

Use Case: iNaturalist Observations

Status: Ready for review

Summary

The Citizen Science platform iNaturalist enables users to submit Occurrence records, optionally supported by one or more instances of digital photos or sounds. User's, including the submitter, may provide identifications and verifications of identifications. Records that fulfill a set of minimum requirements are labelled as "research grade" (see <https://www.inaturalist.org/pages/help#quality>) and shared via GBIF.

Requirements: The usual filtering capabilities for HumanObservations or MachineObservations.
Note: iNaturalist records are incorrectly labeled with *dwc:basisOfRecord* "HumanObservation" when backed by media evidence.

Enhancements covered in this use case include:

- alternative identifications along with the list of specific media used for each identification. Without this enhancement, iNaturalist records can be published adequately with the existing Darwin Core Occurrence Core (<https://tools.gbif.org/dwca-validator/extension.do?id=dwc:Occurrence>) and the Audubon Media Description extension (<https://tools.gbif.org/dwca-validator/extension.do?id=http://rs.tdwg.org/ac/terms/Multimedia>).

Highlights

- no physical evidence
- potentially multiple associated digital media as evidence for an Occurrence
- potential list of Identifications (not currently shared by iNaturalist)
- explicit references to specific digital media used in identification (not currently used by iNaturalist)

Concepts - see [Glossary](#)

Conceptual Model

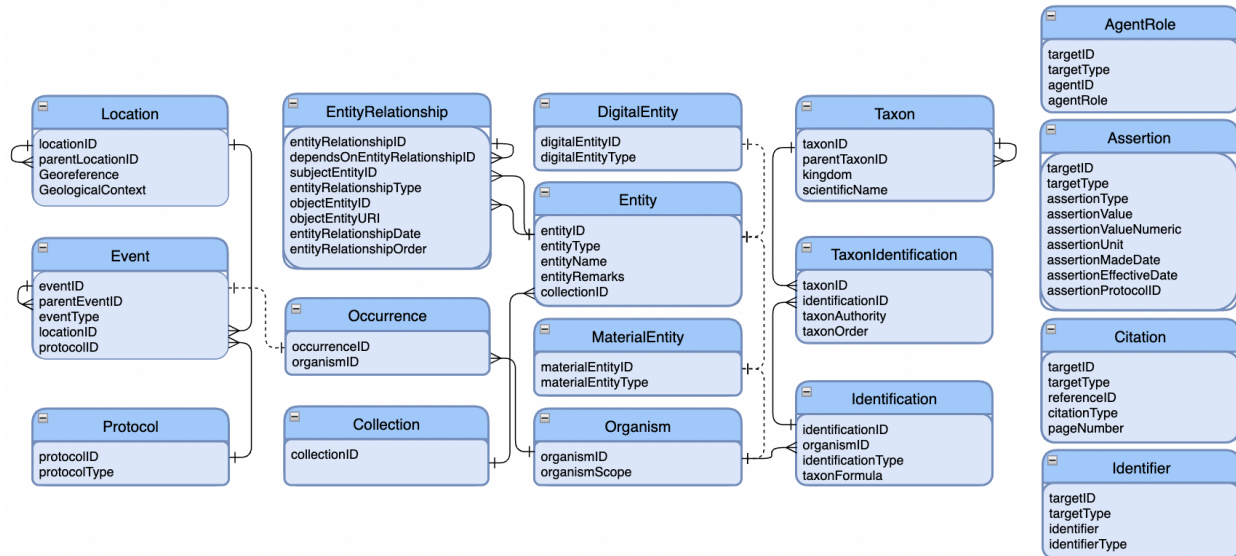


Figure 1: A conceptual model (2024-04-04) covering the activities associated with an iNaturalist observation, in which images provide the evidence for the Identifications of an Organism by multiple people.

An observing **Event** happens at a **Location** at a specific time (included in the **Event**), and may follow a specified **Protocol** with respect to the target **Organism**, (an **EntityOfInterest**). **DigitalEntities** (StillImages and Sound) may be gathered and are added to a **Collection**. The **DigitalEntities** may have derivatives. Multiple **Taxon Identifications** of the **Organism** can be made, each one using a potentially distinct set of **DigitalEntities** as a source of evidence.

Though not presented here, **DigitalEntities** and their derivatives can show parts of an **Organism**. Support for this could be added by following the design pattern in [Use Case: Find Tissue Samples](#). The *derivedFrom* and *partOf* relations should point only to the direct parent. It might be useful in the physical model to include a key to the *parentmostEntity* to shortcut to the highest source in the hierarchy, especially if that **Entity** is an **Organism**.

