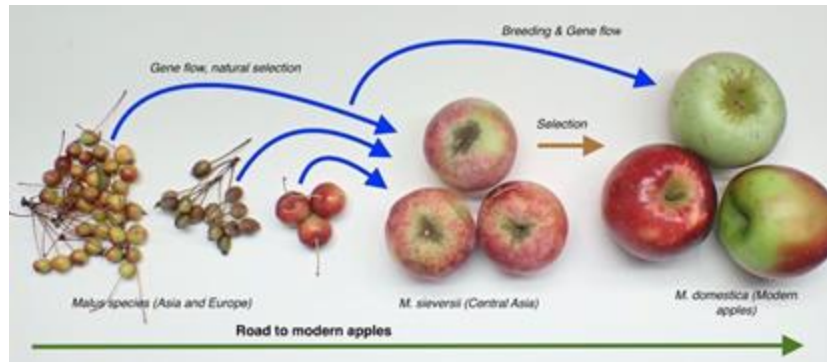
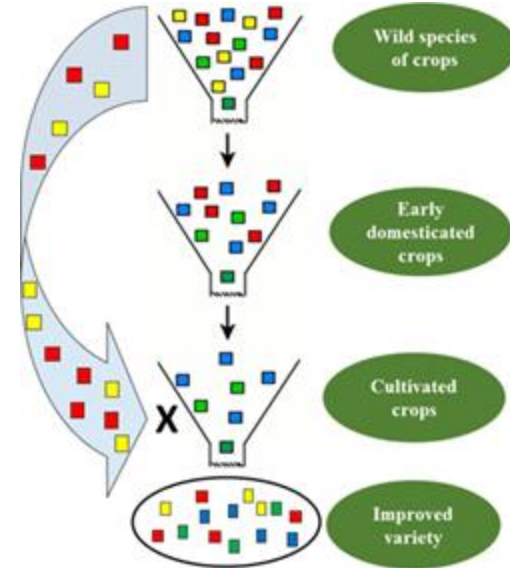


A FAIR workflow platform for Crop Wild Relatives

Daniel Bauer, Erik Kusch, Claus Weiland, Desalegn Chala, Dag Endresen

- Main Objective the CWR pDT

- To facilitate the transfer of traits from CWR to domesticated crops via breeding
 - To enhance crop resilience against climate change driven environmental stresses



Salgotra et al. 2022

Data and models

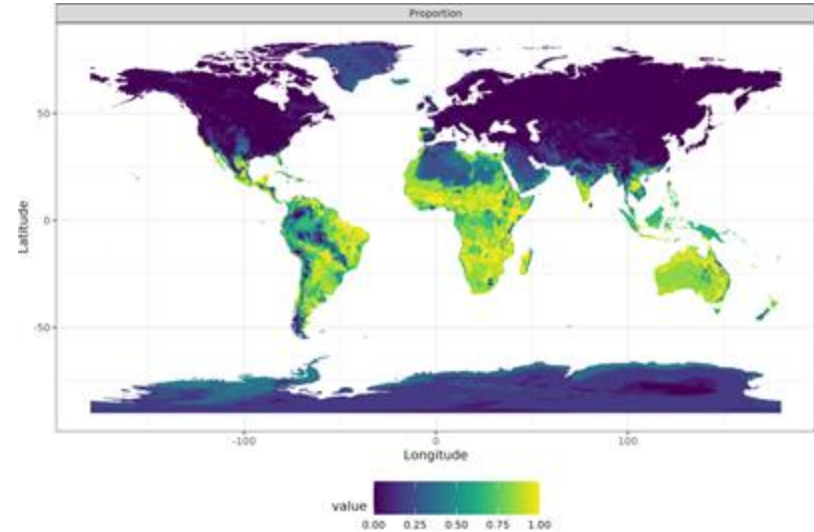
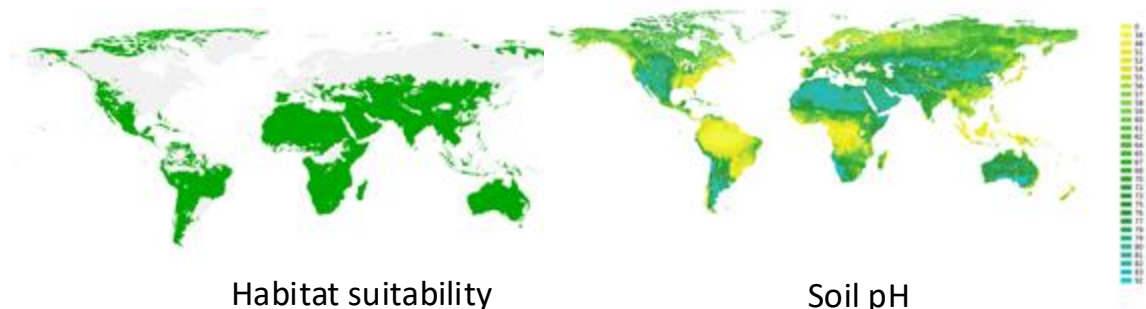
- **Data**

