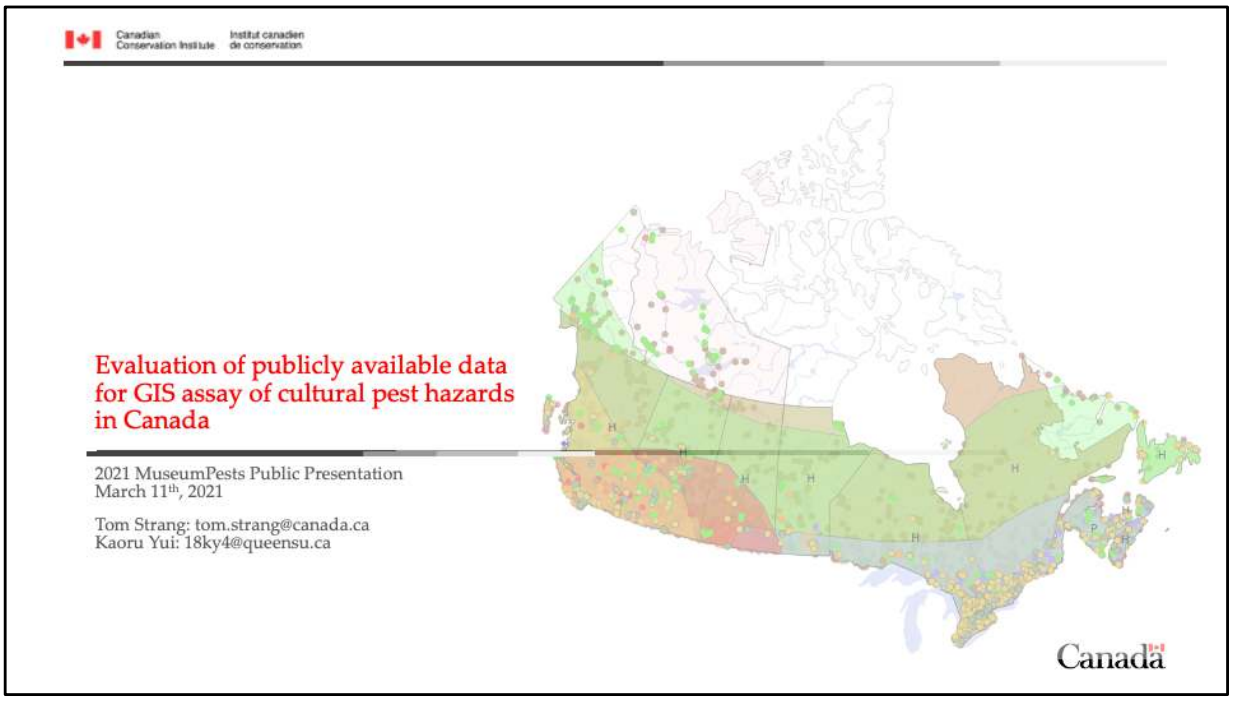
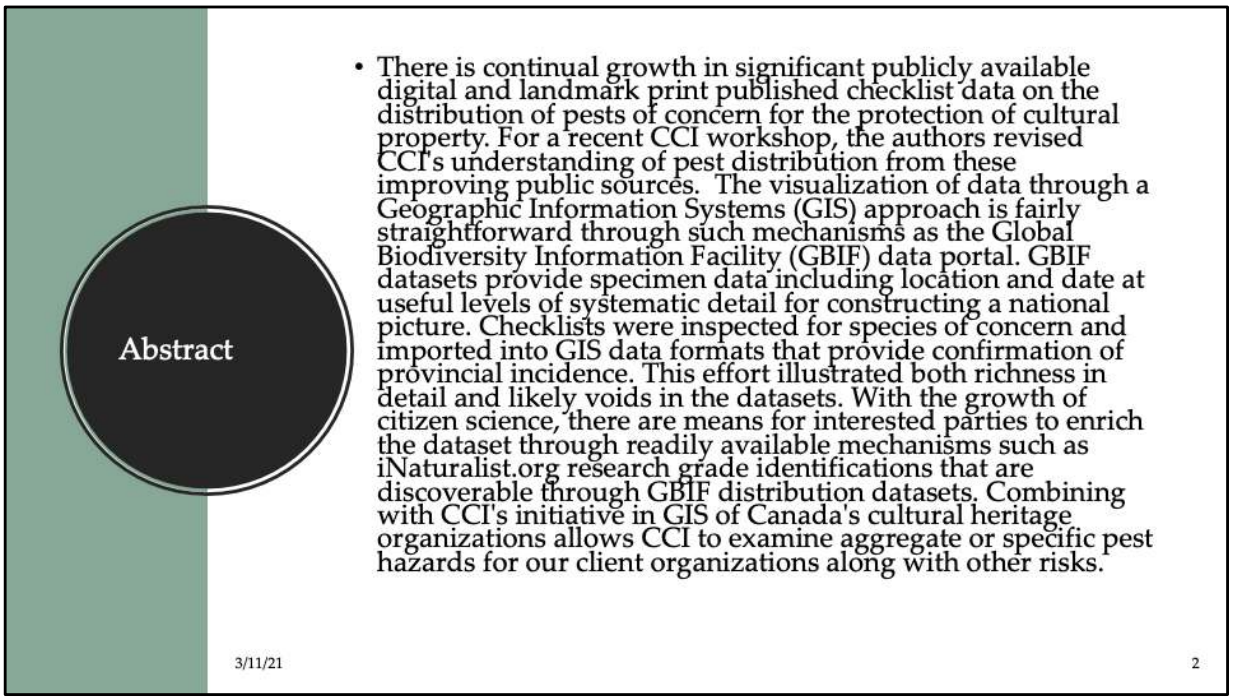
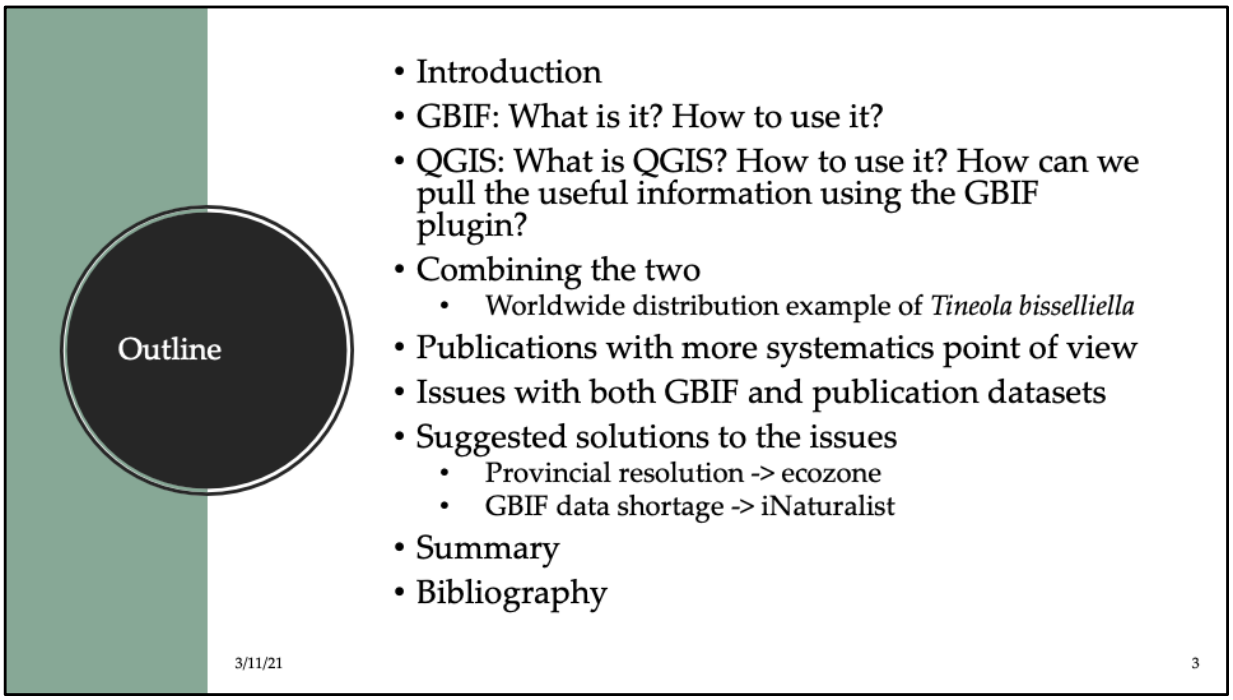


