Prototype PlantFinder - An OGC Web Service

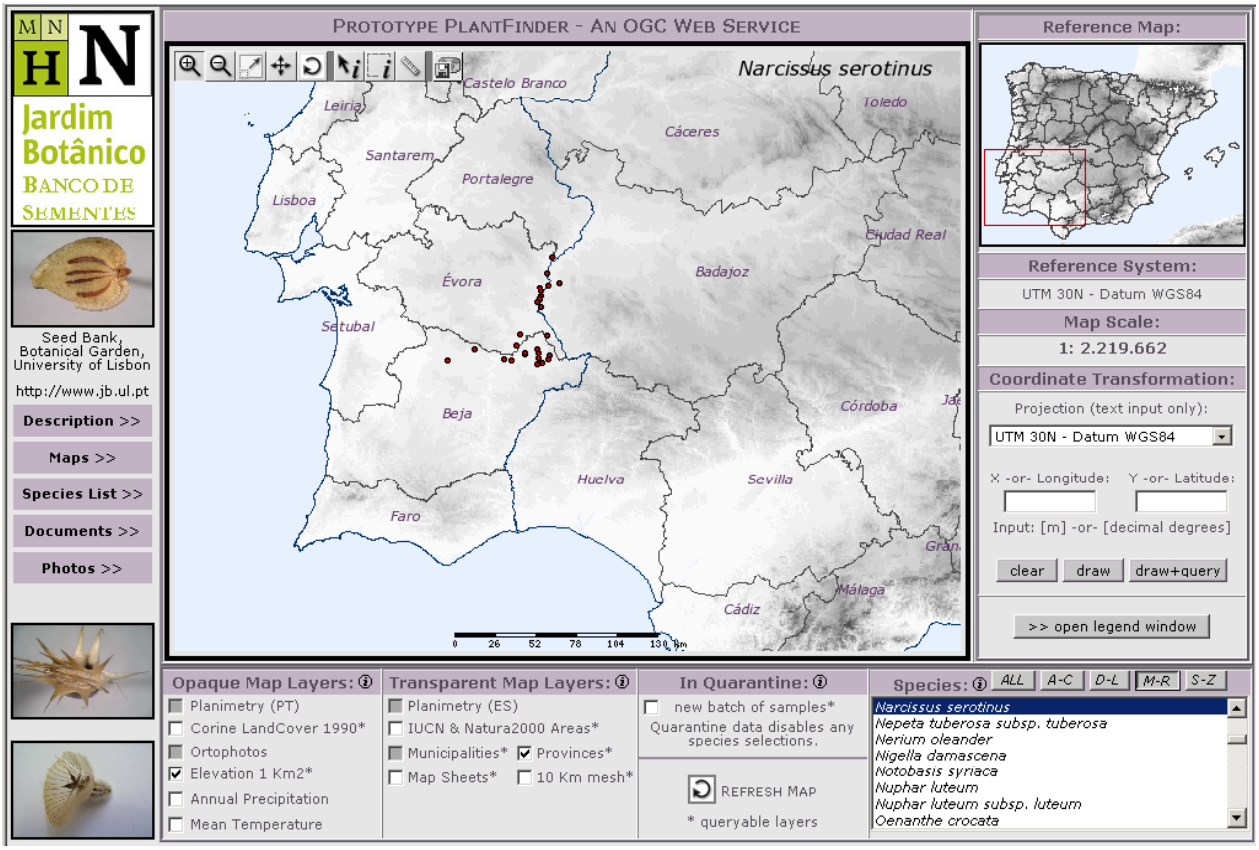
http://jbplantfinder.politecnica.ul.pt/

INTRODUCTION:

PlantFinder constitutes a Web Server application to display the Seed Bank collection sites in an interactive map. Samples and site properties can be queried. It also includes routines for importing data and assigning spatial attributes which allow to identify alternative collection sites. The current version is primarily a 'working version' customised for the Seed Bank at the Lisbon Botanical Garden. Access is presently not limited, so it may be visited by other botanists and the general public.

