Protocol for creating Notes from Nature Transcription Expeditions for the "Keys to the Cabinets" Curator Community

The protocol below assumes that you have made basic decisions about the circumscription of your volunteer engagement project (e.g., the specimens that you will target in collaboration with partner groups like your state native plant society). It makes sense to establish a single project for each partner or group of partners that have the same focus. For example, you might produce a project focused on all specimens collected in your state (irrespective of collection curating them) with your state native plant society and a project focused on your herbarium's specimens (irrespective of location they were collected) for your institution's volunteer corps.

The protocol also assumes that you have successfully uploaded your image files to CyVerse and updated the relevant specimen records in Symbiota with the url at which the corresponding specimen image can be accessed.

STEP A. Register at biospex.org and create a Group.

1) Click Register in the top right at biospex.org.

2) Fill out fields in Register New Account form. The Group Invite Code field permits a user to join a Group to which they were invited. Most of you will not be joining a Group immediately in this way. You would have received your group invite code in an email generated by BIOSPEX.

3) You will receive an account registration confirmation email from BIOSPEX. Click on the link in that email and log in on the page that opens.

4) The "Welcome to BIOSPEX!" page will prompt you to read through the Our Vision, FAQ, and (soon to be created) Resource pages for additional background on how to use BIOSPEX successfully. We encourage you to do so. If you prefer to instead hear it presented to you, you can view Austin Mast's iDigBio Education and Outreach Webinar entitled "A Small-Collections On-Ramp for Participation in the 2016 Worldwide Engagement for Digitizing Biocollections (WeDigBio) Event" at http://idigbio.adobeconnect.com/p2gbkxdsovb/.

5) Click on the green +Create button at the bottom of the "Welcome to BIOSPEX!" page.

6) Enter a name in the Create group form. The group name can be edited later. To start, you could call it "SERNEC's [your state]".

7) If you'd like to immediately invite others to participate in the group, you can click the white Invite button on the Groups page that you see next. We encourage you to add Michael Denslow (michael.denslow@gmail.com) and Austin Mast (amast@bio.fsu.edu)

to your group. This will become important when you would like to onboard your expeditions at Notes from Nature and when you'd like to troubleshoot, respectively.

STEP B. Create a Project at biospex.org.

8) Click Projects at the top of the page.

9) Click on the green +Create button on the Projects page.

10) Fill out the Create Project form and click the dark-blue Create button at the bottom. At a minimum, you will need to fill out the required fields (those with stars). Spend a bit of extra time word-smithing the content, since it will be important in recruitment. For Status, choose Starting. In most cases, you will be the Contact. Contact Title is how you'd like to describe yourself in the context of this project. Some have used "Chief Mobilizer" for that title. Include SERNEC (as "The Southeastern Regional Network of Expertise and Collections") in your list of Project Partners. Include the following in the Funding Source field:

> National Science Foundation under award [your specific award number]. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

For Workflows, choose OCR -> Notes from Nature V2. This will run the images through an Tesseract OCR webservice managed by BIOSPEX and provide you with OCR text strings to search when bundling your specimens into particular expeditions.

You will be asked to provide a logo of maximum size 300 x 200 pixels (width x height) and a banner of size 1200 x 250 pixels (width x height). Select high quality images for these, since they will be prominent features of the public page that BIOSPEX mints for the project (Figure 1). We strongly discourage you from using a banner that is lettering on a solid background.

Examples of wording and images for these fields can be found at the public pages for existing SERNEC BIOSPEX projects, including https://biospex.org/project/wedigflplants, https://biospex.org/project/plants-of-arkansas-discovery-and-dissemination, and https://biospex.org/project/ant-plants-of-virginia.

11) Click on the dark-blue Open button next to your project name in the Projects table.

