

# Guidelines for Managing Unique Resource Identifiers

---

*Prepared by the iDigBio Information Technology Group*

An important outcome of the iDigBio Summit was a request that iDigBio provide guidance in creating and managing unique resource identifiers (URIs) for TCN and institution objects. The following is a proposal that provides unique, persistent, actionable identifiers coupled with a URI resolution service (to be provided by iDigBio) to deliver digital objects and their associated metadata. The strategy described below presents a pattern for identifiers that allows institutions and TCNs flexibility in tailoring identifiers to their needs and capabilities.

The standard for identification advocated by W3C is to use Universal Resource Identifiers (URIs). Each URI is a string that begins with a scheme name (or protocol). Registered schemes include `http`, `https`, `mailto`, `doi`, `ftp`, and `lsid`. Many URI schemes have been registered with the Internet Assigned Numbers Authority (IANA) [<http://www.iana.org/assignments/uri-schemes.html>]. The IANA registry encourages uniqueness of scheme names.

We recommend that providers adopt the `http` URI scheme for all identifiers. It should be noted that although this pattern resembles a URL (Universal Resource Locator), it does not have to be actionable or resolvable directly through a web browser. Details of how to use this scheme for identification are included below. Issues of URI resolution and action are addressed in the Appendix.

Providers may choose to use a different URI scheme but must use a permanent scheme registered with IANA.

Each provider must specify the strategies for URIs and register those strategies with the iDigBio portal. This information will be publically available on the portal.

## Definitions

**Unique Identifier:** a unique, unambiguous, and unduplicated name for an object. An identifier may be associated with a particular physical specimen or with a digital object.

**Persistent:** persistent identifiers are those that are used once, only once, and are associated with a single object. Once assigned to an object, an identifier cannot be assigned to a different object.

**Actionable:** identifiers are actionable when they can be incorporated into a service designed to deliver the referenced digital objects and/or their associated metadata.

**Digital Object:** A digital record of the properties of a thing. An image file is a digital object, as is a metadata record associated with the image file.

## What to Identify

Each specimen should have its own identifier. GBIF has relied on the Darwin Core triple of institution code, collection code and catalog number for specimen identification. There is no guarantee of

## Guidelines for Managing Unique Resource Identifiers

uniqueness for these triples and not every specimen has this information. These triples have not provided the properties that GBIF needs for reliable identification of occurrences over time. GBIF is now advocating identifiers like those described below for all data.

Each distinct digital object should have its own identifier. The primary digital catalog record of a specimen may be identified with the specimen's identifier or may have its own identifier. Each media object should have its own identifier.

### Recommended Unique Resource Identification Pattern

iDigBio recommends the following pattern for TCN and institution URIs. Inherent in this recommendation is a requirement that TCNs and other institutions ensure that all identifiers provided to iDigBio are unique.

The components of the pattern include:

1. Prefix: `http://`.
2. Domain: A TCN or institution domain name that is registered and owned by the TCN or institution: such as `ids.invertnet.org`. It is good practice to choose a name that is not associated with the primary institutional Web server.
3. Collection Identifier: A name for the particular collection, such as `/herb/` for herbaria. This is particularly important for museums or institutions that include more than one collection with potentially duplicated internal object names.
4. Object name: such as a bar code value or unique alphanumeric name.

#### Pattern:

```
http://ids.flnmh.ufl.edu/herb/abcd12345678
|           |           |           |
Prefix     Domain     Collection Identifier Object Name
```

### Summary

1. **Required:** provide a persistent, unique identifier for each digital object shared with the iDigBio portal
2. **Required:** adopt a registered URI scheme for identifiers.
3. **Recommended:** adopt the `http` URI scheme for identifiers.
4. **Recommended:** Use the above pattern for `http` URI identifiers.
5. **Required:** Register every URI scheme and pattern with the iDigBio portal.

## Appendix

### Object Services to be Provided by iDigBio

Standard object services include ones that accept an object identifier and produce a webpage for people to learn about the object, a metadata document about the object, and the digital object itself. The iDigBio portal will provide all of these services for the digital objects accessible from the portal. Provider organizations may choose to provide some of all of these services for their objects.

Each request for metadata about a digital object will return a metadata document in some particular format. iDigBio services will produce metadata documents in RDF, RSS and JSON formats. Other formats will be supported as needed.

RDF metadata documents are of particular interest because of their use in the Linked Data protocol, uses HTTP GET requests to access web pages and metadata. If the header of the request includes the parameter "Accept: application/rdf+xml," the service will return an RDF document containing the object metadata. Without the parameter, a web page will be returned by the web server. iDigBio will use the Linked Data protocol for serving web pages and metadata documents.

### iDigBio Proxy Services

The iDigBio portal will be capable of redirecting object service requests to provider services, in order to assist providers in creating and managing both identifiers and object services. The portal will serve as a proxy for the provider. The portal will include a facility for registering URI patterns and service end points. When a proxy request is received, the portal will use standard http capabilities to redirect the request to the provider service.

In the example below, the herbarium collection of the Florida Museum of Natural History has registered a URI pattern with iDigBio. A request for information about the particular digital object is sent by a user to the proxy server at iDigBio.

```
http://proxy.idigbio.org/?q=http://ids.flnmh.ufl.edu/herb/abcd12345678
```

The proxy server will send the user a response to redirect the request to the following URL for processing by the Museum's object services.

```
http://services.flnmh.ufl.edu/herb/?id=abcd12345678
```

### Version management

Identifiers can be used to represent objects whose content is subject to change. iDigBio intends to provide a service for fetching a particular version of an object by date or version number. If the content of an object changes so much that it can be considered a different object, a different identifier should be attached to the new object.