This is a talk that was given at the MuseumPests Public Presentation on March 11th, 2021.

Tom Strang: tom.strang@canada.ca
Kaoru Yui: 18ky4@queensu.ca

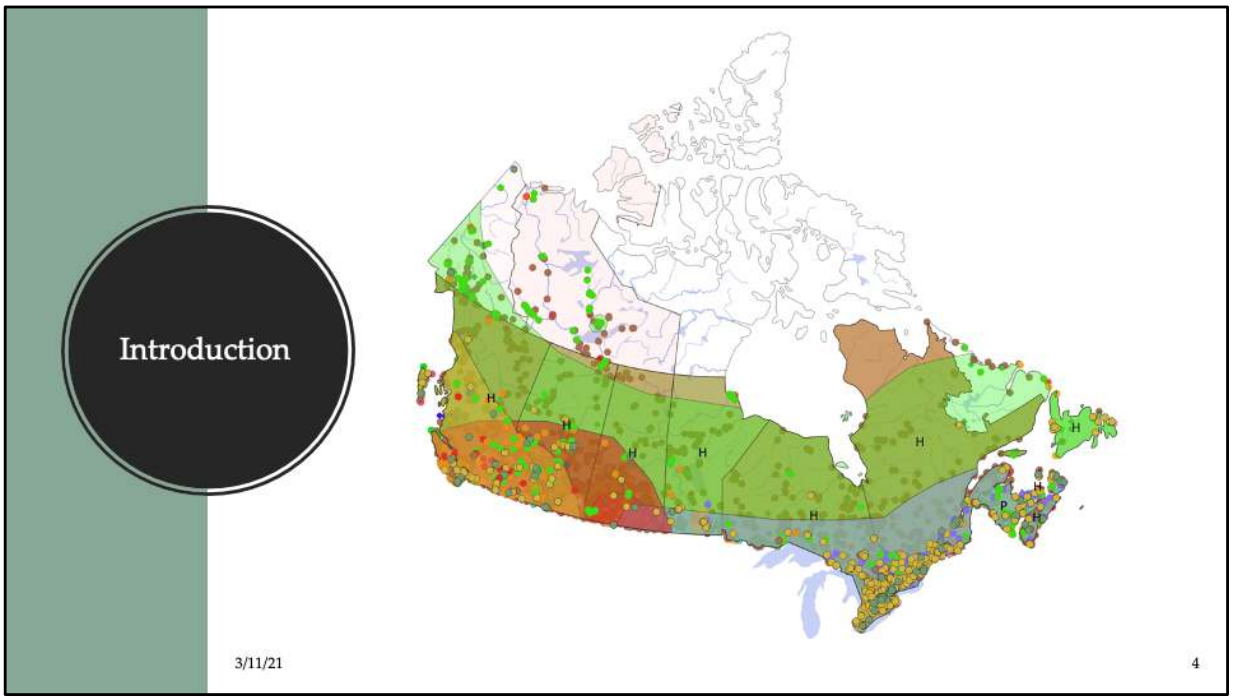


Abstract: There is continual growth in significant publicly available digital and landmark print published checklist data on the distribution of pests of concern for the protection of cultural property. For a recent CCI workshop, the authors revised CCI's understanding of pest distribution from these improving public sources. The visualization of data through a Geographic Information Systems (GIS) approach is fairly straightforward through such mechanisms as the Global Biodiversity Information Facility (GBIF) data portal. GBIF datasets provide specimen data including location and date at useful levels of systematic detail for constructing a national picture. Checklists were inspected for species of concern and imported into GIS data formats that provide confirmation of provincial incidence. This effort illustrated both richness in detail and likely voids in the datasets. With the growth of citizen science, there are means for interested parties to enrich the dataset through readily available mechanisms such as iNaturalist.org research grade identifications that are discoverable through GBIF distribution datasets. Combining with CCI's initiative in GIS of Canada's cultural heritage organizations allows CCI to examine aggregate or specific pest hazards for our client organizations along with other risks.