An arbitrary number of **Assertions** can be made about each class. **Assertions** can be quantitative or qualitative and can have **Assertions** made about them as well. **Agents** can have roles with respect to any class as well, including **Assertions**. Instances of any class may be referenced in **Citations** and have alternate **Identifiers**. These four common aspects of all use cases can be found in [Common Models](#).

Publishing Model

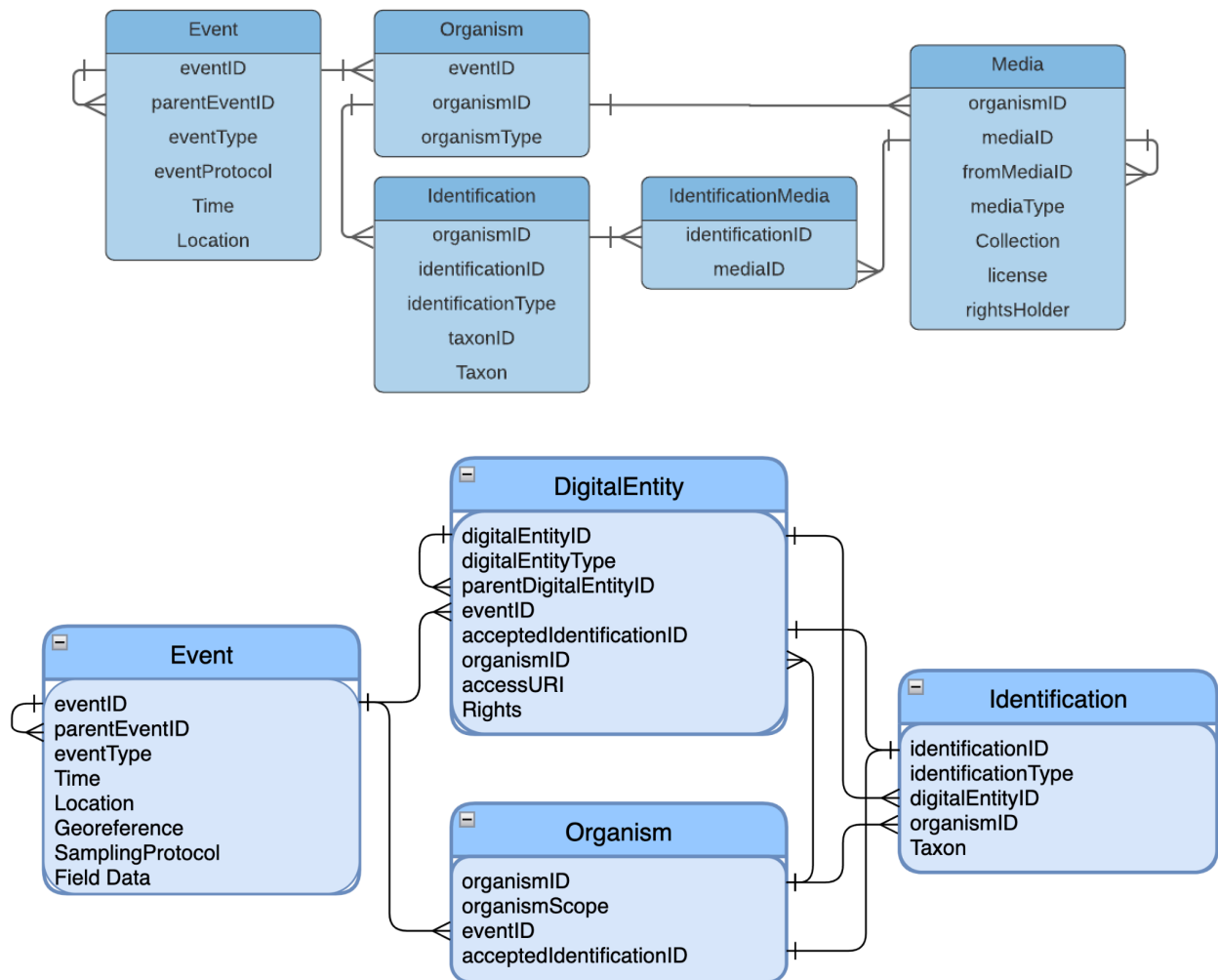


Figure 2. A simplification of the conceptual model for the purpose of publishing **Events** associated with observing **Organisms** (as **MaterialEntities** not necessarily collected), capturing evidence via **Media** (**DigitalEntities**), and identifying the **Taxon** of the source **Organism** from assessment of related **DigitalEntities**. Included fields are not exhaustive. Each class may be accompanied by **Assertions**, **AgentRole**, **Citations**, and **Identifiers**.

