*i*Plants:
THE WORLD'S PLANTS ONLINE

PILOT PROJECT:
FINAL REPORT TO MOORE

1 April -30 November 2004

Version 1.5

Executive Summary

1 *iPlants* aims to produce an accessible index of all the world's plants together with their global distribution and, where possible, an image. *iPlants* will also significantly accelerate the production of Preliminary Conservation Assessments. The pilot phase of the *iPlants* project has involved The Royal Botanic Gardens, Kew (Kew), Missouri Botanical Garden (MO) and The New York Botanical Garden (NY) in partnership.

2 *iPlants* will ensure significant efficiencies and reduce the costs of finding and acquiring botanical information, integrating different sorts of information or aggregating data from diverse sources. Examples of the impact of *iPlants* upon health, science, CITES, forestry, nutrition and conservation are included in this report.

- a A significant body of information already exists about plants. This information, unfortunately, is highly fragmented and impossible to use effectively or reliably since more than one name may refer to the same plant or, a single name may be used to refer to different plants. For the first time *iPlants* will provide a single authoritative list of species linking them to their alternative synonyms and thus serving as an effective gateway to all existing knowledge.
- b *iPlants* has derived a more precise view of the particular demands of a broad spectrum of users working in health, the pharmaceutical industry, sustainable development and publishing as well as conservation. *iPlants* measured the impact of the name index on these particular industries and the costs of continuing without such a list. For example:
  - 85% of possible outcomes when searching GenBank using plant names would be more accurate, more complete or both if the *iPlants* index were included.
  - In species lists used by other initiatives, typically between 25% and 40% of their names are in error thus preventing users from accessing existing information and, for example, distorting conservation initiatives locally, regionally and globally.

3 *iPlants* will accelerate the production of Preliminary Conservation Assessments and increase the accessibility of the baseline specimen data upon which such assessments are based.

- a Only 3% of vascular plants have a global conservation status using current IUCN criteria. From 2003 to 2004 a further 2150 species (less than 1%) were evaluated and published which, since a similar number of new plant species were described during that period, represents little progress.
- b The current IUCN process fails to disseminate the data that underlie the evaluations, with the result that they cannot be assessed (or re-assessed in the light of new data) by end-users. *iPlants* will publish these data.
- c It can take as much as two years between assessment and eventual publication. *iPlants* will both publish Preliminary Conservation Assessments as soon as they have been computed and make the supporting data available to facilitate review and use of the data.

4 Achievement of this ambitious goal requires new ways of working. The *iPlants* pilot project represents a major advance in collaboration in the botanical community: developing methods and synergies to maximise the impact of data on key concerns in conservation and science. Our approach has generated excitement and received endorsement by key players currently providing information services and from conservation practitioners working on the ground.

5 The experience and insights gained during the pilot project have fed into an updated set of technical documentation, including functional specifications for the systems required and an implementation plan for the longer term project.

- a Procures and tools were defined, tested, modified and documented both for building the name index and for production of Preliminary Conservation Assessments.
- b Three major versions of a prototype information service were produced and released. These were used to elicit feedback.

TABLE OF CONTENTS

1. Background

iPlants aims to produce an internet accessible index of all the world's plant species together with their global distribution and, where possible, an image and a Preliminary Conservation Assessment. Achieving such a goal requires a collaborative effort among major botanical institutions and the botanical community at large. *iPlants* will address key challenges facing the successful management and use of plant biodiversity information.

Primarily, *iPlants* is responding to the pressing need for a widely available, consistent and comprehensive working list of all scientific names for plants. It will provide a single stable, authoritative list of "accepted names" each linked to its alternative synonyms. Such a list will enable those that are not trained botanists to access the highly fragmented, but significant and valuable body of information that exists about plants. The Global Strategy for Plant Conservation (GSPC) underlined the need for such a list by identifying this as its first target. The GSPC was approved by 188 governments and comprises 16 targets for completion by 2010 which address the major challenges facing the conservation of plant diversity [1].

People use plants in a myriad of economically and environmentally important ways. They require a single, stable and authoritative reference. Currently they face the confusion caused by the existence of alternative, partially conflicting, published taxonomies. *iPlants* will resolve this confusion.

There are an estimated 400,000 known species of plant and as many as 1,700,000 different scientific names are used in the literature to refer to them. Knowing the correct name of a plant is key to finding out about its uses, distribution and habitat or about its relationship with other organisms and how it can be distinguished from similar plants. Currently, there is no single reference from which to determine, say, how many plant species have been discovered to date, how many occur in a particular country, by what name a species should correctly be known or where it occurs. Without such a comprehensive list, linking accepted names to their synonyms, it is impossible to access all of the information published about a particular plant.

