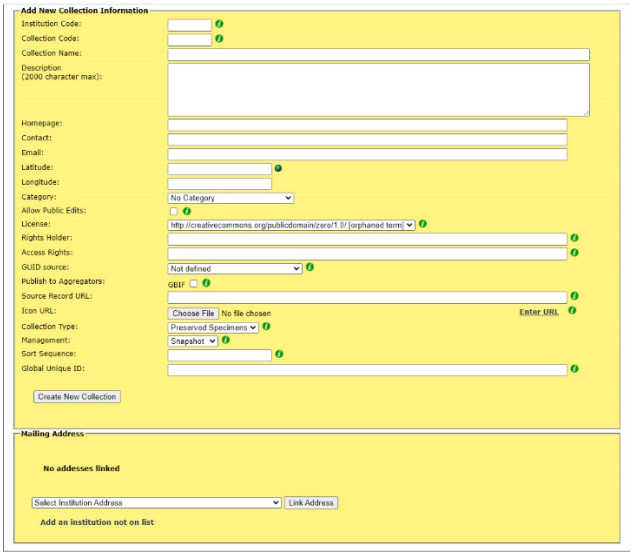
Help document for

Collection metadata



Author: Mary E. Barkworth

Version 1.2, published 30 Jul 2021

Introduction

"Metadata" refers to a set of data that describes other data. Before a collection can be added to a Symbiota or Symbiota2 network or portal[1], the collection's manager needs to provide the portal manager with metadata about the collection. This document explains the fields on the fields on the form (Fig.1) used by Symbiota and Symbiota2 to collect the metadata needed. Some of the metadata for a collection is publicly available. It can be viewed by clicking on the "more info" link by the collection's name (Fig. 2).

