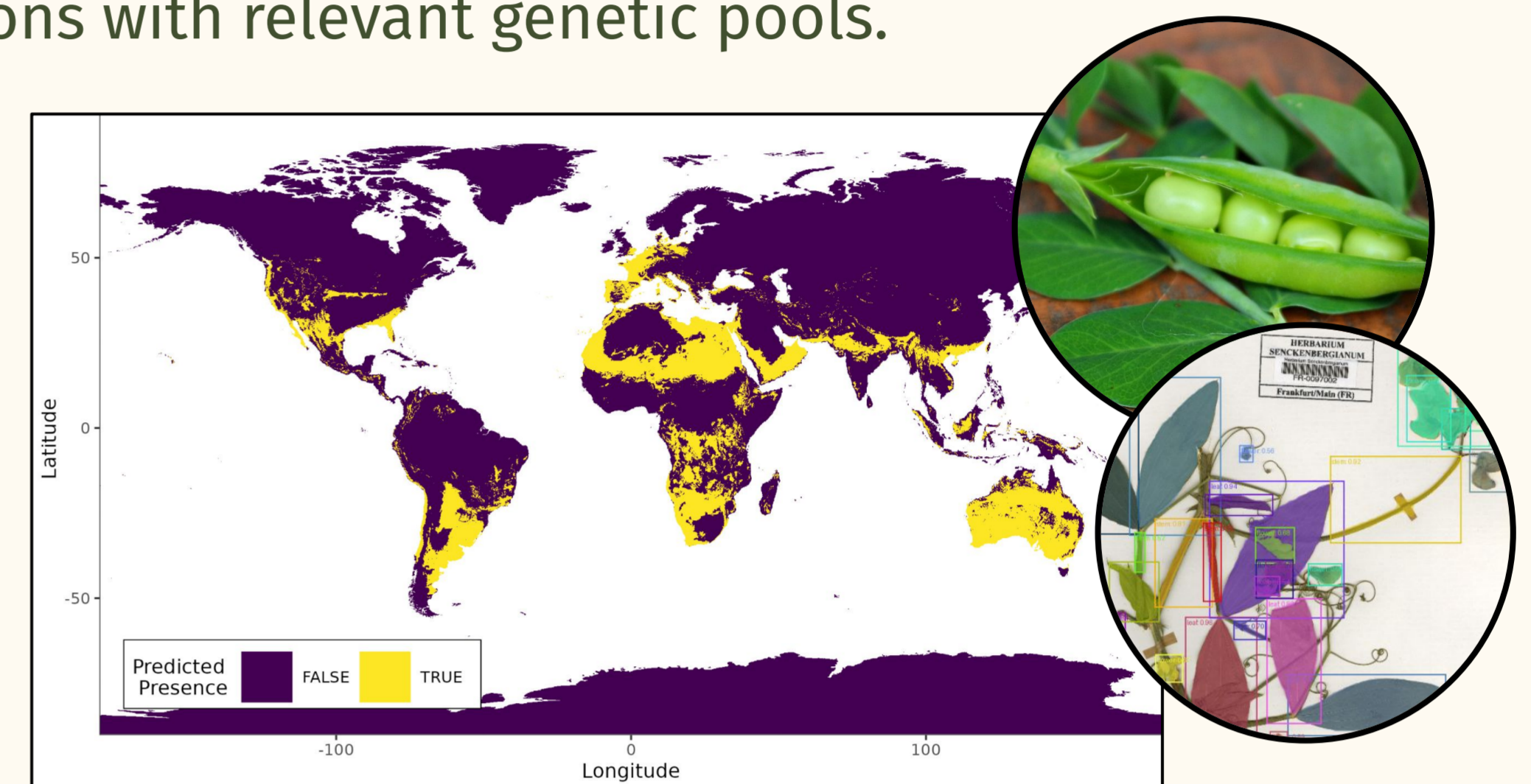

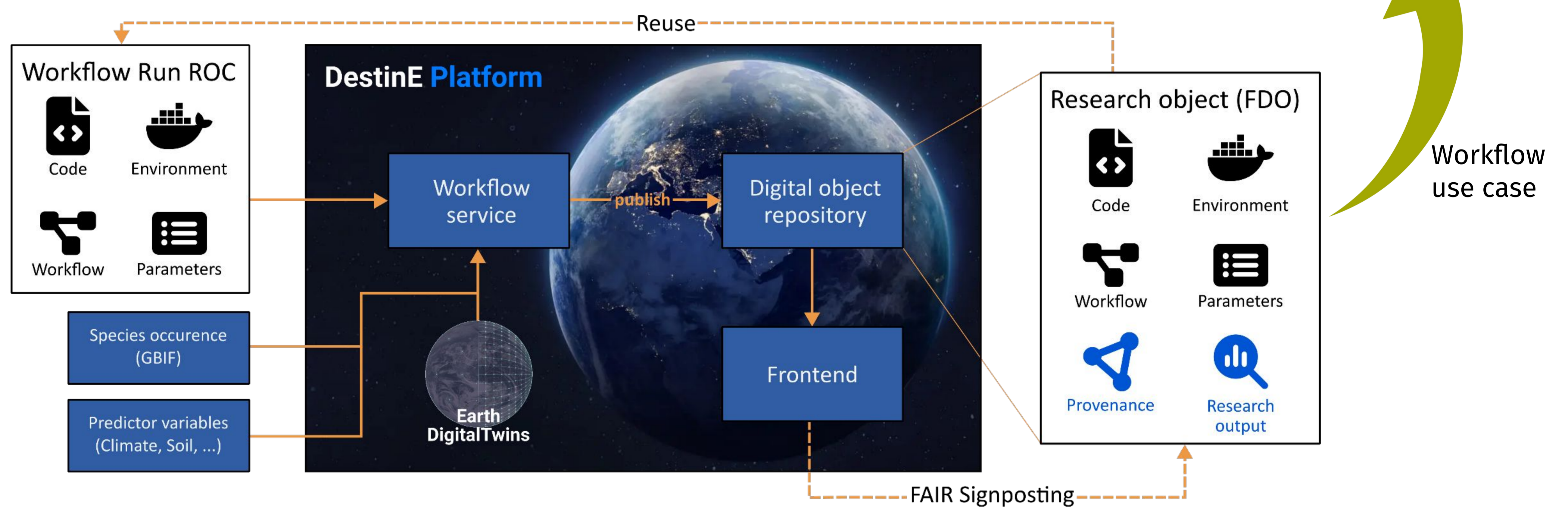**Fig. 2** Predicted habitat suitability map of Lathyrus latifolius (everlasting pea).

Fig. 1 Overview of architecture and data flow for the data space integration of fair workflows.

FAIR Signposting

FAIR Signposting is a technique that allows automated systems to navigate the *Internet* efficiently without human interaction. This is achieved by embedding typed links using known vocabularies into server