Many databases worldwide contain useful information about plants, their use and conservation. These databases are public, private and personal, diverse in size and form and mostly do not link to one another. Every branch of plant science (ecology, evolutionary biology, agriculture, forestry, genetics, etc.) generates different types of information. *iPlants* is exploring how to link effectively to existing data sources so that all key information about a plant species can be assembled by querying on the plant name.

iPlants is also contributing to Target 2 of the GSPC by developing a means to automate the preliminary assessment of the conservation status of a particular species based upon specimen information within existing major collections. Only around 3% of vascular plants have had their conservation status evaluated globally using current IUCN criteria. Between 2003 and 2004 only 2150 additional species were evaluated and published for the IUCN Red List. In order to respond to the needs of biodiversity managers and to meet important Global Targets, such as the World Summit on Sustainable Development's goal of 'a significant reduction in the rate of loss of biodiversity', the rate of data capture and interpretation needs to be dramatically increased. *iPlants* is developing methodologies and tools to accelerate the production of such assessments and to increase the accessibility of the baseline data upon which such assessments are based. The current IUCN process fails to disseminate the data supporting IUCN evaluations and it takes considerable time between assessment and eventual publication after review (as much as two years in some cases). *iPlants* aims to make preliminary assessments and the data underlying them quickly available, increasing the speed of publication, openness of review and use of the data.

[1] The Global Strategy for Plant Conservation approved by the 7th Conference of the Parties of the Convention on Biodiversity, April 2002. Target 1: A widely accessible list of known plant species. Target 2: preliminary conservation assessments for all known plants.

There are a variety of reasons why a list of known plant species and significant numbers of conservation assessments have not been produced previously. The sheer volume of botanical information available requires a consolidated, intensive effort that is beyond the resources of a single organization. Recently, however, the GSPC targets have provided a focus for collaborative activity. New technologies facilitating data transfer, electronic publication and GIS applications now make it possible for data held within institutions to be shared and analysed collaboratively, as well as being disseminated appropriately to meet the needs of diverse users.

iPlants is focusing on collaborative effort to change the ways in which such information is produced to make it faster and better targeted on the needs of end users.

2. The Pilot Project

The pilot phase of the *iPlants* project has involved The Royal Botanic Gardens, Kew (Kew) , Missouri Botanical Garden (MO) and The New York Botanical Garden (NY) working in partnership. Such partnerships are vital to the development, production and maintenance of *iPlants*. The botanical expertise and data necessary to provide validated, authoritative information is scattered across many institutions. The *iPlants* project represents a major advance in collaborative working: developing synergies and formulating system requirements to maximise the impact of data held within the institutions on key concerns in conservation and science.

The purpose of the grant awarded was to support an eight-month pilot project to test the data processing and data gathering procedures for the production and publication of a global plant checklist and associated information of value to the science and conservation communities. The objectives of the pilot project were to produce a prototype information service, to trial, revise and document the procedures for compiling data, and to develop a plan for the longer term project.

3. Improving efficiency and reducing costs

3.1. A Central Index of Plant Names

Most people that are not botanists, including many scientists, assume that a list of the plants of the world must already exist. This is particularly the case for those from temperate parts of the world where there are relatively few plants and a long tradition of studying them. Even scientific and medical journals frequently point to the International Plant Name Index (IPNI in which Kew is a partner) or to the Tropicos System (at Missouri) as if these already were an index capable of solving confusion over the use of plant names. While these two services are indeed the best that are currently available, each has a different purpose from *iPlants,* though both contribute to it.

The provision of a central list of plants with all of their names will permit a quantum leap in the management and use of information about plants across all disciplines. Any attempt to predict what novel (including as yet unconceived of) information products and services might be made possible by *iPlants* will inevitably be imprecise and incomplete. We can, however, measure the savings and efficiencies that *iPlants* will bring to current practices in diverse sectors of society.

The following examples are intended to indicate something of the breadth and type of impact that the *iPlants* list will have on both the precision and efficiency with which information about plants is used in a range of disciplines.

3.1.1. The list as an online service

In the user scenarios undertaken, existing lists of species are used in a variety of ways from identifying species for conservation action (Millennium Seed Bank), to ensuring the safe use of herbal medicines (World Health Organisation - WHO), to defining legislation covering species (CITES). In those lists examined, 25% to 40% of names were erroneous and would prevent users, particularly those without botanical knowledge, from accessing all existing information about the species concerned. Some errors or omissions are clearly more significant than others and can result in the wrong decisions being taken – plants not being recognised as poisonous, or conservation programmes established for widespread, non-threatened plants. The lack of a widely available, consistent and comprehensive working list of all scientific names for plants is therefore clearly a barrier to all of these activities. Please refer to Appendix 1 for more detail.

World Health Organisation – safe use of herbal medicines and detecting cases of poisoning from plant materials

Let us take one example. The Uppsala Monitoring Centre run a WHO programme ("the global intelligence network for benefits and risks in medicinal products" www.who-umc.org). This programme provides a forum for WHO member states seeking to collaborate in the monitoring of drug safety. The Programme records individual case reports of suspected adverse reactions including to herbal medicines and plant materials recorded in cases of poisoning in a common database. This presently contains over 3.1 million case reports. 75 countries are active members of the programme and 11 further "associate" states await achieving compatibility between the national and international reporting formats. Each of the countries participating has designated a National Centre which submits its records to the Centre in Uppsala.