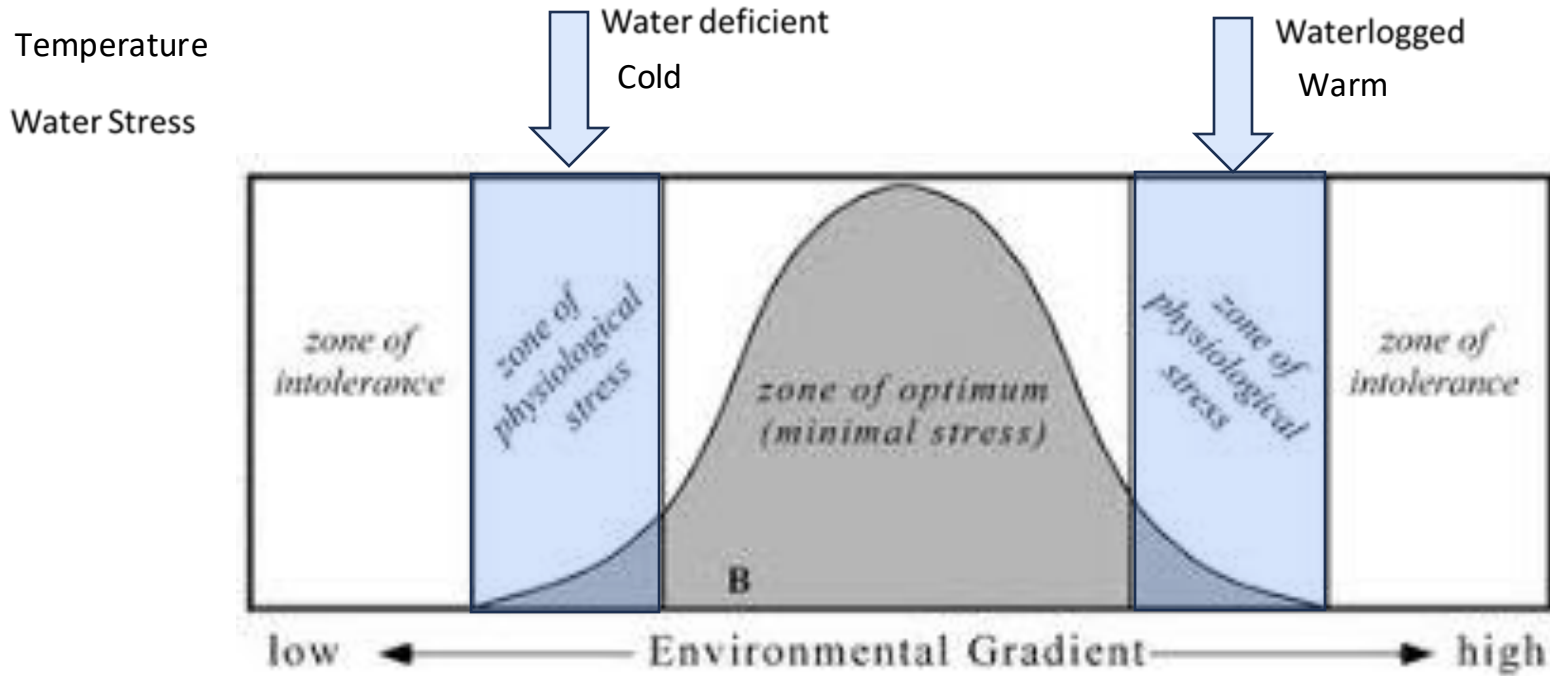
- I. Occurrence data from GBIF
- II. Climate data from ERA5-Land
- III. Soil data from soil grids

