Source Habitat Information: Descriptors and Codes

Revision Date

May 7, 2024

Author

Martin Reisinger

This guide provides details on the set of dataviews in which source collection site/habitat information may be kept. A core set of fields are stored in the accession_source table, but with some additional source/habitat tables, GG now can handle any data that an organization intends to store on the collection site.

Genebank managers can create custom source habitat descriptors (and codes) for an unlimited amount of detail on the collection site. (This set of tables / dataviews is similar to the set of dataviews used to manage crop traits and evaluations.)

NPGS Vetting of Descriptors

In the NPGS, the S/H descriptors will be vetted by the GIS subcommittee, to help maintain some level of standardization.

The Appendix contains change notes pertaining to this document.

Comments/Suggestions:

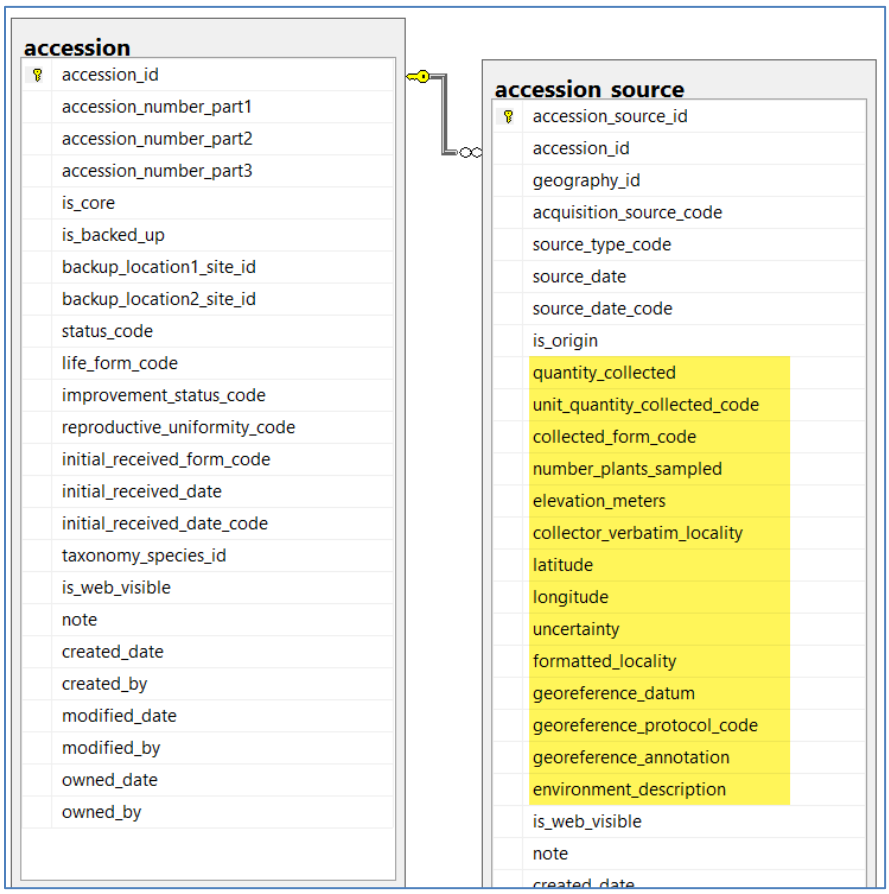
Please contact feedback@ars-grin.gov with any suggestions or questions related to this document. This and other GRIN-Global –related documentation can be downloaded from the GRIN-Global Training page.

Table of Contents

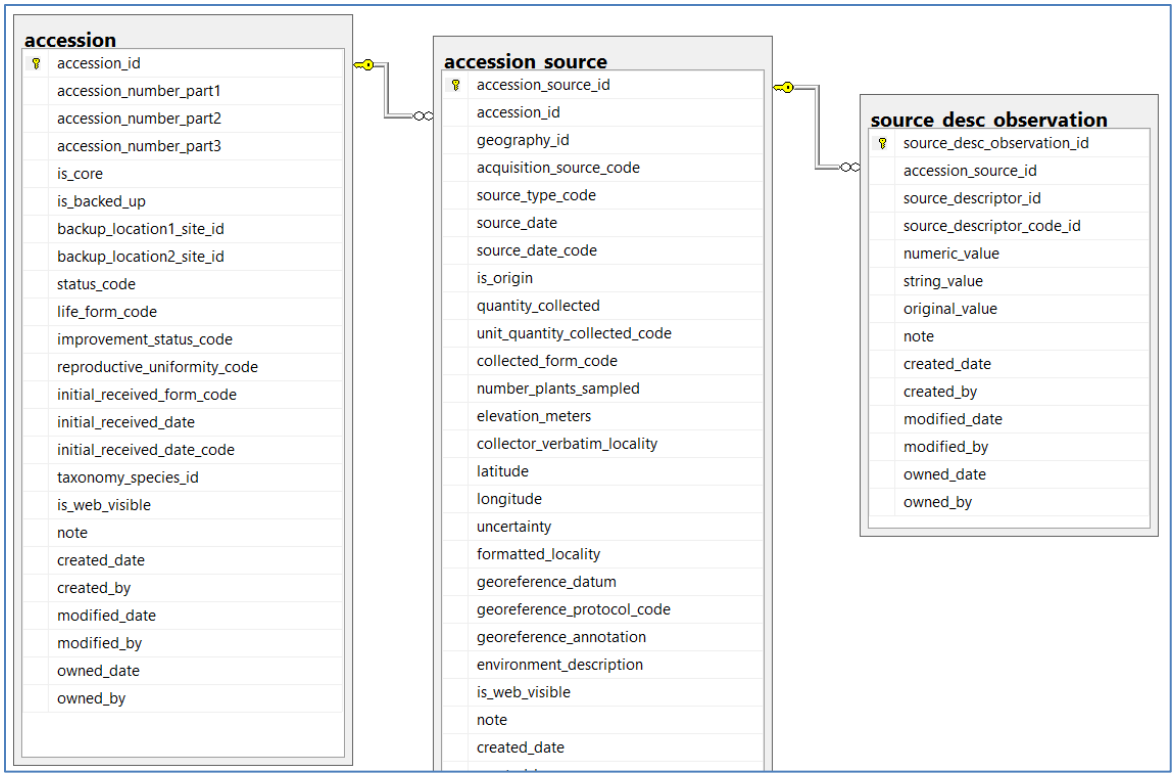
Introduction to the Source Habitat Dataviews

Overview

In the 1.0 GRIN-Global schema, data pertaining to the collection site where the accession was collected was stored in the **accession_source** record. The accession_source table held (and still does hold) data associated with an accession collected in the wild such as latitude, longitude, elevation, general source description, general associated species, etc. The fields in this schema were limited.



