Publishing and Using Biodiversity Information in Canada

Meeting of key stakeholders

Thursday, February 24, 2011
09:00 to 16:30
Salon C, K.W. Neatby Building
Agriculture and Agri-Food Canada Ottawa

Organized by
Canadensys (www.canadensys.net)**;**
The Federal Biodiversity Information Partnership; (www.cbif.gc.ca/fbip/fbip_e.php)
NatureServe Canada (www.natureserve-canada.ca)

Summary

On Thursday, February 24, 2011, a group of 34 representatives of key stakeholders working on collecting, preserving, publishing and/or using biodiversity data (universities, federal government department and agencies, NGOs, museums, private sector) came together to discuss a common Canadian strategy for making these data accessible and facilitate their use.

Biological collections and observation data containing important information needed for Canada to manage its biodiversity are dispersed among different institutions throughout Canada and internationally. For collections, less than 10% of specimen records have been transposed to a digital format, with even fewer of those data available online, nor is there a comprehensive record of the species, taxonomic groups, regions, timeframes, etc. that are contained in each collection. For observational data, many field studies remain in field notes or in personal digital collections that are not published and are at risk of being lost as researchers retire or as computer hard drives fail or are recycled. As a consequence, valuable and needed information remains inaccessible or is being lost. We are not even at the point yet of knowing where the gaps are.

Short presentations were given on past and current initiatives in this area, followed by open discussions on how to coordinate efforts among partners and stakeholders around the table. Participants then identified next steps that could be taken to increase the availability of biodiversity information and facilitate its use. The key conclusions from these discussions are as follows.

- A substantial leap forward is needed in data digitization, publication and use in order to improve our understanding of biodiversity, and to inform decision-making for its preservation and for sustainable development;
- A culture change is underway that favours sharing and publication of data among data holders and users;
- Of the various steps in the biodiversity knowledge chain from field collection to aggregate analysis and usage, digitization and publication of legacy data are critically underfunded; there is no effective infrastructure in Canada to support that effort.
- Efficient tools for data publication, as well as proper incentives and recognition of data publication as a critical deliverable of researchers is essential to justify the time investment required;
- Publication of data and providing efficient access to support use will require cooperation and coordination of all key stakeholder through one coordinating mechanism.
- Data publishing and access must be targeted towards data use, establishing priorities based on the needs of the users;
- Efforts must be coordinated to make the case for data mobilization to governments, funding agencies and the general public;
- Common standards are essential for sharing a broad range of biodiversity data and must be developed and widely implemented;
- A common portal and useful software tools need to be developed (in most cases existing tools can be adapted to fit the needs);

- Efforts concentrating on the publication of metadata, giving information about the contents of Canadian biodiversity collections, will allow users to identify the datasets they need and allow for demand driven digitization efforts.
- The focus of this effort should remain on species level primary biodiversity data from specimens and reliable observations.

As a key first step in the establishment of the appropriate networking and management structure for the publication and use of Canadian biodiversity data by all stakeholders, the three organizing partners were encouraged to apply the input and priorities produced in this meeting to the production of the proposal to the Networks of Centres of Excellence – Knowledge Mobilization (NCE-KM) program.

9:00 - Opening address (Geoff Munro, Chief Scientist & ADM, Natural Resources Canada)

Dr. Geoff Munro stated his unequivocal support for the goals of the meeting, framing it as an effort to "defragment" the way Canadian biodiversity data are managed. Dr. Munro also indicated his expectation that this meeting would help to produce a solid business case for managing these data that can then be used to promote greater efforts, cooperation and data sharing within the Federal Government, as begun through FBIP.

9:45 – Experiences from active networks

As an introduction to this section, Anne Bruneau illustrated biodiversity data with a 3-dimensional plot comprising a "Data Type" axis ranging from genetic to ecosystem data; an "Institutions" axis for the various data-holding institutions; and a "Taxonomic group" axis. For many empty cells, the data may simply be non-existent or so hidden as to be virtually unknown. For others, they may be available but not accessible. Few cells can be tagged with the "accessible" label.

The focus of the stakeholders assembled in this meeting is on the middle region of the first axis: species occurrence data from collection specimens and observations. A goal of the meeting and the NCE-KM application is to make the second axis (Institutions) irrelevant by coordinating among all stakeholders and ensuring that data holders wishing to publish their data are able to do so.