- **Models**

- We rely an ensemble approach of SDM
- **GAM**, **MaxEnt** and **BRT** – replicating these models using different model replication techniques
- Evaluating the models using AUC and TSS

- We do three main things
 - Producing habitat suitability maps
 - Extracting values of environmental stress such as salinity, water stress and so on from suitable pixels
 - Generate maps of geographic areas where stress tolerant populations are growing

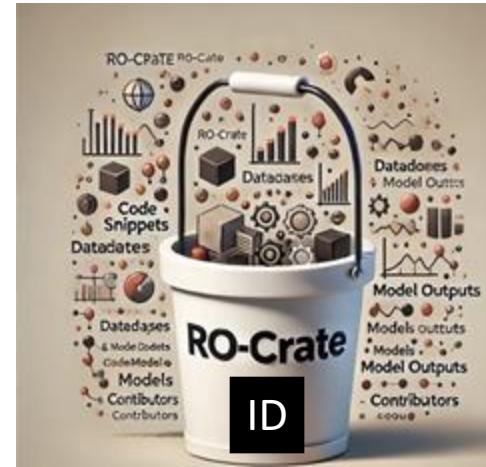




The most extreme the environment a population is adapted to, makes it the best candidate.
The best CWR are those with populations adapted to extreme conditions.

Our today's presentation focuses on the works we have done in connection with:

- WP 5 - FAIRness
- WP 8 - EC Review Recommendation: Collaborations and integration with strategic initiatives/ Programmes.
 - sustainability (afterlife existence)



We highlight about:

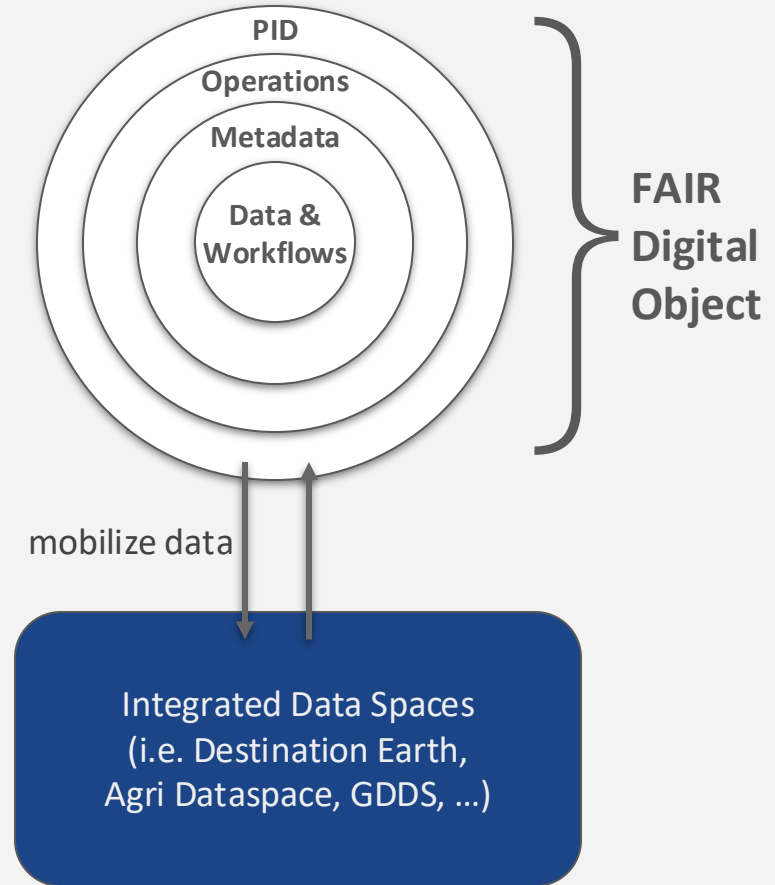
- The FAIR workflow we have developed for the CWR pDT
- The deployment of the pDT to the **Destination Earth Data Lake** as a FAIR digital object via RO-Crate
 - a FAIR-compliant research data container
 - Code, data, model outputs, and metadata, detailing model execution and reproducibility as well as who run the models (ORCID)
- Done in collaboration with **Senckenberg - Claus and Daniel**



Fig produced by ChatGPT after provision of the text: Provide me an illustration of Ro-Crate as a bucket containing various components like code, datasets, models, and outputs in a lightweight, FAIR-compliant manner.