In the post 1.5 schema, this core set of collection site/habitat information still remains in the accession_source table, but new source_ tables have been added to make it possible to associate source habitat observation records with the collected accession. The five new tables provide an extremely flexible method for adding detailed information about the collection site. Genebank personnel can now associate multiple source descriptor observation records with the collection site (accession source) data. This was not possible with 1.0's single accession_source table.



The Source/Habitat descriptors and codes used by the organization will typically be entered by the organization's GRIN-Global database manager(s). For example, an organization can record Source/Habitat descriptors such as Slope, Aspect, Soil pH, and Soil Texture. The genebank personnel will then use those descriptors to enter the Source/Habitat observations (similar to the way they enter Crop Trait observation records).

Refer to the **Appendix IV: List of Standard Descriptors for site environment** section in Bioversity's the document (Technical Bulletin Number 13) "Developing crop descriptor lists, Guidelines for developers" for additional information and examples of standard descriptors.

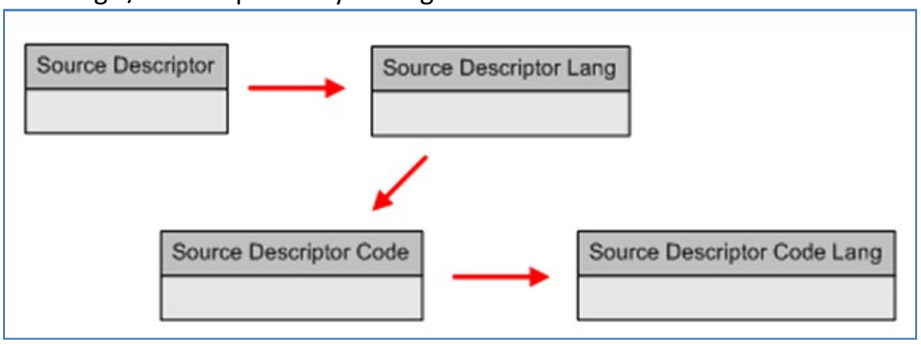
In order to use this expanded functionality, there are three essential steps to follow:

1. The organization determines what descriptor categories, Source/Habitat descriptors, and codes are to be added to the database.

   The GG database administrator will add any source descriptor *categories* to the **Source Descriptor Category** code group.

   The Source/Habitat descriptors are generally added by only one person in the organization, usually the GRIN-Global database administrator (DBA). When a new descriptor is needed, the flow will be similar to the following – the descriptor is added, then the Lang dataview is used to indicate the Title and Description. If the descriptor is a coded descriptor, then the codes must be added. An online spreadsheet contains sample descriptors, codes, and detailed directions for

installing S/H descriptors in your organization's GG database.



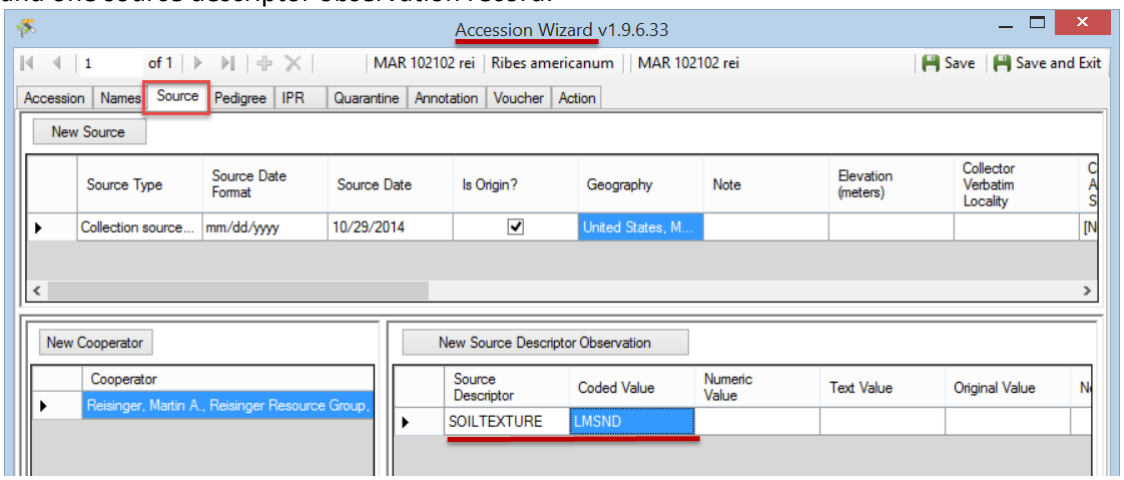
Using the Curator Tool, the GG administrator (typically) enters the Source / Habitat descriptors, titles, and descriptions into the GG database using the **Source Descriptor** and the **Source Descriptor Lang** dataviews.

Similarly, using the Curator Tool, for any descriptors that use codes, someone (again typically the GG administrator), enters the codes, the code titles, and code descriptions into the GG database using the **Source Descriptor Code** and **Source Descriptor Code Lang** dataviews.

This step 1 is repeated only when new descriptors, codes, and descriptor categories need to be added to the database.

2. Collectors visit sites and gather germplasm; they will also record field data about the collection date and the collection site.
3. In the Curator Tool, the accessions are added to the GG database. Accession records can be created using the **Accession Wizard** and while doing so the child **Accession Source** records can be created which can also include the Habitat / Source data. (See Recording Source Habitat Observations.) Typically genebank personnel will handle this task.

The sample Accession Wizard screen below illustrates the recording of the source descriptor observation data. So far, a s shown here, a source record has one cooperator associated with it and one source descriptor observation record:



(Multiple cooperators and additional source descriptor observation records can be added later. There is no limit to the number of cooperators or observations associated with the source record.)

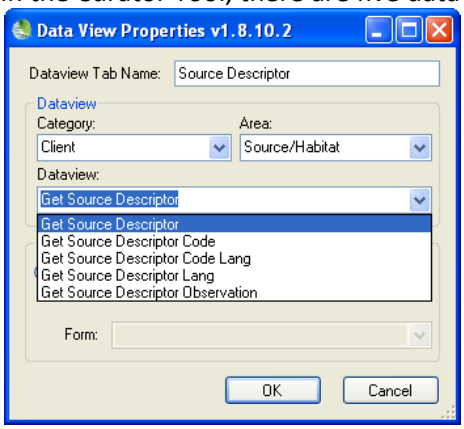
An alternative to entering the source habitat data in the accession wizard is to use the **Source Descriptor Observation** dataview. (The accession records and their related source descriptor records must be in the GG database before any Source Descriptor Observation records can be added.)