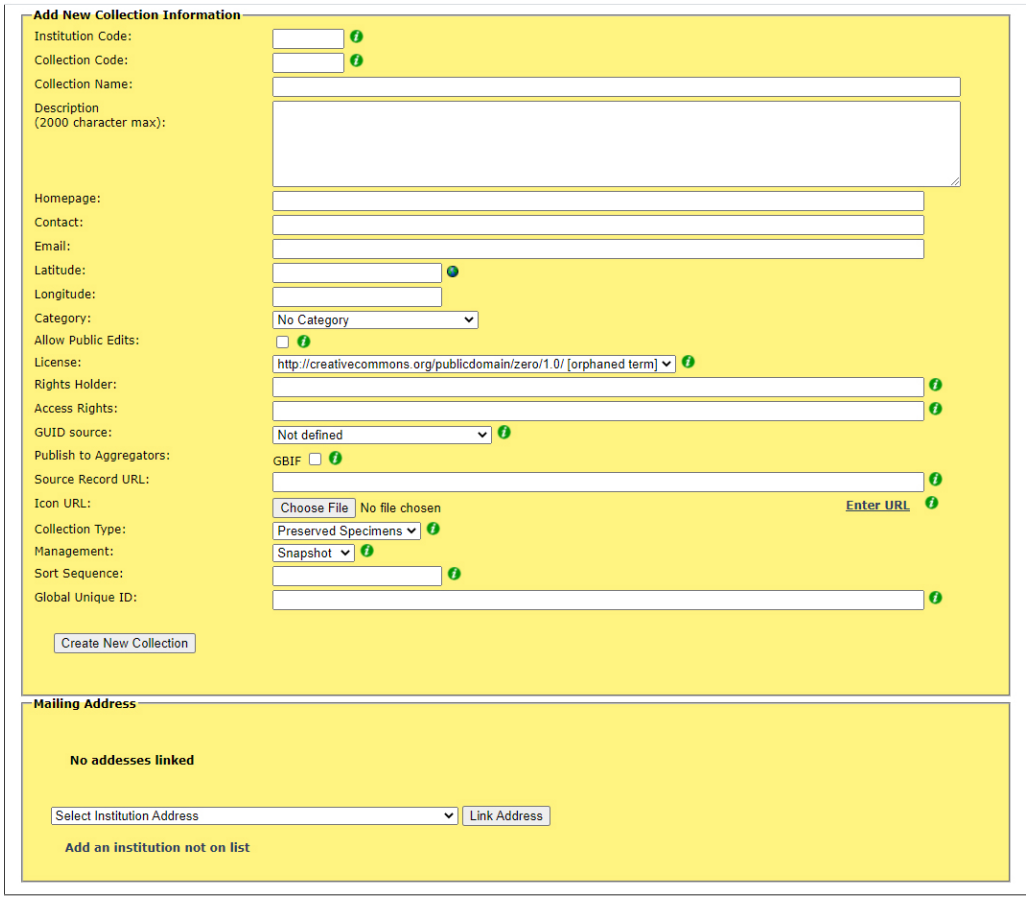


Figure 1 The form for entering and editing a collection's metadata

[1] "Network" refers to all parts of digitally connected entities; "portals" are subsets of networks. Portals may serve a subset of user groups, such as collections within a region or different kinds of specialists. Portals may provide different kinds of access to their users. Symbiota and Symbiota2 are content-management systems for networks. A network may have multiple portals, each with a different focus, that share skills and information across the network, eliminating the need for each portal to develop specialists in all aspects managing the network.

The initial version of a collection's metadata form will be completed by the portal manager, using information provided by the collection manager. Once a collection has added at least one specimen record, it will be included on the list of contributing herbaria revealed when the Search/Search collections link has been clicked (Fig. 2). After that, the collection manager collection is responsible for maintaining the currency of the data shown. Most fields will not change, but some, such as contact information and description, might.

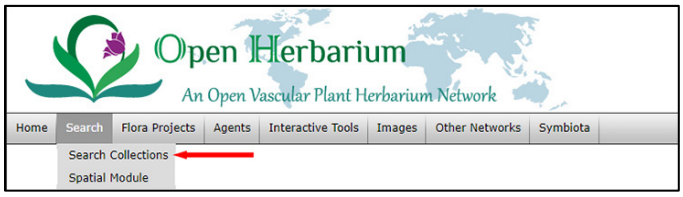


Figure 2 Link for finding the collections contributing to a Symbiota network

Finding the metada for your collection

To view the metadata for your collection, login to your network, click on the "more info´ link beside your collection's name (Fig. 3).

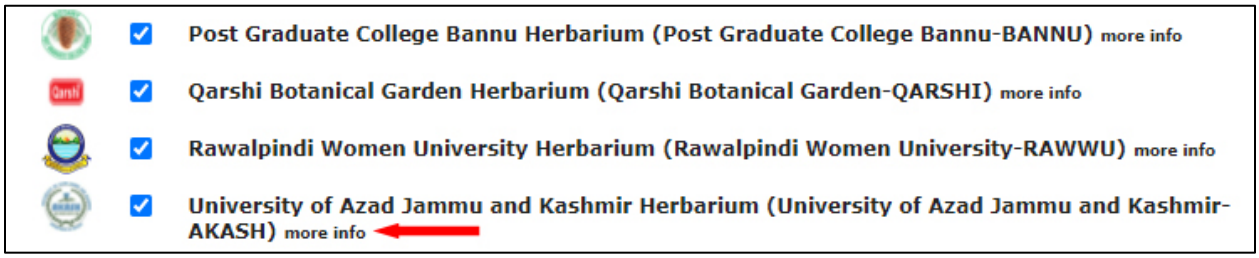


Figure 3 Link for viewing metadata (arrow)

You should see a pencil in the top right corner of the page that comes up (Fig. 4). It is visible only to those with editing privileges for the collection. Click it to see whether you have permission to edit the metadata.