Conscious of the potential confusion that surrounds plant names, the Uppsala Monitoring Centre requested that Kew undertook a short consultancy for them to validate a sample list of 438 plant names from their database. 12 specialists were involved in checking these names (at a cost of approximately 20,000 US$). The consultancy found that 25% of these names were in error. 10%

were synonyms of other plants – duplicating records in their database – and more than 4% were names that were so incomplete or wrong as to be meaningless and consequently useless as a means to link to other published information.

The data do not exist to enable us to calculate the costs of not having the information service that *iPlants* will provide in place today. We can cite further examples, however, such as a conservation programme covering 12 African countries run by the Millennium Seed Bank in collaboration with national conservation agencies. This programme spends $50,000 per annum to employ one full time member of staff and consult 30 different specialists to check IUCN lists for those countries. In another example CITES spent $45,000 in 1999 to validate a checklist of 2,300 species of Orchid without having to pay for the services of the orchid specialist involved. Similar costs will have been encountered, and continue to be encountered, hundreds of times, by many agencies that have a need for authoritative species lists.

3.1.2. The list as a gateway to other data sources

Many information services about plants already exist and aim to meet the particular needs of users from diverse domains. The vast majority of existing information services include plant names and their services would be greatly improved if there was an accessible comprehensive, up-to-date and authoritative index of plant names as proposed by *iPlants*.

iPlants is exploring mechanisms by which its list of plant names could be provided to the suppliers of such information services. During the pilot project we have explored three example cases: one for the scientific community (GenBank), one for the conservation community (UNEP-WCMC) and one for agroforestry and sustainable development community (CNIP - Brazil) (Appendix 1).

GenBank

In one example, metrics were gathered about the accuracy and comprehensiveness of the names underlying the current GenBank Information Service. The records in GenBank were explored for all plants from three families (Araceae, Arecaceae and Orchidaceae) and the names compared with the authoritative index now available for these families in Kew's Monocotyledon Checklist (www. kew.org.uk/monocotChecklist/). The broad conclusion is that (at least for these three families) 15% of the data records within GenBank are linked to names that either do not exist in the scientific literature or are not considered to be the current name for that plant. Genbank are aware of this problem. Their site carries the disclaimer "The NCBI taxonomy database is not an authoritative source for nomenclature or classification – please consult the relevant scientific literature for the most reliable information".

A logical analysis was undertaken of what outcome a user of the GenBank system should expect depending upon what class of name they were searching with and what information was stored within GenBank and where it was stored. The details of this analysis are given in Appendix 1, but in summary 85% of possible outcomes when searching the GenBank system using names would yield an improved response (either more accurate or more complete) by linkage to the *iPlants* name index.

Support for new technologies

Modern technology offer many exciting possibilities for more direct and more flexible ways to access information about plants. DNA Bar-coding is one example. Ambitious programmes are currently being funded to explore the opportunities that this technology offers. One idea is that the precise structure of a short sequence of DNA sampled from within an organism can be sufficient to "identify" that plant from all others.

To be useful, however, the DNA barcode must be linked unambiguously and accurately to a plant name - this is the label by which the plant is known and the link to all available information about that plant including which other plants are most likely to be confused with it. The given name must also be linked to all of its synonyms, as it is this group of names which gives access to all of the information about that plant. *iPlants* will deliver such a list, complete and synonymised.

3.2. Preliminary Conservation Assessments

It is widely recognised that the lack of Conservation Assessments of species presents a major obstacle both to establishing coherent and well focused conservation strategies globally and to planning local conservation interventions. Target 2 of the Global Strategy for Plant conservation is to achieve a Preliminary Conservation Assessment of the conservation status of all plants.

Currently only 3% of vascular plants have a global conservation status using current IUCN criteria. IUCN are publishing conservation ratings at a rate of less than 1 % of all plants per annum. Since a similar number of new plants are discovered as new to science each year this represents little progress toward achieving Target 2 of the GSPC.

It takes considerable time and effort for IUCN assessments to be published since evaluation requires

i) careful ground-truthing including visits to areas of occurrence.
ii) input from plant taxonomists to resolve nomenclatural and taxonomic confusions before any list is published.
iii) formal international committees to approve the status assigned.

The *iPlants* index to plant names will considerably reduce the costs in (ii) above. Nevertheless we can assume that the rate at which new formal IUCN conservation assessments are published will remain lower than required for all 300 – 400,000 plants to be evaluated by 2010.

iPlants therefore has developed a methodology for automating production and publication of "Preliminary Conservation Assessments" based upon analysing specimen distributions (readily available within the partner institution's collections) in a GIS. *iPlants* will thus have a dramatic impact upon the rate at which conservation assessments can be made available through both speeding up the production of these preliminary ratings, and by making them available to the public with little or no delay.

3.2.1. Direct Benefits

Via its online information service *iPlants* will provide, for the first time, reliable statistics of the total number of plants in the world and the % of these which are threatened in any given region. The name index will also enable those searching for information to access all relevant information regardless of what name was used when the information was published.