DATA PRESENTATION AND MAP DISPLAYS:

The Lisbon Seed Bank is a relatively young collection. As the collectors have been conscious about the importance of spatial localization, most collection sites have been referenced by GPS (some coordinates were taken from 1:25.000 and a few from 1:50.000 hard copy maps or from digital maps). In many cases, herbarium specimens have been kept as additional reference samples from the collected populations. These specimens are being incorporated into the Herbarium of the Botanical Garden at Lisbon for later consultation and use. Occasionally, neither seeds nor specimens have been collected but the sites have been recorded as 'species observation'. In general, all original data are conserved, including field notes, localization methods and the reference data sources.



The PlantFinder web server application: To develop the application we have almost exclusively applied licence free open-source software. PlantFinder runs on an ordinary desktop PC, configured as http server. The only commercial component is the operating system (Windows2000 NT). Server software (Apache) and relational database system (PostgreSQL) are open-source software, as well as the map engine (MapServer). MapServer generally runs as a CGI application but here the MapServer API is accessed directly by using PHP/MapScript as scripting interface. Any

web browser supporting HTML 4.0 (such as Internet Explorer, Mozilla or Netscape Navigator), can be used as access point (Web *Client*) to PlantFinder. The user interacts locally with his/her web browser and submits requests to the application server where the PlantFinder application and data are installed. A user can ask the server to perform tasks by submitting requests indicating the map extent and other variables. While this sounds very technical, it is elegant in its simplicity. The parameters that form these requests are actually invisible to the user and thus do not require knowledge on informatics at all. The user sends a request by clicking on the map or on any button or checkbox on the web page. These clicks and the selected options are transmitted over the Internet using a simple communication protocol: HTTP (over TCP/IP). The application server processes these requests and dynamically generates HTML pages which are then returned to the user's web browser.

Geographic area and projection: The seed samples were collected in Portugal but also on Spanish mainland (especially South and South-East Spain), so we had to deal with a variety of different map projections. Other thematic maps, such as provinces, municipalities, UTM grid, IUCN areas, NATURA2000 areas (EC Habitat Directive and Birds Directive), Digital Elevation Model (DEM) and vegetation cover map (CORINE LandCover) have also been acquired in different map projections. Nevertheless, the application is able to display sample sites and other thematic maps simultaneously by reprojecting all map layers on-the-fly to a common projection.

It has also been necessary to add a customized version of the spatial reference system UTM European Datum. Due to the fact that various versions of the European Datum coexist, the spatial reference system codes as published by the EPSG (23029, 23030, 23031) do not specify any distances to the ellipsoid with respect to the World Geodetic System (WGS84). Nevertheless, all public sector cartography produced in Spain implement the same (Spanish) version of the European Datum, while GPS localisations apply the official EPSG datum definitions. The problem arises with map projection changes and coordinates derived from different original map projections – the positional error for Spanish cartography may be up to 200 meter in X or Y due to the datum. In order to correctly reproject the Spanish cartography we use a customized code (Melodensky constants: deltaX, deltaY, deltaZ to WGS84: -87.835, -105.771, -122.5; source: Instituto Geográfico Nacional, Subdirección de Producción Cartográfica, personal communication).

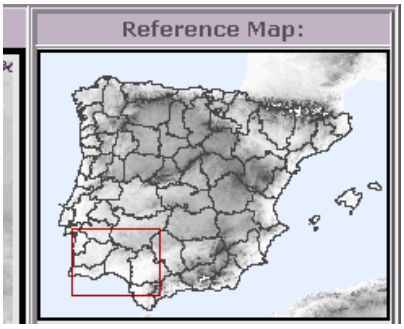
The Geodatabase: PostGIS is an extension to the PostgreSQL database system which allows GIS objects to be stored in the database. Accordingly, all spatial (GIS) data except the raster data layers have been loaded into the geodatabase. Vector layers are converted into database tables with an additional geometry column where the coordinates and spatial reference system are stored. Thus, the database does not only contain sample attributes but also all geometric objects and attributes of the spatial data. PostGIS also adds functions for analysis and processing of GIS objects, such as point-in-polygon searches (see: SPATIAL DATA ASSIGNMENT). MapServer accesses the database in order to draw or query the spatial objects (maps).

