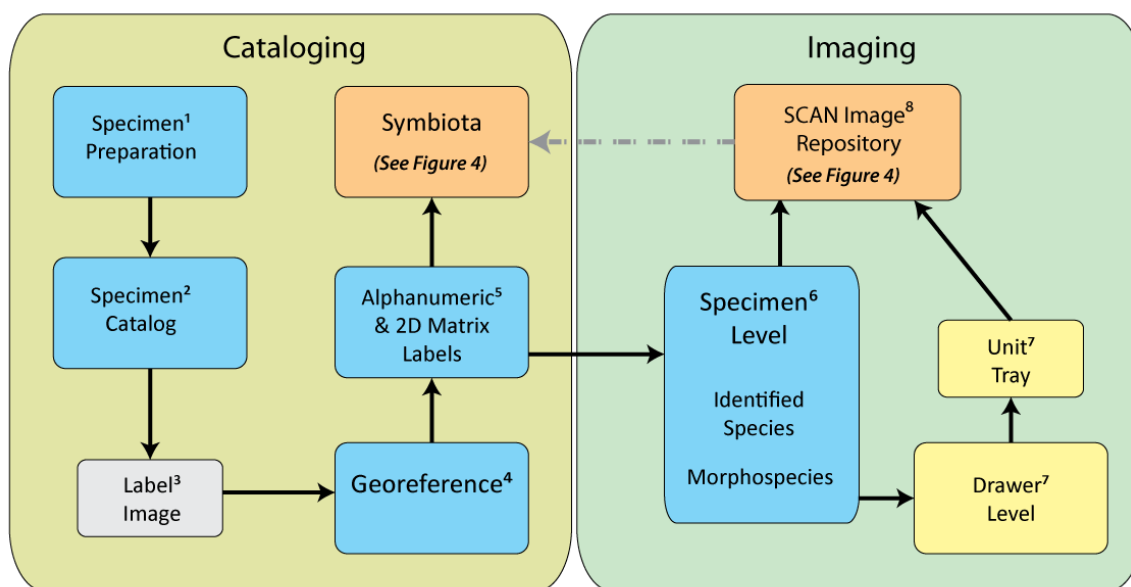




The basic SCAN workflow as proposed.

Digitization workflow (**Figure 1**) steps for *cataloging* and *imaging* specimens (A) Organize collection taxonomically; (B) enter label data & field notes; (C) third party future collaboration for OCR & speech to text solutions; (D) process data through GEOLocate; (E) apply institutional catalog labels and 2D matrix labels; (F) image representative specimens; and (G) image specimen lots. Data is uploaded to the SCAN Image Repository with linkages to Symbiota and local databases. Yellow boxes indicate actions that will occur at a subset of collections.

Figure 1



We will use professional-grade digital imaging systems to obtain montaged images of individual specimens and possibly specimen lots. We will use Bugguide as a model to serve very high resolution online imagery through zoomify file formats to help others identify taxa and allow experts to examine “virtual specimens” for identification.

An overview of how data will flow through SCAN.

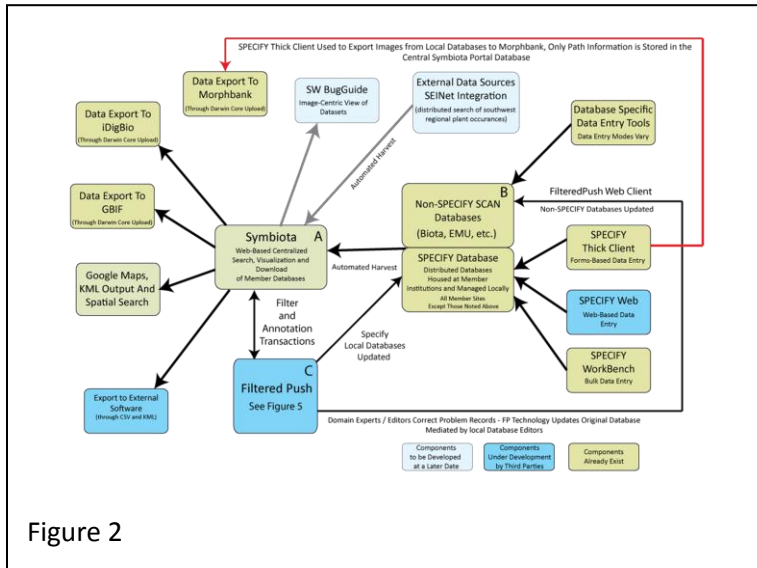


Figure 2

SCAN Cyberinfrastructure (**Figure 2**) comprised of Servers, Software and Processes. (A) Search, aggregated download and limited mapping are provided by the centralized portal system. (B) Data are housed and managed locally. (C) FilteredPush technology at the portal allows efficient and automated discovery and correction of a wide range of data problems at the record level. (D) Images are uploaded to the repository directly and path information is recorded in the local

and central databases. Images are contributed to external repositories from the SCAN repository while catalog data is contributed through the portal.

Integration between Symbiota, Filtered Push and other databases is a central part of SCAN.

Filtered Push (**Figure 3**) deployed in the SCAN framework will specifically; (A) detect records for identification or correction; (B) create annotations describing the problem; (C) route annotated data to domain experts for editing; (D) once edited, update the Symbiota; and (E) update the local database. Local collections managers can then accept or reject these changes or request clarification.

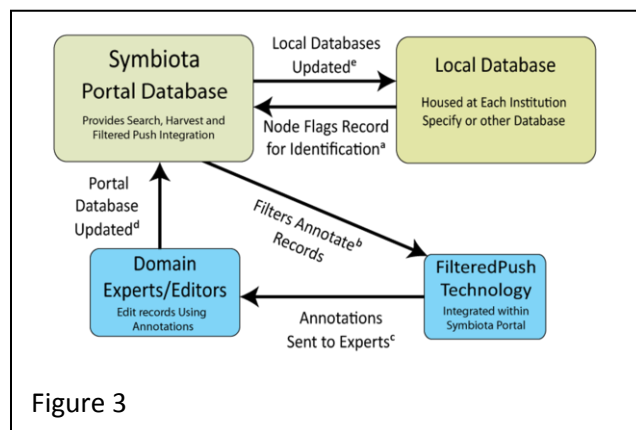


Figure 3

We will be test-driving two online databases at the outset of the project, the AMNH database used by the Tri-trophic TCN, and Symbiota. There is a strong desire to begin SCAN digitization by providing data online in as near real time as possible to aid in QA/QC. It is also desirable to either have data be automatically harvested into the central portal or to empower individual

collections to upload their own data. We have already established a Specify database template that can be used to store collection records locally for each SCAN member collection. This software combination is mature and cross-platform; it can run on Linux, Windows or Mac

workstations or servers, granting maximum versatility for installation configuration to participating institutions. Two of the collections (Denver Museum of Nature and Science (KE EMu) and Texas A&M (in house but may adopt KE EMu), will continue to use their existing database solutions. Specify and Symbiota both store images of specimens and of labels as links to files in the filesystem. The Specify Project has a collaboration underway with the Morphbank Project to integrate Specify image archiving with Morphbank's planned distribute image repositories. Access to the collective holdings of the member institutions will be provided through a centralized data portal housed by the Geospatial Research and Information Laboratory at Northern Arizona University. This portal will be built using Symbiota, which is already in use with SEINet (a large network of herbarium collections). Using Symbiota will allow the automated harvest of records from member node database and storage of those records in the central portal database. This system is intended to allow easy integration into iDigBio's activities through the utilization of appropriate standards and processes.