Preliminary Conservation Assessments will provide an auditable and readily accessible indication of a plant's conservation status. These will become available far more quickly than is possible using current methodologies. This will mean that conservation practitioners and others will have access to a preliminary status for a far higher % of the world's plants sooner rather than later.

The Preliminary Conservation Assessments will be published along with the data underlying them including the georeferenced specimen records and details of the parameters used to calculate the assessments. Such access to the underlying data is not possible with existing published IUCN conservation assessments. This will permit validation of the assessments themselves and also significantly increase the accessibility of these specimen records.

This approach would also permit *iPlants* to subsequently enable users to evaluate the impact of adding in new data records of their own to the conservation assessment and to incorporate feedback mechanisms by which these new data records may contribute to the central pool.

3.2.2. Benefits for those undertaking Conservation Assessments

In addition to making Preliminary Conservation Assessments available for at least 20% of all plants itself, *iPlants* will also greatly facilitate others seeking to generate such assessments themselves.

The name index will reduce ambiguity and error rates in name lists. As seen above between 25 and 40% of plant name lists available to conservationists are erroneous – with significant consequences. UNEP-WCMC estimate that more than 70% of their time, when entering new records to their Threatened Plants Database is spent attempting to validate and resolve the nomenclatural and geographical records.

By resolving nomenclatural confusion and offering a complete list of names, *iPlants* will also impact upon the ease with which those undertaking conservation assessments can access existing data about the plant in the literature. This will ensure that there is quicker access to existing knowledge, more precise and accurate access to existing knowledge and a far better integration of knowledge from the diverse disciplines.

iPlants offers a tested methodology which gives repeatable results from the available data records. Once new data becomes available the conservation assessments can be regenerated automatically.

Another new tool to be derived from the *iPlants* initiative and of immediate use to the conservation community is a single central commonly shared botanical gazetteer. This will combine the existing gazetteers currently maintained within the partner institutions and all of the new geo-location records generated as *iPlants* undertakes Preliminary Conservation Assessments.

4. Achievements

4.1. Information service

4.1.1. Defining the demand

iPlants has used the pilot project to talk with the user community both to obtain a more precise view of the particular demands for such a service that arise from a broad spectrum of users, and to develop the business case (documenting the significance and extraordinary impact that provision of such a service would have in many walks of life) .

One initiative has been to map and categorise the activities or domains upon which *iPlants* will impact either directly or indirectly. Beneficiaries include a broad range of industries, organisations and individuals and each sector comprises users with particular information demands or requirements.

iPlants has documented "Use Scenarios" for a few chosen key sectors (conservation, health and science) based upon existing literature and structured interviews with individuals in the chosen sectors (Appendix 1). Work continues and this approach will be developed for further classes of user.

Within each domain the objectives are

a. to develop and record indicators of the significance and impact of having a central synonymised list of plants
b. to document exactly how these users would employ such a list and their particular requirements for how it should be delivered.
c. to measure the costs of NOT having such a list currently available
d. to find illustrative examples of where this list would have particular impact or of different ways in which it would be used.

This exercise has already informed the design of the Online Service, provided illuminating examples and generated statistics enabling us to better quantify the benefits and impact of implementing the *iPlants* System.

4.1.2. The prototype information service

During the pilot project *iPlants* has designed and built a prototype of the intended Online Service. The purpose of the prototype is to:

a. Demonstrate the functionality that the *iPlants* Online Service will have including the alternative routes which users will be offered to access the information presented.
b. Test user reactions to facilitate their input to further design.
c. Explore and test design concepts.
d. Illustrate example scenarios in which the Online Service would be used.

Our emphasis when developing the prototype has been to focus on modelling and on illustrating the functionality required rather than show-casing the full extent of the data gathered whilst testing the compilation and conservation assessment procedures. Version 3.0 of the prototype has recently been completed and distributed to partners and users alike for yet further appraisal and comment. Feedback from two previous versions has been analysed and used to shape this current release.

Version 3.0 of the prototype online service illustrates how users visiting the *iPlants* website will, once the system is implemented, be able to:

a. Use any scientific plant name (that has been published in the literature) to access many different information sources on the internet directly or indirectly through *iPlants*.
b. Review Preliminary Conservation Assessments based on GIS analysis of the specimen records housed within the partner institutions' collections.
c. Find out how many plants there are in the world, in a country or in a particular family: fundamental questions for which answers are not available to policy makers, scientists or the public.
d. View images of a plant, its habit and distribution around the world and where all of this information comes from.
e. Provide feedback to the *iPlants* Information Service contributing data or opinions upon the distribution, conservation assessments for example or pointing to additional valuable information resources.

Whereas the *iPlants* prototype has been used to mock up functionality and test user reaction, the institutional partners have in parallel implemented several live information systems based around components of the information that would ultimately form part of the *iPlants* System. These live systems illustrate real live implementation of some of the functionality required. Examples include the Monocotyledon Checklist (www.kew.org.uk/monocotChecklist/), the *Flora Zambesiaca* project (www.kew.org/efloras/) and ePIC (www.kew.org/epic/) sites, Tropicos (www.tropicos.org), and the The New York Botanical Garden Virtual Herbarium (http://sciweb.nybg.org/science2/VirtualHerbarium.asp).

