# *i*Plants:
# THE WORLD'S PLANTS ONLINE

## INFORMATION SYSTEMS

APPENDIX 5

***Final Report of Pilot Project: 1 April to 30 Nov 2004***

Version 1.2

**Author(s):** Jackson, M

**Deployment:** ***i*Plants** document library: **www.iplants.intranets.com**

**Directory:** Group documents / Outputs – Final Report/

**Filename:** Appendix 5 - Information Systems

**Approval for Public Release if appropriate:** Formally approved for release by Steering Committee

Date:

## Revision History

| Date | Who | What | Version |
|---|---|---|---|
| 26th Nov 2004 | MJ | Created | 1.0 |
| 7th Dec 2004 | MJ | Amended in light of feedback from CM and AP | 1.1 |
| 23rd Mar 2005 | BA | Feedback from NYBG included | 1.2 |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |

# TABLE OF CONTENTS

# 1. Introduction

## 1.1. The *i*Plants project

***iPlants*** aims to produce an index of all the world's plant species together with, where possible, an image and a preliminary conservation assessment. This index will be made available online.

### 1.1.1. Further Information:

For further information please contact:

The *i*Plants Initiative,
c/o Alan Paton,
Royal Botanic Gardens Kew,
Richmond, Surrey, UK
TW9 3AB
information@iplants.org

## 1.2. Purpose of this document

This document aims to

1) Describe the major Information Systems components required by the iPlants project

2) Provide information on how we would approach the software development

3) Provide an initial look at development schedules and maintenance implications

## *1.3. Outstanding Issues*

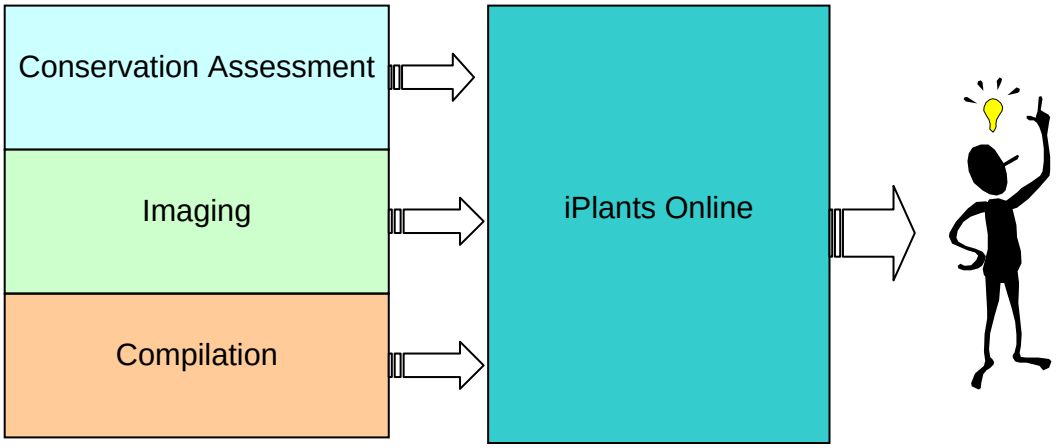
The following issues are raised in this document and have yet to be addressed.

1. When are TDWG distribution maps created?
2. Do we wish to provide online conservation assessment for end-users?

# 2. Components

## 2.1. Overview

The iPlants system consists of four main areas: three workstreams each feeding data into the online system.



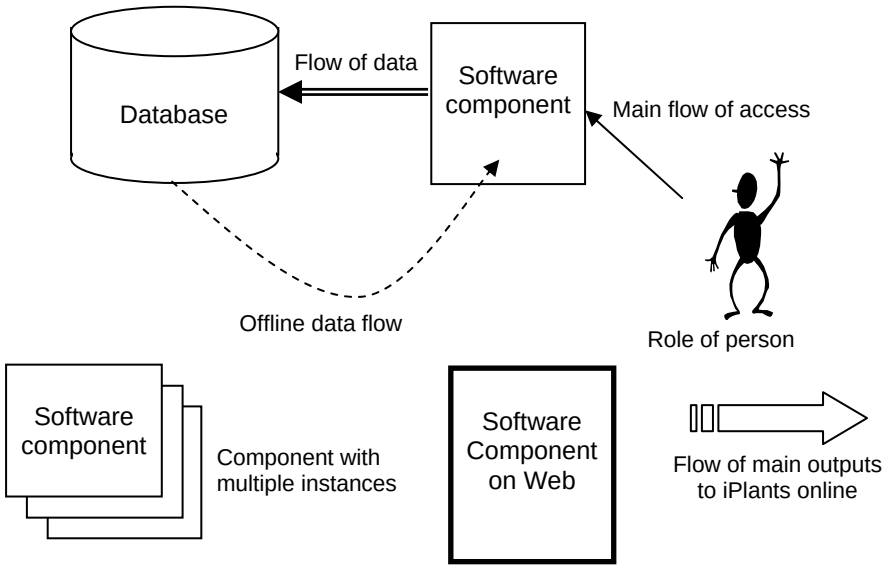

Conservation Assesment is a three-stage process.

In addition to the above, there is a need for systems to help manage the project. Each of these areas is described in more detail in the following sections.

## 2.2. Key to Diagrams

Throughout this document, diagrams are used to provide overviews. See the key below for an explanation of the symbols used. Note that the diagrams are intended to show the major logical components of the system, and therefore many smaller items have been omitted.

Boundary between iPlants system and elsewhere

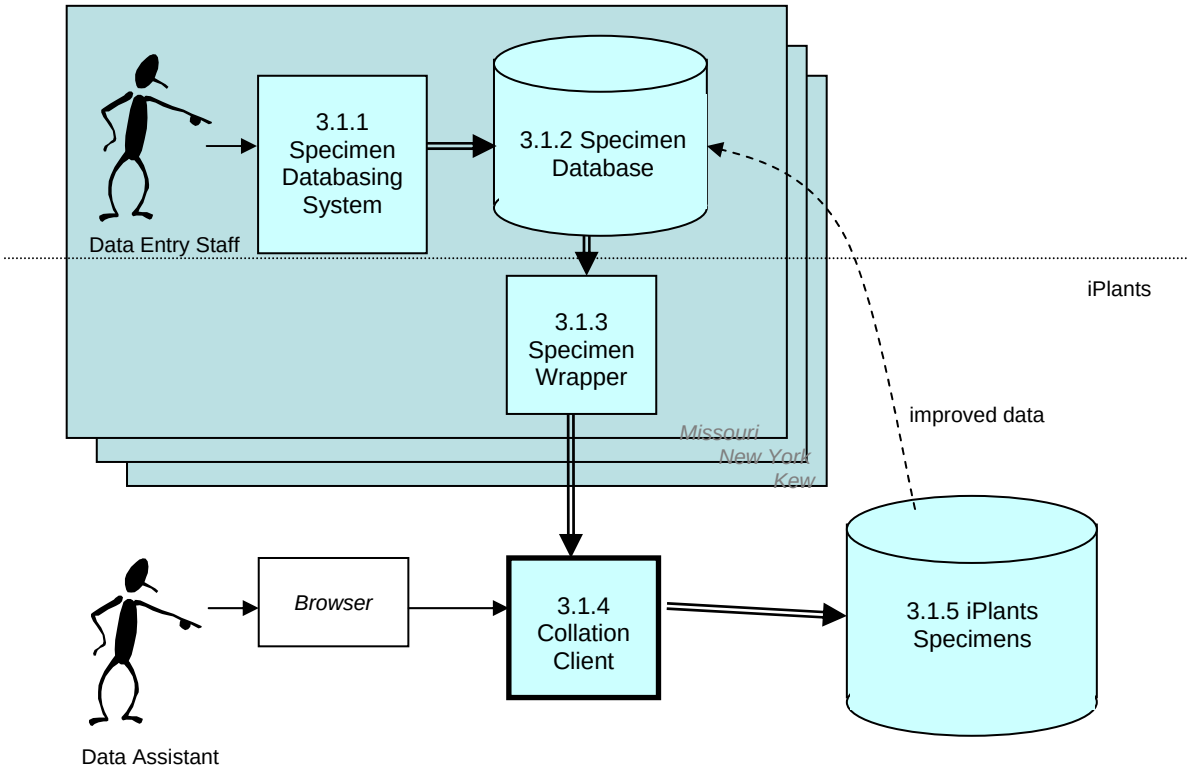


# 3. Conservation Assessment

To produce conservation assessments, several steps are undertaken. Firstly, specimens are databased within holding institutions using institutional software. Secondly, these specimens are exported to a shared web-accessible repository, validated and then georeferenced. Thirdly, the georeferenced specimens are used to produce a preliminary conservation rating. The georefrenecing and conservation assessments could take place by downloading of specimen data to local GIS systems, but it would be more flexible for the project if instead web services were developed. These web services could also be made available to audiences external to the project.

The software components which will be required are described separately below.