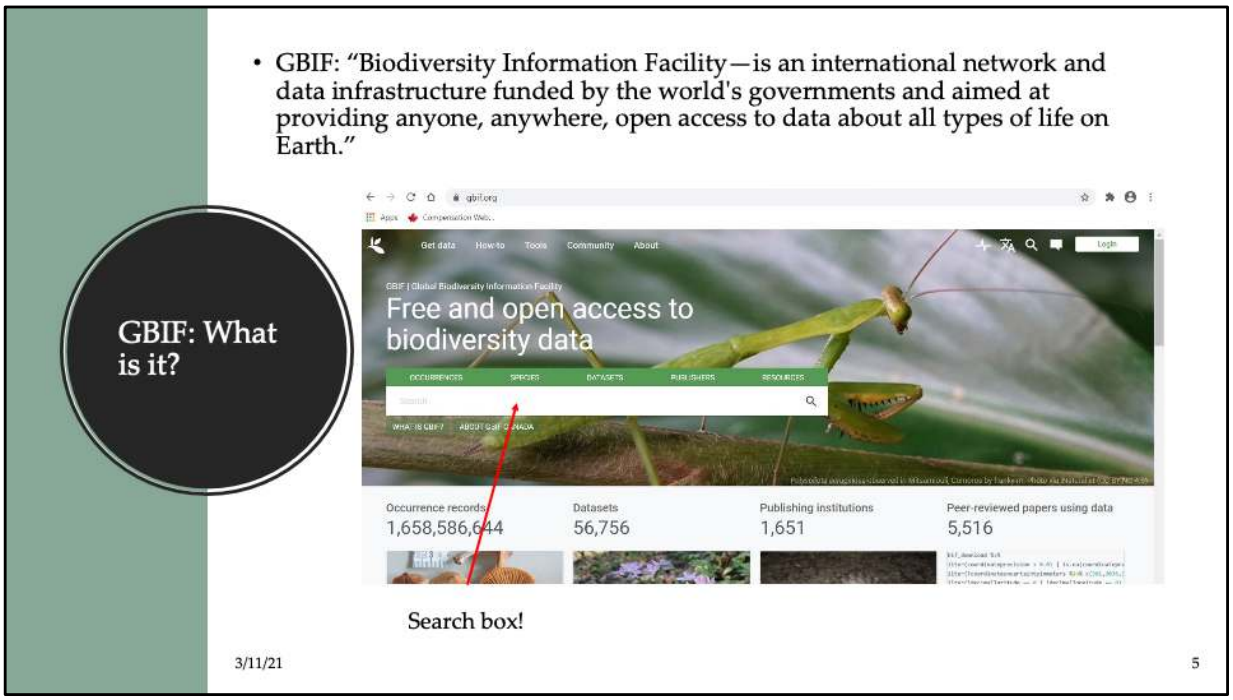
Outline

- Introduction
- GBIF: What is it? How to use it?
- QGIS: What is QGIS? How to use it? How can we pull the useful information using the GBIF plugin?
- Combining the two
    - Worldwide distribution example of *Tineola bisselliella*
- Publications with more systematics point of view
- Issues with both GBIF and publication datasets
- Suggested solutions to the issues
    - Provincial resolution -> ecozone
    - GBIF data shortage -> iNaturalist
- Summary
- Bibliography



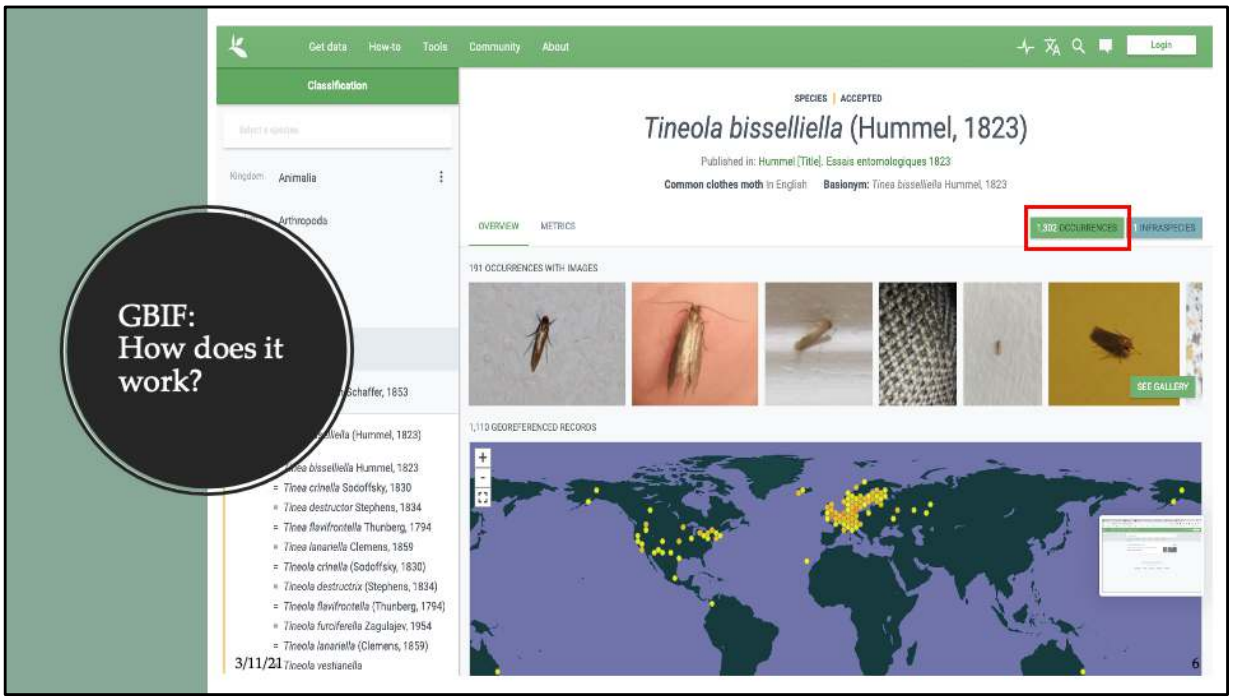
For a recent CCI workshop, CCI revised their understanding of pest distribution of Canada from improving public sources including Global Biodiversity Information Facility (GBIF) data portal and print published checklist data.

This map, built by Tom Strang and Kaoru Yui, shows the distribution of all the museum pest species of concern in Canada. To generate this map, both GBIF data portal (shown as coloured dots) and published checklist data (shown as coloured provinces) were used.
This presentation is to share our findings and the use of these publicly available tools that we have been using to visualize the insect distribution data.