We have initiated exploration of the Prototype Information Service with representatives of the user communities which we intend to serve. This feedback is documented and has been used to feed into modification of the design of later versions of the prototype.

4.2. Defining, testing and documenting procedures

4.2.1. Generating a list of all known plants

The starting point for the creation of a list of all plant names has been the significant reference sources available to and managed within the partner institutions: IPNI, Tropicos and New York Virtual Herbarium. These sources are widely recognised within the scientific and conservation communities as being the major reference sources for plant names. Despite their different objectives, their diverse scopes and lack of comprehensive coverage, these three reference sources together currently represent the most comprehensive resource for a scientist wishing to answer a question about plant names.

The processes required to compile a synonymised index of all plant names from these sources involve identifying, resolving and selecting from among duplicate records, establishing data format standards, and mapping existing records onto these data formats. One of the lessons from the pilot project is that this "standardisation" of name records is most effectively completed prior to any attempt to compile an authoritative list of accepted names with their synonyms and any other data. This is a reversal of the process used previously at Kew and this was one of adaptations made to the existing compilation software. The processes and tools used were documented in a compilation manual and a week long training course for compilers from all three institutions was held at Kew. The families chosen for compilation (Iridaceae, Bignoniaceae, Lecythidaceae and Madagascan endemic families) illustrate variation in size, complexity, in-house knowledge and depth of available information. The compilation system was used at NY, MO and K for standardising name data in the chosen families and then compiling lists of accepted species with synonymy and distribution. Draft compilation of Lecythidaceae at NY and

Madagascan endemic families at MO has now been completed with compilation of Bignoniaceae at MO and Iridaceae at Kew progressing. We are reviewing these checklists and will make them widely accessible in 2005 (see future work below).

The tools and procedures for checklist compilation (including their geographic distribution and other data) have been documented, reviewed and revised in light of the experience of compilers at the three institutions to date (Appendices 2 & 5). Experience suggests that more sophistication is needed in handling the de-duplication process, so as to preserve parsed and/or pre-standardised data where it exists. Also, more can be done to speed up standardisation by the use of automated parsing routines. Integrated access to TL2, BPH and Authors authority files would increase efficiency.

4.2.2. Procedures for specimen based conservation assessments

Herbarium specimens provide an auditable source of data for Preliminary Conservation Assessments of plant species. For many species, herbarium specimens provide the only reliable record of the parameters necessary to calculate such assessments.

Preliminary Conservation Assessments were calculated from georeferenced specimen records. An existing GIS tool was used to calculate parameters and threat criteria as defined in the IUCN Red List (IUCN 2001). The pilot phase focused on procedures both for selecting species for which conservation assessments would be calculated, and for the efficient gathering, combining and analysis of georeferenced specimen records.

It is expensive to database and georeference specimens. With resources limited, databasing and assessment within the *iPlants* project needs to be focused on species which are most likely to be threatened. A series of specimen data gathering exercises were designed and undertaken to identify how to focus data collection more effectively on species which are most likely to be facing a conservation threat. Several taxonomic groups from different geographical areas were selected to represent existing variation in collection frequency and degree of threat. K, NY and MO all collated specimen records for these chosen groups. The data were then combined and analysed.

Species with more limited geographical distributions are more likely to face threat. Draft checklists produced using the procedures outlined above can indicate which species are restricted to a single geographical unit (we used TDWG level 3: "botanical country"). Around 60% of plant species are estimated to have such restricted distributions.

The number of herbarium specimens per species in major global collections may be indicative of threat status. This hypothesis was tested and it was found that species known to be threatened are generally represented by fewer specimens. Clearly it is necessary to have a draft checklist of known species (with their synonyms) available before data capture begins since collections of one particular species may be filed under several synonyms in the collections examined. The draft checklist thus provides a framework for specimen counting across collections. This list can also be used to spot species for which no specimens are found.

Results indicate that if specimen data gathering was restricted to species with a combined total of 10 or fewer specimens across the three collections, then 80% of threatened species and 56% of all species would have been treated, while only 15% of the total collection would need to have been databased.

The combination of restricted distribution and specimen counts can be used to identify a list of candidate species for conservation assessment. Co-ordinating and monitoring data capture and georeferencing is complex. There may be advantages to prioritising data capture by concentrating initially on a few key families which are representative of plant diversity. Candidates include

Legumes, Grasses and Palms. The findings of this work are being prepared for publication in a high impact scientific journal.

The tools and procedures for specimen data gathering, georeferencing and conservation assessment have been documented, reviewed and revised in light of our experience (Appendices 3, 4 & 5). The pilot phase suggests that it is more efficient to count specimens across all institutions first and use this to select the species to database, rather than to database a wider list of species initially and then to reduce this list as specimen count thresholds are reached.