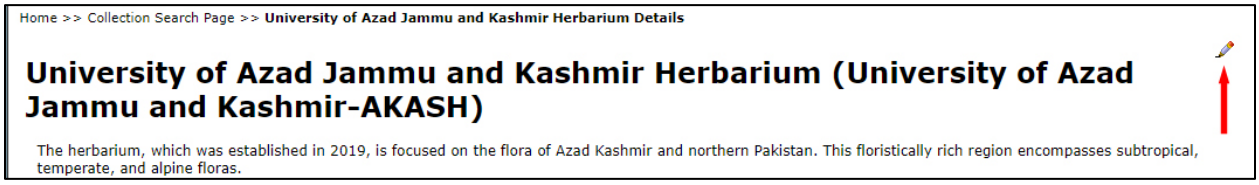


Figure 4 Link to editing panel for the University of Azad Jammu and Kashmir Herbarium

If you do, you will see two boxes (Fig. 5). The link that brings up the page editing version of the metada page is in the second panel. If you only see the Data Editor Control Panel, you do not have permission to edit the metadata. Initially, it is given only to the contact person for the collection. If that person is available, ask them for permission. If he or she is no longer available (retired, changed position, etc.) contact the portal manager and explain the situation.

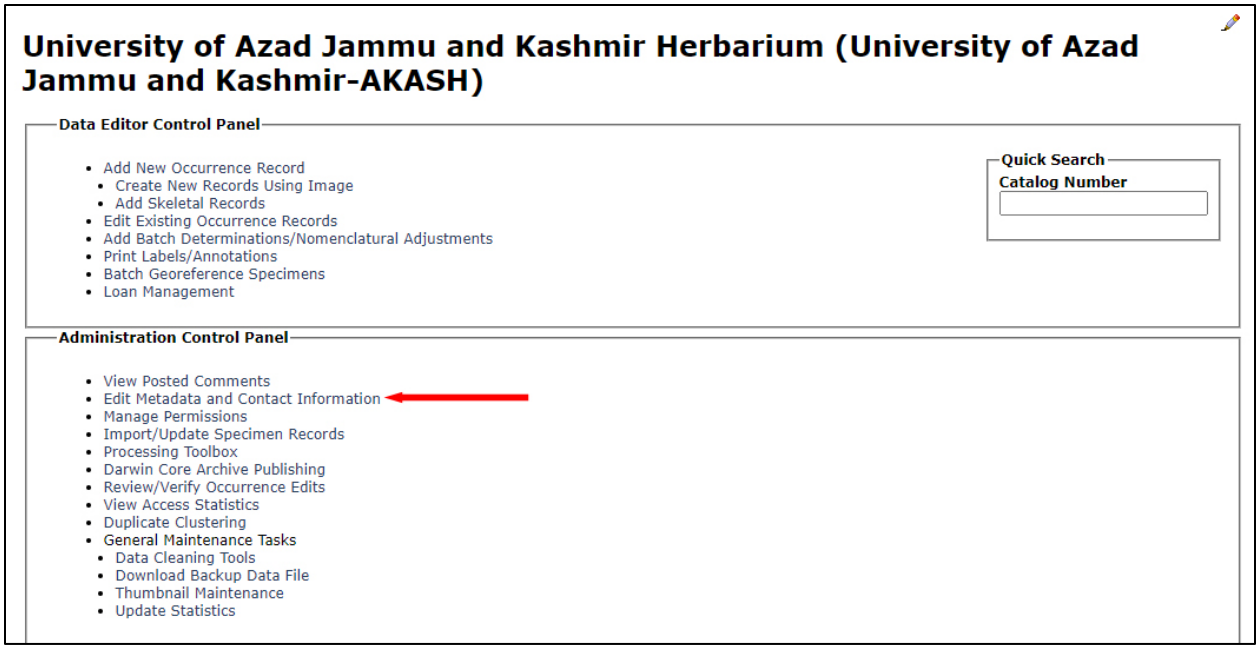


Figure 5 Link to the page for editing a collection's metadata

When a collection agrees to share its specimen data with additional networks such as the Global Information Facility (GBIF), its metadata is part of the shared data. In Symbiota2, portal administrators will be able to send out a such reminder to all in their portal managers on a regular basis. The most important fields to keep up to date are the name of the contact person and the contact email.