NPGS Vetting of Descriptors

In the NPGS, the S/H descriptors will be vetted by the GIS subcommittee, to help maintain some level of standardization.

Source/Habitat Dataviews

In the Curator Tool, there are five dataviews in the **Source/Habitat** area:



- Source Descriptor
- Source Descriptor Lang
- Source Descriptor Code
- Source Descriptor Code Lang
- Source Descriptor Observation

Curator Tool users recording the observations will primarily use the **Source Descriptor Observation** dataview. The other four dataviews are used to set up the descriptors and any related codes. These will often be used by only one person, the GG administrator or someone responsible for establishing the descriptors and codes to be used by the organization.

Curator Tool Habitat Source Descriptor dataviews

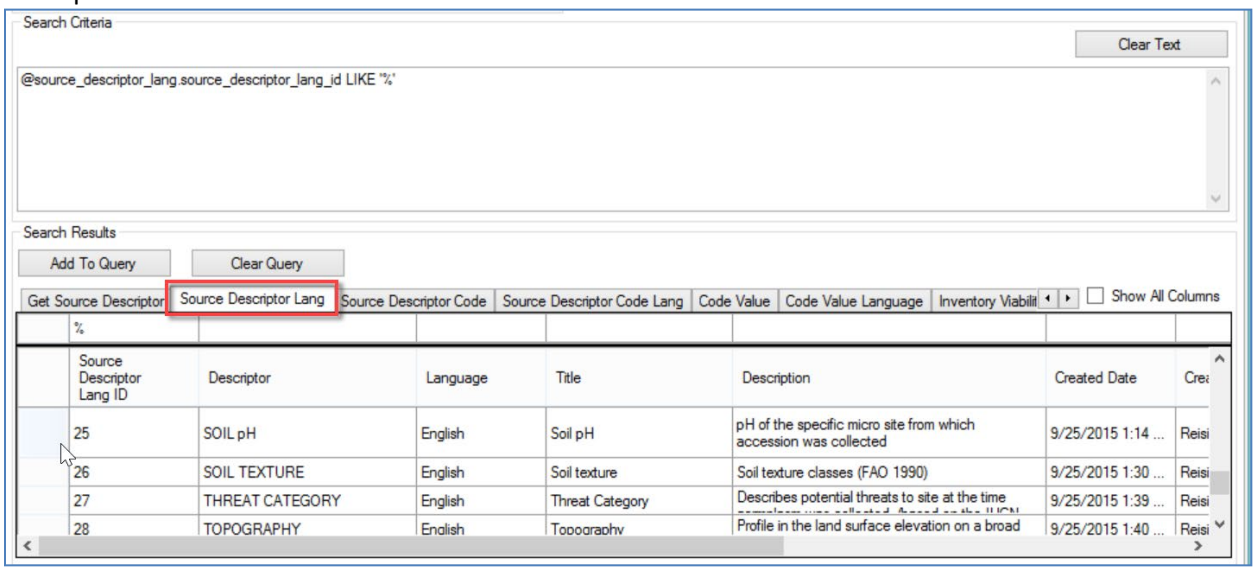
Input Data for the...	Dataview to use...	Records Created By...
Source Descriptor	Source Descriptor Source Descriptor Lang (titles & descriptions)	GG Administrator
Source Descriptor Code	Source Descriptor Code Source Descriptor Code Lang (titles & descriptions)	GG Administrator
Source Descriptor Observation	Source Descriptor Observation (typically accessed via the Source tab in the Accession Wizard)	Genebank Personnel

The ***core*** Global Information System / Georeferencing data is stored in the Curator Tool in the **Accession Source** dataview.

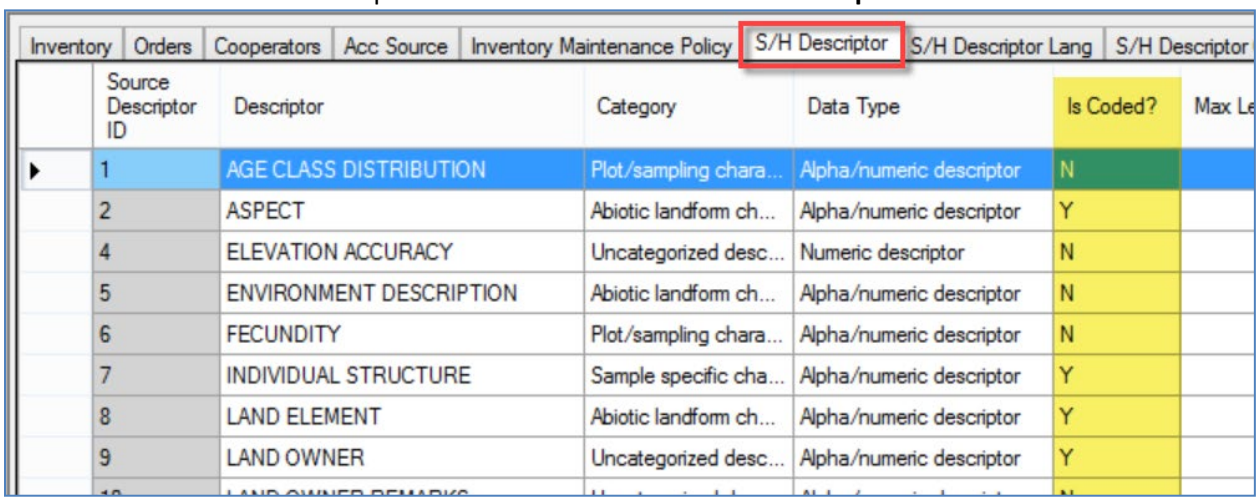
Recording Source Habitat Observations

Understanding What Source Habitat Descriptor to Use

It is important to understand which S/H Descriptor to use when recording the Source Habitat observation data. One way to determine that is to use the Search Tool and display all of the records using the **Source Descriptor Lang** dataview. The **Title** and **Description** field indicate the purpose of each descriptor.