## 3.1. Specimen Databasing




### 3.1.1. Specimen Databasing System

**Purpose**: To facilitate entry of data into Specimen Database.
**Development Status**: None required. Already exist in each institution.
**Deployment**: One per institution.
**Users & access**: Data entry staff within each institution.
**Data**: Specimen data according to institutional standards (see Specimen Database).
**Functions**: Data entry.
**Technology notes**: Varying technology has been used - see table below

| Site | System Name | Type of System | Software |
|---|---|---|---|

| | | | |
|---|---|---|---|
| Missouri | TROPICOS | Custom-made, in-house | C# and Basic |
| New York | NYVH | Package | KE EMU |
| Kew | HerbCat | Custom-made, in-house | Visual Basic |
| Others | Various | | |

### 3.1.2. Specimen Database

**Purpose**: To store specimen data within institutions.
**Development Status**: None required. Already exist in each institution.
**Deployment:** One per institution.
**Users & access**: Data entry staff within each institution.
**Data**: Specimen data according to institutional standards. Varying numbers of specimens, ranging from low hundreds of thousands to low millions.
**Technology notes**: Varying technology has been used - see table below

| **Site** | **System Name** | **Type of System** | **DBMS** |
|---|---|---|---|
| Missouri | TROPICOS | Custom-made, in-house | SQL Server, Pick/D3 |
| New York | NYVH | Package | KE EMU |
| Kew | HerbCat | Custom-made, in-house | Sybase |
| Others | Various | | |

### 3.1.3. Specimen Wrapper

**Purpose**: To export data from Specimen Database in iPlants common format
**Development Status**: To be developed. Potential building blocks are DiGIR and/or BioCASE protocols; Darwin Core and BioCASE data standards. However, some (possibly considerable) further development is likely to be necessary.
**Deployment:** One per institution.
**Users and access:** Collation Client communicating over the Internet.
**Data**: Specimen data according to iPlants Common Specimen Format.
**Functions**: Data export. Needs to filter out any data not intended for iPlants use (including cultivated specimens, specimens identified only to genus level, specimens with doubtful identifications, and hybrids).

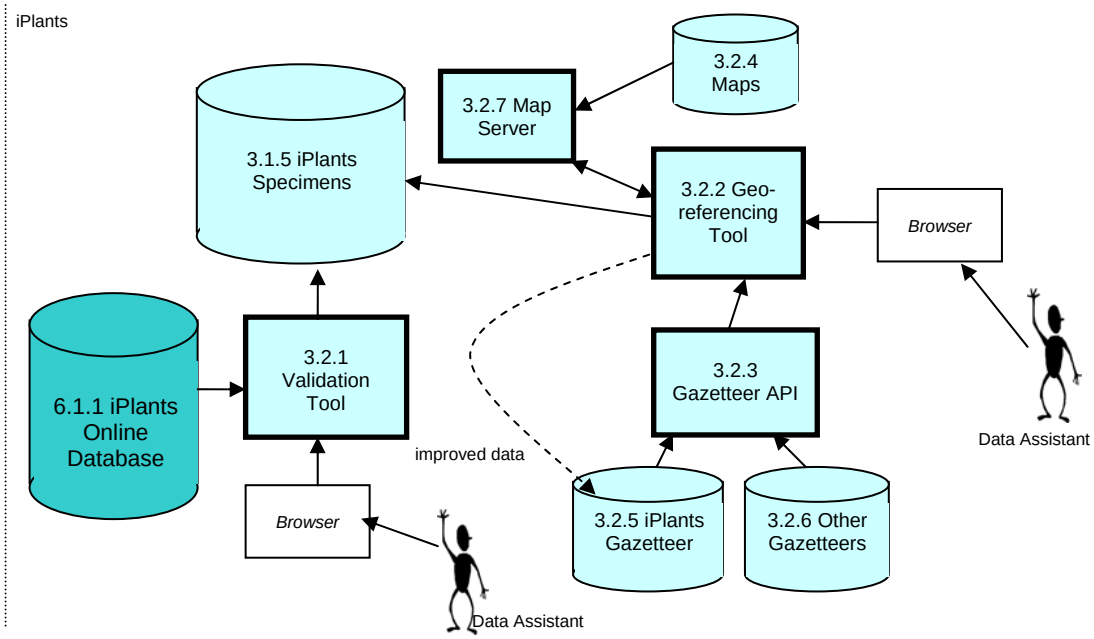
### 3.1.4. Collation Client

**Purpose**: To collate specimen data into iPlants Specimens database.
**Development Status**: To be developed. Potential building blocks are DiGIR and/or BioCASE protocols; Darwin Core and BioCASE data standards. However, some (possibly considerable) further development is likely to be necessary.
**Deployment:** Web-based shared resource. Preferably co-located with iPlants Specimens database.
**Users & access**: iPlants Data Managers.
**Data**: Specimen data (see iPlants Specimens).
**Functions**: Import of data from institutional Specimen Databases into iPlants Specimens database. Should allow selection of target databases (one or more) and taxa (at family and/or generic level, multiple terms allowed).

### 3.1.5. iPlants Specimens

**Purpose**: To hold iPlants specimen records.
**Development Status**: To be developed.
**Deployment:** Web-based shared resource.
**Users & access**: Data Assistants using various software clients over the Internet.
**Data**: A few million specimen records. The specimen data held is according to iPlants Common Specimen Format, and in addition flags to identify duplicate specimens. Also c80,000 species conservation assessments and distribution maps.

## 3.2. Georeferencing




### 3.2.1. Validation Tool

**Purpose**: To validate records, group synonyms, and flag duplicates in the iPlants Specimen database.
**Development Status**: To be developed.
**Deployment:** Web-based shared resource. Preferably co-located with iPlants Specimens database.
**Users & access**: Data Assistants using Browsers over the Internet.
**Data**: Specimen data according to iPlants Common Specimen Format.
**Functions**:
a) Validation of data which will be used during georeferencing and conservation assessment
b) Using the iPlants Online Database, grouping of names into species
c) Identification of duplicate records based on collector name and number.
**Technology notes**: As far as possible, should be an automated process.
**Issue**: It is essential to group names into species before conservation assessment is carried out, and preferable before georeferencing. To do this will require access to the checklist - or at least a table derived from it.

### 3.2.2. Georeferencing Tool

**Purpose**: To apply accurate georeferences to records in the iPlants Specimen database.

**Development Status**: To be developed. Potential building blocks include in-house tools, Biogeomancer.
**Deployment:** Web-based shared resource. Preferably co-located with iPlants Specimens database, and if possible with Gazetteer API.
**Users & access**: Data Assistants using Browsers over the Internet.
**Data**: Specimen data according to iPlants Common Specimen Format, gazetteer and maps.
**Functions**: Access of specimen data in iPlants Specimen database and submission of location strings to Gazetteer API, allowing for specification of geographical filter at regional level. Validation and presentation of results to user, and subsequent application of georeference data to specimen. If new georeference values are entered, return these to iPlants Gazetteer.
**Issue**: When should the distribution map be saved - at this stage or during the Conservation Assessment?
**Issue**: If throughput makes performance sluggish, or management of maps and other issues becomes awkward, then it may be advisable to duplicate this tool plus the Map Server and Maps within institutions.

### 3.2.3. Gazetteer API

**Purpose**: To propose georeferences gleaned from gazetteers, given location and collector data on specimen.
**Development Status**: To be developed. Potential building blocks include in-house tools, Biogeomancer.
**Deployment:** Web-based shared resource. Preferably co-located with iPlants Gazetteer.
**Users & access**: Data Assistants using Georeferencing tool, and end-users using Gazetteer Query Interface, both over the Internet.
**Data**: Specimen location string.
**Functions**: Accept a location string, parse it and check against iPlants Gazetteer and Other Gazetteers. It should be capable of filtering search and results by geographical region. Return potential georeferences (also taking into account offsets from fixed points) with associated validating information such as collector and dates. Submit data as sets of one or more values.

### 3.2.4. Maps

**Purpose**: To provide visual feedback on georeferences and present distributions.
**Development Status**: Some to be acquired.
**Deployment:** Co-located with Map Server.
**Users & access**: Map Server.
**Data**: Digitised maps.
**Functions**: N/A.

### 3.2.5. iPlants Gazetteer

**Purpose**: To aid accurate georeferencing of iPlants Specimens.
**Development Status**: To be developed.
**Deployment:** Web-based shared resource. Preferably co-located with iPlants Gazetteer API.
**Users & access**: iPlants Gazetteer API, and data entry from Georeferencing Tool.
**Data**: Locations, collector information, georeference data.

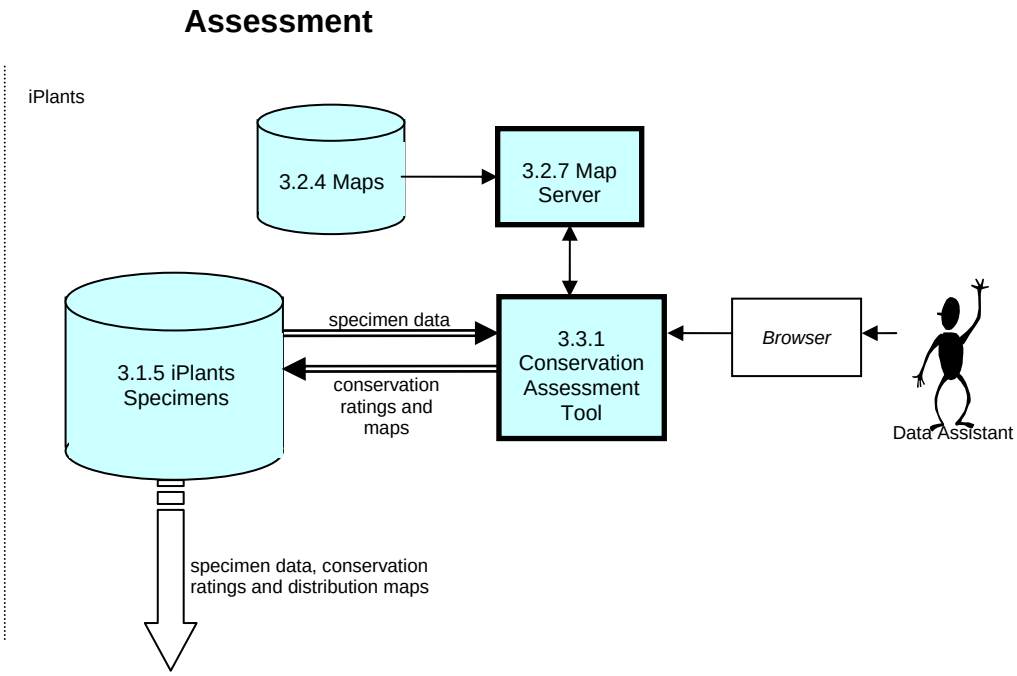
### 3.2.6. Other Gazetteers

**Purpose**: To aide accurate georeferencing of iPlants Specimens.
**Development Status**: Already exist. A list of potentially useful databases can be found at http:/www.kew.org/gis/links/gaz.html. This includes general Gazetteer resources such as the Alexandria Digital Library Gazetteer and the Getty Thesaurus of Geographic Names, as well as specialized botanical resources.