The metadata fields

The remainder of this document explains what should go in each of the fields used on the Collection metada page (Fig. 1). For each field, three comments are provided: a) how the field's data means and how they are used; b) recommendations on deciding what to put in the field; and c) differences or potential differences between Symbiota and Symbiota2+ with respect to the field.

The first priority for Symbiota2 is to ensure it offers all the functionality of Symbiota, but once that has been accomplished, it will be given additional functionality. The "Symbiota2+" comments indicate changes that are planned. Some of them will be in the first release of Symbiota2; when others are incorporated will depend on their perceived priority to users (and funding). This uncertainty is why they are listed as changes to come in version Symbiota2+, not Symbiota2.

Institution code

Explanation: The Institution Code is used by GBIF and other aggregators to report on data coming from different collections belonging to the same institution. All collections belonging to a single institution should have the same information in the Institution Code field even if they operate independently. Existing Symbiota networks vary in how the field is use, some entering a code that is customary in a particular discipline, others using it for the collection code or an acronym for the institution. Changes should not be made without consulting the network manager because they affect the information provided aggregators.

GRSciColl: The Global Biodiversity Information Facility is building a global registry of scientific biodiversity collections, GRSciColl. It took this over from other organizations and is still cleaning the data it inherited. GRSciColl has two fields for institutions, the InstitutionID, which

is its name, and the InstitutionCode. To see how your institution is registered, search for it here. Clicking on the name of the institution will shown the collections that are currently linked to that name. To search for a collection in GRScColl, click here. If you do not find your institution or collection, try alternative names or codes. If you find what you are looking for but wish to edit it, that can be done online. Click the name of the collection or institution to bring up the page for suggesting edits. If you are unsuccessful in finding your institution or collection, click the contact icon at the top of the page (Fig. 6) and or use the contact link near the bottom of the GRSciCOll home page.

Figure 63 Location of GBIFcontact icon

Recommendation 1: *For collections joining a Symbiota network for the first time:* Enter the institution's *name* in this field because it is likely to be globally unique. Examples: "Pakistan Museum of Natural History", "United States Forest Service", "University of Tennessee", "University of Texas", "Utah State University".

Recommendation 2: *For collections already in a Symbiota network*, make no change without consulting your network manager.

Symbiota2+: There will be two fields for entering institution information, "Institution name" and "Institution code". The institution name refers to an upper administrative unit, often a university or a government agency. For "Institution code", enter something that is meaningful locally, such as an acronym (USU for Utah State University). It is highly improbable that two insitutions will have identical names, but highly probable that two or more will be referred to by the same code. Again, allow the network manager to decided what goes where.

Collection code

Explanation: The Collection Code is used to identify individual collections. Ideally each collection would have a unique code, one that is not used by any other biodiversity collection. In practice, this is not easy because many older collections have codes that, although widely recognized within their discipline or region, are not globally unique.

Following the recommendations given here will help ensure *new* collections have a globally unique collection codes by preceding joining a Symbiota network with registration in one of two global registries, *Index herbariorum* (IH; collections of algae, fungi, and plants) and GRSciCOLL (for all other biodiversity collections).

The collection code is often incorporated into the barcodes now used by many collections as the catalognumber for a specimen. Such barcodes have the format CODE0000001, CODE0000002, …. This is not required but it makes it easy to identify the collection to which a specimen belongs plus barcodes can be "read" by computers. This is useful when digitizing a collection.

Recommendations: **Herbaria** should register with *Index herbariorum* before registering with a collection network. Registration can be completed online by clicking the button to the right of the map). One requirement will be to suggest the code desired. It can be up to six letters long. Check that it is not already being used by searching for it in IH. If the search returns the name of a herbarium, the suggested code is already in use and you will have to make a different suggestion. Also, if the herbarium is old but has not been active for some time, use the search feature to check that it does not already have a code by searching for it by location.

If the herbarium is already registered with IH, check the information in it, including the information about staff members, phone numbers, and emails.