responses. These links describe the content in a machine-readable way and allow agents to discover linked resources like FAIR digital objects or metadata.
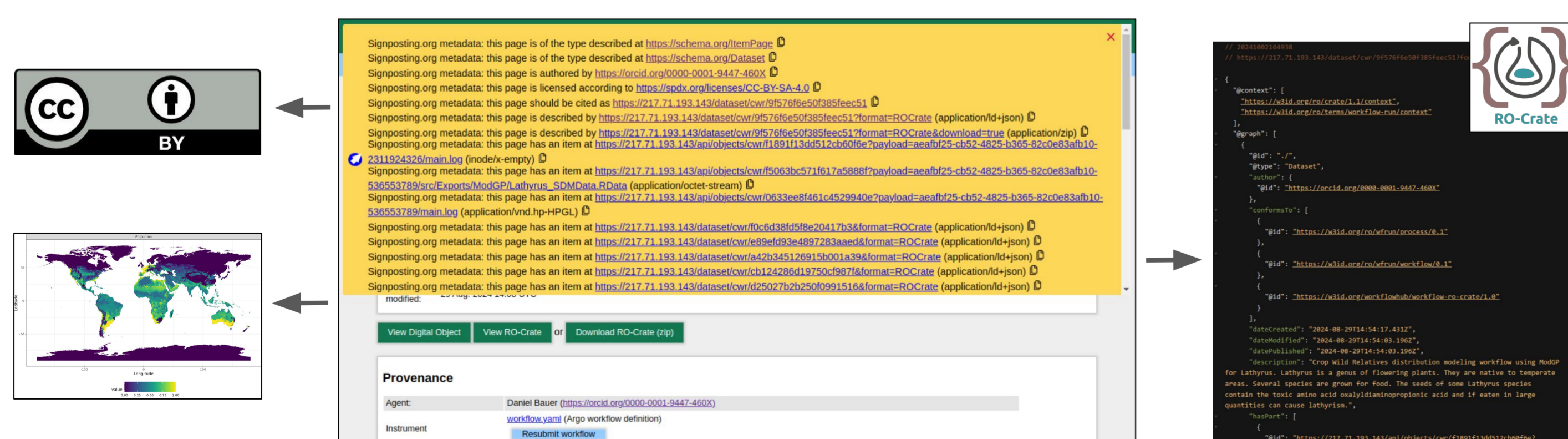

Fig. 3 FAIR Signposting in action: Types links are used to describe the content of a human-readable landing page of a dataset, directing machine-agents to relevant metadata and a machine-actionable representation (RO-Crate) of the dataset. The yellow box is usually not visible to a human.