**Deployment:** Web-based shared resources.
**Users & access**: iPlants Gazetteer API, and data entry from Georeferencing Tool.
**Data**: Minimum of locations and georeference data.

### 3.2.7. Map Server

**Purpose**: To create maps as image files. Used
a) by Georeferencing Tool to visualise specimen distributions
b) by Conservation Assessment Tool to visualise calculations and create species distribution maps
c) by Content Generators to create regional distribution maps for all taxa
**Development Status**: Refers to toolset - need to develop routines within it.
**Deployment:** Web-based shared resource. Co-located with Maps, and preferably Georeferencing and Conservation Assessment Tools.
**Users & access**: Georeferencing Tool, Conservation Asessment Tool and Content Generators.
**Data**: Point data (specimen georeferences), polygons (TDWG regions) and Maps.
**Functions**: To layer point or polygon data with maps and output image files.
**Technology notes**: e.g. ARC/IMS

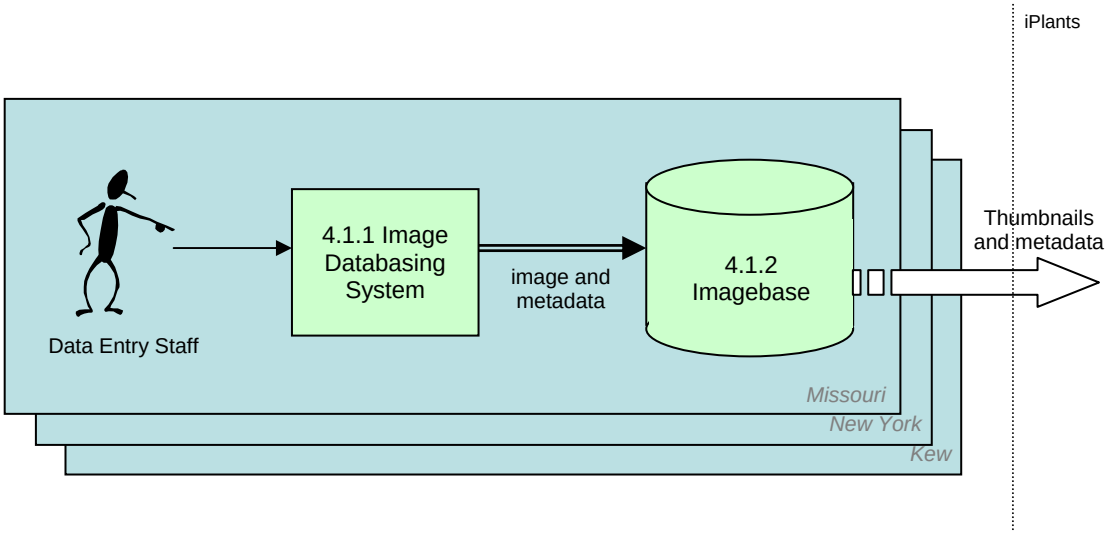



### 3.2.8. Conservation Assessment Tool

**Purpose**: To calculate preliminary conservation ratings based on specimen distributions.
**Development Status**: To be developed. Some research has been done on suitable methodologies and algorithms.
**Deployment:** Web-based shared resource. Preferably co-located with iPlants Specimens database.
**Users & access**: Data Assistants using Browser over the Internet.
**Data**: Georeferenced specimen data, but returns conservation rating.
**Functions**: Takes a set of georeferenced specimen records and calculates conservation rating and intermediate values (AOO, EOO and subpopulation figures). Uses routines developed within GIS.
**Issues**: When should the distribution map be saved - at this stage or during the Conservation Assessment? If end-users were to have access to the tool then it would have to be split into a lower-level API upon which two clients could be built - one for iPlants staff, and one for end-users. This would mirror the gazetteer tool setup. However, since output without visualisation onto a map is unlikely to be acceptable, it would also require the presence of a Map Server with Maps online.

# 4. Imaging

The iPlants website will provide access to images of species. These images are digitised and made available on the Internet using imaging systems already in place in the various institutions.




## 4.1.1. Image Databasing System

**Purpose**: To facilitate entry of image and metadata into institutional Imagebase.
**Development Status**: None required. Already exist in each institution.
**Deployment:** One per institution.
**Users & access**: Data entry staff within each institution.
**Data**: Digital images and metadata including at minimum scientific name according to institutional standards
**Functions**: Image and metadata entry.
**Technology notes**: Varying technology has been used - see table below

| Site | System Name | Type of System | Software |
|---|---|---|---|
| Missouri | TROPICOS, Mr SID | In-house, Package | Perl, Mr SID |
| New York | NYVH | Package & Custom-made, in-house | KE Texpress |
| Kew | Image Server | Package and Custom-made, in-house | TOAD clients (PHP) |
| Others | Various | | |

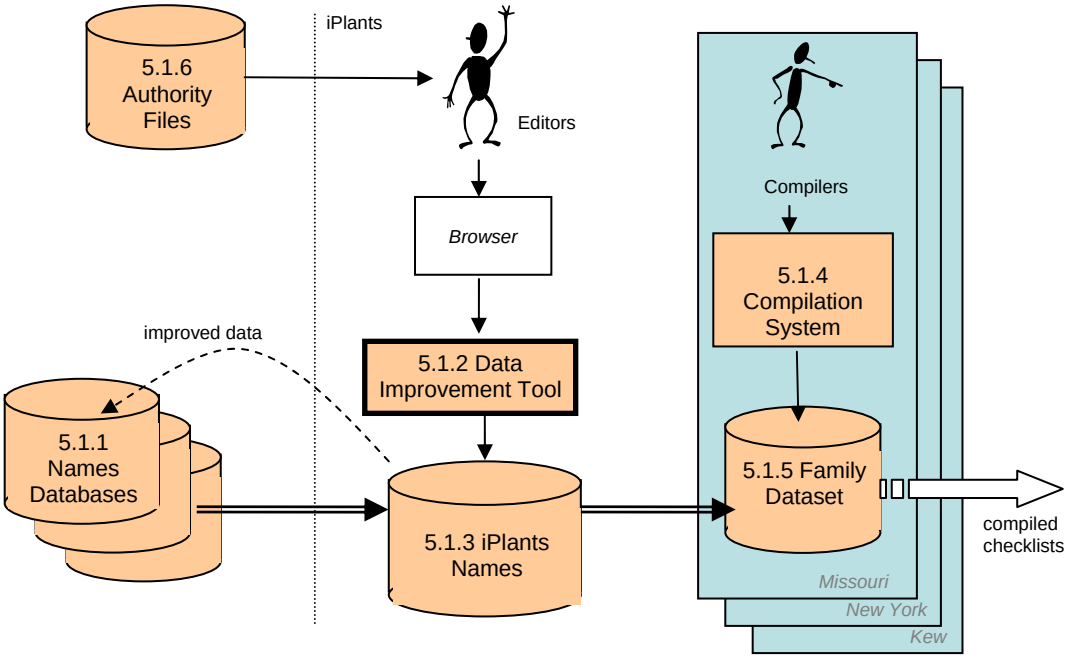
### 4.1.2. Imagebase

**Purpose**: To store digital images and associated metadata.
**Development Status**: None required. Already exist in each institution.
**Deployment:** One per institution.
**Users & access**: Data entry staff within each institution.
**Data**: Digital images and metadata including at minimum scientific name according to institutional standards.
**Functions**: Image and metadata storage.
**Technology notes**: Varying technology has been used - see table below

| Site | System Name | Type of System | DBMS |
|---|---|---|---|
| Missouri | TROPICOS | In-house, Package | SQL Server, Pick/D3 |
| New York | NYVH | Package and Custom-made, in-house | KE Texpress |
| Kew | Image Server | Package | TOAD (MySQL) |
| Others | Various | | |

# 5. Compilation

A core process of iPlants is the production of the list of names with synonymy and distribution. This list will form the backbone of the iPlants online system. The process occurs in two stages. Firstly, name records are collated from various authoritative sources and cleaned up and standardised en masse in an iPlants repository. Secondly, compilers take family recordsets from this repository into a standard compilation system and using the taxonomic literature they apply taxonomic decisions and add species data. The results of this are fed, family by family, into the iPlants website.