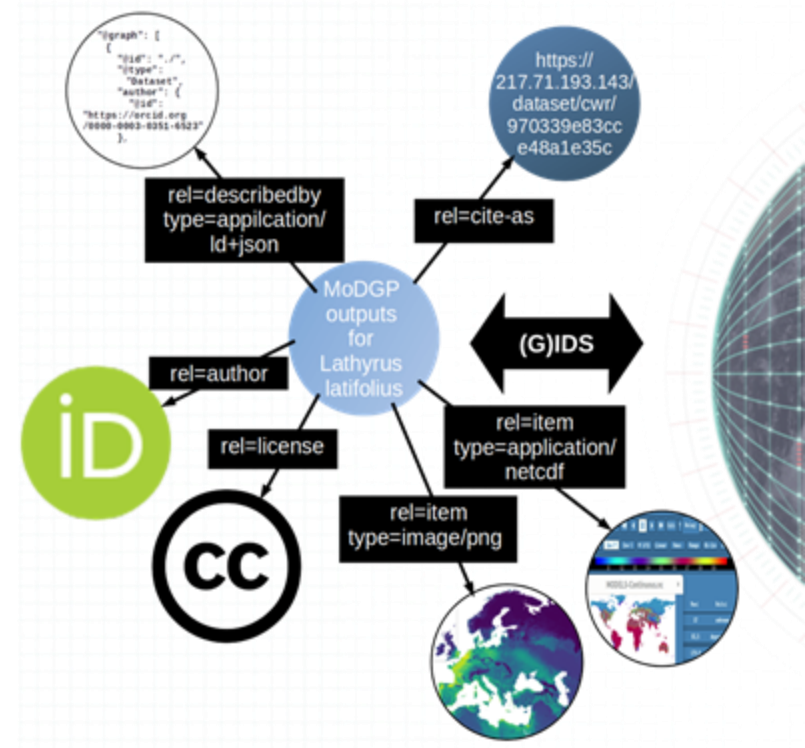
FAIR digital object approach

- Findable, **A**ccessible, Interoperable, Reuseable
- **FAIR Digital Object (FDO)**: Layer-based data model to enable machine-actionability
- **Machine-actionability** requires standardized formats allowing automated processes to discover and operate on data without human interaction



