Here is a link to a .zip file containing all the files I worked with this summer: [[](https://mcgill-my.sharepoint.com/:u:/g/personal/lindsay_trottier_mail_mcgill_ca/EaCudVwOhgdHstxHxflLSfkBAa18Veolg_KsPBdVT-qY_g?e=52pEoy)LivingDataIntern\_LindsayTrottier.zip](https://mcgill-my.sharepoint.com/:u:/g/personal/lindsay_trottier_mail_mcgill_ca/EaCudVwOhgdHstxHxflLSfkBAa18Veolg_KsPBdVT-qY_g?e=52pEoy)

\*\* Please let me know when you've downloaded the .zip file so that I can remove it from my OneDrive and transfer it onto an external hard drive for long-term storage\*\*

Here is a summary of the folders, their sub-folders, and files:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Folder** | **Sub-folder** | **File** | **Description** | **Status / Next Step(s)** |
| BC-CDC\_TexadaList | *n/a* | resultsExport.xls | A spreadsheet containing the search results for all plant species on Texada Island (polygon search option) from the BC Species & Ecosystems Explorer. | Waiting for response from Jenifer Penny with details about accuracy of the search results (some records don't appear on Texada Island according to their observation map) and with information about the source of the data.  *Copy of e-mail chain with Jenifer is attached to this e-mail ("Flora for Texada Island").*  Next Step(s)  - ​Append source data  - ​Convert to Darwin-core format  - Compare to GBIF-CPNWH dataset to identify any unique records for inclusion in the final Flora of Texada Island |
| GBIF\_CPNWH | IntermediaryFiles | *multiple* | Description of intermediary files can be found within the R script. | Data cleaning file, no next steps. |
| *n/a* | Cleaning\_GBIF\_CPNWH.R | Short R script that I used to identify duplicate data between the CPNWH and GBIF source datasets. | Description of intermediary files can be found within the R script.  No next steps. |
| *n/a* | CPNWH\_20240726-121916-distinct-assigned.csv | The CPNWH source data from IMERSS that you provided me with. | Original source file, no next steps. |
| *n/a* | GBIF\_2024\_07\_11\_Tracheophyta-assigned.csv | The GBIF source data from IMERSS that you provided me with. | Original source file, no next steps. |
| *n/a* | GBIF\_CPNWH\_Unique\_Final2.xlsx | The resulting dataset of all records from the GBIF and CPNWH datasets. "Unique" worksheet contains all unique records. "Removed" worksheet contains all duplicate records. | Dataset just needs one final check for potential duplicate records. See record with a note in the "LT\_Notes" column for details about why the record might be a duplicate.  Next Step(s)  - ​Check potential duplicate records indicated by the "LT\_Notes" column in the "Unique" worksheet. For any additional duplicate records, cut the record/row with the least amount of data (usually the CPNWH row) from the "Unique" worksheet, paste it into the "Removed" worksheet, and update the "Why Removed?" column as needed. Clear the "LT\_Notes" cell on the "Unique" worksheet with the now unique record.  - ​Use the "Unique" worksheet as the basis for comparison with future datasets to build the Flora of Texada Island. |
| RProject | *multiple* | *multiple* | A copy of all the folders and files that were forked from Emma's GitHub.  The data > data\_digitization > occurrence\_data is the sub-folder I worked in. Various intermediary data files are contained within these sub-folders. Emma's scripts calls on data from these folders directly. The raw occurrence data I worked with for journal 8 can be found here: data > data\_digitization > occurrence\_data > 1\_raw\_data > HJ-8-occ-entry.xlsx  The scripts sub-folder includes the updated scripts ("\_v2") that I wrote. These have been pushed to Emma's GitHub already. | No next steps. |
| TerryLudwarChecklist | *n/a* | FLORA OF TEXADA [with revisions - JD 19Jun2018]**.docx** | Terry's original Flora Checklist document (as a **Word document**). | Original source file, no next steps. |
| *n/a* | FLORA OF TEXADA [with revisions - JD 19Jun2018]**.pdf** | Terry's original Flora Checklist document (as a **PDF file**). | Original source file, no next steps. |
| *n/a* | TerryLudwar\_TexadaFlora\_Clean.xlsx | Includes  metadata, a list of the sources of observations, a clean copy of Terry's Flora Checklist, and Harvey Janszen's TEXLIST3 species list.  \*\*Terry incorporated Harvey's TEXLIST3 species list into his Flora Checklist, so the records in Harvey's TEXLIST3 species list are duplicates. | Potential need to follow-up with Terry Ludwar to obtain more details about the source of the data.  *Copy of e-mail chain with Terry is attached to this e-mail  ("Re: Flora for Texada Island").*  Next Step(s)  - ​Find and append source data  - ​Convert to Darwin-core format  - Compare to GBIF-CPNWH dataset to identify any unique records for inclusion in the final Flora of Texada Island |
| *n/a* | TerryLudwar\_TexadaFlora\_rough.xlsx | The rough draft of Terry's original Flora Checklist document, converted to a spreadsheet. | Original source file, no next steps. |
| *n/a* | *n/a* | DarwinCoreMetadata.xlsx | Contains a description of all the variables included in the Darwin-core formatted datasets. | There are a few variables that were included in Ben's collection datasets that were not included in the occurrence dataset I worked with. Perhaps Ben can write the descriptions for those variables?  Next Step(s)  - Complete descriptions for any variables missing a description. |

Let me know if you have any questions. As I mentioned in my e-mail yesterday, don't hesitate to reach out if I can help tie up loose ends on this to get the Flora finalized and published.

Thank you for mentorship this summer. This was an interesting project to work on, and I am glad to have had the experience!

It was a pleasure to work together, and I hope our paths cross in the future.

Sincerely,

Lindsay

**Lindsay Louise Trottier**

PhD Student

[Iversen Lab](http://www.lifeinmud.com/) | [MAP Project](http://www.maptraits.com/)

Department of Biology | McGill University

One last note to add: the actual Darwin-core formatted occurrence data from Journal 8 are saved here:  
  
RProject\Harvey\_Janszen\_Legacy\_Project\data\data\_digitization\occurrence\_data\darwin\_core\_data\darwin-core-occurrences\_2024-07-17.csv