## 5.1.1. Names Databases

**Purpose**: Systems storing baseline plant name data.
**Development Status**: None required. Already exist in each institution.
**Deployment:** One per institution.
**Users & access**: One-off copying of data into iPlants Names database. Subsequent feeding back of improved data.
**Data**: Plant names, including place of publication. Several million records taken as a whole (IPNI has 1.5 million names, others will have fewer).
**Technology notes**: Varying technology has been used - see table below

| Site | System Name | Type of System | DBMS |
|---|---|---|---|
| Missouri | TROPICOS | Custom-made, in-house | SQL Server, PICK |
| New York | NYVH | Package & Custom-made, in-house | KE Texpress |
| Kew | HerbCat | Custom-made, in-house | Sybase |
| Others | Various | | |

### 5.1.2. Data Improvement Tool

**Purpose**: To deduplicate and standardise names in iPlants Names.
**Development Status**: To be developed. Some parsing routines exist which could be developed.
**Deployment:** Web-based shared resource. Preferably co-located with iPlants Names.
**Users & access**: Editors using Browser over the Internet.
**Data**: Plant names and place of publication (see iPlants Names).
**Functions**:

a) Formats data from disparate sources into a standard format
b) Allows editor to specify rules for deduplication and parsing.
c) Deduplication of names from multiple Names Databases into new iPlants name record, with links back to source records.
d) Parsing and standardisation of place of publication and author.
e) Other data checks and automated processes.
**Issue**: It would probably be more efficient to incorporate Authority Files into this tool.

### 5.1.3. iPlants Names

**Purpose**: Storage of plant name data.
**Development Status**: To be developed.
**Deployment:** Web-based shared resource.
**Users & access**: Data Improvement Tool.
**Data**: Plant names and place of publication, including duplicates from various Names Databases. Probably 5-10 million records. Contains:
a) Plant names
b) Authors
c) Place of publication
d) If synonym, accepted name (where known)
e) Distribution (where known)
f) Basionym (where known)
g) Notes/remarks
h) Source key
i) iPlants key

### 5.1.4. Compilation System

**Purpose**: The application of taxonomic status and relationships to plant names, and the assembly of distribution data.
**Development Status**: To be developed. A system exists which has been used during the prototype phase but has broader scope. A list of change requests also exists for it. The system could be cut down and further developed. In addition, other potential candidates exist such as the Berlin Model and Tropicos.
**Deployment:** Local to compiler. Co-located with Family Dataset.
**Users & access**: Compilers within each institution, sometimes deployed on laptops and potentially without network access.
**Data**:
A family dataset containing all names at genus and lower level for a given family circumscription (see iPlants Names).
**Functions**:
a) Import of family dataset from iPlants Names database.
a) Assignment of taxonomic status to names.
b) Assignment of relationships between names.
c) Assembly of distribution data.

d) Sourcing of all of the above.
e) Validation of all of the above.
f) Export to iPlants Online Database.

### 5.1.5. Family Dataset

**Purpose**: To store a family dataset during compilation
**Development Status**: To be developed. A database exists which has been used during the prototype phase, and other candidates exist including the Berlin Model and Tropicos.
**Deployment:** Local to compiler. Co-located with Compilation System.
**Users & access**: Compilers within each institution using the Compilation System, sometimes deployed on laptops and potentially without network access.
**Data**: Family dataset of probably no more than 100,000 names at a time. All names at genus and lower level. It will contain:
a) Plant name with standardised author.
b) Place of publication, standardised.
c) Taxonomic status.
d) Linkage to related names (accepted name, basionym).
e) Distribution data (accepted names only).
f) Lifeform (accepted names only)
g) Source references, linked to above.
h) Compiler/reviewer & dates
i) Source keys
j) iPlants key

### 5.1.6. Authority files

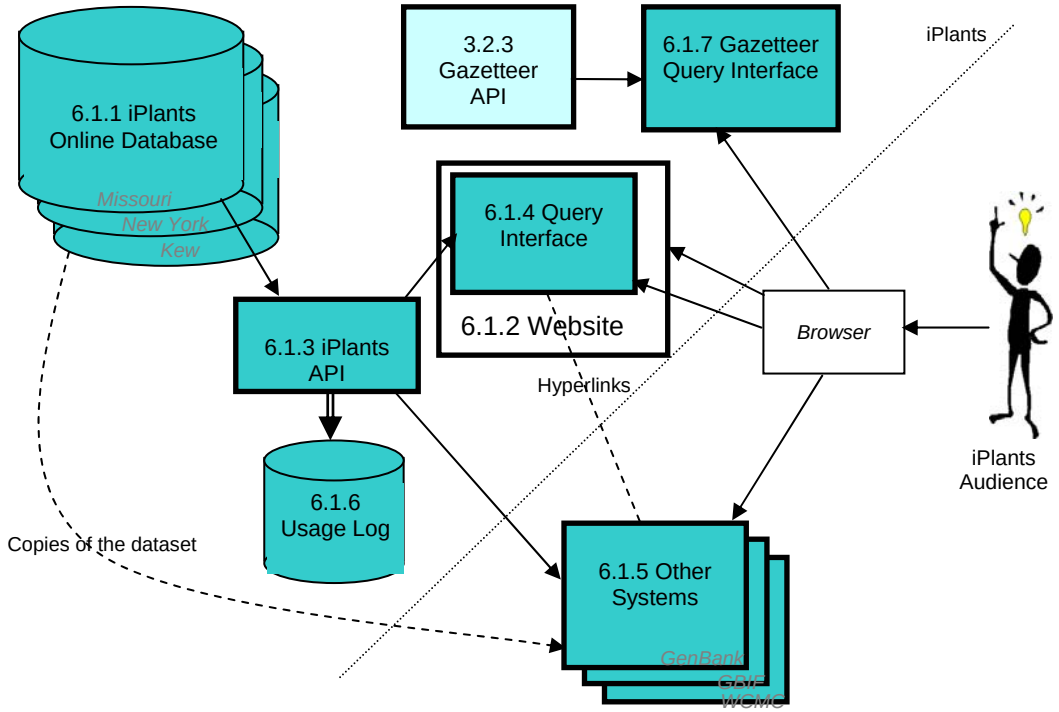
**Purpose**: Provides list of known plant name authors, and publication titles, and standard abbreviated terms which may be used as labels.
**Development Status**: Variable (see table below).
**Deployment:** Currently web-based shared resource with human-computer interface.
**Users & access**: Data entry staff using Browser over the Internet.
**Data**: Each source has a number of references not likely to exceed a hundred thousand or so.

| Authority file | Coverage | Availability |
|---|---|---|
| Authors of Plant Names | Authors | Online at Kew |
| Taxonomic Literature vol.2 (TL/2) | Books | Online, subscription basis |
| Botanico-Periodicum Huntianum (BPH) | Periodicals | Reputedly available electronically |

# 6. iPlants Online

The iPlants website makes use of an API to both underly the web interface which many people will use, and to provide automated access so that others can build iPlants web services seamlessly into their systems. Specialised interfaces will be required for the editors who keep the data up to date, the reviewers who are asked to check the quality of the data, and the system managers whose task it is to maintain the system. Web crawlers will maintain metadata about image resources and hyperlinks to other systems, so that the end-user is presented with context-sensitive links which will always resolve to an answer.

## 6.1. End-user services




### 6.1.1. iPlants Online Database

**Purpose**: The main iPlants data repository. Stores all of the data which will be made available on the iPlants website.
**Development Status**: To be developed.
**Deployment:** Web-based shared resource, replicated to ensure load-sharing and robustness.
**Users & access**:
**Data**: a few million names linked together into c400,000 species records, plus a million or so specimen records. Several release versions of the database will be required, and unknown amounts of feedback per species.

For each name there will be recorded
a) Plant name with standardised author.
b) Place of publication, standardised.
c) Taxonomic status.

d) Linkage to related names (accepted name, basionym).
e) Source references, linked to above.
f) Compiler/reviewer & dates
g) Source keys
h) iPlants key

For each species, there will be recorded
a) Distribution
b) Lifeform
c) where available, IUCN conservation evaluation
d) where assessed, iPlants Conservation Assessment and intermediate values (AOO, EOO, subpopulations)
e) Distribution map (TDWG region-based)
f) A set of 1 or more (usually 1) thumbnail images
g) A set of hyperlinks
h) Review status
i) A history of feedback

For each species for which a conservation assessment has been run, there will also be
a) Distribution map (specimen-based)
b) A time-stamped set of specimen records, probably less than a dozen, each storing the iPlants Common Specimen Format data fields.

### 6.1.2. Website

**Purpose**: Public-facing website giving information about the iPlants project and access to the Query Interface.
**Development Status**: To be further developed. See the iPlants Web Prototype which has already been developed through several iterations.
**Deployment:** Web-based shared resource, replicated to ensure load-sharing and robustness.
**Users & access**: Anyone using a browser over the Internet.
**Data**: Various textual pages.
**Functions**: See the iPlants prototype for details, but in essence:
a) Provision of information about iPlants.
b) Access to Query Interface, email contacts and other services.

### 6.1.3. iPlants API

**Purpose**: To organise and facilitate external access to iPlants data by humans and other software systems.
**Development Status**: To be developed. We may be able to learn some lessons from DiGIR and BioCASE, although these are tailored for distributed resources.
**Deployment:** Web-based shared resource, replicated to ensure load-sharing and robustness.
**Users & access**: Query Interface, and any number of other systems and programs accessing it over the Internet.
**Data**: See iPlants Online Database.
**Functions**: See the iPlants Web Prototype for a visual presentation of much of the required functionality. It will specify a protocol for handling queries, i.e. a definition of the calls which the Checklist database should be able to receive and of the data returned as a result. At minimum, querying by plant name, distribution and conservation status will be catered for, as well as submission of feedback. Also handles passing back of results.