Index herbariorum collects more data than GrSciColl. The additional data provide a more complete description of the collection, provide more guidance as to what can be found in the collection, and is used in generating annual reports about the world's herbaria. The accuracy of the data it contains depends entirely on voluntary editing by collection managers.

Museums and zoological collections: Register with GRSciCol. Registration can be completed online. The individual registering the collection must first register with GBIF. As with *Index herbariorum*, check that the proposed code is not already in use within GRSciColl. If you have questions, contact GRSciColl using the contact link near the bottom of the home page.

Symbiota2+: No change anticipated

Collection name

Explanation: This name is used to alphabetize the collections within a portal. It is also used by GBIF in identifying a collection.

Recommendation 1. If your collection has a name, use that name. If it does not, use a name that starts with a word or phrase familiar to those using it. For example, "University of Someplace Herbarium/Museum", rather than "Herbarium/Museum of the University of Someplace", "Elko BLM herbarium, rather than BLM herbarium Elko"

Recommendation 2. If subsets of a collection are managed separately or are sent to different networks, add a phrase to the collection name to explain which records are being sent to the network concerned. For example, The Intermountain Herbarium (collection code UTC) sends its records to four different Symbiota portals: vascular plants and algae [it has very few algae] go to Intermountainbiota (which feeds into the SEINet network), fungi other than lichens go to Mycoportal; lichens go to the Consortium of North American Lichen, and bryophytes to the Consortium of North American Bryophyte Herbaria. As listed in GBIF they are Intermountain Herbarium (vascular plants and algae), Intermountain Herbarium (fungi, not lichens), Intermountain Herbarium (Lichens), and Intermountain Herbarium (bryophytes), respectively.

Symbiota2+: No change anticipated.

Description:

Explanation: This tells people about the collection. Researchers use the description to decide which collections to visit or where to send loan requests. It appears on the home page for the collection within the portal.

Recommendation. The maximum number of characters permitted (including spaces) is 2000. Most descriptions are shorter, often much shorter. The description of a young collection might, for example, say something like, "The herbarium of XYX University was founded in 2000 to document the plants and fungi sold by food vendors in and around XYX. It has about 1000 specimens".

Suggestions as to what to include are:

- The year the collection was started;
- An estimate of its current size (if it has many non-digitized specimens);
- Its purpose;
- Its taxonomic focus and/or purpose (all fauna, flowering plants, species used for food or medicine in an area, arachnids, ferns);
- Its geographic focus (country, vicinity of a town, a valley);

- Major projects for which the collection has or is playing a lead role;
- the names of well-known collectors for whom the collection was the primary depository during a significant part of their life;
- whether or not collection has any type specimens.
- The description of a young collection might, for example, say something like, “The herbarium of XYX University was founded in 2000 to document the plants and fungi sold by food vendors in and around XYX. It has about 1000 specimens”.

Symbiota2+: No change expected.

Home page

Explanation: if a collection has a home page, this field helps people find it.

Recommendation: Leave empty if there is no home page.

Symbiota2+: No change expected.

Contact

Explanation: Used by 1) individuals wishing to ask a question about the collection and 2) by portal and network managers wishing to alert collection managers about pending updates, outages, requests, or reminders.

Recommendation: The should be someone with administrative privileges on the collection.

Symbiota2+: There will be space for the name, email, and role of two contact individuals. The purpose of the secondary contact is to ensure that portal and network administrators can contact someone responsible for the collection even if the primary contact retires or leaves.

Email

Explanation: This is an email address that will reach the the collection’s contact person. Large collections may have a “Collection email” that can be open by two or more people, for example, by the director and all curators or by the collection manager and the director.

Recommendation: Continue existing practice.

Symbiota2+: Symbiota2 will provide space for two contacts, together with their emails and roles in the herbarium. See comment for “Contact”. One of these could be a collection email.

Latitude and longitude

Explanation: Potential uses include showing a mapf of contributing collections. The data must be in decimal format. It can be determined from Google maps.