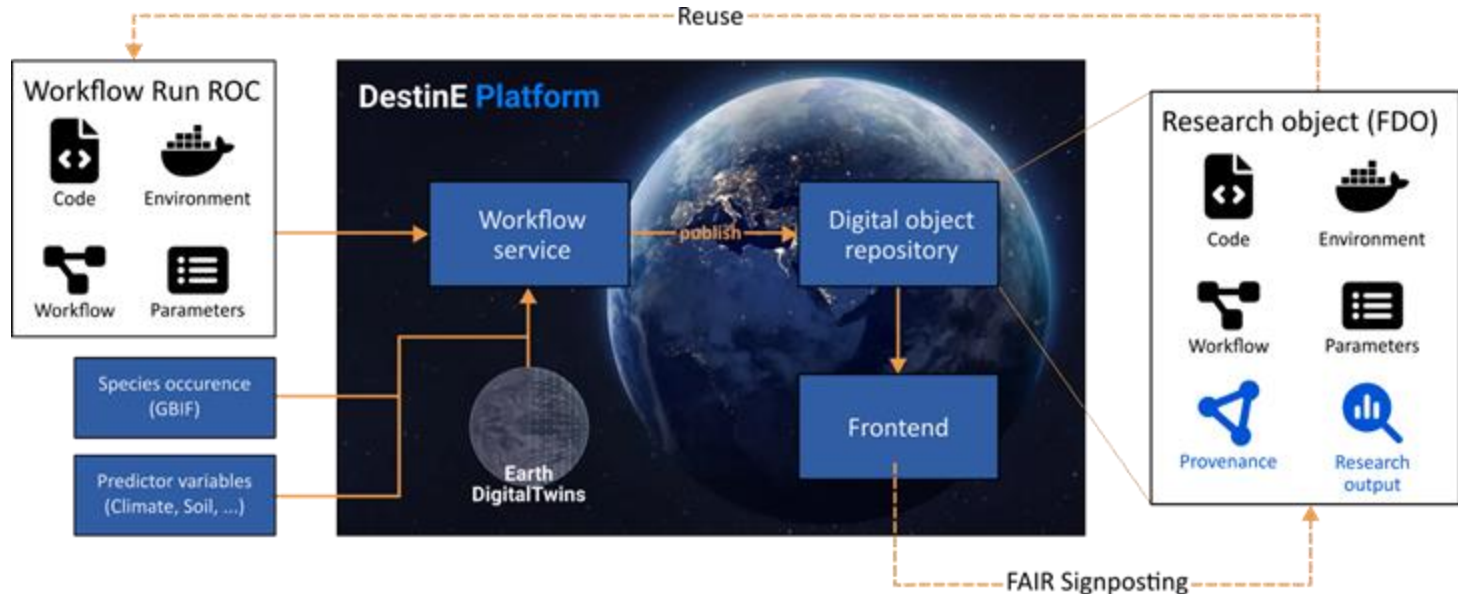
“Webby” FDOs via FAIR signposting

- **“Webby” FDOs:** Integrate FDOs into the existing web ecosystem using common web technologies -> maximize discoverability and integration
- **[RO-Crate](#):** FAIR-compliant container for research data using common vocabularies (schema.org/Bioschemas) for metadata description
- **[FAIR Signposting](#):** Add machine-interpretable links to human-readable web resources using link relation types ([RFC8288](#)) such as item, cite-as, describedBy to present the web topology of a resource



FAIR Workflow Platform concept

- Workflow Run RO-Crate centric platform
- Digital objects managed in Cordra
- Frontend with FAIR Signposting and Provenance visualization



Achievements

A FAIR workflow platform
for crop wild relatives

The screenshot displays a web interface for a workflow platform. The top navigation bar is green with the title 'Crop wild relatives' and links for 'Datasets', 'Workflows', 'Submit workflow', and 'Login'. The main content area is titled 'Species distribution models for Lathyrus'.

Details

PID:	cwr/ba3f7488a84a9060405a
Description:	Species distribution model for Lathyrus
Keywords:	Lathyrus, ModGP, SDM, Crop wild relatives
Published:	2024-08-14T02:22:00.280Z
Author(s):	Daniel Bauer (https://orcid.org/0000-0001-9447-460X)
License:	https://spdx.org/licenses/CC-BY-NC-SA-4.0
Created at:	23 Aug. 2024 11:08 UTC
Last modified:	14 Aug. 2024 02:08 UTC

Below the details table are three buttons: 'View Digital Object', 'View RO-Crate', and 'Download RO-Crate (zip)', separated by an 'or'.

Provenance

Agent:	Daniel Bauer (https://orcid.org/0000-0001-9447-460X)
Instrument:	workflow.yaml (Argo workflow definition) Resubmit workflow
Start Time:	13 Aug. 2024 09:08 UTC
End Time:	14 Aug. 2024 01:08 UTC
Parameters:	species: Lathyrus numberOfCores: 9 modgpSource: https://github.com/BioDT/uc-CWR.git modgpRevision: main

Visualization

The visualization section shows a partial graph with nodes labeled '@type: Dataset' and 'author'.

Crop wild relatives

Submit workflow as RO-Crate

1. Upload Workflow RO-Crate

Note: Upload must be a zip file following the [Workflow RO-Crate](#) specification. The workflow must be in yaml format and be a valid [Kyo Workflow](#).

Choose File No file chosen

2. Edit Workflow

Metadata

Title:

Species distribution models for Lathyrus

License:

Creative Commons Attribution Non-Commercial Share Alike 4.0 International

Description:

Species distribution models for [Lathyrus](#)

Submit workflows



Crop wild relatives

Workflows

Name/Internal name	Submitter	Started	Finished	Duration	Status
Species distribution models for Lathyrus a9d5c2c19f2-402-9889-504e3a6d22	Daniel Bauer 000-0000-9901-8900	26 Aug. 2024 13:08:32 UTC		1 hour, 17 minutes	Running
Species distribution models for Lycopodium 1788330-5011-4381-8111-09000-22880	Daniel Bauer 000-0000-9901-8900	26 Aug. 2024 12:06:20 UTC	26 Aug. 2024 13:06:08 UTC	50 minutes	Successful
Species distribution models for Lathyrus 12203036-487-406-0080-5047104900	Client 000-0000-0000-0000	22 Aug. 2024 13:06:47 UTC	22 Aug. 2024 14:06:27 UTC	48 minutes	Successful
Test Dataset Resubmitted 02303036-487-406-0080-5047104900	Daniel Bauer 000-0000-9901-8900	21 Aug. 2024 14:06:58 UTC	21 Aug. 2024 14:06:49 UTC	0 minutes	Successful
Species distribution models for Lathyrus 00a907b1-9817-a02a-4046-80288a4e9022	Daniel Bauer 000-0000-9901-8900	13 Aug. 2024 09:06:46 UTC	14 Aug. 2024 01:06:04 UTC	16 hours, 16 minutes	Successful
Test Dataset 17723036-1730-4040-5889-404030000000	Daniel Bauer 000-0000-0000-0000	13 Aug. 2024 08:06:46 UTC	13 Aug. 2024 09:06:16 UTC	0 minutes	Successful