### 6.1.4. Query Interface

**Purpose**: Enables querying of the iPlants Online Database suing the underlying iPlants API.

**Development Status**: To be developed.
**Deployment:** Web-based shared resource, replicated to ensure load-sharing and robustness.
**Users & access**: Anyone using a browser over the Internet.
**Data**: See iPlants Online Database
**Functions**:
See the iPlants Web Prototype.
a) Query names by name, geography and/or conservation status.
b) Output results to a series of togglable formats, e.g. list of names, list of species, tabular format. Also perhaps allow selection of output content by data subject (e.g. distribution, synonymy, conservation data).
c) Output formats to screen, file download in multiple formats, printable format (e.g. PDF).
d) Show distribution maps, thumbnail images, and specimen data where available.
e) Option to view fullscale image where available.
d) Links to other systems (e.g. GENBANK, WCMC)

#### 6.1.5. Other Systems

**Purpose**: Other systems which iPlants will wish to link to, and be linked to.
**Development Status**: In general, none required. In some cases, where significant benefit for end-users can be acheived, it may be appropriate to deploy collaboration funds to enable closer links.
**Deployment:** Varying but mainly deployed to Internet.
**Users & access**: Varying, but generally web-based and open access, so anyone using a Browser over the Internet.
**Data**: Varying, but must include a plant name.
**Functions**: Varying.

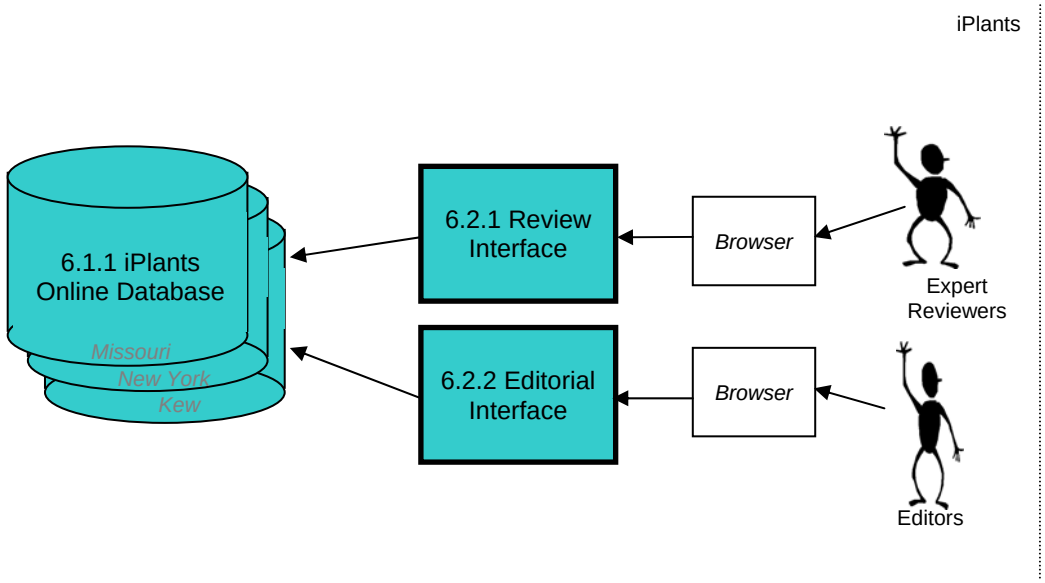
#### 6.1.6. Usage Log

**Purpose**: Record of mainly end-user access to iPlants Online Database. Useful to monitor usage and identify problems.
**Development Status**: To be developed.
**Deployment:** Preferably co-located with iPlants Online Database, replicated to ensure load-sharing and robustness.
**Users & access**: Editors.
**Data**: Information available from Browser sessions and iPlants API, e.g. IP address of user, dateand time, queries submitted, response times, results returned, and errors. Logs grow quite quickly and a strategy of recycling and archiving will be necessary.

#### 6.1.7. Gazetteer Query Interface

**Purpose**: Enables interaction by end-users with the iPlants Gazetteer API.
**Development Status**: None required. Already exist in each institution.
**Deployment:** Web-based shared resource. Preferably co-located with the Gazetteer API.
**Users & access**: Anyone using a Browser over the Internet.
**Data**: See Gazetteer API.
**Functions**: Submission of location strings and return of georeferences.

### 6.2. Editorial



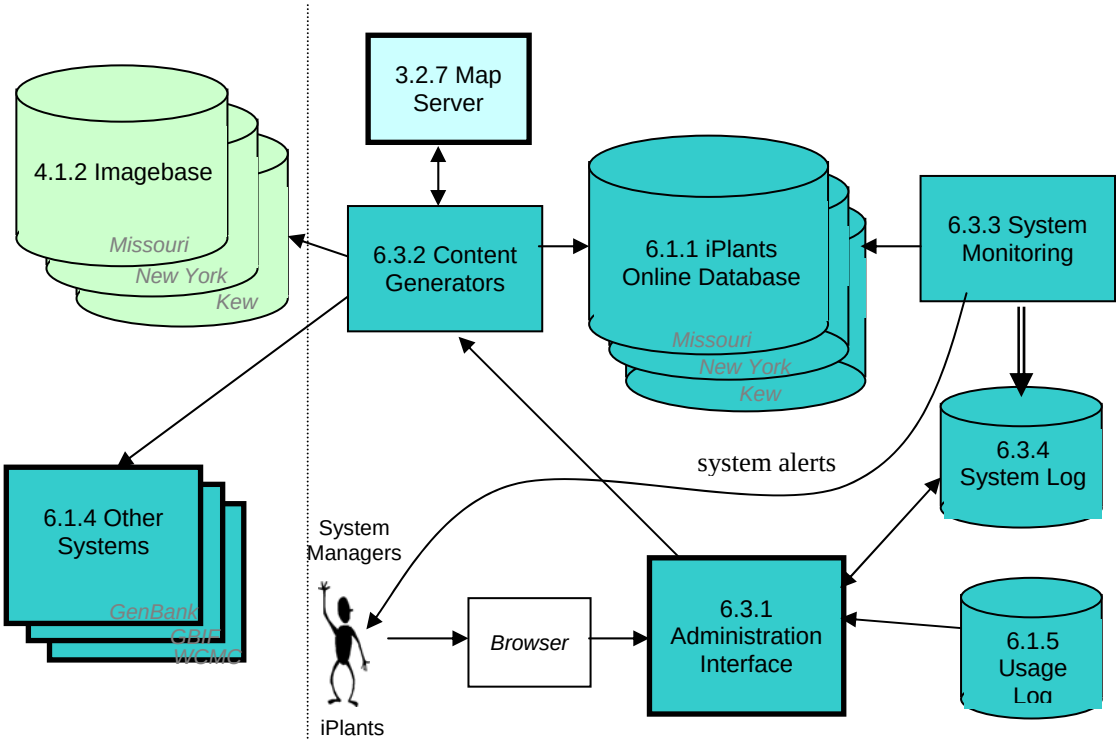

### 6.2.1. Review Interface

**Purpose**: Enables expert reviewers to review datasets within the iPlants Online Database.
**Development Status**: To be developed.
**Deployment:** Web-based shared resource.
**Users & access**: Expert reviewers using a browser over the Internet.
**Data**: All data in iPlants Online Database.
**Functions**:
a) To view and print a dataset 'checked-out' for review.
b) To record agreement or disagreement with individual names or groups of names.
c) To record comments to back up disagreements.
d) To indicate completion of review.

### 6.2.2. Editorial Interface

**Purpose**: Enables editors to manage data within the iPlants Online Database.
**Development Status**: To be developed.
**Deployment:** Web-based shared resource.
**Users & access**: Editors using a browser over the Internet.
**Data**: All data in iPlants Online Database.
**Functions**:
a) To add a new name.
b) To edit an existing name, e.g. by adding new distribution records.
c) To read and review feedback and expert review comments.
d) To reject and/or delete feedback.
e) To respond to feedback.
f) To create and administer users, roles and passwords (e.g. expert reviewers).
g) To set access restrictions. For example, to allow or deny query access to lists under review.

## 6.3. System Management




### 6.3.1. Administration Interface

**Purpose**: Enables system managers to oversee the correct technical functioning of the iPlants online system.
**Development Status**: To be developed.
**Deployment:** Web-based shared resource, replicated to ensure robustness. Preferably co-located with the iPlants Online Database, System Log and Usage Log.
**Users & access**: System managers using a browser over the Internet.
**Data**: All data in iPlants Online Database.
**Functions**:
a) To produce statistics on usage and data content.
b) To check on operational status of online components.
c) To report and perhaps diagnose errors recorded by the system.
d) To manipulate configuration parameters, and activate or deactivate components and/or sites.
e) To activate and/or schedule tasks, e.g. ad-hoc backups, indexing, etc.

### 6.3.2. Content Generators

**Purpose**: To visit Internet sites and refresh knowledge of data content which will be accessed by iPlants. Potentially a series of separate components.
**Development Status**: To be developed.
**Deployment:** Co-located with iPlants Online Database
**Users & access**: Automated, but capable of initiation by system managers.
**Data**: Image thumbnails, Hyperlinks, metadata caches.
**Functions**:
a) Web crawler visits imagebases from which iPlants gets images, updates cache of species for which site has images, and refreshes thumbnails and metadata for all plant taxa.
b) Web crawler visits sites to which iPlants links, and updates cache of species for which site holds data, and hyperlinks necessary to access that data.

