World Flora Online – Technical Report on Use Cases

Version 7

26 June 2014

Introduction

The formulation of use cases for the World Flora Online (WFO) goes hand in hand with defining the scope and content of the WFO. The broad scope of the WFO sets the foundation for deciding which use cases are appropriate, and in turn, the use cases help define and prioritise which data are stored in the WFO and how they are managed. Defining the scope of the WFO and documenting the use cases is important in both the technical design of the WFO, but also in managing expectations of future users.

This use case report concentrates on the 'contributors' of data to the WFO and the 'consumers' of WFO data. Although it is recognised that there will be other actors involved the WFO workflows/data pipelines, and that extra functionality will be needed to support the use cases for these additional actors - these will be detailed later.

It is recognised that it is highly unlikely that all of the use cases listed in this report will be achievable in the short term, due to the coverage of data and the lack of highly atomised data needed to fulfil the more advanced use cases. However, these advanced use cases are recorded for future use.

Broad Scope of WFO

The WFO will be an information discovery portal, bringing together floristic data on all known plant species that are available in various electronic formats. WFO will include baseline information on plant names, distributions, descriptions, etc. It will provide a single consensus classification, and give the user expert guidance on reliability, accuracy and completeness. The WFO aims to primarily be a reference for conservationists, especially those working towards GSPC Targets and those involved in other CDB activities (CITES, Ramsar, etc.).

Assuming that *'we will take what we can get*'...

In the long term, WFO will:

- be a free, open access[1], online web-based resource, enabling access to floristic data on a global scale;
- provide a global overview of the diversity of plant species on the planet;
- be comprehensive and authoritative, and include vascular plants and bryophytes;
- be a synoptic Flora[2] with a defined data set on all the world's plant species;

[1] **Open Access**, as defined by the *Budapest Open Access Initiative* (2001, www.budapestopenaccessinitiative.org), is that users must be able to "read, download, copy, distribute, print, search, or link to the full texts of these articles, crawl them for indexing, pass them as data to software, or use them for any other lawful purpose without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself. The only constraint on reproduction and distribution, and the only role for copyright in this domain, should be to give authors control over the integrity of their work and the right to be properly acknowledged and cited."

[2] **A Synoptic Flora** is considered to contain information on:

- Nomenclature (accepted scientific names, synonyms, references)
- Description (morphology and sometimes habitat)
- Identification (keys, illustrations, photographs)
- Distribution (country occurrences, altitude)

- publish future versions of *The Plant List* - static, citable updates to the working list of all plant species, *www.theplantlist.org*[3]
- largely contain pre-existing data from national/regional Floras (in the absence of Flora treatments checklist information may be used) and monographs;
- primarily provide information at the rank of species, although information at other ranks will be accommodated;
- include geographic distributions for each species, and thus provide baseline data[4] on plant diversity in each country;
- primarily be in English, at least initially, although data will be maintained in the languages in which it was originally provided.

WFO *will not*:

- be a critical, monographic revision of each of the world's plant species
- be a detailed local Flora with vouchered distributional data
- include algae, fungi, or plants only known as fossils
- be comprehensively multi-lingual – at least in the first phases

Actor definitions

Initially, during the creation of the WFO, the main users (actors) for WFO will be those contributing, managing and editing data. Use cases based on their requirements will shape the system design and functionality. However, use cases for consumers of the data also need to be identified at this stage to shape which data are stored in the WFO, and interfaces to those data. Thus the actors are grouped according to these main classes, although it is expected that individuals will have several roles in these classes.

A. Consumers

1) Conservationists

Especially those who are working on GSPC Targets and other CDB areas.

- Conservation Scientists
- Conservation Planners/Policy Makers
- Invasive species researchers
- Sustainable use researchers
- Economic Botanists/Ethnobotanists

2) Plant Taxonomists

Those involved in producing floristic and monographic treatments.