Simultaneous display of remote cartography: MapServer is also enabled as Web Map Server (WMS) *client* and may include map layers from remote WMS servers. In this case the Web *Client* is not a user with a web browser but the PlantFinder server that sends and receives requests to and from another server. This enables the visual overlay of distributed geographic information, over the Internet. The Spanish Spatial Data Infrastructure (*Infraestructura de Datos Espatiales de España* IDEE www.idee.es) provides topographic maps. The same technology also allows us to incorporate the annual mean temperature and annual mean precipitation maps which are retrieved from the Digital Climate Atlas of the Iberian Peninsula (*Atlas Climático Digital de la Península Ibérica*, Autonomous Univ. of Barcelona, Spain; http://opengis.uab.es/wms/iberia/). Orthophotos are not yet nationwide available neither for Portugal nor for Spain but may be incorporated as online service in the future, as well as the Portuguese topographic map.

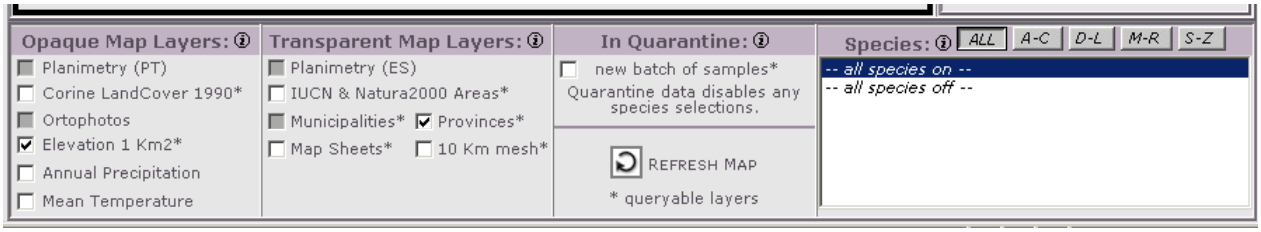
Operation of PlantFinder:

Toolbar: The user may navigate on the map, pan, zoom-in and out using the toolbar buttons located on the upper edge of the map area. Zoom-in may be done by a simple click as well as by drawing a rectangle on the map.

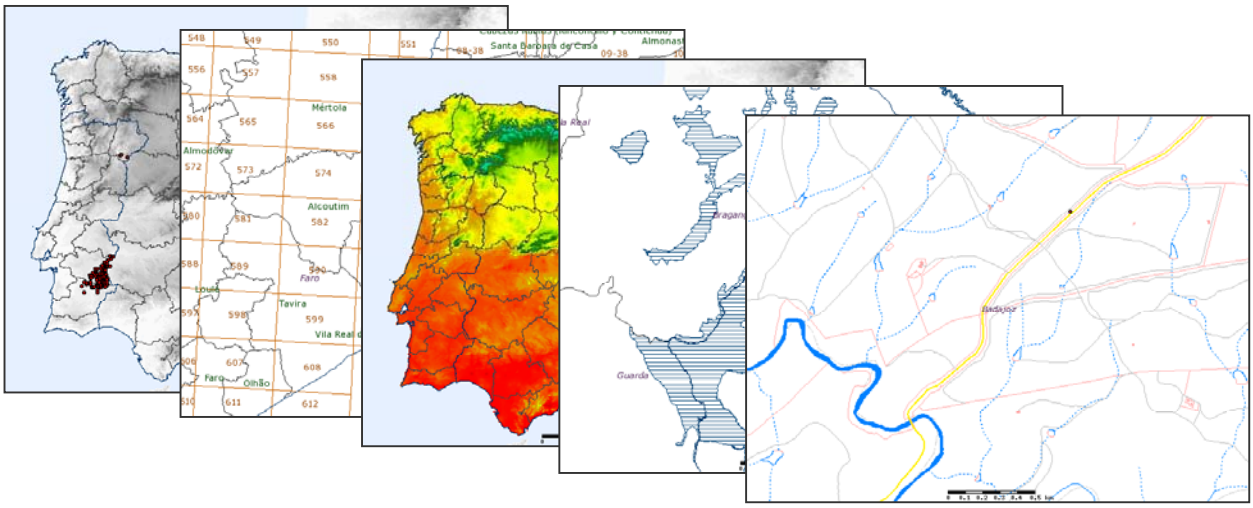
The currently visible extent of the map is indicated as a red rectangle on a small reference map to the upper right. Note that the reference map itself is also interactive – the user may drag the rectangle on the reference map to a different place and the big map will be automatically updated accordingly. There are also toolbar buttons for point and rectangular queries and a distance measurement tool. Their functionality will be explained later. The last button opens a new window with the map graphic so that the user can save it to a file.



