

Guidelines for Managing Persistent 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 persistent identifiers and services for TCN and institution objects. The following is a proposal that provides unique, persistent, actionable identifiers based on a URI scheme (see Section 2 of “Architecture of the World Wide Web” <http://www.w3.org/TR/webarch/#identification>), together with a resolution service to be provided by iDigBio. Adoption of this recommendation will allow iDigBio and specimen repositories to acquire, maintain and 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 2 most important reasons for creating and maintaining identifiers for objects are:

1. An identifier provides a handle that can be used for keeping track of all of the characteristics of an object, including its primary properties, commentary on those properties, and relationships with other objects,
2. An identifier provides a handle that can be used to provide services for the object, including those that deliver the data and metadata of the object, and those that link the object to services for other objects.

The properties and services of an object are dependent on the ability of the provider of the object to maintain the mapping from the identifier to the object. In this document we recommend keeping track of the mapping by creating identifiers that include primary properties of the object, such as institution code, collection code and catalog number.

It is important to realize that these properties may change, for instance if a collection is moved from one institution to another, or a specimen moves from one collection to another. Each provider must make a commitment to maintain the mapping from identifier to object as those properties change.

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 `urn`. 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 publicly available on the portal.

Definitions

Persistent: persistent identifiers are those that are assigned 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.

What to Identify

Each specimen and each distinct digital object should have its own identifier. In particular, 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   |         Object Name
                  |
                  Collection Identifier
```

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.