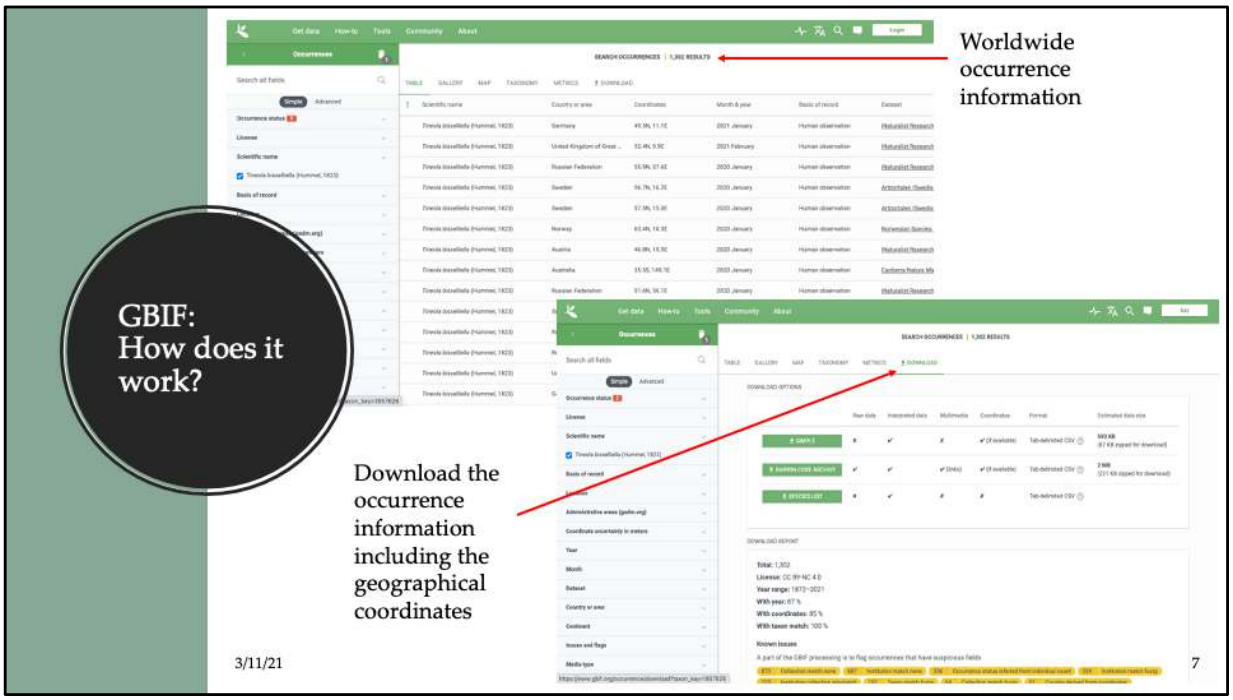


GBIF—the Global Biodiversity Information Facility—"is an international network and data infrastructure funded by the world's governments and aimed at providing anyone, anywhere, open access to data about all types of life on Earth" (GBIF.org, 2021).

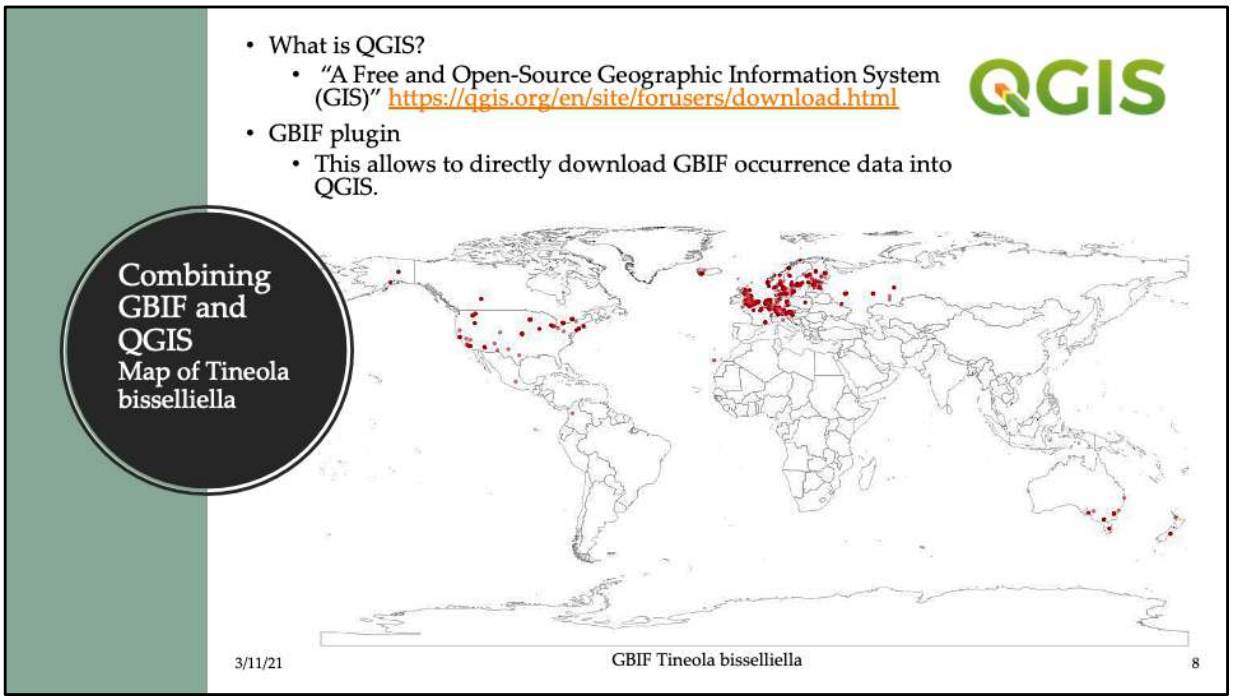
GBIF website: https://www.gbif.org/



When you search a species of your interest, GBIF website comes back with occurrence data that includes the location of where the record is collected and images. The website generates metrics page which a user can see the number of occurrences by month, area, year, source of dataset, and basis of record.

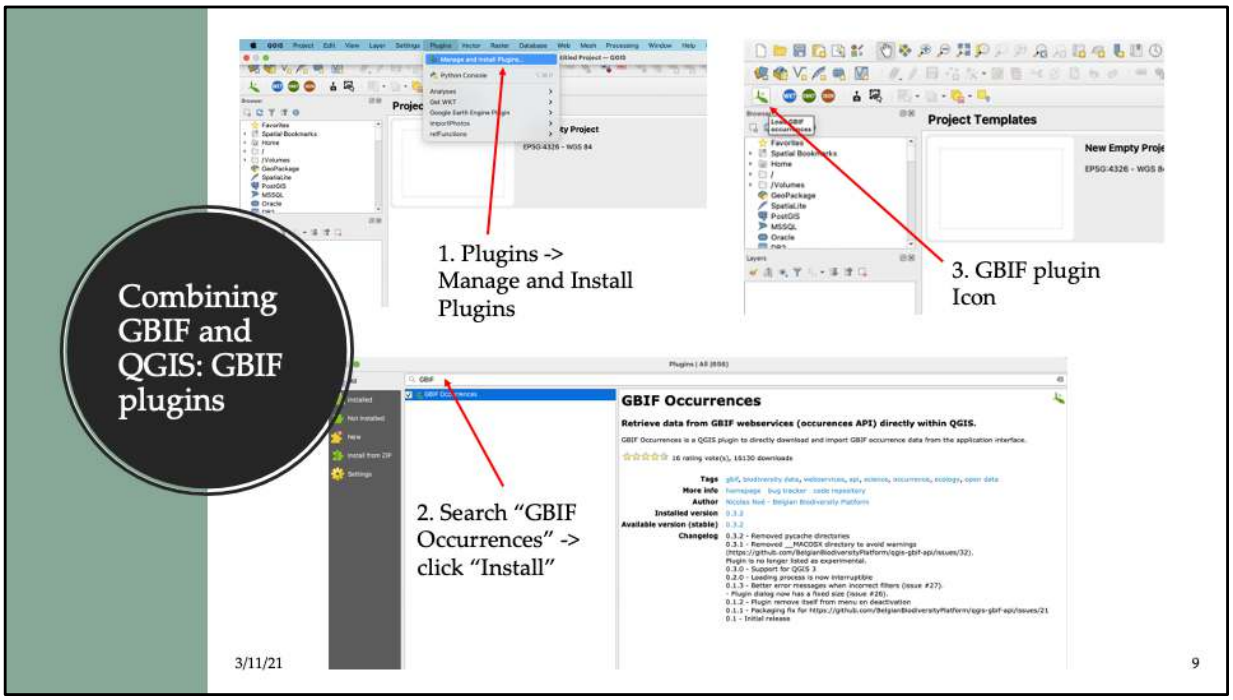


GBIF website allows a user to download the worldwide occurrence information by clicking occurrence, then download icons. In the case of *Tineola bisselliella*, the oldest occurrence data can be dated back to 1872. Each occurrence record has country, coordinates, moth and year, basis of record, and dataset information.