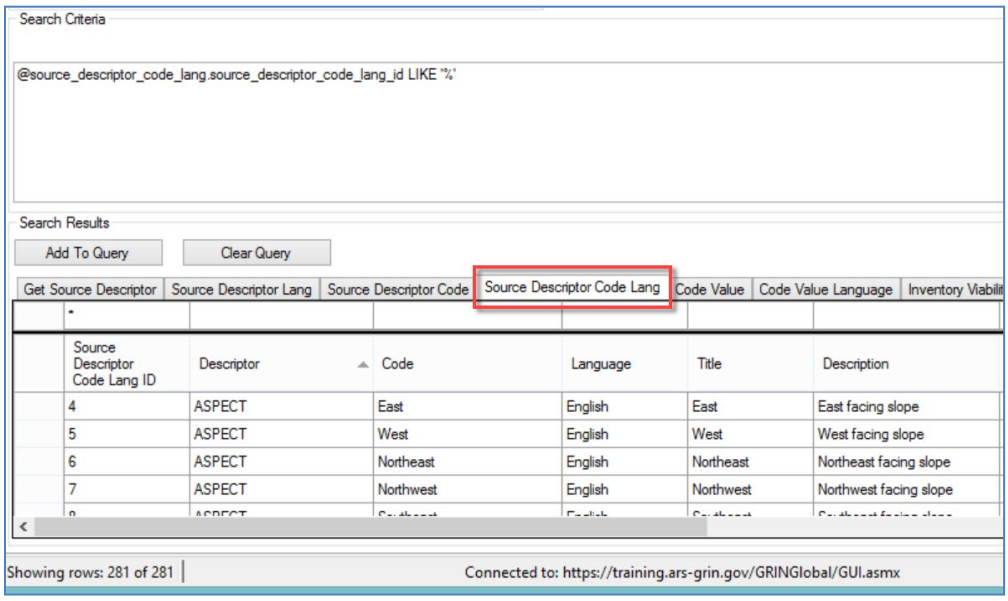


You can determine if the descriptor is coded or not in the **Source Descriptor Code** dataview:



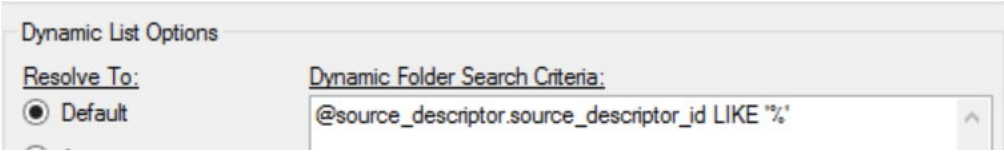
Determining the Source/Habitat Descriptor Code Values

Similarly, you can display the codes using the Search Tool and the **Source Descriptor Code Lang** dataview:

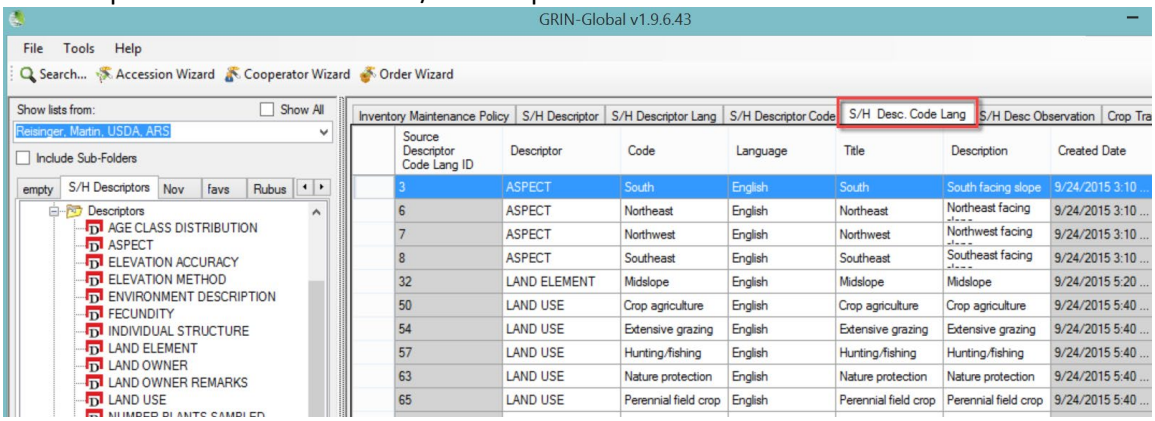


Highly recommended: In the Curator Tool, create a dynamic folder so that you can refer later to the descriptors and codes without needing to return to the Search Tool. Switch back and forth between the four S/H dataviews to determine the descriptors' descriptions and code values.

The Dynamic Folder criteria: **@source_descriptor.source_descriptor_id LIKE '%'**

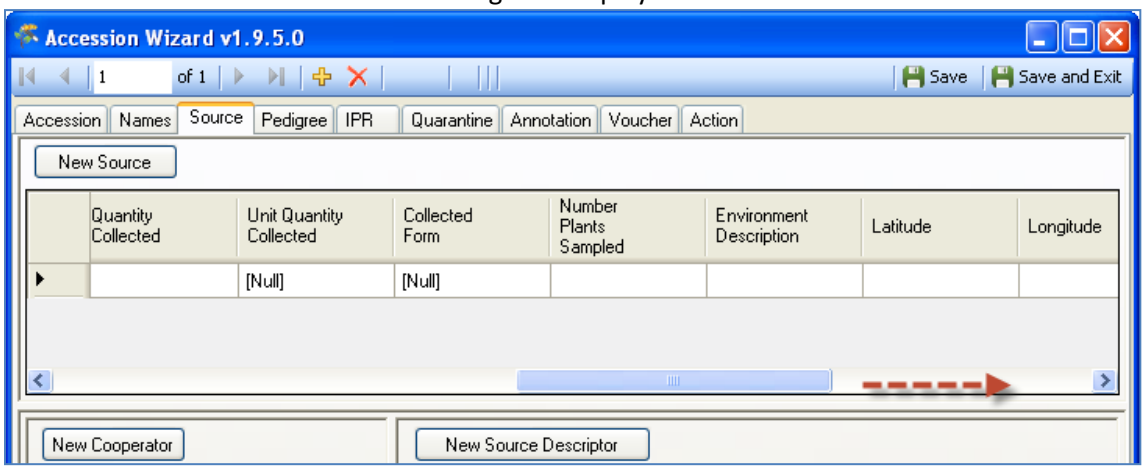


…or set up a static folder with the S/H Descriptors:



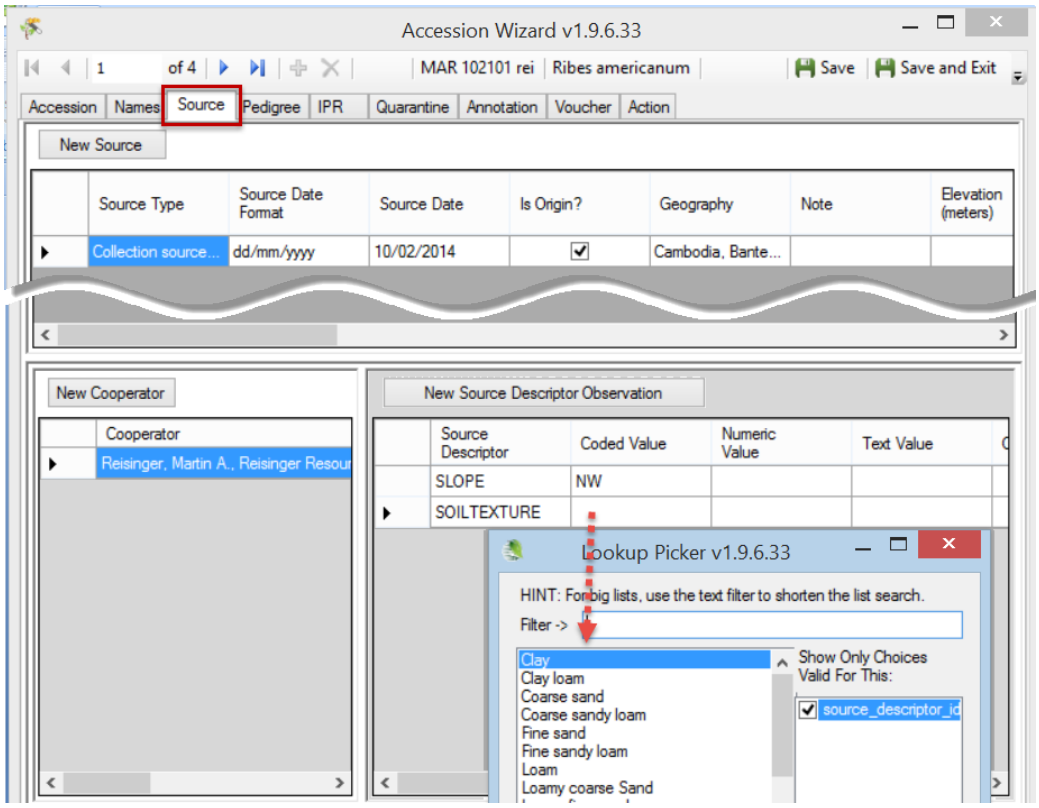
Using the Accession Wizard

The easiest way to enter Habitat Source Observations is via the Accession Wizard's **Source** tab. In the **Source** window, work from top down, left to right. First indicate the **Source Type**. Since these source observations are designed to describe the collecting site, in most cases the Source Type will be "Collection source event." As mentioned earlier, some of the basic geo data is stored directly in the **Accession Source** record – scroll to the right to display these fields:



In the bottom left panel, click on the **New Cooperator** button and select a cooperator form the lookup list; include each cooperator that participated on the collecting trip.

Use the New **Source Descriptor Observation** panel in the bottom right and input as many descriptors as desired.



Source Descriptor Observation: Coded Value / Numeric Value / Text Value

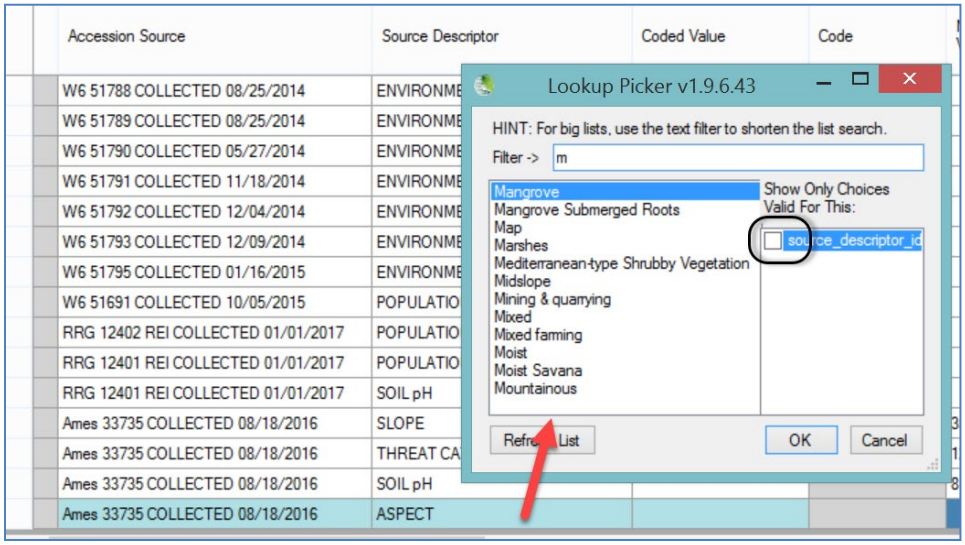
These three fields are mutually exclusive. Unfortunately, when recording the observation and using the **Source Descriptor Observation** dataview, there is no interface clue to indicate which field to use – you must know how the descriptor was defined.

Enter data in only one of the three fields – **Coded Value**, **Numeric Value**, or **Text Value** – the other two fields should not be filled. For example, if the descriptor uses a coded field, select a value for the **Coded Value** field from the lookup; do not input in the **Numeric Value** or **Text Value** fields.

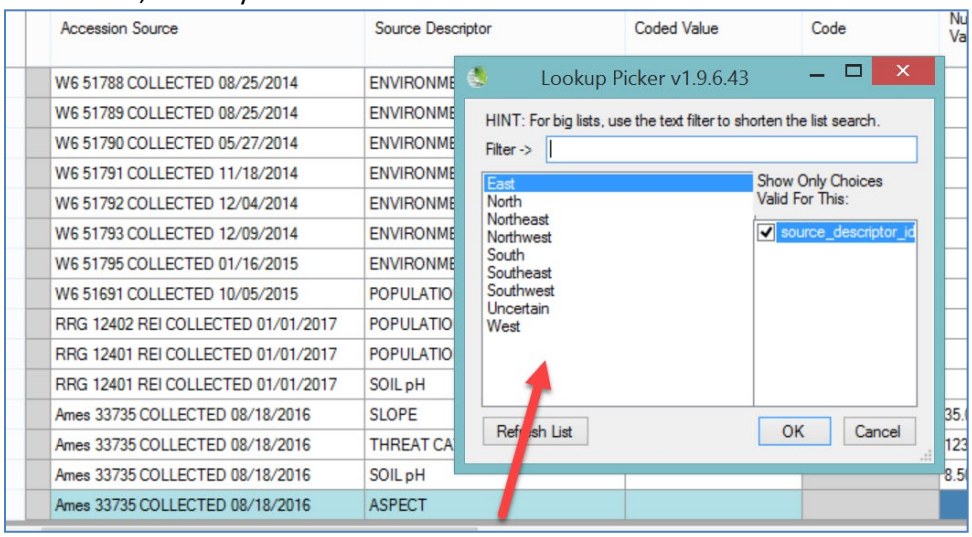
Not sure which descriptor to use? See Understanding What Source Habitat Descriptor to Use See Coded or Not? section to determine if the Source/Habitat Trait is coded or not.