**Technology notes**: Should be a largely automated process.

### 6.3.3. System Monitoring

**Purpose**: To monitor and log status of iPlants online system components, and react accordingly.
**Development Status**: To be developed.
**Deployment:** Co-located with iPlants Online Database, replicated to ensure load-sharing and robustness.
**Users & access**: Automated.
**Data**: See System Log.
**Functions**: Receiving errors and status checks and storing them in the System Log. Periodically checking status of components and storing results in System Log. Monitoring performance and taking (preferably preventative) action and/or notifying system administrators if major problems develop.
**Technology notes**: An automated process running continually.

### 6.3.4. System Log

**Purpose**: To store data about the functioning of the online components of the iPlants system.
**Development Status**: To be developed.
**Deployment:** Co-located with iPlants Online Database, replicated to ensure load-sharing and robustness.
**Users & access**: System Monitoring system.
**Data**: Errors and status checks output by system components. Downtime of system components. Configuration changes and other actions carried out using the Administration Interface. Logs grow quite quickly and a strategy of recycling and archiving will be necessary.

# 7. Project Management

During the Prototype Phase our hypothesis that there would be a need for systems to help manage the project was confirmed. A starting point for these requirement is documented under Group Documents / Project Management Documents.

## 7.1. Intranet

One of the successes of the iPlants has been the iPlants Intranet or online web-based project portal. This has proven to be a useful mechanism for

a) hosting and sharing project documents
b) exchanging datasets
c) storing project tasks and software change lists
d) maintaining a project calendar
e) listing project staff with contact details

We did not use the discussion lists feature as this was accomplished more directly using email, but we may wish to use this in future in order to create and maintain a project archive.

There will be a continuing need for this facility.

## 7.2. Project Management Software

In the Production Phase, the Project Manager, IT Manager and Site Managers are likely to need some form of project planning software to plan and manage their activities. Standard requirements are to

a) identify tasks
b) identify dependencies between tasks
c) agree target dates for tasks
d) assign tasks to people

The Task Tracker within the iPlants Intranet was easy to access and use, but lacks some more advanced features such as the ability to link tasks (item b above), and so could not be used in itself to create a schedule. It also did not have a visualisation tool and so couldn't, for instance, draw a Gantt chart. Therefore, it is likely that there will be a need for a more advanced project management tool, e.g. Microsoft Project. However, it is noted that these tools not to be well-suited to presenting or sharing information on the Web.

## 7.3. Project Master List

The iPlants Production Phase project will also be characterised by a series of processes which are being carried out simultaneously on multiple taxonomic datasets. At any one time, for instance, there will be several databasing strands under way, one or more in each institution, as different families are tackled. Simultaneously there may be several imaging initiatives, and several families being compiled. It may be possible to keep an overview of these as separate tasks by using the Intranet or Project Management Software described above, but it may be more intuitive to maintain instead what has been termed a 'Master Switchboard' showing, for each family

a) its current status (i.e. the process it is currently undergoing and where)

b) who it was assigned to and when
c) state of progress
d) estimated date of completion

If maintained actively, this could avoid the need to constantly communicate and check progress, especially where processes must await completion of previous processes. It would also provide a useful record of progress, and if allied to volume figures, can provide the raw material for future throughput estimates.

There have also been some indications that there may be a need for recording management information at the genus, or even species level. For instance, a list of family circumscriptions couched as genera will be necessary to ensure that all names are ultimately dealt with; and since specimen counts are to take place for each name, it makes sense to store these alongside the names for future reference. This may lead to there being a store of 'administrative' data alongside each name in the checklist.

Whatever the requirements of the project, this software will almost certainly need to be developed specifically for the project.

# 8. Types of Component

## 8.1. Shared Production Systems

The iPlants production process will create and use two significant shared databases, the iPlants Specimens and iPlants Names, and their associated data manipulation interfaces. These groupings have the following characteristics

- usage will be modest and restricted to iPlants staff
- access will usually be during working hours
- access may be required from several locations, and so may be considered Internet-based (preferably, for maximum flexibility, using standard Web Browsers)
- limited simultaneous usage may occur
- modest downtime or sluggishness can probably be tolerated
- data storage will be needed for a few million records

## 8.2. Compilation System

Compilation will take place using a Compilation System devised for that purpose. It has the following characteristics

- the system will be single-use
- access will usually be during working hours
- access is required locally, and potentially from locations which are not networked or accessible to the Internet
- modest downtime or sluggishness can probably be tolerated
- data storage will be required for less (usually considerably less) than 100,000 records

## 8.3. Online System

The iPlants website will offer a variety of interfaces to the iPlants Online Database. The core function of iPlants is to provide query facilities to an Internet audience. An API will be provided which will enable flexible automated query access to the database by software systems, and in addition there will of course also be a web interface (which may or may not also use the same API) for use by humans. The following characteristics are expected

- usage will be high and unpredictable, especially where automated access is concerned
- access will be around the clock
- access will be required from anywhere with an Internet connection (and should be achieved using only a standard Web Browser)
- high simultaneous usage can be expected
- downtime cannot be tolerated, and response times must be short
- data storage will be needed for in the region of ten million records

Several other interfaces are also needed to enable management and maintenance of the system. These will probably have similar characteristics to the Shared Production Systems.

## 8.4. Gazetteer

The gazetteer is another example of a Shared Production System, but in addition will also offer a public interface. The characteristics of the latter can therefore be expected to be

- usage will be middling
- access will be around the clock
- access will be required from anywhere with an Internet connection (and should be achieved using only a standard Web Browser)
- some simultaneous usage can be expected
- downtime should not be tolerated, and response times should be reasonable
- data storage will be needed for several hundreds of thousands of records

### 8.5. Map Server

The map server has the specialised task of creating the various maps that the system needs. It will have characteristics similar to other Shared Production Systems.

# 9. Approach to Production Phase Development

## 9.1. Methodology

The Information Systems Manager is responsible for delivering the project's information systems. He/she will liaise with an information technology (IT) panel comprising representatives from the Consortium and report to the Steering Committee.

The software will be developed according to Kew's in-house methodology, which is based upon the Dynamic Systems Design Methodology (DSDM – see www.dsdm.org). This development method is based on the following principles:

- Active user involvement is essential
- The team should be empowered to make decisions
- Software must be delivered frequently
- Fitness for business purpose is the criteria for success
- Development should be iterative and incremental
- All changes are reversible
- High level requirements are baselined
- Testing is integrated throughout the development lifecycle
- Collaboration & co-operation between developers and users is essential

Development will be broken down into modules, timeboxing and prototyping will be utilized, and feedback will be frequently sought. Defined software development tasks will be assigned to each programmer, but the entire IT team will function as one through regular team meetings, peer review of design specifications, cross-testing of software and so-on. Development work will take place on a shared server so that all code is accessible to the entire team. Active involvement from users and the IT panel will be required.

Since the project is a joint effort of three geographically separated institutions and various potential collaborators, open communication will be essential to the project's success. An email list for the IT development will be established, with open membership from across the Consortium. Full use will be made of an iPlants Intranet site to store project documents (including Communications, Risk, Configuration Management, and Quality Assurance) which will enable all collaborators to have access to common information. Feedback will be encouraged.

Quality will be attained by documented testing against design specifications throughout the development lifecycle. All source code will be managed through a system configuration tool that maintains versions and allows rollback. Documentation will be produced to enable hand-over of development tasks and the installation or configuration of delivered components. A procedure for final acceptance of the systems by the Steering Committee will be agreed with them.

## 9.2. Re-use of Existing Work

Before developing a component, a check will be conducted for existing examples to see whether any of these might be suitable for adoption, enhancement or simply as a source of ideas. We know of several areas where this might be so, and these are annotated in this document under each component. For example, during the prototype phase we enhanced an existing Kew compilation tool, made use of DiGIR software and based our specimen exchange standard on the Darwin Core.

## 9.3. Data Standards

iPlants will use established Data Standards wherever they can be applied. Several useful standards have been endorsed by TDWG, and we also intend to take careful notice of solutions adopted by international bodies like GBIF. We will use the following

- Authors of Plant Names
- Botanico-periodicum-huntianum
- Taxonomic Literature edition 2
- World geographical scheme for recording plant distributions
- IUCN Red List Categories for Conservation Ratings
- International Code of Botanical Nomenclature

# 10. Technology Choices

## 10.1. Industry-Standard Technology

The iPlants system will be built using widely available and used, industry-standard software. Niche and 'bleeding-edge' products will be avoided. The expected benefits to iPlants include

- tried and tested software is more likely to contain awide feature set, and less likely to contain bugs
- appropriate skills and support services will be more widely available
- a wide userbase encourages active product development and the provision of suitable upgrade paths

At the time of writing, technologies which meet these criteria include

- operating systems such as Windows 200/2003 and Unix
- relational databases such as Oracle, Sybase, SQL Server, MySQL, MS Access
- programming languages such as Java, C#, Visual Basic and PHP
- web servers such as Apache and IIS
- Servlet technology such as JSP and ASP
- HTML and XML

## 10.2. Preferred Technology

### 10.2.1. Familiar Technology Set

During the development and deployment phases we expect to use software components already in use by the Kew in-house team, such as Unix, the Apache web server, MySQL, Java/JSP, PHP technology and Microsoft Access. These carry the following advantages

- they offer the range of functionality required
- they are technologies known to Kew, which therefore has in-house expertise
- they are low-cost
- they offer a high degree of platform independence (see below)

One possible exception to the above is the deployment of the Online System. See below for details.

