PlantCollections™: linking national and international living plant databases

Christopher P. Dunn*[1], Boyce Tankersley[2], & Pamela Allenstein[3]

1 Lyon Arboretum, Hawaii, USA 2 Chicago Botanic Garden, USA, 3 American Public Gardens Association, Delaware, USA

Introduction

Easily accessible lists of, and curatorial information about, living plant collections in the world's botanic gardens is essential for educational, scientific, horticultural, and other uses. To date, many gardens maintain databases to serve their institutional needs. With mounting global concerns about loss of biological diversity, it is essential that gardens develop ways to share essential information about their holdings without compromising any proprietary information. *PlantCollections*™, a US federal-funded project intends to serve such a purpose. At present, 16 US institutions have entered into a collaborative project that will use the latest open-source technology to permit easy uploading of data to a common portal and searching of data by any user. The initial phase of the project will make available critical information on 47,000 accessions and 300,000 plant taxa. Ultimately, plant images can be accessed (via Morphbank) and distribution maps generated (via Google Map). When complete, such a widely-accessible database will go far to meeting the goals of Target 8 of the Global Strategy for Plant Conservation, particularly as we attempt to expand the project to include international botanic gardens and arboreta.

The Need for Accessible Databases

A first step toward inventorying the rare plants in *ex situ* living collections, so that they may support reintroduction efforts in their countries of origin, is knowing in which gardens plants are housed and curated. Accessible databases of living plant collections in botanic gardens contain essential information for conservation scientists, horticulturists, taxonomists, educators, and the general public. With increasing concerns regarding worldwide loss of biodiversity, such data are critical to meeting the goals of the *Global Strategy for Plant Conservation* (CBD Secretariat, 2003), particularly Target 8.

If Target 8 (which calls for 60% of threatened plant species to be in accessible *ex situ* collections) is to be realized, then those charged with making such plants accessible must have the means to share information among gardens and databases. One major challenge to achieving this goal is that gardens use a wide variety of databases, operating systems, and hardware. Furthermore, the data fields recorded and maintain by gardens varies widely, with some gardens recording a fairly limited set of data, while others record hundreds of fields.

In addition, anyone needing to search these databases must generally contact each garden to make a request. A few gardens do provide limited quantities of data via the internet; however, this is more the exception than the rule. Even if all gardens had all their data readily available, the time required to search them all, one by one, would be enormous and inefficient.

Pooling Data

The more effective way to provide as much data to end-users as possible is facilitate the capacity of gardens to make data available via the World Wide Web (WWW). This approach can be achieved in one of two major ways. In one model, data from each garden are stored on a central server, allowing the end-user to quickly search on large database. This "centralized query model" requires that each garden agree to a standard file format. In the second model, an end-user's search goes to each participating institution's database individually. This "distributed query" model can be slower (although not always perceptibly), but allows each participating institution more latitude with respect to its data storing and handling preferences. There are merits to each.

One very successful centralized query system is vPlants™, the online herbarium of the Chicago region (Schaub and Dunn 2002). Here, each participating institution (Chicago Botanic Garden, The Field Museum, and The Morton Arboretum) provides data to a central server and allows the user to search for taxa across the entire dataset or to pick and choose which herbarium to search. However, preparing the data for export to the server is more involved and time-consuming, with each record needing to be "translated" into, in this case, XML.

With the distributed query model, the demands on each garden's database (or database manager) are somewhat less challenging. With the availability of new open-source data management systems such as DiGIR, the ability to build "communities" of databases (and thereby to build communities of plant professionals) is unlimited. In addition, each garden can retain its preference database management system and software (BG-Base, Access, FileMaker, etc.). The only community-wide decisions that really need to be made are to agree on the data fields that will be made searchable on the WWW. The major sensitivity is whether to reveal the collection locations of rare plant taxa. Generally speaking, such data would not be made available electronically, but could be made available upon request from an appropriate authority.

The *PlantCollections*™ Project

The main goal of *PlantCollections*™ is to build a "community" of plant collections data users. We aim to share plant information more widely, enable collaborative research, improve availability of germplasm, improve plant collections management, and advance technology.