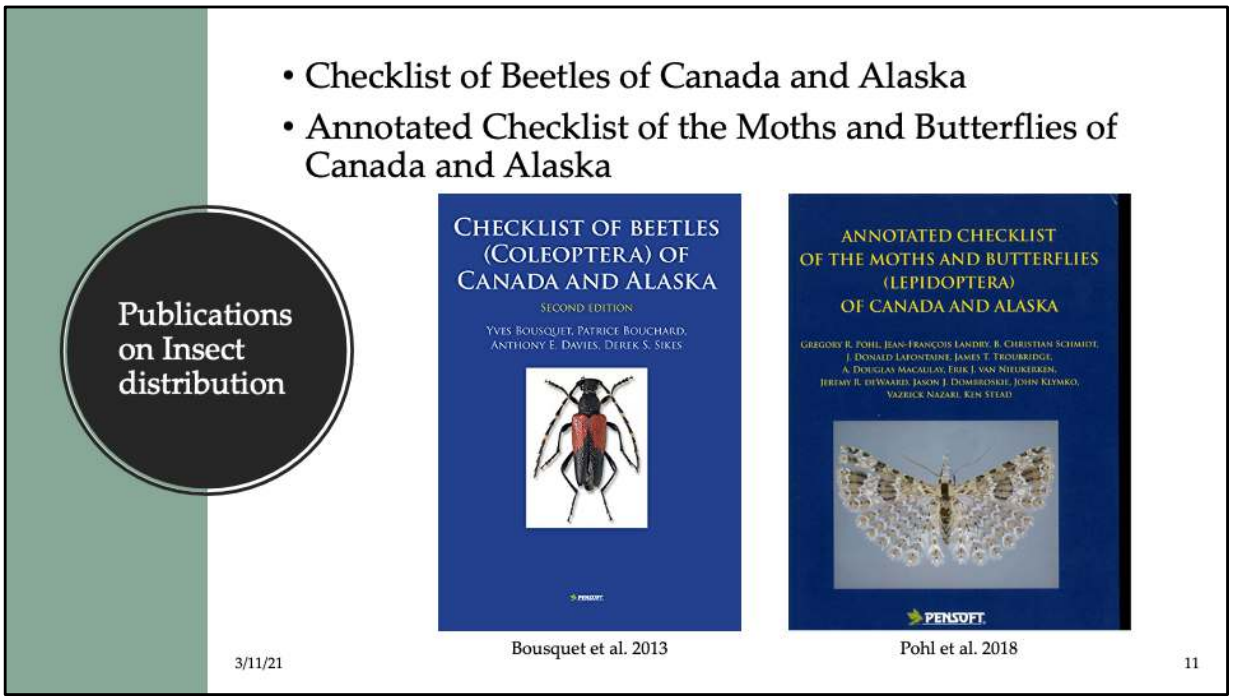
With this downloaded information, you can create a map easily by using QGIS, which is an open-source Geographic Information System (GIS) software. It's free and very powerful tool. It's available for both mac and windows users. Here, the world map shows the occurrences data of *Tineola bisselliella* that are currently obtained through GBIF portal (March 2021).

QGIS download: https://www.qgis.org/en/site/

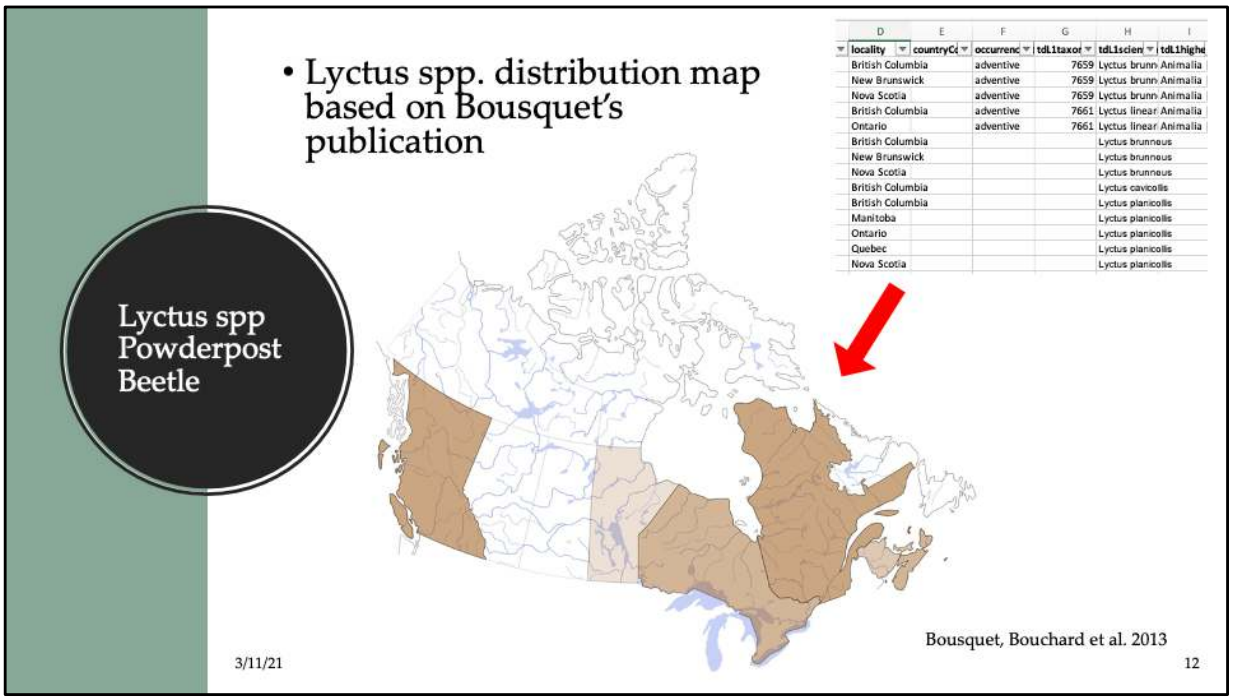


Alternatively, GBIF Occurrences plugin can be installed on your QGIS. It is a QGIS plugin to directly download and import GBIF occurrence data from the application interface. Once you have QGIS downloaded and open, it is very straightforward process to install the GBIF plugin. The instruction is on the slide.