The map layer list is located right below the map area. Here, the user may customise the map composition and choose the thematic maps to be displayed by setting the corresponding marks in the checkboxes. While some layers are mostly used as backdrop image by adding a spatial reference to the map, others may be queried additionally.

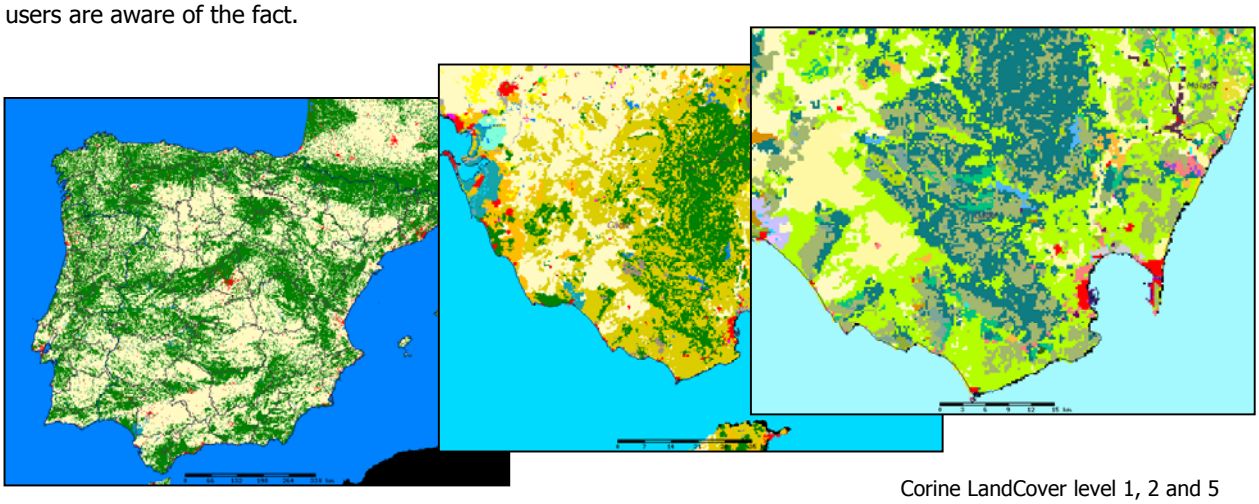


Opaque and transparent map layers: The Portuguese topographic map, vegetation cover, elevation, annual mean temperature and annual mean precipitation are opaque (raster) map layers and should therefore be displayed alternatively. Provinces and municipalities, IUCN and NATURA2000 areas as well as UTM grid and map sheets are transparent (vector) layers which may be displayed and combined with any other layers.



Scale dependent display: The current map scale is indicated on the panel to the right of the map area. The Spanish Spatial Data Infrastructure provides topographic maps as Web Map Service (WMS) with details depending on the scale, reaching a maximum at about 1:25.000. Local thematic vector layers (municipalities, grid system, map sheets etc.) are also labelled according to the map scale. However, some maps are not visible at every scale. Their checkboxes are disabled (grey colour) if the current scale is not recommendable for presentation. Our intention was to prevent that the map composition was visually overloaded, which also would slow down the display. For that we