### 10.2.2. Platform Independence

Platform independence means that software components will be as little tied to specific underlying software layers as possible. Platform independence gives iPlants the flexibility to swap underlying platform (e.g. when it is phased out by the manufacturer, or when an alternative platform shows significant benefits) without impacting unduly on higher components.

Platform independence also confers the freedom to deploy components to a variety of platforms, so as to facilitate local solutions and choice of platform, and to enable wider uptake. This is important for iPlants, as each of the partner institutes has a different installed IT infrastructure, and cannot lightly take on the support of new and unfamiliar IT platforms, even if the initial costs of doing so were fully funded.

Unix is ubiquitous on hardware systems, and all of the other software runs on both Unix and Windows (and other systems). MySQL uses SQL and ODBC-compliant and can thus be swapped for another relational database should the need arise. Microsoft Access will only run on Windows, but since this is the only likely deployment environment this is not seen as a problem.

#### 10.2.3. Shared Production Systems

For the Shared Production Systems we would deploy MySQL databases and Java/JSP servlets running on Unix with Apache web servers. These systems are proven and capable of supporting the level of use we expect them to receive. They could probably be located on a single server with Internet access.

#### 10.2.4. Compilation System

For the Compilation system we would expect to use Microsoft Access, a powerful and flexible environment for PC users. It provides both a useable database and a rich interface with useful editing capabilities out of the box, together with the ability to program a sophisticated data entry system. It is also has good connectivity and integration with other Microsoft Office components.

#### 10.2.5. Online System

It is the expectation of iPlants that the Online System will attract large-scale usage, especially through automated access, and as an authority system it must exhibit robustness and be available on a 24x7 basis. During development we expect to use the same tools as for the Shared Production Systems. However, we will also need to review available technologies and strategies for their ability to handle the expected throughput. and provide the required robustness. If necessary, we will migrate the systems to platforms tailored for this type of operation.

At this stage, we have several strategies for dealing with these issues.

Firstly, we expect that usage will start relatively slowly and gradually scale up as the iPlants dataset is added to, knowledge of iPlants spreads and other systems build in links. We do not expect initially to hit levels of demand or usage beyond the capabilities of our existing technologies.

Secondly, we will profile likely demand, and compare with figures from other biodiversity systems, so that we have a better idea of what to expect.

Thirdly, we will add accessibility features in progressive stages and may well pursue a strategy of 'soft releases', releasing systems without undue publicity so as to allow demand to slowly accrue while we gain experience of usage levels and patterns.

Fourthly, we will design the systems so that scalability and flexibility is allowed for. For example, by avoiding usage of any proprietary features of MySQL and connecting to it through industry-standard interfaces, we should be able to migrate to a higher-throughout relational database if that becomes necessary.

Fifthly, we will plan to replicate the system across multiple sites. Platform independence will be an important consideration here, as we will wish to have as few barriers as possible to establishing new sites. Initially the focus may be on the sites of consortium members, but we

should seek replication sites which give a 24-hour coverage, i.e. that any time of the day there will be one site operating during office hours. The advantages of replication are likely to be

a) that when one site suffers unexpected downtime the others should remain available. Also, individual sites can be taken down for scheduled maintenance without interrupting service
b) it should be possible to route requests away from Internet bottlenecks
c) load sharing will occur
d) provided sites are selected around the globe, there will be fewer 'blindspots' when there are no staff available during normal office hours to attend to system availability and maintenance issues

The iPlants API (and also the iPlants Gazetteer API) will provide Web Services. The eXtensible Markup Language (XML), Simple Object Access Protocol (SOAP), and Web Services Description Layer (WSDL) should prove a useful industry-standard technology for achieving this.

### 10.2.6. Map Server & GIS

All three institutions employ GIS software from ESRI, which offers a complete set of industry-leading GIS-oriented software, so it makes sense to standardise on this. The main components required by iPlants are a Map Server and algorithms to work within the Conservation Assessment Tool. The Map Server is software situated on a central server which can visualise point or polygon distribution data onto maps and supply the results to a browser. ESRI supply ARC/IMS to fulfil this role, and it would be used extensively by project staff during georeferencing and conservation assessment, as well as in the production of regional distribution maps created from the checklist distribution data.

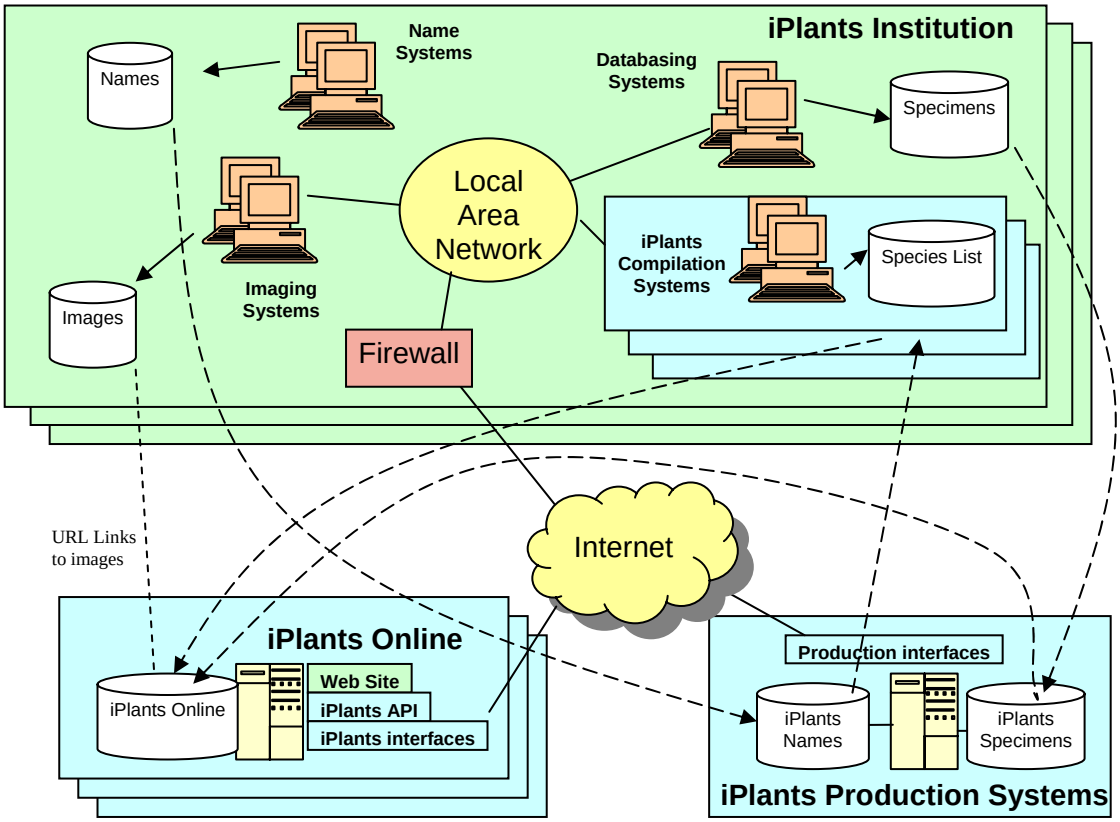
Creating distribution maps is a resource-intensive process. Although it would be possible to deploy ARC/IMS to end-users, and so make distribution maps on demand, better response times are likely to be achieved by creating the maps during the production process and storing them ready-made on the iPlants website as image files.

# 11. System Architecture and Maintenance

## 11.1. System Architecture

Each institution would deploy a variety of existing internal systems to capture specimen, name and image data, as well as using one or more copies of the Compilation System. iPlants staff in each institute would access the iPlants Production Systems and iPlants Online over the Internet. Major data movements are shown as dashed arrows.

The figure below illustrates how the main iPlants system components would be deployed.




## 11.2. Maintenance and Support

During production, IT and IS support would be required at each institute for the following:

a) desktop systems, network services, internet access and image storage for 14-19 project staff
b) the iPlants Compilation System, several copies of which may be located in each institute
c) the Specimen Wrapper which will give access to institutional specimen data
d) potentially, numerous data exports and other ad-hoc tasks and enquiries

In addition, support would be required at Kew for

e) the iPlants Production Systems and Map Server, situated on a web-accessible server
f) the iPlants Online System, initially installed at Kew only

Once the online system is established, there will come a need to replicate it to other sites. Maintenance of each site will probably need to be shared between an iPlants support team with knowledge of the iPlants applications, and local staff. The former would predominantly use remote access to monitor the health of the replicated systems and remotely configure and support them as far as possible, while the latter would be responsible for ensuring uninterrupted service of the hardware and Internet connectivity.