Selecting a Coded Value

When using coded descriptors, ***do not uncheck*** the **Show Only Choices Valid** options box as shown here:

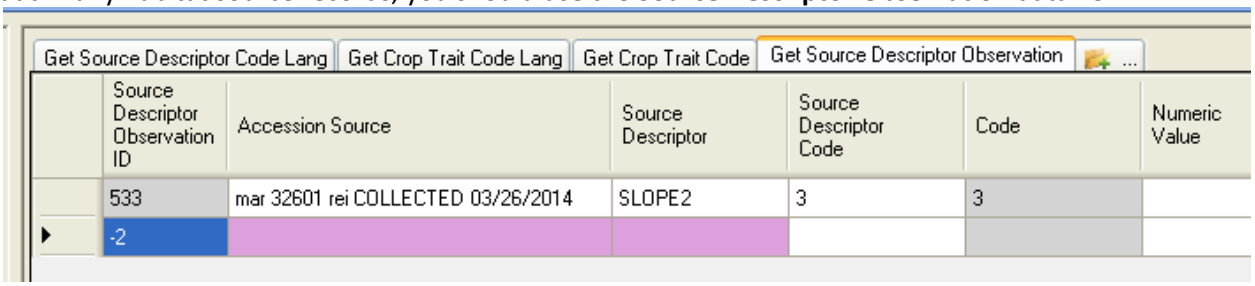


For **ASPECT**, the only valid codes are:



Bulk Adding of Source Habitat Records

The accession wizard works well when updating accessions manually; however, when you intend to bulk add many habitat source records, you should use the **Source Descriptor Observation** dataview:



You can drag and drop data from a spreadsheet into this dataview.

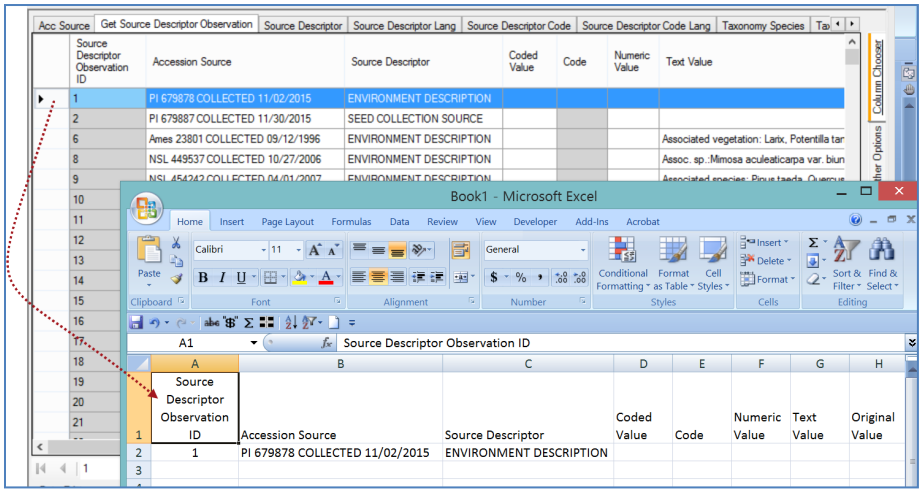
Setting Up a Spreadsheet for Bulk Adding S/H Observations

In setting up a spreadsheet, you need to know the descriptor name, and if the descriptor is a coded descriptor, you will need to know what codes are valid.

Also remember that you do not update the gray read-only fields – that will be done for you after you successfully add the records.

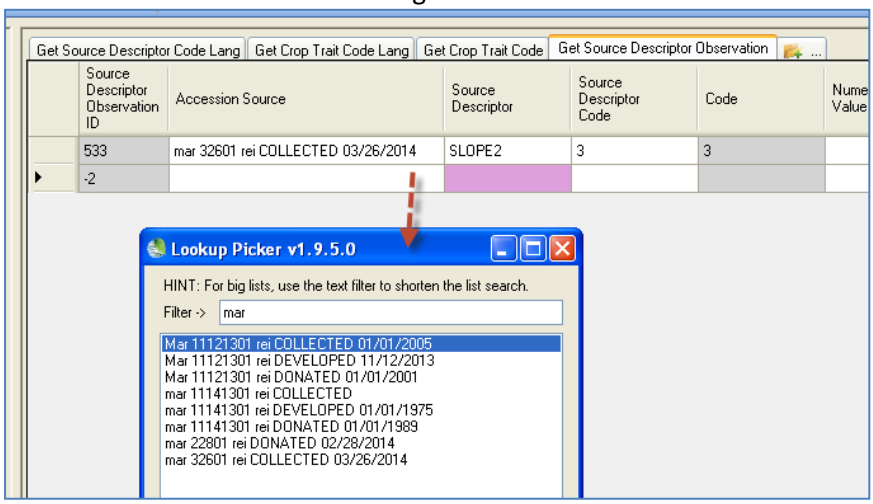
Copy the Curator Tool S/H Observation Grid

Select the first row and then drag to a blank spreadsheet:



You will need to know the full accession key (prefix, number, and suffix), the event code (typically “COLLECTED” and the event date. You can see in the example below how the

Accession Source field is combining that data.



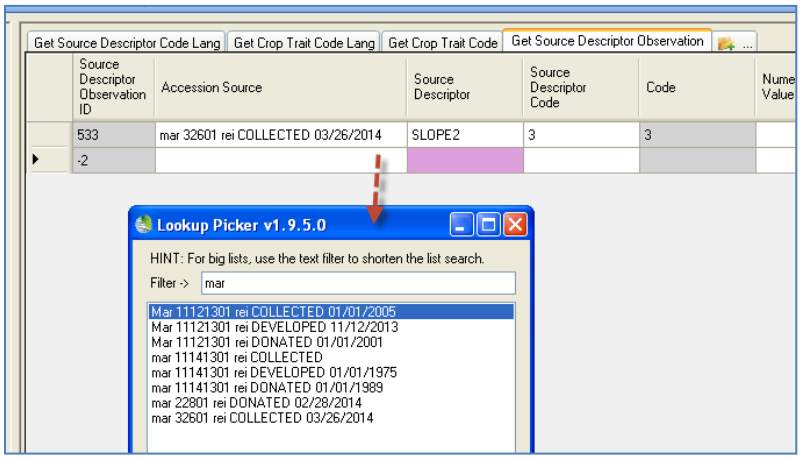
You will then need to edit the spreadsheet, for example, records being created are new records so they should have column A blank (keep the heading **Source Descriptor Observation ID,** but the rows for the new records will have column A blank).

