

# DiSSCo related output

This template collects the required metadata to reference the official Deliverables and Milestones of DiSSCo-related projects. More information on the mandatory and conditionally mandatory fields can be found in the supporting document 'Metadata for DiSSCo Knowledge base' that is shared among work package leads, and in Teamwork > Files. A short explanatory text is given for all metadata fields, thus allowing easy entry of the required information. If there are any questions, please contact us at [info@dissco.eu](mailto:info@dissco.eu).

## Title

Natural Science Identifiers &  
CETAF Stable Identifiers

**Identifier of the author(s)**

## Author(s)

## Affiliation

## Publisher

**Identifier of the publisher**

## Resource ID

**Publication year**  
2020

## Related identifiers

## Is it the first time you submit this outcome?

Yes

**Creation date**

**Version**

1

**Project reference**

DiSSCo Prepare (GA-871043)

**Project output**

Tech Blog

**Deliverable/milestone number****Dissemination level**

Public

**Rights****License**

CC0 1.0 Universal (CC0 1.0)

**Resource type**

Text

**Format**

PDF

**Funding Programme**

H2020-INFRADEV-2019-2

**Contact email**



H2020-INFRADEV-2019-2  
Grant Agreement No 871043

The DiSSCo Technical Team gets asked a lot about Natural Science Identifiers (NSId). What are they? Why do we need them in addition to CETAF Stable Identifiers? Are they just for DiSSCo/Europe or are they global? In this post we answer those questions.

#### Q1. What is a Natural Science Identifier (NSId)?

A Natural Science Identifier (NSId) is a universal, unique persistent identifier for digitised natural science specimens (i.e., Digital Specimens) and other associated object types. An NSId will help you unambiguously refer to a specimen you are working with or will help to find a specimen that someone else has told you about by giving you the NSId e.g., as a reference in a journal article.

NSIds are intended to be long-lasting – on timescales familiar to collection managers, curators and scientists working with natural science collections. The design goal for NSIds is 100 years or longer with the intention to adapt and survive changes in their underlying technical implementation and support.

Technically, an NSId is a unique alphanumeric name string registered in the Handle System that acts as an opaque abstract reference to the thing that is identified; in this case, a Digital Specimen. Administration of the Handle System globally is a shared responsibility overseen at its top-level by the DONA Foundation. At the sub-global (Europe and other continents) level, DiSSCo is presently (mid-2020) analysing different options for the technical implementation.

#### Q2. What is the role of a Natural Sciences Identifier (NSId) compared to the CETAF Stable Identifier?

In Darwin Core (DwC) terms, a CETAF Stable Identifier (CSI) acts as a unique identifier for the occurrence (i.e., `dwc:occurrenceID`) represented by (for example) a preserved specimen in a natural science collection. In practice, because a CSI is often portrayed as, and is a valid URL, it references and represents a specific digital record for the specimen in an institution's data portal, and/or in a Darwin Core Archive package. In Europe, CSI are increasingly adopted by CETAF members, many of whom are also participants to DiSSCo as the means to provide human- and machine-readable access to the institution's available information about the corresponding specimen.

A Digital Specimen referenceable by its unique Natural Sciences Identifier (NSId) represents the sum of knowledge about a specimen object on the Internet. The NSId acts as the anchoring point for all known data and information about and/or derived from a specimen. From it (provided such information has been linked) you can find all that is known. The Digital Specimen acts as a surrogate for the physical specimen. It is a mutable digital object that can be manipulated; a common curation space where experts should be able to contribute above and beyond the local collection data.

#### Q3. How does an NSId function differently and complement the CETAF Stable Identifier?

When a specimen moves from one institution to another (e.g., change of ownership), or when institutions merge or form alliances the CETAF stable identifier can change. However, the NSId of a Digital Specimen never changes. It remains the same and always points (resolves) to where the Digital Specimen data is currently stored on the Internet.