3) Other Scientists

- Ecologists
- Anthropologists
- Archaeologists
- Pharmacologists

[3] RBGE Kew has offered to lead in the production of an **Interim List** to use as the start for the **WFO Taxonomic Backbone** (Consensus Classification) – see footnote on next page. This would build on experience gained from producing **The Plant List** (TPL) versions 1.0 and 1.1, but take a different approach by starting with names and add taxonomic concepts/classifications to this. The Interim List would be built using more taxonomic databases than those used for TPL 1.1. The Interim List will then be updated within WFO with the expectation that, if proven successful, future versions of TPL would be produced from the WFO Taxonomic Backbone.

[4] **Baseline data** for conservation is considered to be information on what species occur in what countries, what they look like and what are their scientific names.

4) General Interest Groups
Natural historians, citizen scientists, etc.

B. Contributors

1) **Primary data providers**
   People who hold floristic, monographic or checklist information, at least at country level.
2) **Information converters**
   People who convert unstructured information, like OCR text and PDFs, into structured data or who convert structured data into uniform, standardized data for further review and classification.
3) **Taxonomic curators**
   People who collate, synthesise and update the taxonomic backbone consensus classification[5] and maintain the taxon concept relationships[6] between alternative classifications[7].
4) **Expert taxonomic reviewers**
   Experts in a particular taxonomic group or region who peer review, or otherwise comment on quality and coverage of data, from primary data providers and information converters, which is provided in the WFO.
5) **Technical data/system managers**
   Informatics experts managing the WFO systems - harvest, store and standardize data.

C. Other stakeholders

1) **Institutional interests**
   People representing institutional interests of the institutions or organisations that have committed resources to provide sustainable technical or editorial support for the production and maintenance of the WFO. The interests are primarily in the areas of organisation, partition of work, acquisition of funding, and governance of the WFO.
2) **Taxonomic research planners and evaluators**
   People involved in formulating floristic and monographic research programmes, setting research priorities and evaluating outputs.

A. Consumer Use Cases

A minimum set of Use Cases common to all kinds of consumer would be, I would like to...

- have free, open web-based access to the WFO with multiple browsers
- be able to print and/or download data, regardless of size

[5] **Taxonomic Backbone** (Consensus Classification) will be a fully synonymised classification hierarchy of all scientific names of plants included within WFO. The Taxonomic Backbone will represent the consensus classification of accepted names, and synonyms, as agreed by the WFO taxonomic experts, and wherever possible, taxon concepts and the relationships between them.

[6] **Taxon concepts** are taxon names with an associated taxonomic circumscription. **Taxon concept relationships** are statements of equivalence or otherwise between taxon concepts.

[7] **Alternative Classifications** are alternative names, and alternative classification hierarchies, that have been linked to data within WFO. In most cases they are the names, and classifications, given by primary data providers when they supplied data to WFO, which do not agree with accepted names in the WFO Taxonomic Backbone.

- search for plant species according to one or more of:
    - scientific plant name
    - country (region, i.e. including range)
    - common/vernacular name
- be advised on a single, accepted scientific name for a species
- be advised on synonyms of the accepted name
- be advised on the scientific name, and synonyms, submitted by the data provider
- see descriptive information of a plant (text)
- see images of a plant (photos, drawings, etc.)
- find identification tools (keys) to the plants in a country/region
- be advised on the most appropriate description where multiple descriptions are available
- see identification tools associated with a plant (keys)
- be advised on the countries for which this plant is recorded within WFO (list, map)
- see the source of the data (with links back to original data)
- be advised on the reliability of the data in data sources
- be advised on the conflicts/discrepancies between data sources when there are more than one available **[Not Phase 1]**[8]
- annotate/rate data sources which I have found most useful **[Not Phase 1]**
- annotate data with my comments **[Not Phase 1]**
- (machine to machine users) enable interaction of data with other systems (e.g. Map of Life, GBIF, EOL, VertNet, etc. – i.e. conform to standards such as DarwinCore)
- be informed added value from data available in source datasets accessed by following the hyperlink (e.g. structure descriptions, taxon concept relationships, specimen voucher data).of the range of additional information that is available in the contributing datatsets. Could be described in text as in eMonocot, or categorized (icons, keywords).

