

Proposed changes to Darwin Core for invasive species research

What is Darwin Core?

Darwin Core (DwC) is a key data standard used for the dissemination of biodiversity observation data, most notably by the Global Biodiversity Information Facility (GBIF) . As such, its structure acts as a filter through which data are passed and it can have a profound influence on the availability and quality of shared data. For this reason it is critical that DwC is made suitable for the data needs of invasion biology and conservation more broadly.

The challenge

However, in practise the process of converting observation data to DwC can seem rather procrustean. Although DwC is a versatile standard, it is still in the process of evolution and lacks some finesse. A balance needs to be found between flexibility and standardization. On the one hand, flexibility helps data providers find data fields to map their data, but on the other hand, too much freedom dilutes the usefulness of data and the benefits of standardization.

Data from all sorts of biodiversity projects and sources must be reshaped into the data fields and structures provided by DwC, optimally, without losing valuable information in the process. The potential data sources are vast and include data from the whole earth, every living organism and all habitats. Furthermore, these data are collected for various reasons and those reasons have an influence on the formats and types of data collected. Therefore, DwC has a difficult role to play in providing a standard suitable for all observation data. Providing data for research on Invasive Alien Species (IAS) is only one of these roles, so improving DwC for IAS research needs to be done in such a way as to not reduce its usefulness for other research areas and to encourage contribution from a wide variety of data providers, not just those interested in biological invasions.

Data needs in invasion biology

There are many types of information that are needed for science and policy on IAS and many of these are already covered within DwC, including specific details of taxonomy and location. However, there are three basic pieces of information that are regularly collected but lack adequate provision in DwC. These are (1) whether the organism is native or alien to the location; (2) whether it currently exists at the location and (3) how was it originally transported to this new location.

Existing DwC vocabulary and suggested changes

[establishmentMeans](#)

This term already exists within Darwin Core and is defined as “*The process by which the biological individual(s) represented in the Occurrence became established at the location.*”

The suggested controlled vocabulary includes the terms “native”, “introduced”, “naturalised”, “invasive” and “managed”.

Our view is that this definition and the suggested vocabulary are mismatched and that the concepts of nativeness, management, invasion pathway and invasiveness should be in separate fields as they represent different concepts. We advocate adopting the invasion pathway categorisation terminology recommended by the [Convention on Biological Diversity](https://www.cbd.int/doc/meetings/sbstta/sbstta-18/official/sbstta-18-09-add1-en.pdf) for the `establishmentMeans` field (see <https://www.cbd.int/doc/meetings/sbstta/sbstta-18/official/sbstta-18-09-add1-en.pdf>). This will provide a richer terminology for data on introduction vector and a greater degree of standardization.

Furthermore, the need to make this change also highlights the concept of nativeness and the need for a dedicated data field for this within Darwin Core (see following section?).

origin

An expression of whether something is indigenous to an area is currently only expressed in DwC in the current controlled vocabulary of [establishmentMeans](#). However, by separating the concepts of whether something is indigenous from how it came to be at the current location enables an important and clear distinction of the organism’s origins at a site. Nativeness is considered an important criterion for the conservation of organisms when considering their importance to the biodiversity of a region.

We propose adding a new `'origin'` field to DwC adopting the [origin](#) field of the [IUCN](#). This simple vocabulary is already used by the conservation research community and aligning DwC and the IAS research community with this vocabulary has many advantages for standardization.

The IUCN has the following terms for the field origin, “native”, “reintroduced”, “introduced”, “vagrant” and “unknown”. One issue that remains to be resolved is whether there should also be a distinction between introductions that occurred before and after the modern era. Certainly in Europe this distinction is frequently made and has conservation value. In Europe the modern era is at the beginning of the Columbian exchange in 1500. For other parts of the world, such as New Zealand, this date maybe later.

[occurrenceStatus](#)

This field already exists in DwC and has a suggested vocabulary of “present”, “absent”, “common”, “irregular”, “rare” and “doubtful”. Though this information is useful it can equally well be expressed in DwC using the field `individualCount` or a combination of `organismQuantity` and `organismQuantityType`. However, in the particular case of species checklists there is no current means to express if something still exists in the area, or has become extinct. For this reason we suggest changing the controlled vocabulary of `occurrenceStatus` to the vocabulary used by the IUCN term ‘presence’. This term has

the simple vocabulary “extant”, “possibly extinct”, “extinct”, “extinct post 1500”, “presence” and “uncertain”.

We also recommend that the DwC [documentation](#) is updated to make it clear that this term is intended to be used for species checklists, rather than observations. This clarification of the terms of this field is not required specifically for invasive species research, however, combined with the addition of the origin term and the `establishmentMeans` it does enable a much clearer expression of the status of native and alien organisms in a region or at a site.

Example

The [Alien Plants of Belgium website](#) contains a full list of alien plants for Belgium and details of their introduction and status. To publish this checklist on GBIF the fields need to be mapped to Darwin Core. Below is a small section of the checklist and how I have interpreted it using the proposed vocabularies and new `origin` term. Source, Verloove F. (2016). Manual of the Alien Plants of Belgium. Botanic Garden of Meise, Belgium. At: alienplantsbelgium.be, accessed 25/08/2016.

A section of the checklist of the Alien Plants of Belgium

Taxon	Mode of introduction	First Record	Most recent record	Origin	Degree of naturalization	Means of introduction
<i>Acanthus spinosus</i>	Deliberate	2016	2016	E AF AS-Te	Cas.	Hort.
<i>Amaranthus clementii</i>	Accidental	1939	1949	AUS	Cas.	Wool
<i>Amaranthus albus</i>	Accidental	1857	N	NAM	Nat.	Grain, wool,
<i>Oenothera angustissima</i>	Accidental	1860	1884	NAM	Ext.	Ore

Darwin Core interpretation of the checklist illustrating the loss of relevant information and level of accuracy during translation from the checklist to DwC

Taxon	dwc:origin	dwc:occurrenceStatus	dwc:establishmentMeans
<i>Acanthus spinosus</i>	vagrant	extant	horticulture

<i>Amaranthus clementii</i>	vagrant	extinct	container/bulk
<i>Amaranthus albus</i>	introduced during the modern era	extant	seed contaminant container/bulk ¹
<i>Oenothera angustissima</i>	introduced during the modern era	extinct	container/bulk

1. Note that fields in Darwin Core that require multiple entries take a pipe (|) delimited string.

Other fields in Darwin Core perhaps relevant to alien species issues

MeasurementOrFact

measurementID
measurementType
measurementValue
measurementAccuracy
measurementUnit
measurementDeterminedBy
measurementDeterminedDate
measurementMethod
measurementRemarks

Occurrence

organismQuantity
organismQuantityType
lifeStage
reproductiveCondition
Behavior
establishmentMeans
occurrenceStatus
associatedTaxa
occurrenceRemarks

Organism

associatedOccurrences

ResourceRelationship

resourceRelationshipID
resourceID