# Preston

a biodiversity dataset tracker

Jorrit H. Poelen

https://github.com/bio-guoda/preston

Biodiversity datasets are increasingly digital, integrated, online and openly available . . . thanks to an immense transnational effort in standardization, digitization and community building.
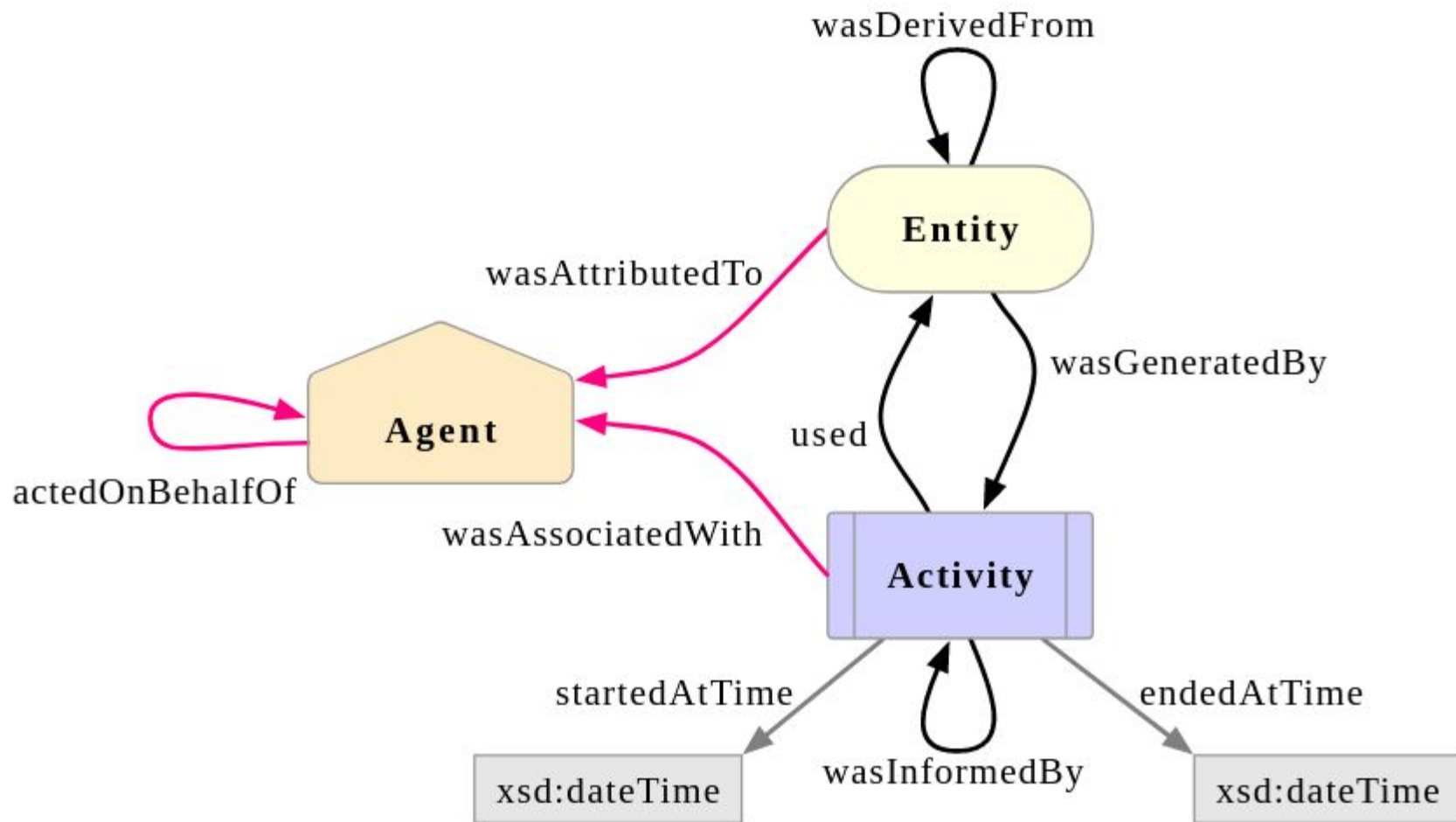
and . . .

Change is hard.

The controlled distribution, integration and archiving of ever changing datasets in an international consortium of institutions continues to be a challenge.

# Embracing Change.

Preston can track the change in biodiversity datasets and their registries using the Provenance (POV) and Provenance, Authoring and Versioning (PAV) ontologies and content-addressed hashes.

wasDerivedFrom

**Entity**

wasAttributedTo

wasGeneratedBy

**Agent**

used

actedOnBehalfOf

wasAssociatedWith

**Activity**

startedAtTime

endedAtTime

xsd:dateTime

wasInformedBy

xsd:dateTime

# Keeping track.

Preston is a command line tool that can perform carefully recorded visits to data registries (e.g., GBIF) and their partner institutions to store local versioned content in a biodiversity dataset graph.

Try it @ https://github.com/bio-guoda/preston

```
$ java -jar preston update "http://api.gbif.org/v1/dataset/suggest?q=Amazon&amp;type=OCCURRENCE"
<https://preston.guoda.org> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://www.w3.org/ns/prov#Sof
<https://preston.guoda.org> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://www.w3.org/ns/prov#Age
<https://preston.guoda.org> <http://purl.org/dc/terms/description> "Preston is a software program that finds
<83e11589-3579-47d8-ad5b-126e640112cd> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type> <http://www.w3.org/
<83e11589-3579-47d8-ad5b-126e640112cd> <http://purl.org/dc/terms/description> "A crawl event that discovers
<83e11589-3579-47d8-ad5b-126e640112cd> <http://www.w3.org/ns/prov#startedAtTime> "2018-09-05T09:42:39.614Z"^
...
```

Curious to hear your thoughts.