# 12. Development Schedule

## 12.1. Priorities

The major development activities of iPlants are shown below, together with a first-cut assessment of priority.

| Component to be created | Significance | Work-arounds | Needed By |
|---|---|---|---|
| Conservation Assessment | | | |
| iPlants Specimens | Necessary before centralised set of validation, georeferencing and conservation assessment tools can be deployed. | Use manually created and collated data, and ad-hoc tools. Less efficient and more complex. | As soon as possible. |
| Specimen Wrapper & Collation Client | Necessary for efficient exporting of data to iPlants Specimens. | It would be possible but not efficient to manually export data into iPlants Specimens. | As soon after iPlants Specimens as possible. |
| Validation Tool | Allows centralised validation, detection of duplicates and assignment of synonymy. | Individual routines. Less efficient and more complex. | As soon after iPlants Specimens as possible. |
| Georeferencing Tool | Allows centralised georeferencing. | Local georeferencing of downloaded datasets within GIS. Less efficient and more complex. | As soon after Validation Tool as possible. |
| iPlants Gazetteer & Gazetteer API | Aids georeferencing. | Do without (possibly ad-hoc swapping of gazetteer data). Less efficient and prone to error. | As soon as convenient. |
| Conservation Assessment Tool | Allows centralised conservation assessment. | Local conservation assessment within GIS. Less efficient and more complex. | As soon after Georeferencing Tool as possible. |
| GIS algorithms | Allows automated preliminary conservation assessment. | Manual routines within GIS. Less efficient. | Will need refinement with experience. With Georeferencing Tool. |
| Map Server | Allows visualisation during georeferencing and conservation assessment, and production of maps for iPlants online. | Local georeferencing and conservation assessment within GIS. Less efficient and more complex. | As soon after Georeferencing Tool as possible. |

| | | | |
|---|---|---|---|
| Imaging: no components to be developed | | | |
| Compilation | | | |
| iPlants Names | Necessary before bulk Data Improvement Tool can be deployed. | Name standardisation takes place within existing compilation tool on family by family basis. Inefficient. | As soon as possible (note that maximum efficiency benefit will accrue from creating this as early as possible) |
| Data Improvement Tool | Standardises names. | See above. Could perhaps deploy some ad-hoc routines. | Ditto |
| Compilation System & Family Dataset | Compiles family (e.g. assigns taxonomic status and distribution). Assumes standardised names. | Can use existing tool developed for prototype, manually importing standardised data. | As soon after iPlants Names as possible. |
| iPlants Online | | | |
| iPlants website | Public presence of iPlants. www.iplants.org. | None. | A skeletal 'place-holder' advertising the project and indicating what will be coming would be sufficient to start. The bulk should be released with the Online Database. |
| iPlants Online Database | Holds the data to be made available to end-users through iPlants. | None, except that other resources might be offered in its place to begin with. | As soon as iPlants decides it has enough data to go live with. |
| Query Interface & Usage Log | Web interface to query Online Database, and log of these queries. | See above. | See above. |
| iPlants API | Provides a standard automated interface which will allow other software to access iPlants. | Other sites could implement URL links. | As soon as possible after the Online Database. |
| Gazetteer Query Interface | Allows end-users access to the iPlants gazetteer service. | iPlants could supply gazetteer data en masse to interested parties. | As soon as convenient. |
| Review Interface | Allows reviewers to work online direct to Online Database. | Manual procedures. Not as efficient. | As soon as convenient. |
| Editorial Interface | Allows editors to maintain Online Database. | Behind-the-scenes ad-hoc editing; or no active editorial until interface available. | As soon as possible after the Online Database. |
| System Monitoring & System Log & | Records system status, takes action to minimise problems | Manual checks and actions. Not integrated, and less | After the Online Database is released, and at least 6 months |

| | | | |
|---|---|---|---|
| | and alerts system management staff. | efficient. | before replication is in place. |
| Administration Interface | Allows system managers to monitor and administer the Online Database. | Manual ad-hoc interrogation of the System Log. | As soon as possible after System Monitoring and at least 6 months before replication is in place. |
| Content Generators | Maintains image and link content in a cache. | Linkages are not 'intelligent', i.e. they do not know if they will resolve to results, and/or manual construction of cache. | As soon as possible after the Online Database is released. |
| Management | | | |
| Intranet | Web-accessible shared repository for project documents, tasks, contact details, etc. | Emails, ad-hoc methods. Less efficient, complex and likely to adversely affect project performance. | As soon as possible. |
| Project Management Software | Tools to allow project staff to carry out individual project scheduling and monitoring activities. | Widely available - workaround probably not required. Excel and other packages can provide basic features. | As soon as required. |
| Project Master List | Shared system to assign and monitor tasks, and note status of taxa. | Local recording methods. | Need to refine requirements through experience. Begin as soon as possible. |

## 12.2. Potential Schedule

A first attempt at setting a development schedule based on the

| Task | Year 1 | Year 2 | Year 3 |
|---|---|---|---|
| **Conservation Assessment** | iPlants Specimens, Specimen Wrappers and Collation Tool, and Validation Tool. GIS algorithms developed. | Enhance as required. Georeference and Conservation Assessment Tools, Map Server. | Enhance as necessary. |
| **Compilation System** | iPlants Names, Data Improvement Tool. | Enhance as required. Compilation System. | Enhance as necessary. |
| **Online Checklist** | First Cut of Website, iPlants Online Database, Query Interface and Query Log. | Second Cut, also Editorial Interface, System Monitoring and Administration Interfaces. | Third Cut, also Review Interface and Content Generators. |
| **Gazetteer** | - | iPlants Gazetteer and Gazetteer API. | Enhance as required. Gazetteer Query Interface |
| **Management** | Intranet, project management software, and First Cut of Project Master List (very simple). | Enhance as required. | Enhance as required. |
| **Hardware** | Project staff operational. Server for iPlants Production Systems and Online Database. | Enhance and support as necessary. | Enhance and support as necessary. Online servers replicated to 2 other sites. |
| **Summary** | First Cut systems: Online Checklist operational at Kew | Refined Systems: Online Checklist with API operational at Kew | Refined Systems: Online Checklist operational from 3 sites |

# Appendix: Glossary

| | |
|---|---|
| 24x7 | A service that is available without interruption at any time. |
| API | Application Programming Interface – an interface which other software can connect to. |
| Berlin Model | A data structure designed by the Berlin Botanical Garden and Museum to store taxon concepts, together with some associated software which uses it. See http://www.bgbm.org/biodivinf/docs/bgbm-model/ |
| BioCASE | Biological Collection Access Service for Europe protocol and associated schema for collection data. See www.biocase.org. |
| Biogeomancer | Web-based system which returns georeferences for string locations. See www.biogeomancer.org |
| Compilation System | The software, people and procedures used to compile the *i*Plants online list of the plants of the world |
| Darwin Core | Darwin Core data structure (an agreed set of data elements for exchanging Natural History collections data) |
| DiGIR | Distributed Generic Information Retrieval project which has implemented an XML-based API to access specimen data based on the Darwin Core |
| DIVERSITAS | An international initiative aiming to promote integrative biodiversity science, linking biological, ecological and social disciplines in an effort to produce socially relevant new knowledge. |
| ESRI | The de facto world leader in supplying GIS software. See www.esri.com. |
| GBIF | Global Biodiversity Information Framework, whose aim is to make the world's biodiversity data freely and universally available. GBIF works cooperatively with and in support of several other international organizations concerned with biodiversity, and is active in setting up online biological information resources, and the iPlants members play an active part. See www.gbif.org. |
| GenBank | Online database of sequence data at the US National Center for Biotechnology Information |
| GSPC | The Global Strategy for Plant Conservation.<br>Convention on Biological Diversity adopted the Global Strategy for Plant Conservation (decision VI/9), including 16 outcome-oriented global targets for 2010. |
| GTI | The Global Taxonomic Initiative.<br>The GTI was established by the Conference of the Parties to the Convention on Biological Diversity to address the lack of taxonomic information and expertise available in many parts of the world, and thereby to improve decision-making in conservation, sustainable use and equitable sharing of the benefits derived from genetic resources. |
| ICBN | International Code of Botanical Nomenclature. The internationally-accepted set of rules regulating plant names used and applied by taxonomists worldwide. See www.bgbm.fu-berlin.de/iapt/nomenclature/code/SaintLouis/0000St.Luistitle.htm |
| IOPI | International Organization for Plant Information. |

| | |
|---|---|
| | Manages a series of cooperative international projects that aim to create and link databases of plant taxonomic information. |
| *i*Plants | The ***iPlants initative*** |
| iPlants Common Specimen Format | Data definition for specimen data agreed within iPlants and expressed as an XML schema. See Group Documents / Databasing Documents / DIGIR-Data Formats / IPlants-Digir Schema V. 2, and for further detail on georeference fields Group Documents / Databasing Documents / georeferencing guidelines |
| iPlants Web Prototype | Prototype non-functional website set up to demonstrate potential functionaility of the iPlants system. See www.kew.org/data/iPlantsv3/ |
| IPNI | International Plant Names Index.<br>An internet accessible listing of all published plant names with their authors and place of publication. Additional nomenclatural information such as basionym, date of publication and type collections are supplied for some names where available. |
| IT IS | Integrated Taxonomic Information System.<br>Designed to supply authoritative taxonomic information on plants, animals, fungi, and microbes of North America and the world. |
| IUCN | International Union for the Conservation of Nature |
| K | See Kew |
| Kew | The Royal Botanic Gardens, Kew, London, UK |
| LUCID | Knowledge management tool for diagnosing biological organisms |
| MBG | The Missouri Botanical Garden, St. Louis, MO, USA |
| MO | See MBG |
| NatureServe | A US non government agency networking science to conservation |
| NY | See NYBG |
| NYBG | The New York Botanical Garden, New York, USA |
| NYVH | The New York Botanical Garden's Virtual Herbarium |
| RBG Kew | See Kew |
| Sp2000 | The Species 2000 initiative<br>Has the objective of enumerating all known species of plants, animals, fungi and microbes on Earth as the baseline dataset for studies of global biodiversity. |
| TDWG | International Taxonomic Databases Working Group, a voluntary body of taxonomists with a mission to encourage and adopt standards relevant to taxonomic computing. All three iPlants institutions are members. See www.tdwg.org. |
| Tropicos | Online Botanical Database of the Missouri Botanical Garden |
| UNEP | United Nations Environment Programme. |
| WCMC | World Conservation Monitoring Centre (Cambridge) |
| Web Services | An API which is acccessible on the Internet. |