Online access to specimen and taxon data

- In the last 20+ years the internet has had a huge effect on the accessibility of specimen-based museum data
- These data relate to individual specimens, and to taxa, notably species
- These data may be inherent in the specimens themselves, or in the way (-s) in which the specimens have been labeled and annotated

Roles of herbaria

- herbaria house and curate specimens
- specimens document the flora
- specimens may also document vegetation units
- specimens are, of course, also central to taxonomic research in which taxon concepts are refined, and (phylogenetic) relationships are clarified
- *What do these statements mean operationally?*
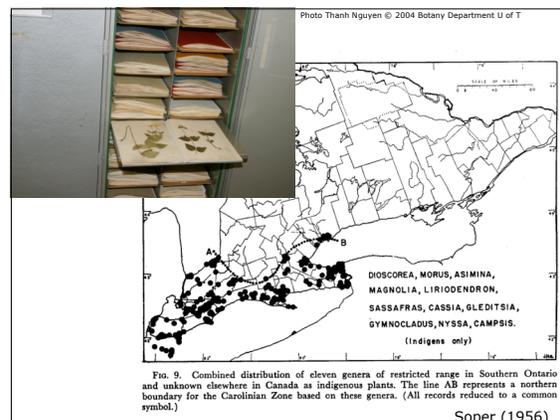
Opportunities for data suppliers

- Online visibility; the ability to be seen as the custodian of valuable data
- Accountability to funders (taxpayers, granting agencies, etc.)
- Outreach to new and existing user communities
- Creativity, innovation, and shameless self-promotion (e.g. floras, species pages, etc.)
- Greater specimen security, labor savings, when making loans can be avoided

Opportunities for data users

- Specimen data may be obtained more quickly than through traditional loans
- Specimen data may be obtained even in situations where loans could or would not be made
- Online specimen data could be highly linked, e.g. to publications, other databases, images, etc.
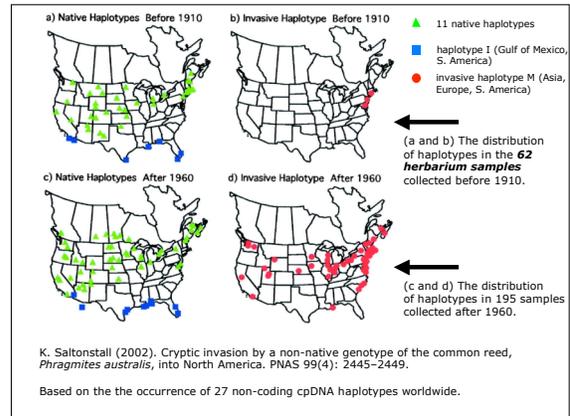- Online specimen data could be linked to from taxon databases, pages, etc.

But…

- Online availability of specimen data requires the existence of digital data in the form of text and/or images
- Deployment of specimen data not trivial with respect to expertise and time requires for success
- Garbage in - garbage out

Photo Thanh Nguyen © 2004 Botany Department U of T

DIOSCOREA, MORUS, ASIMINA, MAGNOLIA, LIRIODENDRON, SASSAFRAS, CASSIA, GLEDITSIA, GYMNOCLADUS, NYSSA, CAMPSIS. (Indigens only)

Fig. 9. Combined distribution of eleven genera of restricted range in Southern Ontario and unknown elsewhere in Canada as indigenous plants. The line AB represents a northern boundary for the Carolinian Zone based on these genera. (All records reduced to a common symbol.)

Soper (1956)

Soper and Fox publications

- Print only; pre-microcomputer, pre-internet
- Data capture by humans, probably already skilled, or trained to be so
- Map production by humans, thus error detection built in
- Publication itself is the repository for the data in accessible form

a) Native Haplotypes Before 1910   b) Invasive Haplotype Before 1910

▲ 11 native haplotypes

■ haplotype I (Gulf of Mexico, S. America)

● invasive haplotype M (Asia, Europe, S. America)

(a and b) The distribution of haplotypes in the *62 herbarium samples* collected before 1910.

c) Native Haplotypes After 1960   d) Invasive Haplotype After 1960

(c and d) The distribution of haplotypes in 195 samples collected after 1960.

K. Saltonstall (2002). Cryptic invasion by a non-native genotype of the common reed, *Phragmites australis*, into North America. PNAS 99(4): 2445–2449.

Based on the the occurrence of 27 non-coding cpDNA haplotypes worldwide.

Saltonstall publications

- Print and electronic, internet-accessible
- Data capture by humans, probably already skilled, or trained to be so
- Map production by by computer; error detection orobably by humans; complete?
- Repository for the data may be private (author or institution) or accessible as supplementary material associated with publication (hence possibly only available at whim of publisher)

Database issues

- Data entry quality control?
- Standardized fields?
  - Consider the data you catalogued while perusing herbarium specimens last week.
  - What kinds of data did you encounter?
  - How could these kinds be standardized?
- TDWG development of data standards and best practices. See *CANADENSYS* website, *HERBARIA* listserver, Darwin Core, etc.
- *How do existing online resources deal with these issues?*

Evaluating resources

- See discussion on *TAXACOM* listserver; use search function on MarkMail site to find threads relating to database issues
- And what about data-aggregating projects, e.g. Catalogue of Life, Encyclopedia of Life, Wikispecies, etc.?