Monitor progress

Crop wild relatives

Species distribution models for Lathyrus

Details

ID: [c9bfba3f1688a4a0900405a](#)

Description: Species distribution model for Lathyrus

Keywords: Lathyrus, ModGP, SCAM, Crop wild relatives

Published: 2024-08-14T02:22:00-280Z

Author(s): Daniel Bauer (<https://orcid.org/0000-0002-9447-8900>)

License: <https://dx.doi.org/licenses/CC-BY-NC-SA/4.0>

Created at: 23 Aug. 2024 13:08 UTC

Last modified: 14 Aug. 2024 02:08 UTC

or

Provenance

Client: Daniel Bauer from Senckenberg Society for Nature Research

Reuse workflows

Crop wild relatives

Datasets

Datasets found: 10

Lycopodium deuterodensum
[c9a1050b0000d071d8303001](#)

Description:

Files: 12

License: <https://dx.doi.org/licenses/CC-BY-SA-4.0>

Has workflow? no

Has Provenance? no

Created: 26 Aug. 2024 13:08 UTC, Last modified: 26 Aug. 2024 13:08 UTC

Lycopodium densum
[c9a1050b0000d071d8303001](#)

Description:

Files: 12

License: <https://dx.doi.org/licenses/CC-BY-SA-4.0>

Has workflow? no

View results

Human- and machine-readable research data

Crop wild relatives Datasets Workflows Submit workflow Logout

Species distribution models for Lycopodium

Details

PIO: cwr059ec70feae551253061

Description: Crop Wild relatives distribution modeling workflow using ModGP for Lycopodium. Lycopodium is not not actually a crop species, but serves as a good test genus due to the limited amount of species in this genus (and thus the short workflow execution runtime). Lycopodium is a genus of clubmosses within the family Lycopodiaceae. These perennial, evergreen plants are characterized by their small, needle-like or scale-like leaves arranged in spirals or whorls along the stems. Lycopodium species reproduce via spores rather than seeds and often form dense mats in forests, particularly in moist, shaded environments.

Keywords: Lycopodium, ModGP, SDM, test, CWR

Published: 2024-08-26T13:19:36.870Z

Author(s): Daniel Bauer (<https://orcid.org/0000-0001-5447-4800>)

License: <https://odp.org/licenses/CC-BY-SA-4.0>

Created at: 26 Aug. 2024 13:08 UTC

Last modified: 26 Aug. 2024 13:08 UTC

[View Digital Object](#) [View RO-Crate](#) or [Download RO-Crate \(zip\)](#)

Provenance

Agent: Daniel Bauer (<https://orcid.org/0000-0001-5447-4800>)

Instrument: [workflow.yml](#) (Argo workflow definition)

Start Time: 26 Aug. 2024 12:08 UTC

End Time: 26 Aug. 2024 13:08 UTC

Parameters: species: Lycopodium
numberOfCores: 8
modgpSource: <https://github.com/BioDTU-CWR-gt/modgpRevisions>: main