Accession Source

As mentioned previously, this field is combining three elements:

- accession identifier
- the source event code
- the date

In working with an Excel sheet to bulk add the S/H observations, consider using the CT's lookup picker to manually select valid accession source data.



Use the Online File Which Has the Pages (and Codes) for each Descriptor

Review the section Understanding What Source Habitat Descriptor to Use for determining the descriptors' intended usage and their codes.

Alternatively, an online workbook (http://www.grin-global.org/docs/Source_Habitat_Descriptors.xlsx), was designed primarily for creating the GG Source Habitat Descriptors; the intended audience is the GRIN-Global administrator who is responsible for implementing these descriptors. However, each worksheet is useful to Curator Tool users who can review the Descriptor's Description. Each worksheets include the descriptor's code values and codes and when the descriptor is a coded descriptor.

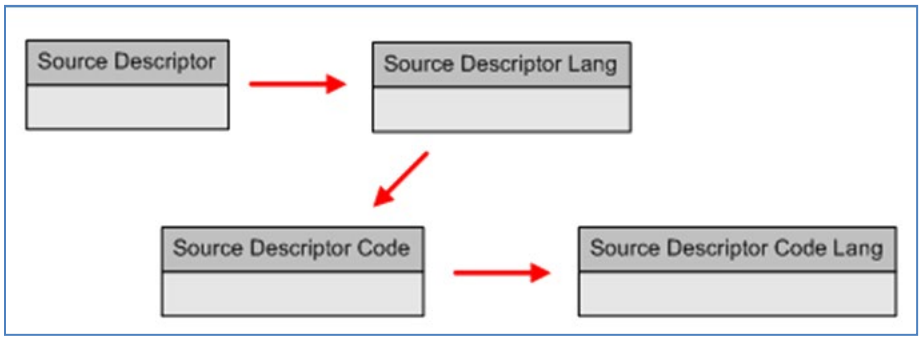
Dataviews Used to Define the Source Habitat Descriptors

Before any Source/Habitat observations can be recorded, the Source/Habitat descriptors must be added (typically by only one person in the organization, usually the GRIN-Global database administrator). (In the NPGS, the S/H descriptors will be vetted by the GIS subcommittee, to help maintain some level of standardization.)

When a new descriptor is needed, the flow will be similar to the following – the administrator adds the descriptor via the **Source Descriptor** dataview, then uses the **Source Descriptor Lang** dataview to indicate the Title and Description.

If the descriptor is a coded descriptor, then the codes must be added, using first the **Source Descriptor Code** dataview, then the **Source Descriptor Code Lang** dataview.

An online spreadsheet [https://www.grin-global.org/docs/Source_Habitat_Descriptors.xlsx] contains sample descriptors, codes, and detailed directions for installing S/H descriptors in your organization's GG database.



Source Descriptor Dataview

Get Crop Trait | Get Source Descriptor | Get Source Descriptor Lang | Get Source Descriptor Code | Get Source Descriptor Code Lang | Get Crop Tr

Source Descriptor ID	Descriptor	Category	Data Type	Is Coded?	Max Length
294082	SOILDRAINAGE	Abiotic soil characteristics	Alpha/numeric descriptor	☑	
-2		[Null]	[Null]	☐	

The **Source Descriptor** record has three required fields:

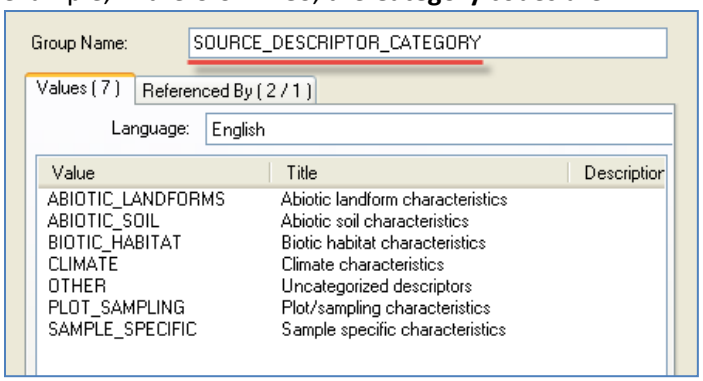
- Descriptor
- Category
- Data Type

Descriptor

Descriptor is a descriptive name for the descriptor, inputted when the new descriptor record is created.

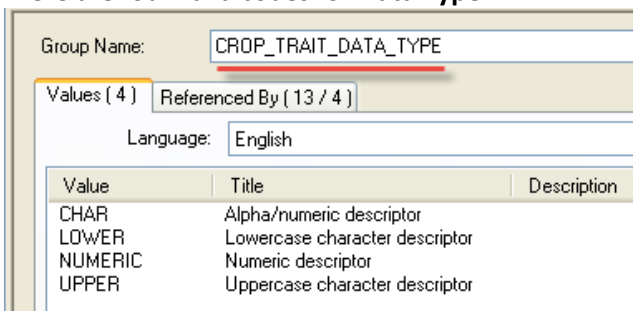
Category

Category and **Data Type** use drops downs. All GG drop downs use codes entered by the GRIN-Global database administrator. If a value is needed for either dropdown, contact the GG administrator. For example, in the U.S. NPGS, the **Category** codes are:

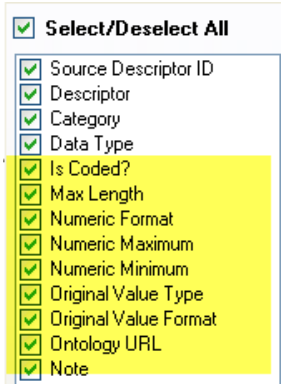


Data Type

There are four valid codes for **Data Type**.



Optional Fields



Is Coded?

When the **Is Coded?** field is selected, the descriptor will use coded values to denote the level of the descriptor. (Qualitative characteristics can be measured using nominal, ordinal or binary scales.)

Refer to the online data dictionary for descriptions of the other fields.

Source Descriptor Lang Dataview

Ensure the **Source Descriptor Lookup** is updated before adding the related **Source Descriptor Lang** record.

The **Source Descriptor Lang** record has two required fields:

- Descriptor
- Language

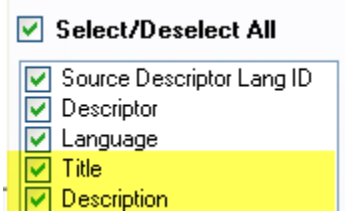
Descriptor

Descriptor is a descriptive name for the descriptor; selected from the **Source Descriptor Lookup** table.

Language

Language is selected from the **System Language Lookup** table.