Each institution has a different set of geographical expertise and resources. It was found that georeferencing specimen data is best done in bulk, with each institution concentrating on specimens from its area of geographic expertise rather than on the specimens from its own collections or on species for which it has taxonomic expertise. This avoids all institutions trying to build the expertise and resources necessary to georeference collections globally. Particularly significant have been the conclusions drawn from these exercises about the significance of a central shared botanical gazetteer to serve not only the immediate purpose of facilitating Preliminary Conservation Assessments within *iPlants* but also the broader needs of our users. This will be an enormously valuable tool for the conservation community in its own right.

Another finding has been the overriding benefit of developing software tools, available over the web, to enable partners within *iPlants* to record and validate georeference records and to automate the GIS analyses used to derive Preliminary Conservation Assessments. Providing these on the web will, in the longer term, facilitate feedback and participation by a much broader community. A key element of the *iPlants* project is to dramatically increase the production of Preliminary Conservation Assessments. Increasing accessibility to the supporting data records and automated calculation of preliminary assessments will considerably shorten the time taken currently by IUCN to publish conservation assessments and ensure that much needed data are made available to the conservation community in a timely fashion.

4.2.3. Specimen API and piloting the use of DiGIR

The distributed Generic Information Retrieval (DiGIR) framework has been investigated and the Darwin Core field set examined and modified for *iPlants*. Specimen data from K, MO and NY have been merged using these modified Darwin Core fields to test the conceptual model. The DiGIR fields were further modified as a result of problems identified when trying to georeference the combined specimen records. DiGIR is now implemented at MO and NY. DiGIR has not yet been implemented at Kew as it does not recognise Sybase. This problem is still being investigated. BioCASE software may prove to be a more flexible candidate for the future, and indeed GBIF is currently seeking to merge the DiGIR protocol with BioCASE.

4.2.4. Digital image management and links to further information

Existing digital images of species of the Palm family have been located and catalogued at each institution and combined on the *iPlants* intranet site. This has enabled analysis of the overlap between the institutions and the cost of providing an image for each species. Alternative procedures for the co-ordination of image capture of all species at all institutions have designed, tested and documented (Appendix 3).

An important feature of the envisaged *iPlants* Online Service is to provide users with links to further sources of information about each plant where appropriate. Key reference sources have been targeted and opportunities for linking and sharing data with these systems are being explored. Alternative procedures for detecting and evaluating other digital information sources and electronic texts have also been designed, tested and documented (Appendix 3).

4.3. System specification

We were able to achieve our technical objectives, and in general our original technical vision has been confirmed. Naturally, we have identified many areas for improvement and uncovered some complexities. We have a catalogue of the various software components required to build, deploy and manage the *iPlants* System. Overall, more shared work will be necessary, and the Internet has a larger role to play in this than previously expected. The experience and insights gained have fed into an updated set of technical documentation, including functional specifications for these tools which will form the basis for going forward (Appendices 2, 3, 4, & 5)

4.4. Project management

An excellent working relationship has been built up between team members working on the *iPlants* project at different institutions which has resulted in a greater awareness of the particular strengths of the current partners and where the *iPlants* agenda interacts with other initiatives within the institutions.

An *iPlants* intranet site was established and used extensively to monitor progress of the project, share and review documents and to discuss issues. The basic model of a project manager based at Kew with site managers at each of the institutions worked successfully to deliver the pilot project outputs.

Exploring alternative procedural routes to achieve tasks during the pilot phase has enabled us to detect bottlenecks, reduce unnecessary communication and to define the critical points from a management perspective and the key pieces of information required to monitor these processes.

In the production phase of *iPlants* there will be a need to manage several tasks in parallel. Work in the pilot project demonstrated the need to develop an on-line tracking system to allow allocation of tasks, monitoring of the status of work in particular plant groups, responsibilities for particular tasks and deadlines for their completion.

Data capture rates and effort expended were recorded during the pilot project. This information will be used in producing more accurate cost estimates for the production phase.

5. Plan for the longer-term project

5.1. Embedding *iPlants* within the community

Long term development and maintenance of the *iPlants* project requires:

i. strong institutional commitment from the key partners
ii. effective participation by other botanical institutions and the broader botanical community
iii. a close relationship to be built and maintained with target users of *iPlants* to ensure that we respond to real demands.

The Pilot project has enabled us to move forward strongly on all of these fronts.

A Memorandum of Agreement between the initial partners in *iPlants* for long term work has been signed. The Institutions have committed significant core resources to this goal, are seeking to raise matching funds for particular aspects of the work and are committed to the long term maintenance of *iPlants* [see below].

Discussions with other botanical institutes have initially focused on those whose collections and expertise best complement the existing partners including South African National Biodiversity Institute, Pretoria; Komarov Botanical Institute, St Petersburg, Russia and the Museum Nationale d'Histoire Naturelle, Paris, France. Each of these has expressed a strong interest in collaborating in this initiative.

Feedback from users of *iPlants* will be strongly encouraged, recorded and displayed where appropriate. This will broaden participation in data gathering and help engender a strong link with users and potential collaborators.