Visualization

○ @Type: Dataset

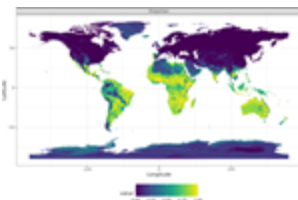
author @

```
// 30299515176130
// https://117.71.193.143/Dataset/cwr/059ec70feae551253061?format=RO-Crate

{
  "@context": [
    "https://w3id.org/ro/crate/1.1/context",
    "https://w3id.org/ro/terms/workflow-run/context"
  ],
  "@graph": [
    {
      "@id": ".",
      "@type": "Dataset",
      "author": {
        "@id": "https://orcid.org/0000-0001-5447-4800"
      },
      "conformsTo": [
        {
          "@id": "https://w3id.org/ro/wfrun/process/0.1"
        },
        {
          "@id": "https://w3id.org/ro/wfrun/workflow/0.1"
        },
        {
          "@id": "https://w3id.org/workflowhub/workflow-ro-crate/1.0"
        }
      ],
      "dateCreated": "2024-08-26T13:19:36.870Z",
      "dateModified": "2024-08-26T13:19:36.870Z",
      "datePublished": "2024-08-26T13:19:36.870Z",
      "description": "Crop Wild Relatives distribution modeling workflow using ModGP for lycopodium. Lycopodium is not not actually a crop species, but serves as a good test genus due to the limited amount of species in this genus (and thus the short workflow execution runtime). Lycopodium is a genus of clubmosses within the family Lycopodiaceae. These perennial, evergreen plants are characterized by their small, needle-like or scale-like leaves arranged in spirals or whorls along the stems. Lycopodium species reproduce via spores rather than seeds and often form dense mats in forests, particularly in moist, shaded environments.",
      "hasPart": [
        {
          "@id": "https://117.71.193.143/object/cwr/feae7af8a147964a209?overlayed=76480330-9051-43ff-9112-c06abc1fd998"
        }
      ]
    }
  ]
}
```



Machine-Actionability: RO-Crates and FAIR Signposting



Signposting.org metadata: this page is of the type described at <https://schema.org/ItemPage>

Signposting.org metadata: this page is of the type described at <https://schema.org/Dataset>

Signposting.org metadata: this page is authored by <https://orcid.org/0009-0009-0544-5117>

Signposting.org metadata: this page is licensed according to <https://spdx.org/licenses/CC-BY-SA-4.0>

Signposting.org metadata: this page should be cited as <https://217.71.193.143/dataset/cwr/2b408313359f2b6f18c2>

Signposting.org metadata: this page is described by <https://217.71.193.143/dataset/cwr/2b408313359f2b6f18c2?format=ROCrates>

(application/ld+json)

Signposting.org metadata: this page is described by <https://217.71.193.143/dataset/cwr/2b408313359f2b6f18c2?format=ROCrates&download=true>

Signposting.org metadata: this page has an item at <https://217.71.193.143/objects/cwr/066270d30e33a451e64?download=91c23560-7799-4064-bc3f-95d62e722154/manifest.json> (application/json)

Signposting.org metadata: this page has item at <https://217.71.193.143/objects/cwr/0077a98ac0644e020?download=91c23560-7799-4064-bc3f-95d62e722154/main.log> (text/plain)

Signposting.org metadata: this page has an item at <https://217.71.193.143/objects/cwr/5956015a098ef326246?download=workflow.yaml> (text/yaml)

[View Digital Object](#) [View RO-Crate](#) OR [Download RO-Crate \[zip\]](#)

Provenance

Agent:	Daniel Bauer (https://orcid.org/0009-0544-5117)
Instrument:	workflow.yaml (Argo workflow definition)
	Resubmit workflow
Start Time:	13 Aug. 2024 08:08 UTC
End Time:	13 Aug. 2024 08:08 UTC
Parameters:	text: last parameter