As a Conservationist, I also want to...

- search according to one or more of the following additional criteria:
    - altitudinal range
    - flowering time
    - fruiting time
    - conservation status (IUCN Category)
    - conservation status (IUCN Category) through time **[Not Phase 1]**
    - CITES listing
    - kind of known threat (included in IUCN Category) **[Not Phase 1]**
    - occurrence within a predefined polygon (country level and above)
    - occurrence within a polygon the user submits (country level and above)
- be advised on the altitudinal data on a plant species
- be advised on the ecological/habitat data on a plant species **[Not Phase 1]**
- be advised on the IUCN conservation status for a plant species
- be able get statistical data (quantities, percentages, charts) for:
    - taxon[9] name – how many genera within a family, how many species within a genus
    - geographic region – how many families/genera/species (world, continent, country, user-defined polygon)
    - habitat/ecosystem – how many families/genera/species **[Not Phase 1]**
    - conservation status – how many species
    - endemic status - how many families/genera/species

[8] Use cases indicated as '**Not Phase 1**' are deemed not to be a priority for the first phase implementation of the WFO Portal – although they may be implemented if resources permit.

[9] A **Taxon** is a taxonomic group at any rank, e.g. family, subfamily, genus, subgenus, species, subspecies (pleural taxa).

As a Plant Taxonomist, I also want to...

- be able to search for and retrieve floristic/monographic accounts using the following additional criteria:
    - content type (checklist, flora, or monograph)
    - format (atomised data, unstructured text, pdf-text, pdf-scan) **[Not Phase 1]**
- be able to view multiple regional floras and multiple monographic information for a taxon on the same page
- be advised on taxon concept relationships between alternative classifications **[Not Phase 1]**
- be able to show different taxa on the same map with different colours, add overlay layers for elevation, climate, etc. **[Not Phase 1]**
- discover classification/nomenclature knowledge gaps to help direct future research
- discover description knowledge gaps to help direct future research **[Not Phase 1]**
- discover floristic/distributional knowledge gaps to help direct future research
- be able use data in WFO to help fill known knowledge gaps:
    - be able to create a 'checklist' for a country/region and download
    - be able to create a 'proto-flora' for a country/region and download
    - be able to create a 'proto-monograph' for a taxon and download
- find out how I can be involved in the WFO global community
- register my area of expertise/ongoing floristic/monographic work with the WFO
- be informed about on-going efforts to create floristic/monographic accounts **[Not Phase 1]**
- be notified when a floristic/monographic account is added to the WFO
    - for a taxon (family/genus/species) **[Not Phase 1]**
    - covering a country/region **[Not Phase 1]**
- apply for access as a Taxonomic Curator or Expert Taxonomic Reviewer
- be able to submit feedback to the WFO Taxonomic Curators/Expert Taxonomic Reviewers

As an Ecologist, I also want to...

- search according to one or more of the following additional criteria:
    - ecological trait **[Not Phase 1]**
    - habitat/ecosystem **[Not Phase 1]**
- be advised on ecosystem information **[Not Phase 1]**
- be advised on ecological trait information **[Not Phase 1]**
-

As an additional user, I also want to...

- be able to search by name assigned in pharmacology [pharmacologist] **[Not Phase 1]**
- be advised on the use of species as free text or categorized data [anthropologist/archaeologist] **[Not Phase 1]**
- be advised on where and in what form plants are trafficked [ethnobotanist] [**Not Phase 1]**
- be able to search for chemical constituents [pharmacologist, phytochemist] **[Not Phase 1]**

B. Contributor Use Cases

As a Primary Data Provider, I want to...