The *iPlants* plant list (with names, synonyms and distribution) will be a fundamental reference for any system which seeks to acquire, manipulate or disseminate plant data. Without this list, it will be impossible for these other systems to locate all relevant data; to omit misleading data; or to communicate results with an unambiguous taxonomic scope. The *iPlants* system will play a dual role. It will be:

a) as **a direct online service providing**, fundamental information about plants, their synonymy , distribution, conservation rating image and links to key other sources.

We are committed to continuing to develop the "user-focus" groups and developing the dialogue with these groups throughout the build and maintenance phases (see future work plan).

b) as **a gateway to botanical information** to be embedded within other information systems.

Discussions have taken place with two example target systems: GenBank and UNEP-WCMC. These organisations are key providers of science and conservation data and potential users of *iPlants*: both require a synonymized list of known plant species so that data held under different synonyms can be brought together and both offer data of great potential value to those seeking information concerning a particular species. As a first step in developing a closer relationship with GenBank, Kew intends to offer GenBank the use of part of the Monocot Checklist and discussions on how best to incorporate this information into the GenBank Taxonomy Server will continue.

Discussions with UNEP-WCMC are exploring the mutual value of linking and sharing data. UNEP-WCMC welcome the opportunity to divest itself of the responsibility of maintaining its own list of names and to include the *iPlants* name index into its own Threatened Plants List. The

enthusiasm shown by UNEP-WCMC for adopting the *iPlants* name index underlines the relevance and significance of *iPlants* to implementation of international conventions such as CITES.

Another key partner for *iPlants* is GBIF. GBIF aims to make the world's primary data on biodiversity freely and universally available via the Internet. However, GBIF is dependent on organisations adding content to databases and then making their data available through the GBIF network. *iPlants* will contribute to the Electronic Catalogue of Names (ECAT) programme of GBIF. Each institution plans to make specimen data available to GBIF through DiGIR interfaces to their specimen catalogues.

Kew, as part of their role in facilitating consultation on Target 1 of the GSPC, co-sponsored a workshop with GBIF, Species 2000 and BBSRC to determine for what families checklists existed or were in preparation, the major gaps in coverage, who has expertise to help fill these gaps and what were the main barriers to progress. The workshop was attended by 28 delegates from 15 countries. The *iPlants* project was well received and it was felt that such a project could fill a vital co-ordinating role. Many of the limiting factors preventing broader participation in assembling a working list of known plant species could be overcome with relatively small and well targeted funds for items such as enabling specialists time to review a draft or visit collections or to sponsor meetings to resolve particular issues preventing consensus. The workshop was important in helping to identify future collaborators in the *iPlants* project. The outputs of the meeting: a table of families, current coverage and significant gaps are being made available on-line as part of GBIF's ECAT meta data base.

Discussions have also taken place with Kevin Thiele (CBIT- LUCID) exploring possible linkages between LUCID and the *iPlants* project.

5.2. Development of a sound management structure

A Project Manager based at Kew will take overall responsibility for the management of the project, delivery of the project outcomes and liaison with the Moore Foundation. Co-ordination of the project across the partner institutions will be facilitated by a Steering Committee comprised of senior representatives from each of the project partners. The Project Manager will be responsible for liaising across the institutions of the partnership, managing the project budgets, monitoring progress, preparing reports, and conducting related business. Each project partner will have a Site Manager responsible for making sure that the work goals for his/her institution are met and submitting reports to the Project Manager. The pilot project indicates that this structure will work and that the Site Manager posts will be critical to the success of the Production Phase.

5.3. Matching funding

The prototype phase of *iPlants* was funded by GBMF and resources from the partner institutions. The financial report of the pilot project will be completed by January 2005. The proportion of funding during the pilot project is anticipated to be roughly 33% GBMF, 67% institutions.

All institutions involved in the *iPlants* project have a good track record for attracting additional funding for related and complimentary activities. While the pilot project has been running Kew has secured $50,000 from GBIF to complete production of a checklist of Rubiaceae; 940,000$ from the Andrew W. Mellon foundation for digitising African type specimens; NY has secured $180,000 for an investigation into web-based rapid digital specimen image and data capture in a collaborative proposal to the NSF with Yale University and $320,000 for a type specimen index for macro fungi again from the NSF; $650,000from the Andrew W. Mellon Foundation and $200,000 from the Gilbert & Ildiko Butler Foundation, Inc. for collections computerization; MO

have received $623,000 for digitising Central American Specimens from the Taylor Family Foundation and $450,000 from the Andrew W. Mellon Foundation for digitising African types.

Arising from work done in the prototype, MO is preparing a grant application to NSF to further develop design concepts and our agenda for collaborative checklist production. The institutions intend to continue to seek funding for particular components where opportunities arise. These include i) approaching the Mellon Foundation for funding specimen image capture from parts of the world other than Africa, ii) approaching GBIF ECAT and DIGIT programmes for checklist production and specimen imaging and databasing, and iii) approaching ESRI for development of GIS tools. Since 2003, on average over $3 million per annum (excluding GBMF funding) has been raised by the three institutions for activities related to *iPlants* (Appendix 6).

iPlants will, of course, also continue to receive long term core support from the three sponsoring institutions Kew, MO and NY (Appendix 7).