Optional Fields



Title

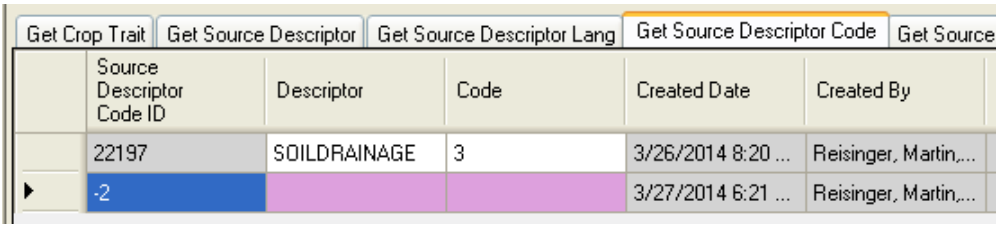
The source descriptor's title.

Description

The source descriptor's description.

Source Descriptor Code Dataview

When a **Source Descriptor Trait** is defined as "Coded," the codes for the trait must be defined and must be unique. If you are unsure what codes have already been defined, search on the **Source Descriptor Code Lang** dataview to list the existing codes.



Descriptor

Descriptor is a descriptive name for the descriptor; selected from the **Source Descriptor Lookup** table.

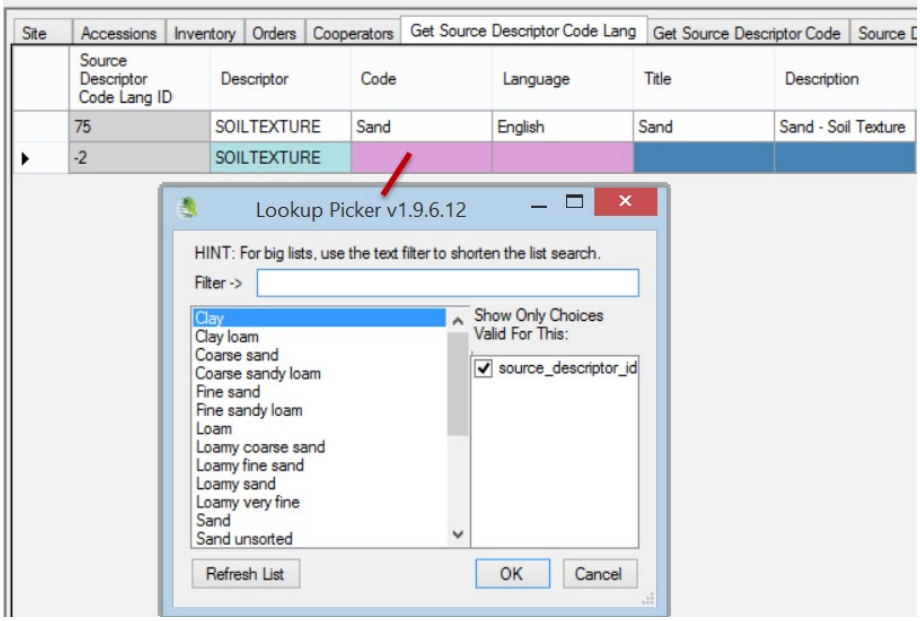
Code

The value which will indicate the scale level.

An online spreadsheet contains sample descriptors, codes, and detailed directions for installing S/H descriptors in your organization's GG database.

Source Descriptor Code Lang Dataview

Ensure the **Source Descriptor Code Lookup** is updated before adding the related **Source Descriptor Code Lang** record.

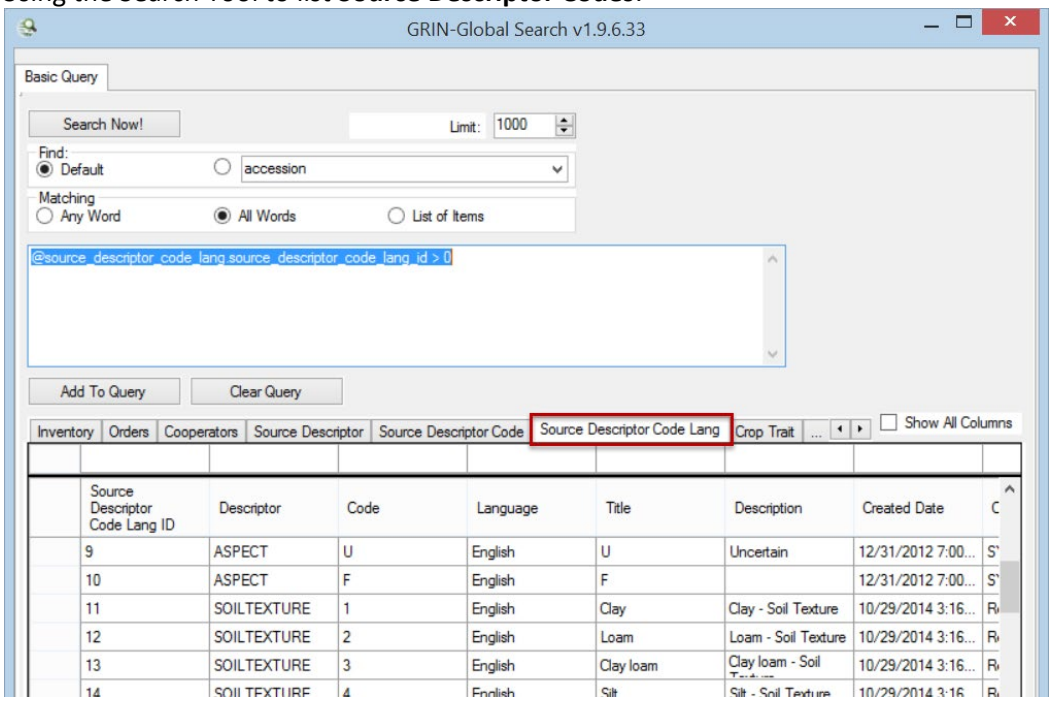


The **Source Descriptor Code Lang** record has three required fields:

- Descriptor
- Code
- Language

The primary function of this language dataview is to associate a **Title** and **Description** with a **Source Descriptor Code** record, so the typical ...Lang record will have those two fields filled as well.

Using the Search Tool to list **Source Descriptor Codes**:



Appendix

Changes in this Document

– May 7, 2024

- changed links to spreadsheet's source location

– September 17, 2020

- changed links to spreadsheet to .xlsx instead of .xlxm

– January 23, 2017

- added NPGS note about adding descriptors; also added bulk adding of observations directions

– January 5, 2017

- major rewrite of the introduction and background information

– November 3, 2014

- captured screens to reflect the current dataview heading names
- edited overall text