- understand the aims and scope of the WFO
- know what kinds of data are stored/cached by WFO
- know what content is stored/cached by WFO
- know if data on my groups/region is already stored/cached within WFO
- know how my data will visually appear online
- know that the accepted name, and synonymy, that I provide will be shown with my data
- understand the creation/purpose of the taxonomic backbone consensus classification

- know the data elements/structure needed to contribute my data
- know what WFO will change or do with my data
- know the technical process and format to contribute my data
- register my dataset with the WFO
- upload my data with metadata
- access the contribution/upload system via the web using multiple browsers
- know how to contribute updates to my data
- know the use policy for data I am contributing
- know statistics on the use of my data
- know how my data will be attributed
- know how to provide attribution data[10]
- have links back to my original data, e.g. online Floras
- have my data visible to others online after a defined release date **[Not Phase 1]**
- be informed when comments/feedback are made on my data

As an Information Converter, I want to...

- access the contribution system via the web using multiple browsers
- store information sets submitted by Primary Data Providers
- open information sets, manipulate them, and store work-in-process
- work with a variety of information sets – unstructured, semi-structured, well-structured
- work with regional flora, monographic (generic or family), checklists, occurrence lists, conservation status lists, habitat/native/endemic/etc lists, and common name lists.
- work with a variety of web services which can be used for data structuring, tagging, data cleaning, linking, etc.
- see a list of all my and others' work-in-process and work completed **[Not Phase 1]**
- have tools and a "workbench" to perform data refinement and merger tasks/steps
- have standardization data and tools to apply them to the data refinement and merger process
- have tools to assess the quality, standardization, conformance of work-in-process
- merge standardized, conformant data to the WFO master datasets

As a Taxonomic Backbone Curator, I want to...

- access the data management system via the web using multiple browsers
- change the taxonomic status[11] of names in the taxonomic backbone consensus classification and provide required reference(s)
- discover unresolved names that need further work
- add new names with status and reference(s) into the taxonomic backbone consensus classification
- make or revise the relationships between names in the taxonomic backbone consensus classification
- compare alternative classifications and identify conflicts
- make or revise relationships between taxa in alternative classifications
- record attribution for the Taxonomic Curator
- see a log of taxon classification change activity, by whom and when

As an Expert Taxonomic Reviewer, I want to...

[10] **Attribution Data** is considered to include information on: data provider as source, original dataset publication and date, acceptance source/publication and date, synonymy source/publication and date, author, editor, Illustrator.

[11] **Taxonomic Status** of a name is the opinion that it is either an accepted name, a synonyms of an accepted name, or an unresolved name. It is not the Nomenclatural Status of a name, e.g. invalidly published, illegitimate, etc.

- access the data management system via the web using multiple browsers
- annotate data sources for a taxon or taxonomic group to indicate reliability of the data
- record attribution for the Expert Taxonomic Reviewer
- see a log of taxon change activity, by whom and when

As a Technical Data/Systems Manager, I want to...

- access the data management system via the web using multiple browsers
- ingest data from *Information Converters* workbenches
- harvest/ingest data from *Primary Data Providers* who have been established as updating data sources
- provide the system tools[12] for use by those involved in WFO
- provide single-sign-on system controls and security to enable access by contributing users to the WFO systems
- administer users
- operate the servers, disks and network access

C. Other stakeholders Use Cases

As a person representing institutional interests, I want to…

- influence decisions made with respect to WFO
- be informed, and possibly part of, funding initiatives aiming at the establishment and maintenance of the WFO
- be able to contribute and enhance my institution's research interests through collaboration in the WFO
- enhance my institutions standing through collaboration with the WFO
- develop infrastructural, methodological and research synergies with other partners in the WFO consortium.
-

As a person planning and evaluating taxonomic research programmes, I want to...

- discover taxonomic knowledge gaps to help direct future research
- discover description knowledge gaps to help direct future research
- discover floristic/distributional knowledge gaps to help direct future research
- promote research projects aimed at data creation to fill gaps in taxonomy, descriptions, and distributions

[12] **System Tools** include; markup tools for Information Converters; classification tools for Taxonomic Curators; editorial tools for Expert Taxonomic Reviewers