12) Click on the Project Url on the upper right and proofread your content on the project's public page. You might want to ask your project collaborators to proofread the content at this point.
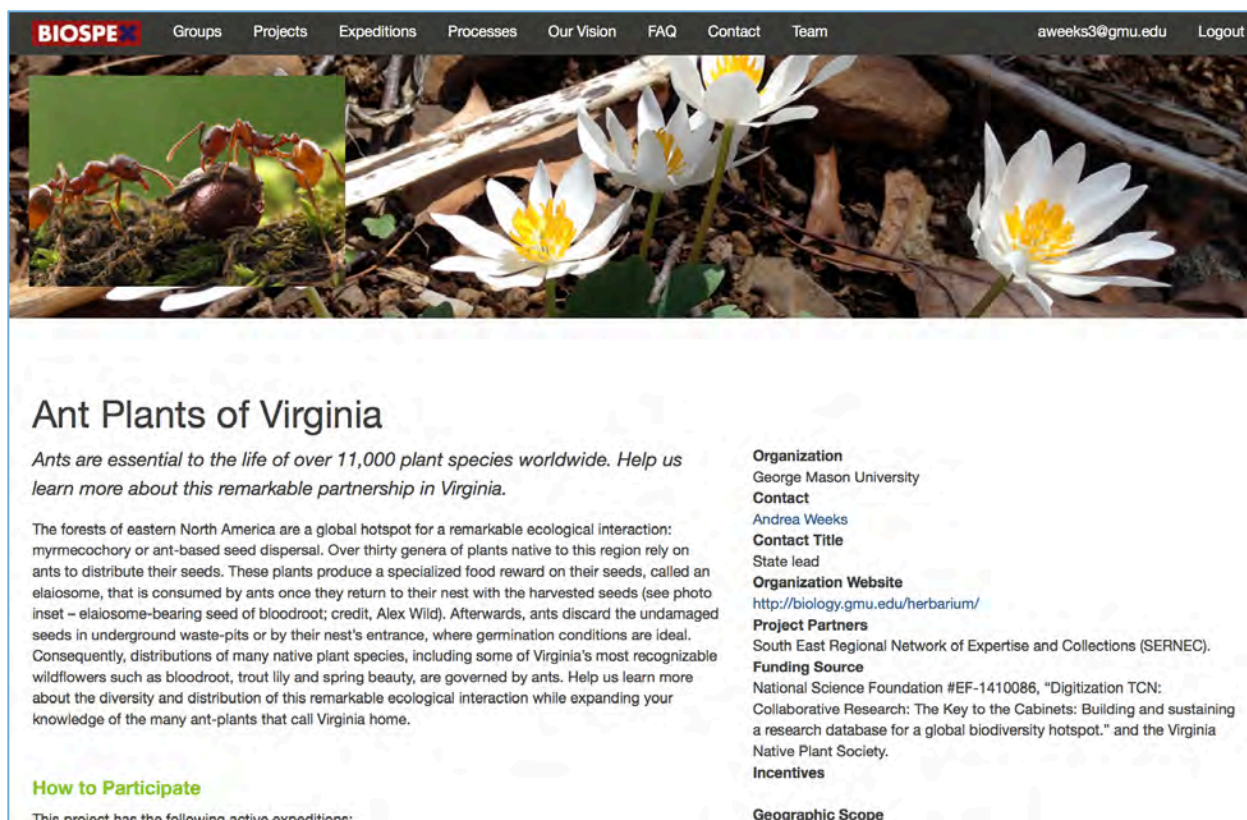
Figure 1. This is what a public-facing Project page looks like.

STEP C. Export a Darwin Core Archive from Symbiota and upload it to your BIOSPEX project.

13) Set preferences in your web-browser to keep folders zipped on download (e.g., in Safari, go to Safari->Preferences->General then Uncheck "Open safe files after downloading" in that window). BIOSPEX expects zipped Darwin Core Archives.

14) Go to http://sernecportal.org/portal/ and login.

15) If you have not already done so, check that your collection's GUID Source is correctly set. Click My Profile, click on the Specimen Management tab, select the relevant collection in the Collection Management section, select Edit Metadata and Contact Information in the Administration Control Panel section, and double-check the selection next to "GUID Source". If you have questions about this, see Herrick Brown's email with the subject "SERNEC portal collections in need of metadata updates" from June 16, 2016. If questions remain, contact Herrick at herrick.brown@gmail.com.

16) Click on Specimen Search->Search Collections in the top menu.

17) Select the relevant subset of SERNEC institutions and click the Next button. At these early stages of SERNEC's deployment of expeditions, we encourage you to limit your selection to those institutions for which you have administrative privileges. This

could be all institutions in your state, if you are state lead. We are still figuring out the etiquette for including specimens from other SERNEC institutions in a project's expeditions.

18) Search for the relevant subset of specimens and click the Search button. For example, if your first expedition is planned to be focused on Lauraceae of your state, enter Family or Scientific Name=Lauraceae and State/Province=[your state] and check the radio button in front of "Limit to Specimens with Images Only". Checking that radio button is important, since it is not possible for volunteers to transcribe in the absence of an image.