5.4. Long-term maintenance

The institutions involved have a strong track record of maintaining information services from core resources or obtaining additional funding where necessary. Examples of this include Tropicos at MO, IPNI at Kew and New York Botanical Garden Virtual Herbarium at NY. IPNI incorporates *Index Kewensis* which began in 1885 with five years funding from Charles Darwin but has been developed and maintained continuously ever since, largely from core Kew funding. IPNI (itself a successful collaboration between Kew, Harvard University and the Australian National Herbarium) went live online in 2000. Tropicos is seen as the source of nomenclatural and other information about plants by the botanical community particularly in the Americas. It has been supported and available continually since 1996. The New York Botanical Garden's Virtual Herbarium is the largest resource for people seeking access to a specimen collection, and has been available online since 1996, including a searchable version of *Index Herbariorum*.

Kew, MO and NY are excited about the significant advance that *iPlants* would offer. *iPlants* will be a fundamentally important, outward-looking service provided by botanists to the wider community of those who need access to information about plants. It will be fundamental to resolving many of the conservation and biodiversity problems faced by the modern world. The three institutions commit to underwriting the maintenance of the *iPlants* system and data in the long term using core resources and will also seek to attract additional funding and collaborators.

The resources necessary for long term maintenance of the *iPlants* system and data include:

- The assimilation of 10,000 changes to name records every year as a result of ongoing scientific research and the description of new species. This is estimated to require two full time posts.
- Ongoing conservation assessments and imaging of plant species as part of the institutions core activity
- A database editor will be required to assimilate and respond to feedback and to assist with the administration of the review process.
- Maintenance of the IT infrastructure. The system will need, at every site running the system, the following:
    a. ongoing third-party maintenance costs for the infrastructure (hardware, operating systems, software, Internet connection, etc)
    b. Replacement cycle for the above components
    c. Ongoing technical support and management for the infrastructure (support people to notice and diagnose problems, do backups, manage storage, servers and databases, and generally keep things working). It may be possible to provide some of this remotely.
    d. Continuing development of the system as new opportunities come along and new technologies emerge. If the system is successful then we should anticipate having to

respond to new ways of merging it into the wider bioinformatics network and of presenting information just to keep it relevant and useful

6. Future work to July 2005.

The Kew, NY and MO will complete several key tasks initiated during the prototype phase using institutional resources during the first half of 2005 for example:

1. Complete draft checklists of Iridaceae, Lecythidaceae and Bignoniaceae
2. Explore approaches to delivering these checklists to the public using existing vehicles
3. Investigate possible standardisation methods for IPNI data to speed up subsequent checklist compilation during the production phase.
4. Remain active in the International Community's initiatives to develop and adapt appropriate data standards
5. Work together to evaluate existing software tools and develop our joint plan for the development of new tools necessary for the production phase.

The *iPlants* project would also like to hold further user focus workshops in order to test reaction to the final version of the prototype and to continue to refine how the information gathered in the project can best be tailored to meet user needs. Some additional funding will be required to host such workshops, though the greater part of the cost will be shared among the partner institutions.

The institutions will also be continuing their current research programmes, which include many projects which make a direct contribution to the *iPlants* goals.

7. Conclusion

iPlants aims to produce an accessible index of all of the world's plant species together with, where possible, an image and a Preliminary Conservation Assessment. Achieving such a goal requires a collaborative effort among major botanical institutions and the botanical community at large.

iPlants will ensure significant efficiencies and reduce the costs of finding and acquiring botanical information, integrating different sorts of information or aggregating data from diverse sources. The impact will be felt across a very wide range of sectors of society including health, sustainable development, agroforestry and nutrition as well as in conservation and science.

During the pilot project we have developed a joint understanding on how to obtain this vision. However, to build the systems and scale up data capture for delivery of the *iPlants*' goals, we need funding at a level well beyond that possible from core resources.

8. List of Appendices attached to this report

Appendix 1 Defining the demand

Appendix 2 Compilation procedures

Appendix 3 Procedures for specimens, conservation, images and links

Appendix 4 Manual for georeferencing and Preliminary Conservation Assessments

Appendix 5 Information systems

Appendix 6 Matching funds raised externally

Appendix 7 Ongoing institutional commitments with outputs relevant to iPlants

Appendix 8 Glossary

Appendix 8 : Glossary

Compilation System	The software, people and procedures used to compile the *iPlants* 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.
GBIF	Global Biodiversity Information Framework. Making the world's biodiversity data freely and universally available. GBIF works cooperatively with and in support of several other international organizations concerned with biodiversity.
GenBank	Online database of sequence data at the US National Center for Biotechnology Information
GSPC	The Global Strategy for Plant Conservation. 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. 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.
IOPI	International Organization for Plant Information. Manages a series of cooperative international projects that aim to create and link databases of plant taxonomic information.
iPlants	The ***iPlants initative***
IPNI	International Plant Names Index. 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. 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 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.
Tropicos	Online Botanical Database of the Missouri Botanical Garden
UNEP	United Nations Environment Programme.
WCMC	World Conservation Monitoring Centre (Cambridge)