Recommendation: Do not provide more than 3 decimal places; extra places are Explanationless. A difference in latitude of 0.001° corresponds to a length of 11 m. The distance in meters corresponding to a longitude difference of 0.001° varies because lines of longitude come together at the poles. At the equator, a difference in longitude of 0.001° corresponds 11 m; at 45° north or south, it corresponds to 8 m; at 80° north or south, it corresponds to 2 m.

Symbiota2+: Symbiota2 will use it to generate maps of networks and their portals. There will be an option enabling their display on the home page of portals and networks.

Category

Explanation: Networks with many contributing collections may divide them into categories based on some criterion, such as geographic location. By default, the search page lists categories

alphabetically (and collections alphabetically within a category). The category is determine by the network manager.

Recommendation: Leave as is, determined by the network manager.

Symbiota2+: No change expected.

Allow public edits

Explanation: By default, only people with explicit permission can edit the records in a collection; their edits are applied immediately. Checking this button enables anyone with an account on the network to edit a collection's records, but the changes are not applied until approved by the collection manager.

Recommendation: Leave box unchecked unless there is someone who can be assigned to checking any edits submitted.

Symbiota2+: No change expected.

License

Explanation: This is a legal statement about who and under what conditions others can use the data and images provided by a collection. Individual networks may vary in their license policies. For OpenHerbarium and OpenZooMuseum, the default license is Public Domain (CC 0) which ["… allows creators to give up their copyright and put their works into the worldwide public domain. CC0 allows reusers to distribute, remix, adapt, and build upon the material in any medium or format, with no conditions." The most restrictive license acceptable to these networks and GBIF is the Creative Commons license CC BY-SA. , "This license allows reusers to distribute, remix, adapt, and build upon the material in any medium or format, so long as attribution is given to the creator. The license allows for commercial use. If you remix, adapt, or build upon the material, you must license the modified material under identical terms."

If there is a list of "species of concern" for an area, access to "Locality" information and specimen images for those species from that area will require special permission even if the default license for the collection is "Public domain." " Species of concern" should include species on endangered species lists but also species that are not formally listed as endangered but are the focus of poaching. If such a list exists, ask the network administrator to include it in the network's checklists. It is also possible to put similar blocks on all species growing in specific geographic areas such as private land.

Recommendation: Use the CC 0 license because few collections have the resources for pursuing those who do not comply with the conditions of a more restrictive license.

Symbiota2+: No change expected.

Rights holder

Explanation: Name of the institution that owns the collection.

Recommendation: If you are not sure, ask the administrator above you in the chain of command. Before doing so, find out whether there is a list of "Species of concern" for your country in case you are asked about protection from biopiracy.

Symbiota2+: No change expected.

Management

Explanation: This refers to how records are made available to the network. There are three methods. The method used by a collection should be selected in consultation with portal or network manager.

- Live data entry: record information is added directly into the network database through the appropriate data entry form. The collection does not maintain a separate database. Edits made online appear immediately.
- Upload from a file generated by an internal collection database system such as BRAHMS or Specify and mapped to Symbiota. This is referred to as "Snapshot" data entry. If any of the records need to be modified, the modifications must be made in the collection's internal database. They will appear online when the data from the internal database are next uploaded to the network.
- Spreadsheet data entry: When data are entered into a spreadsheet for subsequent upload, data entry is said to be live. Corrections to a record should either be made online or by uploading a spreadsheet with, where appropriate, revised data for one or more records. When the a spreadsheet with revised data is uploaded, incoming data will be matched to existing data by the CatalogNumber [barcode] and *all* the old data for that record replaced.

Recommendation: It there is reliable, fast internet and no existing database, use live data entry. If there is an existing, well-designed relational database, work with the portal/network manner to develop an export/upload routine. This only needs to be done once, after which uploading data will be a matter of 2-3 clicks. If there is no existing relational database and internet access is slow or unreliable, use a spreadsheet. Ask the network manager for a spreadsheet designed for use in your country.