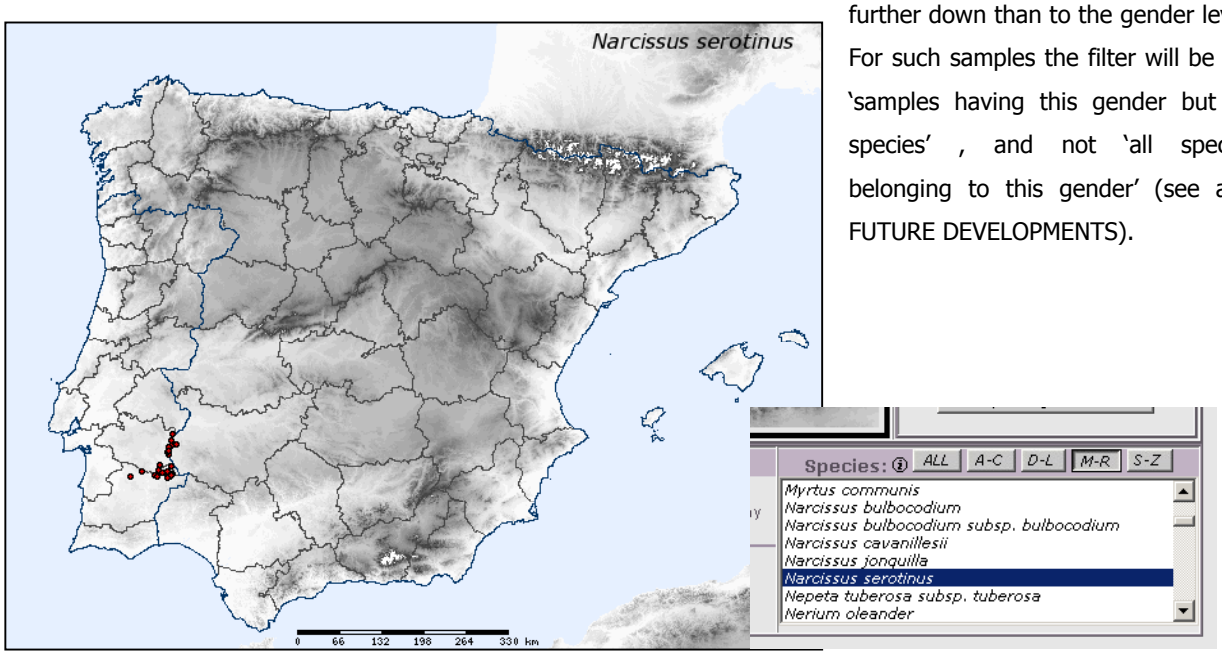
have been somewhat generous at setting the upper and lower scale limits, providing some of the maps beyond the scale at which they were originally produced for. However, the CORINE LandCover map can be displayed at all map scales. As CORINE maps have 3 classification levels, we show the more detailed classification levels at small scales. Classification precision and spatial precision, although both are related, are not quite the same but we hope that most users are aware of the fact.



Corine LandCover level 1, 2 and 5

Samples in 'Quarantine': In order to apply checks for inconsistencies on any new sample site data, we first introduce the registers into a 'quarantine' table which can be displayed separately (note that the quarantine data layer remains empty after the verification and until new data has been introduced).

Species maps: Once the registers in 'quarantine' have been verified and combined with the seed collection sample set, the user may either display all point locations or may select a determined species in order to display the geographic origin of the samples of this species held at the Seed Bank at Lisbon. The list of scientific names has been subdivided into 4 groups (A-C, D-M, L-R, S-Z) in order to make the search and selection easier. Note that only those species already included in the PlantFinder database appear in the selection list. The list will thus uptake automatically in the future as new samples are incorporated. We also decided to list samples which have not been determined further down than to the gender level. For such samples the filter will be the 'samples having this gender but no species' , and not 'all species belonging to this gender' (see also FUTURE DEVELOPMENTS).