Event - one row for each **Event** in which an **Organism** is observed. The Darwin Core Event Core could be used as is.

Terms - all terms in the latest version of Event Core (<https://github.com/gbif/rs.gbif.org/tree/master/core>).

Organism - one row for an **Organism** observed in the **Event**.

Terms - *organismScope* is *dwc:organismScope*.

DigitalEntity - one row for each digital media entity (*StillImage* or *Sound*, iNaturalist does not support *MovingImages*) derived from an **Organism**. This extension could have **Collection**

information in it, or a **Collection** table could be used separately and linked. The Audubon Media Description extension could be used for this case

(<https://github.com/gbif/rs.gbif.org/tree/master/extension/ac>).

Terms: *organismID* (foreign key to the **Organism** that was the subject of the media), *digitalEntityID* (identifier for the media item), *parentDigitalEntityID* (identifier for a media item from which the **DigitalEntity** was derived), *digitalEntityType* (~*dc:type*, e.g., "StillImage", "Sound"), *license*, *rightsholder*.

Identification - one row for each **Identification** of an **Organism**. The terms in the Darwin Core Identification History extension (<https://github.com/gbif/rs.gbif.org/tree/master/extension/dwc>) could be used for this.

Terms: *organismID* (identifier for the **Organism** that was the target of the **Identification**), *identificationType* (e.g., "field identification", "expert identification", "identified by DNA"), and all of the appropriate terms of the Darwin Core **Identification** and **Taxon** classes.

Publishing Model using Darwin Core Archives

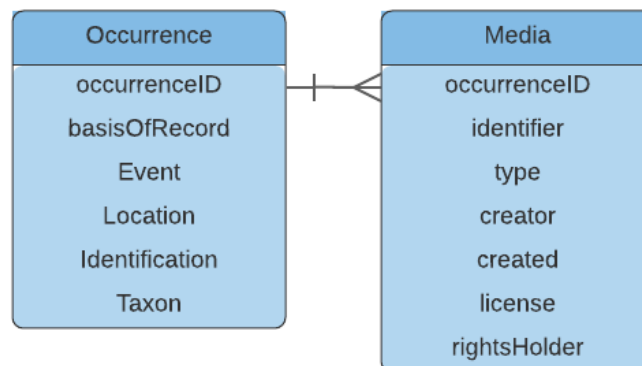


Figure 3. The highly simplified model for the publishing iNaturalist Occurrence in a Darwin Core Archive using an Occurrence Core and an Audubon Media Description extension. Included fields are not exhaustive. Occurrences could optionally be accompanied by an MeasurementOrFact extension, though this is not done in practice. Similarly, the Audiovisual Core is shown here but extension definitions that predate the Audiovisual Core - such as the GBIF multimedia extension - are often still used in practice for this kind of case.

References

[Common Models](#)

[Use Case: Find Tissue Samples](#)

<https://github.com/gbif/rs.gbif.org/tree/master/core>

<https://rs.gbif.org/extension/>

<https://tools.gbif.org/dwca-validator/extensions.do>
<https://tools.gbif.org/dwca-validator/extension.do?id=dwc:Occurrence>
https://github.com/gbif/rs.gbif.org/blob/master/core/dwc_occurrence_2020-07-15.xml
<https://tools.gbif.org/dwca-validator/extension.do?id=dwc:Event>
https://rs.gbif.org/core/dwc_event_2016_06_21.xml
<https://tools.gbif.org/dwca-validator/extension.do?id=dwc:MaterialSample>
https://github.com/gbif/rs.gbif.org/blob/master/sandbox/core/dwc_material_sample.xml
<https://tools.gbif.org/dwca-validator/extension.do?id=dwc:Identification>
<https://rs.gbif.org/extension/dwc/identification.xml>
<https://tools.gbif.org/dwca-validator/extension.do?id=http://rs.tdwg.org/ac/terms/Multimedia>
https://github.com/gbif/rs.gbif.org/blob/master/extension/ac/audubon_2020_10_06.xml

Acknowledgements