19) Check that the thumbnails in a few of the results pages look correct. Specimens should be oriented in the proper way, and there should not be any broken links. If there is a problem with orientation, you will need to correct it at CyVerse. If there are broken links, email Herrick Brown (herrick.brown@gmail.com). Both of those problems will need to be fixed before progressing with that set of specimens.

20) If your search result was <500, consider returning to step 18 and experimenting with a broader circumscription for your expedition. Though there is not currently a minimum size on expeditions, all expeditions take the same amount of time for staff to onboard at Zooniverse. We would prefer to have expeditions that are 2,000–3,000 specimens in size. The maximum size for an expedition is 5,000 specimens. NOTE, you can reach these target numbers by combining the results of several separate searches that have each been taken through to step 23. For example, to produce the Ant Plants of Virginia expeditions, Andrea Weeks did several separate searches for records of species known to be myrmecochorous.

21) Once satisfied, click the small spreadsheet-icon in upper right panel of the returned search results.

22) Leave the default choices in the Download Specimen Records window (Figure 2) and click on the Download Data button. Symbiota Native is a Darwin Core Archive with a bit of additional information in the file. BIOSPEX does not have any trouble ingesting it.

Figure 2. These are the default settings for the Symbiota download that are known to work well in for the BIOSPEX ingestion.

23) Place your zipped Darwin Core Archive in a place that you can find it again (e.g., your Desktop).

STEP D. Ingest the Darwin Core Archive at BIOSPEX.

24) Click on the Projects link in the top menu of BIOSPEX.

25) Click the gray +Import button on the line of the relevant project.

26) In the Import Darwin Core File box click on the Choose File button, choose the zipped Darwin Core Archive produced in STEP C, and click the dark-blue Upload button. The upload process is quite rapid (a few seconds). You will receive an email telling you that it is completed or if there were errors in the process, such as duplicate records.

27) If you chose the project workflow that involved the OCR webservice, that process will probably take tens of minutes to a few hours. When you click on Processes in the top menu, you can view the progress of the OCR webservice. You might not see the OCR process start immediately. If circumscription of your expeditions requires searches on the OCR text string, you will need to pause until OCR processing is done. You will receive an email when that processing is complete. If you aren't going to be relying on the OCR text string in downstream steps, you need not pause to wait until it is completed.

STEP E. Produce your Expedition at BIOSPEX.

28) Click on the Expeditions link in the top menu of BIOSPEX (Figure 3). If you have just completed Step D you will already be on this page.
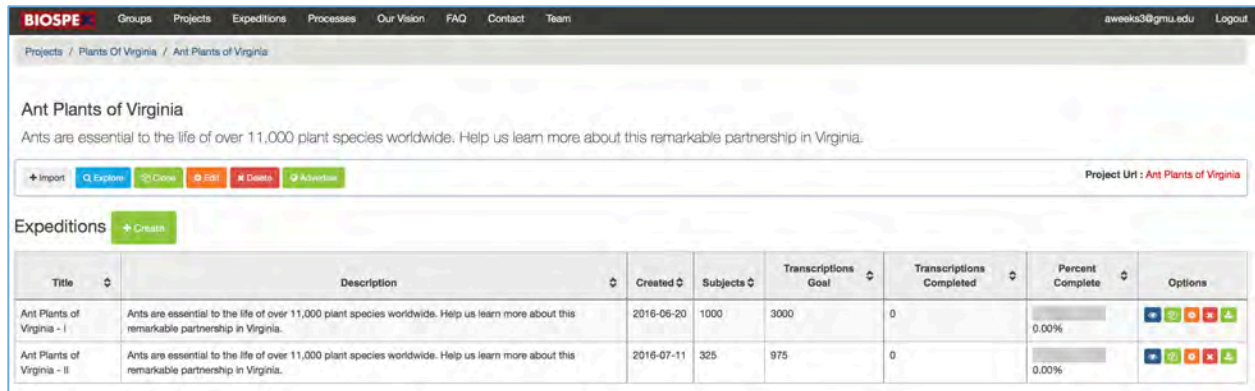
Figure 3. This is what the Expedition page looks like.

29) Click the green +Create button.