DATA QUERY AND QUERY OUTPUT PRESENTATION:

In addition, tools are provided to query the samples which we expect to be useful for seed bank and herbarium management. They also allow to query and determine environmental characteristics of the sample sites. Such parameters are required for plant-site suitability and species distribution modelling. There are several different ways to obtain information on the collected samples, their site and spatial attributes, which are given below:

A simple point query on the map outputs the location coordinates, the map projection and the map scale at which the click-point has been produced. At broad map scales, the coordinates are rounded up in order to prevent the impression of unrealistic spatial precision. Next, all the features of the reference maps that are currently visible and coincide at this point are plotted. Note that not all the reference maps are queryable. Most of the WMS maps are just a graphic representation of the underlying, remote cartography (i.e. topographic maps) and do not represent quantitative or qualitative values.

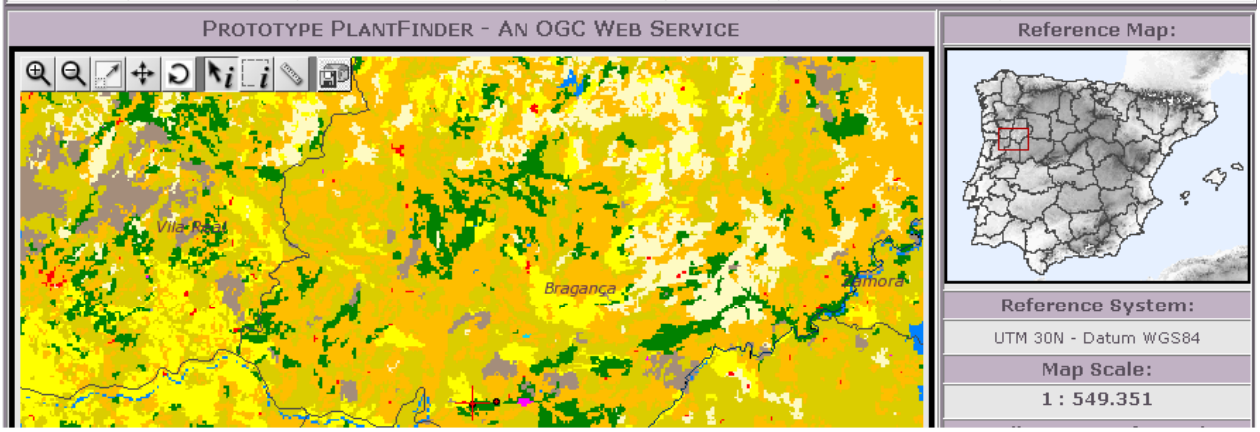
Point Query Location:

Map Projection:	X location:	Y location:	Current Map Scale:
UTM 30N - WGS84	161247	4565149	1 : 549.351

Country	Portugal
Mean Altitud [m]	813
CorineLandcover 1990	Forests
District/Province	Bragança

Samplesite:

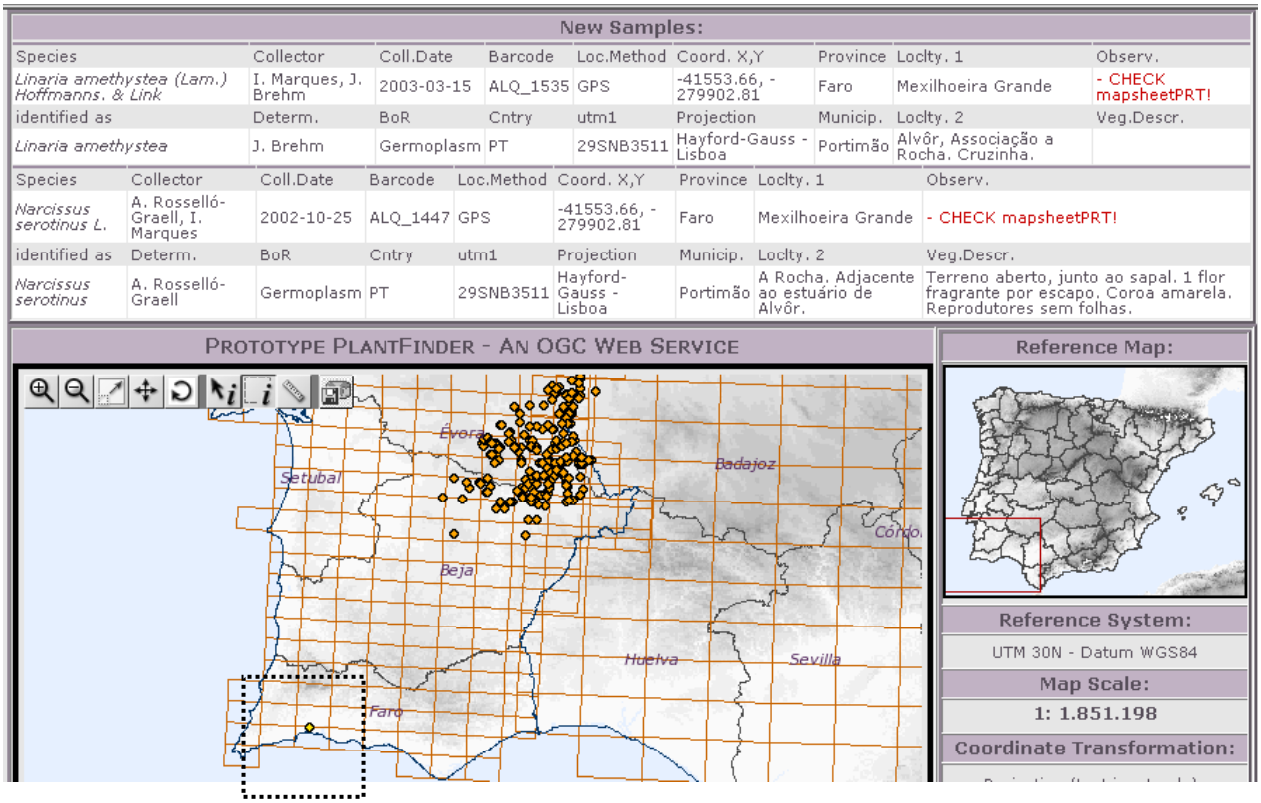
Barcode	AccessionType	Taxon	Coll.Date	Locality	Vegetation Description
ALQ_2446	Specimen	*Ornithogalum pyrenaicum L.*	2004-05-31	Torre de Moncorvo.	
ALQ_2447	Specimen	*Luzula lactea (Link) E.H.Meyer*	2004-05-31	Torre de Moncorvo.	
ALQ_2445	Specimen	*Linaria triornithophora (L.) Willd.*	2004-05-31	Torre de Moncorvo.	Flores rosas.
ALQ_2444	Specimen	*Linaria*	2004-05-31	Torre de Moncorvo.	Flores amarelas.



If a sample site is also present at the click-point, attributes for all samples i.e., seed, herbarium specimen and/or observation, are shown. Since the sample site data table consists of more than 50 columns (not counting the relational tables columns), a reasonable subset of data fields are chosen. This subset varies according to *what* and *how* it was queried: If the sample site has been queried as point, the query output is shorter and rather oriented to traditional herbarium tasks and provide information that one would expect to find on a specimen's label.

Rectangular queries permit to select several sample sites at once and also return more detailed information. Note that, during rectangular queries, all reference maps are ignored and no results are shown. Depending on whether quarantine data or collection data is currently on display, the information which is returned is somewhat different. In order to better control the quality of the recently introduced (quarantine) data, those fields which permit identification of possible, mostly spatial, inconsistencies are shown. If collection data has been subject to the query, the corresponding UTM grid code is returned but the coordinates are not shown. Instead, institution and collection codes, plant family, more detailed vegetation and locality descriptions are provided. This information may be of special interest to other botanists.

.



New Samples:

Species	Collector	Coll.Date	Barcode	Loc.Method	Coord. X,Y	Province	Loclty. 1	Observ.
Linaria amethystea (Lam.) Hoffmanns. & Link	I. Marques, J. Brehm	2003-03-15	ALQ_1535	GPS	-41553.66, -279902.81	Faro	Mexilhoeira Grande	- CHECK mapsheetPRT!
identified as	Determ.	BoR	Cntry	utm1	Projection	Municip.	Loclty. 2	Veg.Descr.
Linaria amethystea	J. Brehm	Germoplasm	PT	29SNB3511	Hayford-Gauss - Lisboa	Portimão	Alvôr, Associação a Rocha. Cruzinha.	