RO-Crate

An RO-Crate (Research Object Crate)[4] is a lightweight implementation of a FAIR digital object and serves as a container to share data in a machine-readable format. Each crate includes metadata describing its content in JSON+LD format with vocabularies from the Semantic Web, i.e. Schema.org and Bioschemas. External resources can be incorporated by linking them by their persistent identifiers. Workflow Run RO-Crates extend the RO-Crate profile to capture prospective and retrospective workflow provenance.
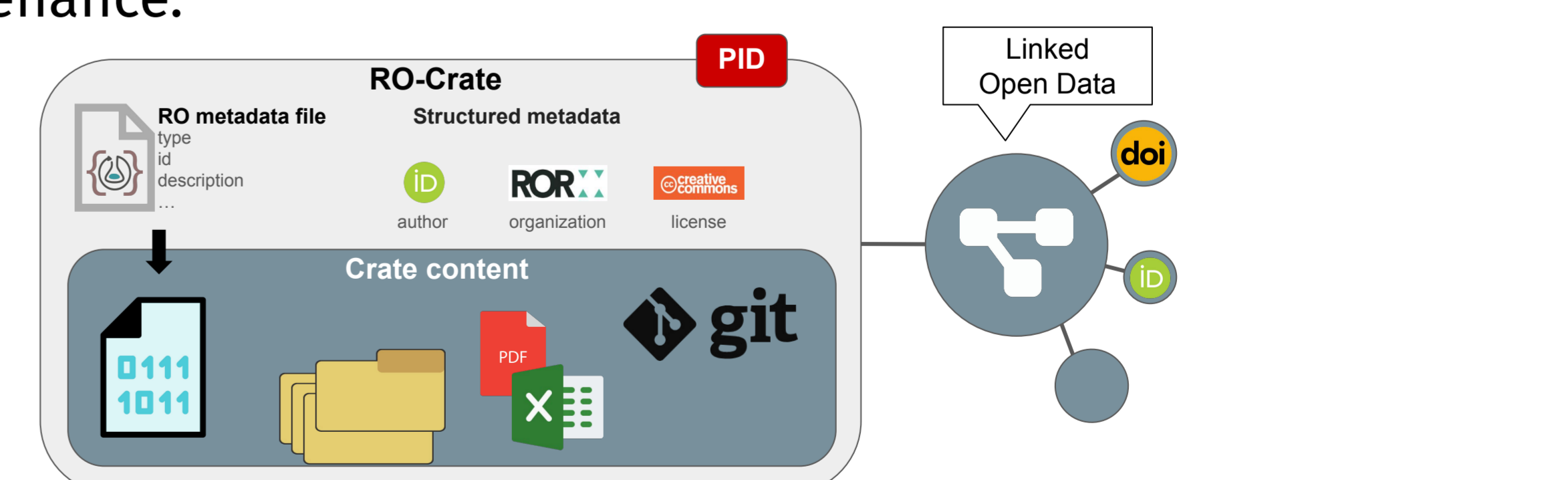

Fig. 4 Schematic representation of an RO-Crate linked to open data. Adapted from [5].

References:
1. Soiland-Reyes S at al. (2022), *Packaging research artifacts with RO-Crate.* https://doi.org/10.3233/ds-210053.
2. Simone L et al. (2024), *Recording provenance of workflow runs with RO-Crate.* https://doi.org/10.1371/journal.pone.0309210.
3. Soiland-Reyes S et al. (2024), *Enabling FAIR Digital Objects with RO-Crate, Signposting and Bioschemas.* https://doi.org/10.37044/osf.io/gmk2h.
4. Chala D et al. (2024), *Prototype biodiversity digital twin: crop wild relatives.* https://doi.org/10.3897/rio.10.e125192.
5. Soiland-Reyes S (2024), *FAIR Research Objects and Computational Workflows - A Linked Data Approach.* https://doi.org/10.5281/zenodo.13359051.

SENCKENBERG
world of biodiversity

UNIVERSITY OF OSLO
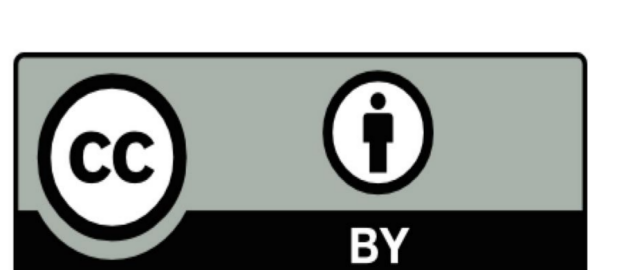
Destination Earth | Funded by the European Union | Implemented by ECMWF esa EUMETSAT