Symbiota2+: It will make it relatively simple for collections to intall a local version of the online database that is synchronized with the online database when internet access is available. Installing the local version and providing training on its use will require someone with IT skills, but this could be someone associated with the lead institution.

GUID source

Explanation**:** This is a globally unique identifier for an occurrence record. For live data and spreadsheet data entry (see "management"), it is provided by the network. For data imported from a database managed by the collection, that database should provide a GUID for each record. The GUID format is a randomly generated string of 128 integers.

Recommendation: Dp not change the setting used when first joining the collection unless the mode of data entry changes. Changes should be made by the network's IT manager.

Symbiota2+: No change expected.

Publish to Aggregators

Explanation: Aggregators aggregate records from different collections. Symbiota portals and networks are aggregators, but there are many other aggregators, such as GBIF, which aggregates data from many different kinds of database located in any part of the world. Symbiota makes it very easy to share data with GBIF. Collections need to apply to become a GBIF publlisher using the online form. When approved, they will be sent a password. Once received, the collection's administrator should check the GBIF box in the collections metadata page, and click the Darwin Core Archive Publhing Link in the Administrative toolbox. This will ask for the password. Once inserted, the data will be published in a format that GBIF can ingest. The data will appear in GBIF within 48 hours.

Recommendation: **Do not check** the box if you are already sharing the colletion's records with GBIF. If the collection's records are not currently being shared, apply to become a GBIF publisher once you have digitized 100 specimen records and are able to start a regular program of

digitization, even if only of 5 specimens a week (of course, more would be better), until it is completely digitized.

Symbiota2+: It will become easier for a collection to share its records, or subsets of records, with multiple specialist aggregators, such as Mycoportal (for fungi excluding lichens) or Pteridoportal (for Pteridophytes).

Source Record URL

Explanation: The source record field is used when a collection maintains its own database on a separate server from that of the Symbiota network.

Recommendation: The IT person in charge of the locally maintained database should provide the necessary information to the portal manager.

Symbiota2+: No change expected.

Icon URL

Explanation: The icon is displayed whenever a record from the collection is displayed. It helps draw attention to the source of the record. The maximum size is 350 x 350 pixels.

Recommendation: Review the icons displayed on the Search page of a Symbiota network to help you decide on an icon to use for your collection. They should be square and show well when displayed. Icons are always small, so should be simple. Color helps make them memorable.

Symbiota2+: No change expected.

Collection Type

Explanation: This field indicates the basis for the majority of the occurrence records in a collection. There are five recognized terms: PreservedSpecimen, LivingSpecimen, FossilSpecimen; HumanObservation, MachineObservation, or MaterialCitation. For natural history museums, the appropriate choice is probably "PreservedSpecimen". Symbiota also allows for recording image-based voucher occurrence records. For these, the appropriate collection type is "MachineObservation." For field notes, the BasisofRecord is "MaterialCitation".

Recommendation: Use appropriate recognized terms.

Symbiota2+:

- Enable overriding the default collection type setting for individual records, for example, to allow for occasional image-documented records provided by a collection's staff.
- Enable incorporating more explicit descriptions from the Dublin core as recommended for use in DwC. The more explicit types are "Still image", "Moving image", and "Sound". This will require additional modification of the backend which makes it a more significant change than enabling some records to override a default collection type.

Sort sequence

Explanation: Determined by the portal manager. By default, the order within a *category* is alphabetical, using the *CollectionName*. This is one reason the *Collection name* should start with words people use, not "herbarium", "Fish collection".

Recommendation: Leave empty so that sorting is alphabetical.

Symbiota2+: No change

Globally Unique ID

Explanation: This is a GUID for the collection. If the collection already has a GUID (assigned by a local database such as BRAHMS, Specify, or a data aggregator, that value should be entered here. If not, a value will be assigned by the Symbiota network.

Recommendation: If using a local database, ask its manager whether it assigned the collection a GUID. If so, contact the portal or network manager about entering it here. If not, let the network set the value.

Symbiota2+: No change expected.

Security key

Explanation: Set by the network; No change allowed.

Symbiota2+: No change expected.