30) Fill out the three fields Title, Description, and Keywords. The Expedition title will be the same as used for Notes from Nature (Figure 4), so it is worth a little extra time word-smithing at this point to make your title interesting. Example entries from WeDigFLPlants' Mints of Florida—More than Mojitos Expedition are "Transcribe historical label data from Florida specimens from the plant family Lamiaceae." for Description and "Mints, Lamiaceae, Florida, #WeDigFLPlants, Transcription" for Keywords.
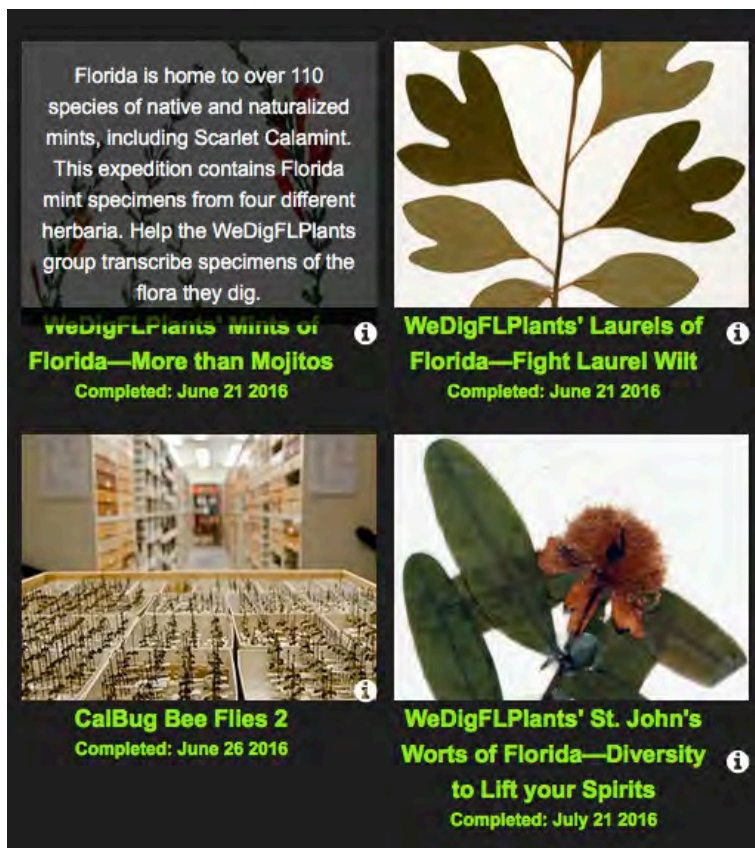

Figure 4. Expeditions at Notes from Nature.

31) Assign subjects to the expedition using the workbook below the three fields. There are five default columns shown when creating a new expedition: the radio-button, assigned, id, accessURI, and ocr. All of the subjects uploaded for your project and unassigned to an expedition are shown in that workbook. If this is your first expedition, you should have "No" in the assigned column for all rows in the workbook. You can view the thumbnail image or a larger-sized image of the subject by clicking on links in the accessURI column. If the OCR process has completed, you will see OCR text strings in the ocr column.

If you have only uploaded subjects that you would like to assign to this expedition, you can change the workbook view to 500 subjects at the bottom of the workbook (it is set by default at 10) and click on the radio-button at the top of each page to add them all to the expedition. If you have more than 500 subjects, you will need to progress through the additional pages and click the radio-button at the top of the first column on each page. **NOTE:** If you go back to a page that you have already done this on, assigned will still read "No". However, they will have been entered for that expedition. This is a known issue that BIOSPEX is working to fix.

If your uploaded, unassigned subjects for this project represent more than one expedition, you will need to use some of the workbook tools (Figure 5) to narrow down the subjects for your expedition.

Figure 5. Tools in the lower left of the BIOSPEX workbook.

Let's start with the icon that looks like a bulleted list. That opens a window that allows you to choose the columns that you are seeing in the workbook. The available column headers are from the Darwin Core Archives that have been uploaded for your project. There will not be data available to view yet for many of those columns. A few that we have found useful in the context of the SERNEC project (in addition to the BIOSPEX defaults) are occurrence.institution, occurrence.collectionCode, occurrence.catalogNumber, occurrence.family, occurrence.scientificName, occurrence.year, occurrence.month, occurrence.day, and occurrence.reference. The last is a link to the record in the SERNEC portal. If you choose to view all of these, you

will need to increase the size of your browser window. You can also drag the columns into a new order at the column heading and resize the columns at the column edges.