NatureServe Canada (Douglas Hyde) NatureServe is a Non Governmental Organization that serves as a focal point for contributions of data from provincial and territorial Conservation Data Centres (CDCs) and, as a steward of information. NatureServe has the largest database for rare elements in Canada.

NatureServe's report ''*The state of biodiversity information*'' (www.natureserve.org/publications/natureserve_canada_SOBI_2010.pdf) makes clear that biodiversity data are mostly inaccessible and that a national partnership going beyond institutional boundaries is needed for their effective management.

Canadian Biodiversity Information Facility (CBIF) (Guy Baillargeon) The Canadian Biodiversity Information Facility (CBIF, www.cbif.gc.ca/home_e.php) works under the Federal Biodiversity Information Partnership (FBIP), a partnership between all federal departments and agencies involved in biodiversity. CBIF is the Canadian node for GBIF, which gathers and distributes biodiversity data through its portal at an international level.

Canadensys (Anne Bruneau) Canadensys brings together university and botanical garden collections with the purpose of serving their data through a shared portal, focussing initially on plants, insects & fungi. Canadensys works with partner institutions to streamline digitization and develop informatics tools. Its Canada Foundation for Innovation (CFI) funding covers the establishment of the infrastructure as well as digitization and publication efforts to make approximately 20% of the available data accessible to all.

NatureServe Canada, CBIF and Canadensys have been working together to coordinate Canada's participation in international efforts, and plan for future developments, such as this meeting. The three institutions are also complementary in that they may access different sources of funding.

11:00 – Discussion 1: The need for biodiversity information (why?) – Felix Sperling (University of Alberta)

Why are we here? Why do we need to publish biodiversity information? Why now?

In addition to losing biodiversity, Canada is also continuously losing knowledge and data that it has amassed about biodiversity. Canada is falling behind on its international commitments as well as its responsibilities to future generations.

Patchiness of the available data (regional, taxonomic, etc.) limits the possibility to aggregate and analyse them effectively. However, inputting legacy data is time consuming, costly and poorly funded in Canada.

With general agreement on the importance of preserving data and enhancing their accessibility and use for knowledge advancement, conservation, policies, regulations, sustainable development, environmental assessments, etc., the focus of the discussion shifted towards the obstacles to the publication of the data.

Attribution / funding: Funded biodiversity projects often lack incentives to publish the data generated in addition to the articles. Such publication should be considered a key deliverable, as important as more traditional scientific articles. In addition to emphasizing the importance of capturing and sharing the data, the following points will be essential in effecting a positive change in the funding culture. Tools are needed that track the provision and use of data, and that ensure accurate attribution of credit. This would also alleviate the obstacles caused by perceived ownership of the data and the tendency to hoard them for hypothetical future research and traditional publications. To assist in developing project budgets, we also need to determine the cost of getting such data. Finally, the link between biodiversity preservation and data management should also be explained to the public.

While resource companies wish to avoid local encroachment by direct competitors, they are also interested in clustering with other firms to improve site appeal and local services. This creates an incentive for sharing environmental assessment data that can benefit an open biodiversity data network.

13:00 – Discussion 2: Strategies for moving forward (how?) – James Macklin (Agricuture and Agri-Food Canada)

The technical question ‘how is the data going to be shared’ is not a great challenge. An overview of the need for standards and tools and how they may be adapted from existing resources was briefly provided but the main focus of the discussion was social and political issues, whichare more difficult and require the development of a Canadian model adapted to our reality.

Metadata about a wide range of large and small collections, whether they are already digitized or not, can be compiled and published relatively quickly. They can then be used to identify what primary data may be available, missing, needing to be digitized, etc. For a company potentially interested in a particular area a meta-database may allow the identification of relevant datasets. If a dataset is as yet inaccessible, funding for targeted digitization may be put in place (demand-driven digitization). Publication and accessibility for all to the newly-digitized information will need to be determined on a case by case basis but, generally speaking, publicly-funded digitization should result in publicly-available data.

Clearing-house model: If such a structure were put in place, in which the data user is linked directly with the provider, without further intervention of the connecting institution (e.g. GeoConnection), it would be very important to develop effective mechanisms to keep track of uses and attribution information.