This is like the prevailing situation for published journal articles, where Digital Object Identifiers (DOI) are

frequently used these days. Each journal publisher has their own hierarchical and location-specific structuring of URLs that point to journal articles they published. Constructed within registered Internet domain names that often contain elements of publishers' market branding, they tend to look like this example: [https://academic.oup.com/bioscience/advance-article/...](https://academic.oup.com/bioscience/advance-article/). The first part of this URL tells you it refers to an advance article to be published in the journal BioScience, published by Oxford University Press. When the article moves, for example from being an advance or early online publication article to one that appears in a specific issue and volume of a journal, this URL will change. A DOI, on the other hand provides a more abstract and neutral mechanism designed for the long term. The DOI for the previous example article is doi: 10.1093/biosci/biaa044, although even from this you can infer information (the journal name, at least). This DOI does not change, even when the actual location of the article changes.

The best DOIs (and other kinds of Handle, including NSId) are opaque ones that carry no information that could potentially become out of date and incorrect. In ideal scenarios using truly opaque Handles, events such as renaming of journals and publishers' mergers/acquisitions are invisible as far as the DOI is concerned. Behind the scenes, the DOI (Handle) record is updated to point to the new location of the article and most users remain none the wiser.

NSIDs perform the same function described for specimens. A Digital Specimen (with its NSId) will always contain a reference to the 'occurrenceID' (DwC) or 'Unit/GUID' (ABCD) of a specimen, allowing the physical specimen and its local digital record to be precisely located.

Q4. How does this work for collections that use GUIDs or Darwin Core triplet URNs instead of CETAF Stable Identifiers?

In many collections RFC 4122 globally unique identifiers (GUID) and other constructs, such as uniform resource names (URN) formed from Darwin Core Triplets (i.e., `institutionCode:collectionCode:catalogNumber`) are used instead of CSIs as unique identifiers for the occurrence. As with CSIs, these act as specific references for specimens in an institution's data portal, and/or in a Darwin Core Archive. Note, however that in the case where GUIDs are used, these on their own contain no information. Neither the location of a physical specimen nor the institution or collection to which it belongs can be deduced from a GUID.

Nevertheless, as explained above a Digital Specimen (with its NSId) will always contain a reference to the 'occurrenceID' (DwC) or 'Unit/GUID' (ABCD) of a specimen, which means that records in institutional data portals can still be linked to the NSId via the Digital Specimen. Often there will be enough other information present in the Digital Specimen, such as `institutionCode` and `collectionCode` to allow accurate dereferencing.

Q6. When is an NSId assigned to a Digital Specimen?

An NSId is assigned to a Digital Specimen (DS) when the DS is first created as part of a digitization process. It remains forever the unique and persistent identifier of that Digital Specimen throughout the entire lifecycle of the DS.

Q8. Can a Digital Specimen have more than one NSId?

The simple answer to this question is 'no'. Each Digital Specimen has a single NSId assigned by the authoritative institution for the specimen and associated with it throughout the Digital Specimen's entire lifecycle.

However, sometimes it is necessary to preserve and reference a snapshot of a Digital Specimen at a specific moment. In this case, the snapshot can be assigned an additional NSId that can be used to retrieve the state and content of the DS as it was at that moment. This identifier will not be an NSId in the range of the authoritative institution but another in the namespace of identifiers for attribution and provenance events. This identifier will include a timestamp as part of its suffix (i.e., the suffix is a RFC 4122 version 1 GUID from which the generation timestamp can be decoded).

Q9. Is this NSId mechanism just for Europe/DiSSCo or is it global?

The Handle System is global. That means Handles can be created for digital and/or physical objects anywhere in the world. By resolving these Handles the identified object can be found from anywhere else in the world.

Science based on specimens in natural science collections is global science. Scientists working with specimens are anywhere in the world. They collaborate with their global colleagues. They often work with specimens that can be in any of the thousands of collections housed in museums, universities and other institutions around the world.

Ideally, Natural Science Identifiers should be applied and used worldwide to make it easy to refer to and find what is needed. Proliferation of multiple similar identification mechanisms would only lead to confusion for scientists, without making their working practices easier and more efficient.

END.