To build a community of living plant databases and users, the Chicago Botanic Garden, along with 15 other US gardens, the American Public Gardens Association (APGA), and the Natural History Museum and Biodiversity Research Center at the University of Kansas formed a partnership to develop, test, and implement an open source software application for data sharing. This application, DiGIR (Distributed Generic Information Retrieval) has been widely successful in organizing similar partnerships with disparate databases (Figure 1).

With funding provided by the Institute for Museum and Library Services, each participating institution receives a server (with DiGIR installed), software support from the University of Kansas, and staff training. Each institution will supply as much data as possible to populate 161 fields of data. These fields were agreed to by all participants and form what is referred to as a "federated schema." Not all gardens maintain 161 fields, but each supplies what it is able to.

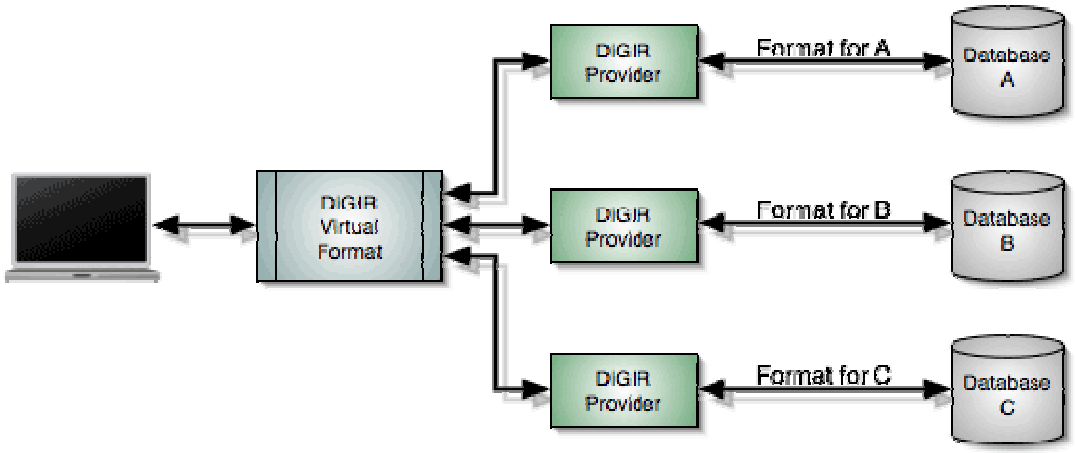

Figure 1. General representation of a distributed data network, in which each database (institution, in this case) provides data via a DiGIR provider to the end-user.

Once PlantCollections is available online, anyone with internet access will be able to search the 161 fields that describe 47,000 accessions and 300,000 plant taxa. As the project matures, images of plants will be made available via another US funded project, Morphbank, and maps generated with Google Map.

Types of information included in the federated schema include taxonomic data, source of the plant material, climatic data, hardiness, morphological features, number of individuals in a garden's collections, contact information, among many others. Two types of searches, "standard" and "expert" will be available. In addition, all institutions can be searched, or a subset chosen.

Future Developments

Although PlantCollections is not yet fully-functional, we are considering the next steps to broadening its scope and enhancing its value. First, we will test the practicality of adding a similar network to the PlantCollections one. This is more challenging than simply adding another garden to the community. Second, we will attempt to add collections from other parts of the world, some of whose data might be in completely different character sets. To be of value worldwide, a single source of data is preferable.

Conclusions

The need to efficiently access data from institutions around the world is paramount if progress is to be made in reaching the GSPC targets. With respect to achieving Target 8, ready access to plant information requires an online registry of living plants in botanic garden collections. *PlantCollections*™, a US-based project is one example of an emerging technology to just that end.

References

Schaub, M., and C.P. Dunn, 2002. vPlants: a virtual herbarium of the Chicago region. First Monday 7(5): May 2002. (peer-reviewed online journal; www.firstmonday.dk)

CBD Secretariat, 2003. *Global Strategy for Plant Conservation.* CBD Secretariat, Montreal, Canada.

http://www.cbd.int/programmes/cross-cutting/plant/default.asp