Citizen science might be very useful for certain aspects of digitization such as geo-referencing of specimen collection points.

Prioritization: Digitization and publication efforts must be prioritized through different lenses for different users (species-at-risk, invasives, climate change, regional assessments, etc). A comprehensive metadata inventory, combined with accessibility of previously-digitized information is essential for the various prioritization exercises. A well-structured governance model is essential to coordinate these.

14:00 – Discussion 3: Responsibilities, leadership and coordination (who?) – Douglas Hyde (NatureServe Canada)

In addressing how to govern data and who plays what role, the pros and cons of various models were discussed: Wikipedia (easy to update but difficult to vet); GenBank (difficult to update and not always well curated); Google; Cornell's eBird model; Intergovernmental Panel on Climate Change-model; Etc.

As part of this section, the issues of **governance of the data and long-term viability of the data repositories** were discussed. The Federal Government has open-ended responsibilities mandated by Acts of Parliament. Universities and museums also typically take their responsibilities as knowledge repositories very seriously. Data holders (scientists, institutions, government departments, etc.) must retain ownership of their data as well as the curatorial responsibility to authenticate them and keep them updated & corrected (GBIF was criticized for holding data with errors that they could not get corrected by the owners).

This brought the participants back to the issue of incentives for researchers to submit their data and the various possibilities for fees for certain data users or data uses, subscription, etc., which conflict with the principle of maintaining publicly paid-for data in a publicly accessible format. The user interface was seen as critical to address the needs of the users since scientists/government, companies and the public would need to use different interfaces adapted to their requirements.

NGOs are also critical stakeholders in the biodiversity information chain since they can leverage funds (donations) and resources (volunteers) that other groups may not always be able to access effectively. NGOs can also lobby government.

As mentioned previously, industry and business partners are also essential users of biodiversity data, essential partners in sustainable development and can help fund digitization, data cleaning, publication and other key efforts.

15:30 – Discussion 4: Priorities for moving forward (what?) – Anne Bruneau

Key suggestions for moving forward were discussed, including:

- The need to develop the "coalition of the willing" evident in the meeting into a collaborative, coordinated effort with participation and support from different quarters, able to provide access to much more data than at present and to develop the tools to analyze them.
- The need for advocacy and search for support at different levels: federal government, granting agencies (support grants and inclusion of data publishing requirements in research grant criteria), media/public outreach, etc.
- The development of a policy paper for Canadian biodiversity data publishing, including case studies, cost-benefit analyses, relevance for specific policy issues (climate change, invasives, species-at-risk, etc.), gaps and priorities.

Regarding specific tools, the following suggestions were made:

- Metadata harvesting tools.
- Standardized tools for capturing observation data (NatureServe's various CDCs use different tools and they don't have the capacity to quickly share data.
- Tools to capture/evaluate the uses of the data.
- Tool for community geo-referencing.
- Legal tools for dealing with liability clauses.
- Tools allowing further quality control of the published data.

Note: the opportunities for informal networking during the meeting (breaks and lunch period) confirmed the high level of interest and excitement of the participants with the prospect and opportunity to move forward in a coordinated manner with the publication and use of digitized Canadian biodiversity data

Conclusions

The meeting concluded with thanks to the participants for constructive input and closing comments from the meeting convenors….

Follow-up

This meeting was followed the next day by a working meeting of the NCE-KM application group (G. Baillageon, A. Bruneau, S. Dupont, F. Ensink, D. Hyde, J. Macklin, P. Desmet and F. Sperling). The various points and suggestions made at the workshop were reviewed and evaluated in the context of writing that application. The workshop confirmed the need to proceed with the application and to establish the Canadensys+ Network as an opportunity for key networking and for coordinating the publishing and access to primary biodiversity information in Canada. Various specific points (metadata registry, holder and user interactions, tool suggestions, etc.) were integrated in the application. Other points were kept for future discussions within the established Network and at other venues (lobbying, changes in grant evaluation criteria, interactions with the public, production of the policy paper). The need to adequately fund digitization efforts in Canada could not be addressed in the NCE-KM application but will be a priority of the Network. Finally, whether or not the Canadensys+ Network is funded by the NCE-KM program, it is clear that the stakeholders present at the first meeting will keep working to the best of their ability towards the establishment of a suitable structure to address all the needs identified.