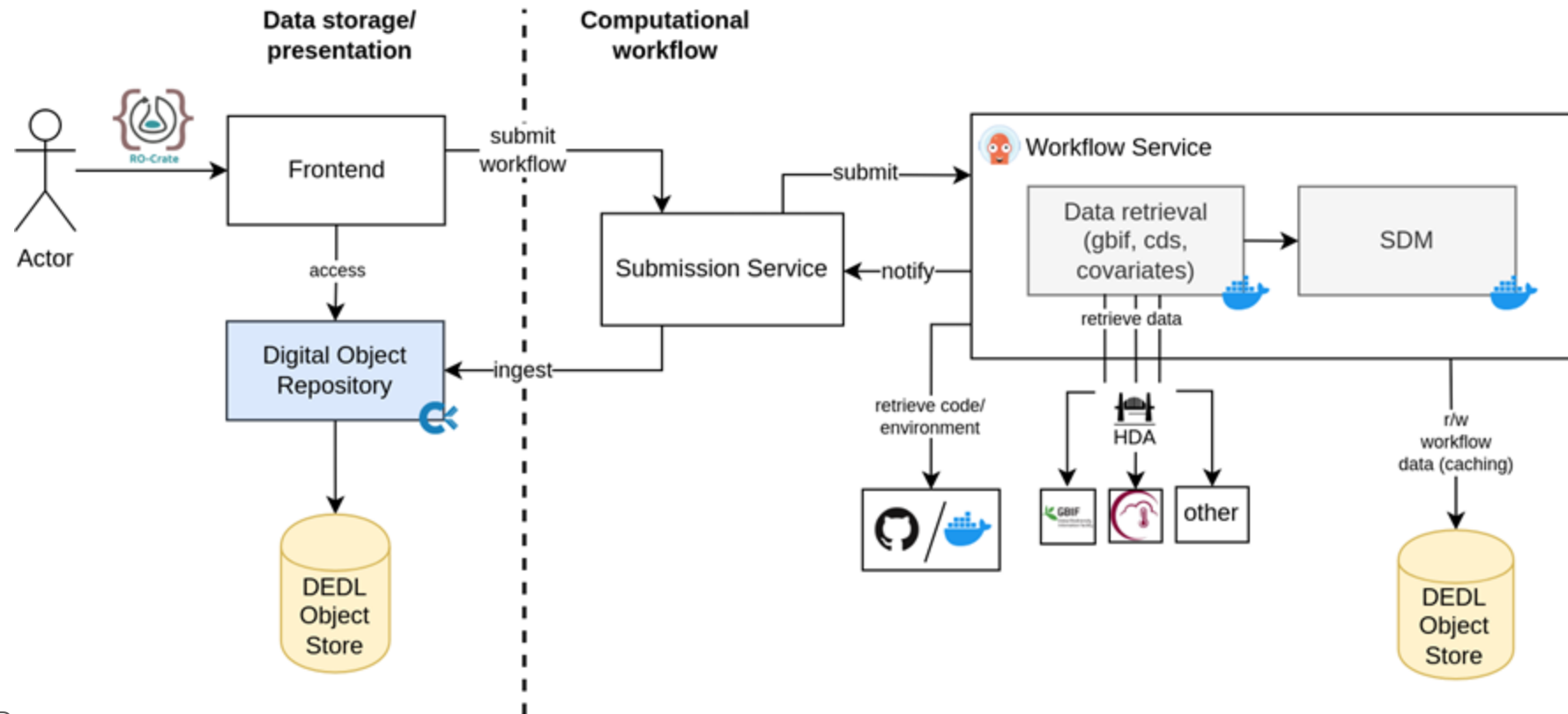
```
{
  "@context": [
    "https://schema.org",
    "https://orcid.org/terms/works/cv/credits"
  ],
  "type": "Dataset",
  "author": "https://orcid.org/0009-0009-0544-5117",
  "license": "https://spdx.org/licenses/CC-BY-SA-4.0",
  "describedBy": [
    "https://217.71.193.143/dataset/cwr/2b408313359f2b6f18c2",
    "https://217.71.193.143/objects/cwr/066270d30e33a451e64?download=91c23560-7799-4064-bc3f-95d62e722154/manifest.json",
    "https://217.71.193.143/objects/cwr/0077a98ac0644e020?download=91c23560-7799-4064-bc3f-95d62e722154/main.log",
    "https://217.71.193.143/objects/cwr/5956015a098ef326246?download=workflow.yaml"
  ],
  "download": "https://217.71.193.143/dataset/cwr/2b408313359f2b6f18c2?format=ROCrates&download=true",
  "item": "https://217.71.193.143/objects/cwr/066270d30e33a451e64?download=91c23560-7799-4064-bc3f-95d62e722154/manifest.json",
  "mainLog": "https://217.71.193.143/objects/cwr/0077a98ac0644e020?download=91c23560-7799-4064-bc3f-95d62e722154/main.log",
  "workflow": "https://217.71.193.143/objects/cwr/5956015a098ef326246?download=workflow.yaml"
}
```

```
> curl -Ik https://217.71.193.143/dataset/cwr/2b408313359f2b6f18c2 | grep link
link: <https://schema.org/Dataset> ; rel="type" ,
<https://orcid.org/0009-0009-0544-5117> ; rel="author" ,
<https://spdx.org/licenses/CC-BY-SA-4.0> ; rel="license" ,
<https://217.71.193.143/dataset/cwr/2b408313359f2b6f18c2?format=ROCrates> ;
rel="describedBy" ; type="application/ld+json" ,
...
```

Open Topics / Outlook

- Adding metadata from **within** the workflow to the output
- Credentials/Secret management still a challenge

Components



Deployment