If you like the columns that you are seeing, in the order you are seeing them, click the Save Grid State button at the bottom of the workbook. Otherwise, when you return to the workbook you will start again with the default BIOSPEX columns.

Now you can sort by column, if you click on the column header. The up-down arrow heads to the right of the column heading indicates that you are sorting on that column. If you want to sort by another column, you will need to click the column heading a second or third time to make the up-down arrow heads disappear before going to your next column to sort.

You can filter the subjects that you are seeing by clicking on the icon that looks like a magnifying glass. The left-most dropdown in the pop-up gives you the ability to select any column that you are currently viewing in the workbook. The center dropdown gives you a lot of options for filtering. The right-most bit is either a dropdown or a field, depending on what you selected in the dropdowns to the left of it.

Let's look at an example. If you are defining expeditions based on family membership of the subjects, and you've uploaded subjects from multiple families into the project, you would first click the bullet-list icon and choose to view occurrence.family. Then you would click the magnifying glass and select that column name in the right-most dropdown, choose equal in the center dropdown, and type the name of the family in the field to the right of it. The workbook will then only show those records meeting that criteria, and you can view 500 and click the radio-button at the top of the first column on all of your pages.

If you think that your subjects are a mix of records that are skeletal and records that are complete, here's a useful hint. Show the column occurrence.year in the workbook, then filter your records in the following way. Click the + button at the top of the filter window. Choose occurrence.year "does not contain" "1" in the first line and occurrence.year "does not contain" "2" in the second line. Collection date was not something recorded in the SERNEC skeletal record creation, so the presence of a year in that field suggests further work to create a completed record.

32) Once you are satisfied with the subjects included in the expedition, click the green Process button on the view page for that expedition. NOTE: Once the expedition has been processed there is not an easy way to add or remove subjects from it. If you later find additional subjects that fall into that same topic, you can make a "Part 2" expedition. BIOSPEX is creating a zip that includes the images resized to Notes from Nature's specifications and a manifest file that tells Notes from Nature about the images in the zip. You will receive an email from BIOSPEX once processing for the images is completed and the zip is available for download at BIOSPEX.

STEP F. Give Michael Denslow what he needs to launch the expedition at Notes from Nature.

33) Send an email to Michael Denslow (michael.denslow@gmail.com) with the following content: (1) The BIOSPEX project name and expedition name. He will need to be a part of the group managing the project to get to the page where he can download the zip file created in Step E. (2) An image that will be used to illustrate the expedition "tile" at Notes from Nature (Figure 4). This should be 320 pixels wide x 220 pixels high. (3) A two-sentence appeal to volunteers to participate in the expedition. This appears when the volunteer hovers over the tile (see the mints-focused expedition in Figure 4). (4) The URL of your public project page at BIOSPEX. This will be the link that opens when you click on the information icon in the lower right of a tile at BIOSPEX. (4) Notice of whether or not you need any of the SERNEC skeletal-record fields completed. These include Scientific Name, Country, State, and County. If you skipped any of these fields during the imaging step, you he will need to include them for transcribers in your expedition. (5) Two short blogs posts, one introducing your expedition and one for when it is done. There are lots of examples at blog.notesfromnature.org.

STEP G. Advertise your new project and expedition widely.

34) Use email, social media, your institution's communications office, etc. to tell everyone about the important work they can contribute to by participating in your expedition.

STEP H. Participate as a scientist on the Notes from Nature forum.

35) Check in on Talk (https://www.zooniverse.org/projects/zooniverse/notes-from-nature/talk) twice a week for a few minutes each to answer volunteer's questions about your expedition. Alert Michael Denslow to any issues that you don't feel comfortable answering.

Further Resources

1) Robert Bruhn in the BIOSPEX developer, and he is responsive to questions posed using the Contact form at that site.

2) If you have big-picture requests for BIOSPEX development, send them to Austin Mast (the BIOSPEX PI) at amast@bio.fsu.edu.

3) Austin gave a recorded webinar in the iDigBio E&O series that covered much of the content of this document. It is available to watch at http://idigbio.adobeconnect.com/p2gbkxdsovb/

4) WeDigBio 2016 is scheduled for October 20–23. Consider planning an onsite transcription blitz in association with WeDigBio using resources at wedigbio.org.