The plugin allows you to search your interested species data from QGIS software by pulling out the data from the GBIF data portal.
GBIF datasets provide specimen data including location and date at useful levels of systematic detail for constructing a national picture. The map generated by QGIS can be exported as an image file.

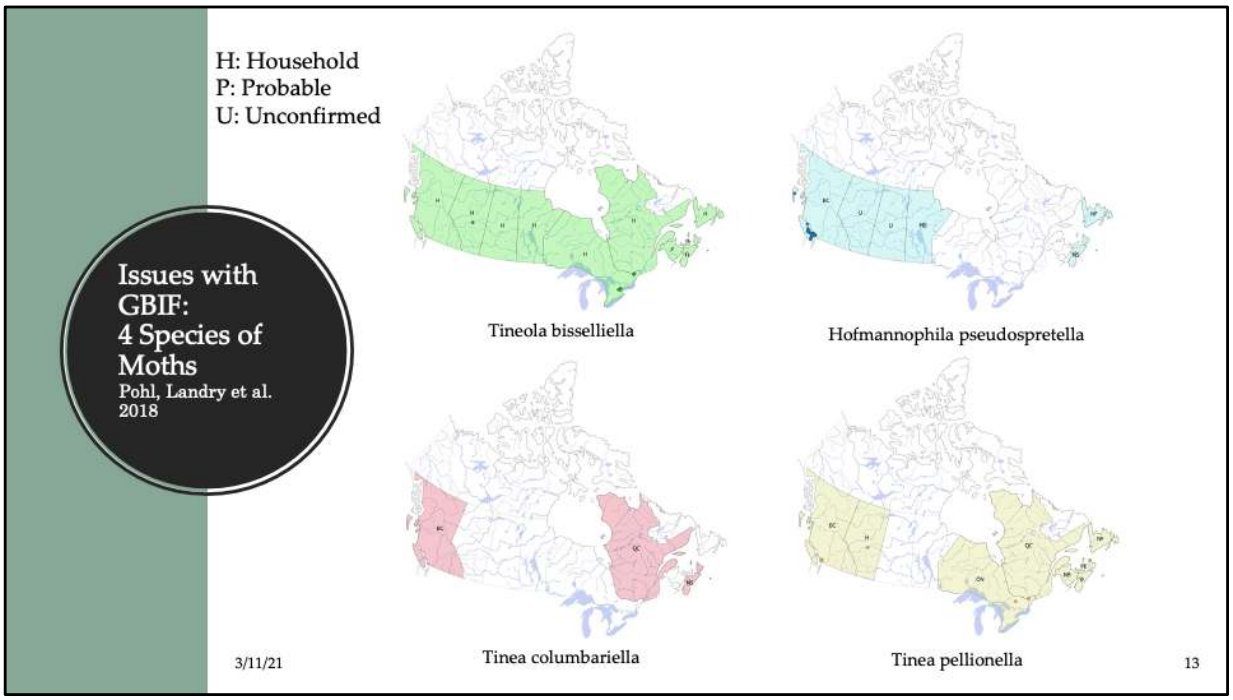


Aside from the GBIF, information from other sources were collected as well. Both *Checklist of Beetles of Canada and Alaska* and *Annotated Checklist of the Moths and Butterflies of Canada and Alaska* were used. They both represent systematic reviews of the coleoptera and lepidoptera that are known to occur in Canada and Alaska. The coleoptera checklist book has references of over 8000 species. In the lepidoptera checklist has over 5000 species listed. Both are recorded by province, territory or state in a classification framework.

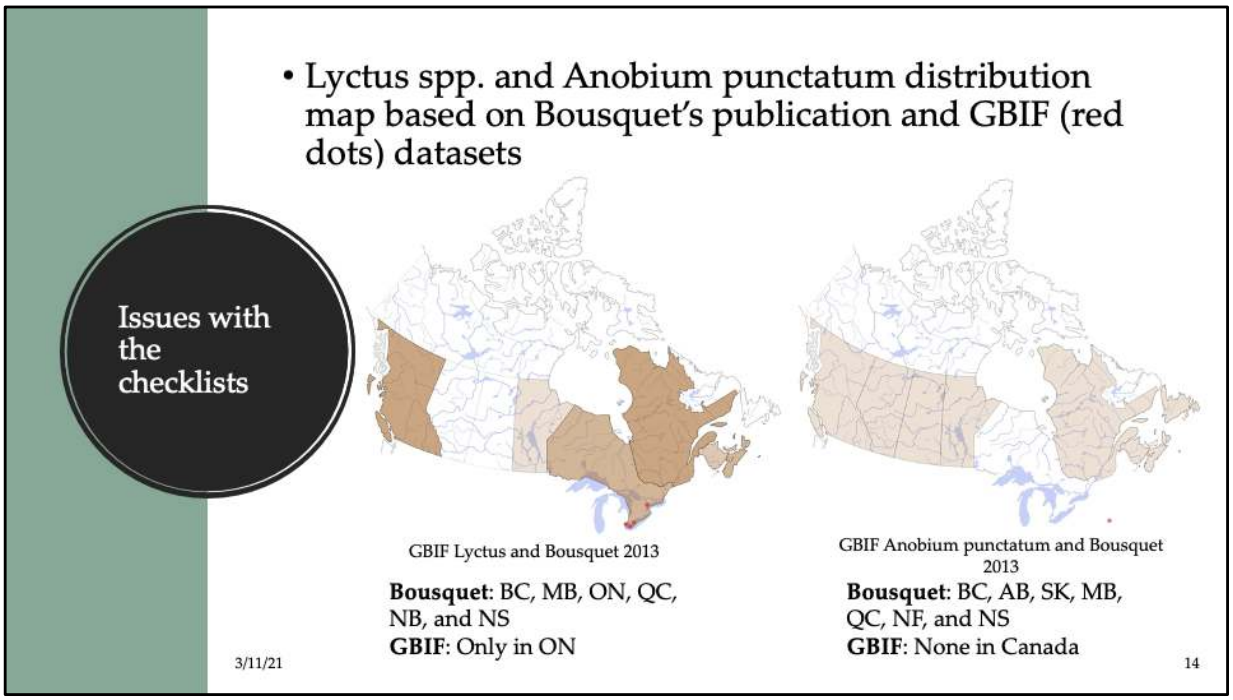


locality	countryC(	occurrenc	tdL1taxot	tdL1scien	tdL1highe
British Columbia		adventive	7659	Lyctus brunn	Animalia
New Brunswick		adventive	7659	Lyctus brunn	Animalia
Nova Scotia		adventive	7659	Lyctus brunn	Animalia
British Columbia		adventive	7661	Lyctus linear	Animalia
Ontario		adventive	7661	Lyctus linear	Animalia
British Columbia				Lyctus brunneus	
New Brunswick				Lyctus brunneus	
Nova Scotia				Lyctus brunneus	
British Columbia				Lyctus cavicollis	
British Columbia				Lyctus planicollis	
Manitoba				Lyctus planicollis	
Ontario				Lyctus planicollis	
Quebec				Lyctus planicollis	
Nova Scotia				Lyctus planicollis	

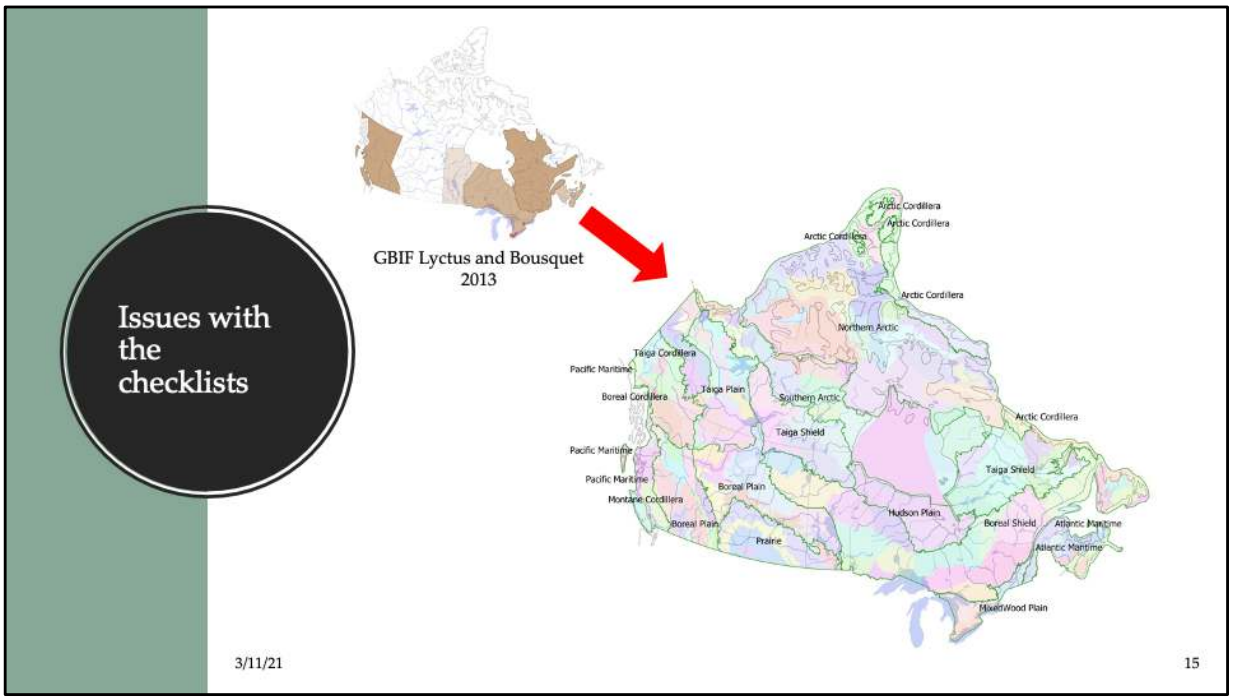
Using the records published in these books, we created excel spreadsheets pulling out only the museum pest species with locality information. Then, the spreadsheet dataset was combined together with the base map of Canada using QGIS and SQL commands. Further power of SQL can be obtained when QGIS is combined with data stored in PostgreSQL with PostGIS spatial extensions.
(https://www.postgresql.org/ https://postgis.net/)



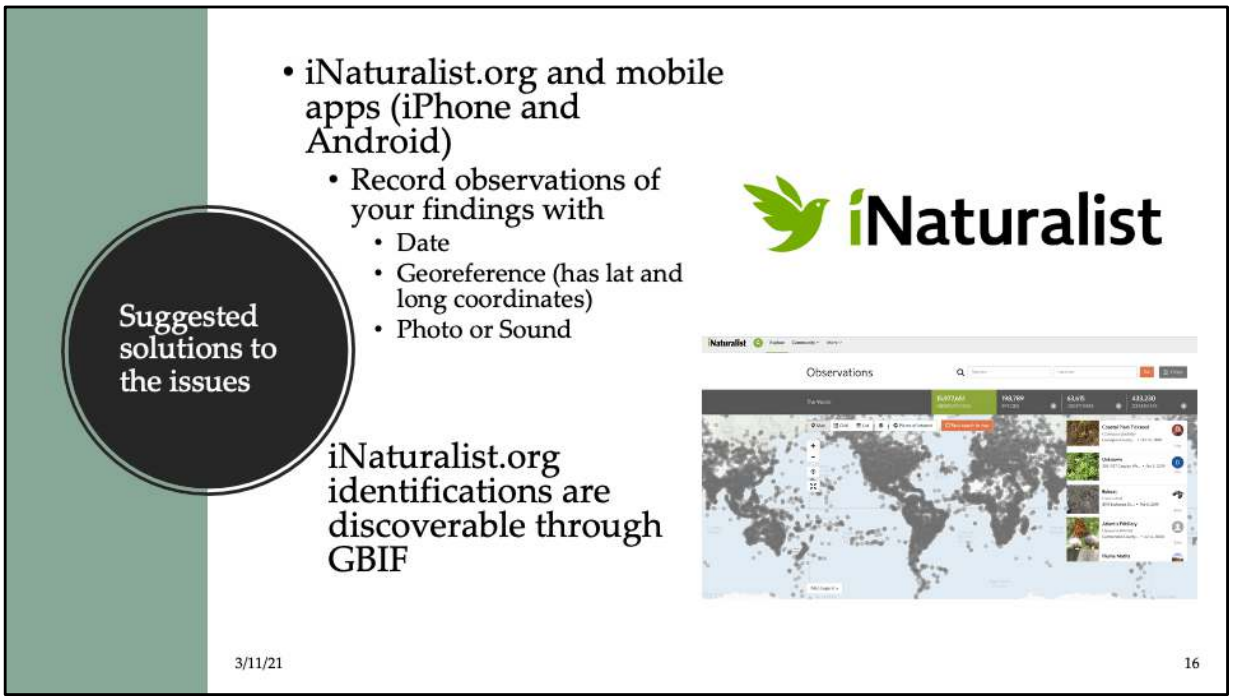
These are other example maps that show distribution of 4 kinds of moths that digest keratin. Provinces that ever had an occurrence record of these species are highlighted. The upper-case abbreviation on the map including, H, P and U, represents the followings.
H = “in human environments only” (Pohl, Landry et al., 2018, p. 24).
P = “probable occurrence meaning that these species are expected in the region, but they have not yet been found there” (Pohl, Landry et al., 2018, p. 24).
U = “unconfirmed presence in the region; these are either plausible old literature records for which no vouchers are known, or known collection records for which the determinations are uncertain or unverified” (Pohl, Landry et al., 2018, p. 24).



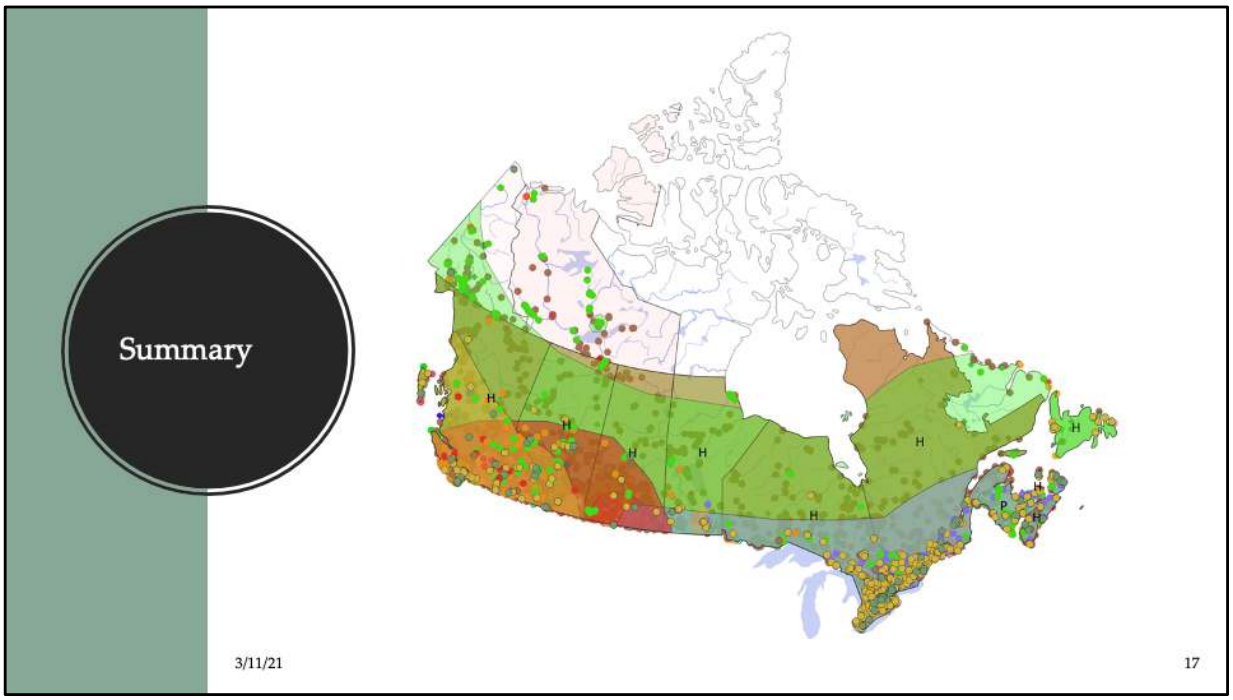
This effort of comparing the two datasets from GBIF and the checklist publication illustrated both richness in detail and likely voids in the datasets. These are the maps of both *Lyctus* species on the left and *Anobium punctatum* on the right. As you can see on the map the GBIF datasets for *Lyctus* species that are shown as red dots only appear in Ontario, but not in other provinces that were recorded on the checklist. For *Anobium punctatum*, there is no GBIF occurrence data entry in Canada. This means that GBIF doesn't yet have enough data.



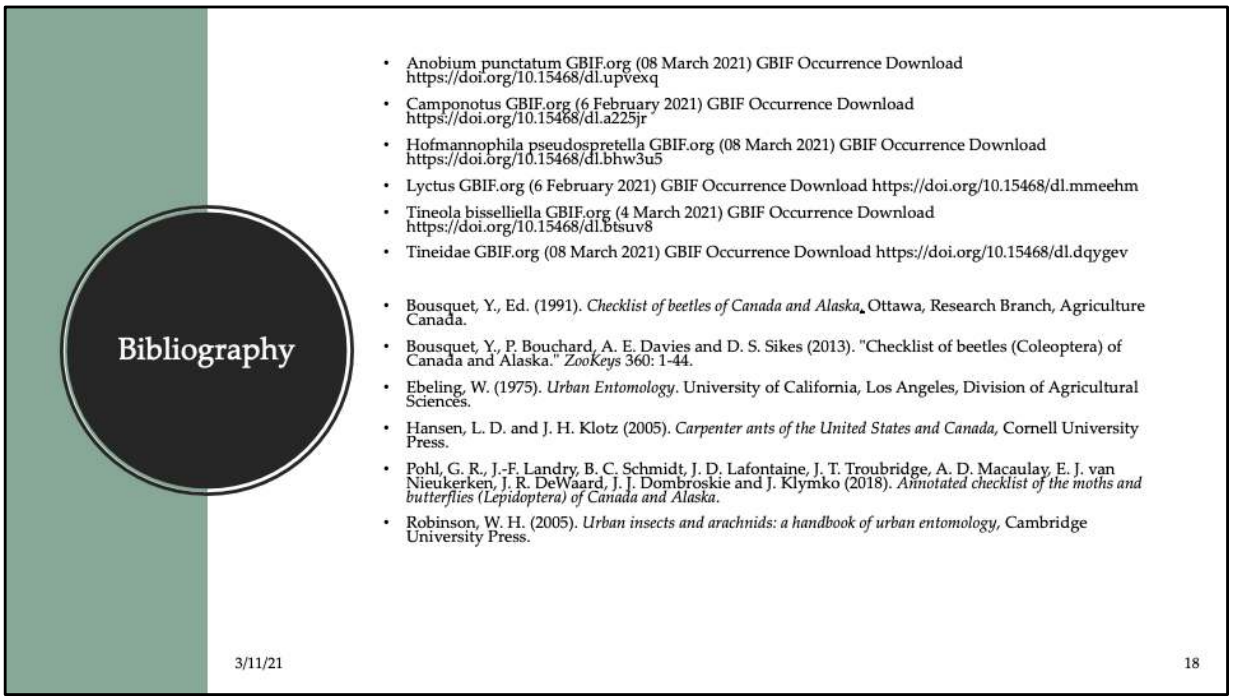
On the other hand, the checklists represent the records by provincial resolution which is also problematic. To better understand the distribution of museum pests in Canada, ecozones or ecoregions boundaries need to be considered instead of provincial border.



To enrich the dataset, there is readily available mechanisms such as iNaturalist.org which is a joint initiative by the California Academy of Sciences and the National Geographic Society. Research grade data added to the iNaturalist.org website will be discoverable through GBIF and can be used to improve the distribution data.



Distribution map like this can be a powerful visual communication tool and can be easily generated by using publicly available software and websites as mentioned. With the growth of citizen science, there are means for interested parties to enrich the dataset and improve our understanding of worldwide museum pest distribution.



Lastly, GBIF generates a reference for each occurrence record that is downloaded. Therefore, the traceability of data is ensured.