Species	Collector	Coll.Date	Barcode	Loc.Method	Coord. X,Y	Province	Loclty. 1	Observ.
Narcissus serotinus L.	A. Rosselló-Graell, I. Marques	2002-10-25	ALQ_1447	GPS	-41553.66, -279902.81	Faro	Mexilhoeira Grande	- CHECK mapsheetPRT!
identified as	Determ.	BoR	Cntry	utm1	Projection	Municip.	Loclty. 2	Veg.Descr.
Narcissus serotinus	A. Rosselló-Graell	Germoplasm	PT	29SNB3511	Hayford-Gauss - Lisboa	Portimão	A Rocha. Adjacente ao estuário de Alvôr.	Terreno aberto, junto ao sapal. 1 flor fragrante por escapo. Coroa amarela. Reprodutores sem folhas.

DATA INCORPORATION:

Currently, data entry into the database is done by means of a password protected entry mask, designed for loading batches of spreadsheet records. Most of the import functions and spatial search are performed by the geodatabase management system (PostgreSQL/PostGIS) in the background. The entry mask has been developed as separate application to the map application and can be accessed by staff members only. Presently, preference to incorporation of the already existing data is given (see also FUTURE DEVELOPMENTS).

SPATIAL DATA ASSIGNMENT:

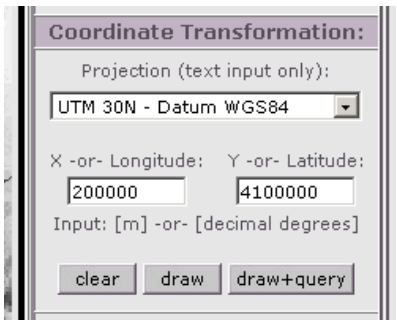
During the development of the database it was found that additional tools for data registration, automatic assignment of spatial parameters (such as altitude, municipality etc.) and data quality controls were needed. While in the past all spatial attributes –and sometimes the coordinates themselves- have had to be looked up one by one in digital and hard copy maps, we developed import routines that do the spatial searches on all reference maps and assign the values automatically to all new accessions. This includes country, province and municipality as well as UTM grid codes and map sheets. The spatial search on all vector map layers is completely done by the spatial database system in a number of steps. The first step is a coordinate transformation of the sample locations from their original reference system (any reference system that has a valid EPSG code, plus our customized codes for Spanish cartography) into four different output reference systems: While UTM30 datum WGS84 is used to display all data in PlantFinder, geographic coordinates (in decimal degrees), Hayford-Gauss datum Lisbon Militar and UTM30 datum ETRS89 are generated for external use.

The spatial database processes about 6 records/sec. A second routine then assigns data values derived from the two raster maps (Digital Elevation Model and CORINE LandCover map). Here the spatial search is done by the map engine (MapServer) and the results are written to the database. This process is even faster.

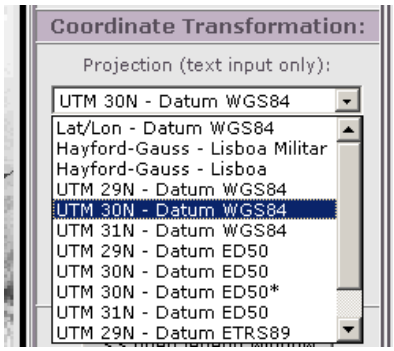
Those records to which spatial information had been assigned to prior to the import, are completed and verified. The assignment routines do not overwrite the previous references but compare the values with those found by it's own spatial search. In case of inconsistencies the routines automatically add a comment into a data field specially designed to receive one or more alerts. This 'comments' field is also shown in the query output of quarantine data, aiding the staff identifying and verifying any suspicious records by visual inspection of the map location and comparison of field notes. Any post-import data modification that affects the coordinates or reference system, triggers these assignment routines again and thus automatically updates the register.

In order to consider a possible optimisation of the spatial attribute assignments, we register in addition to the data values the sources of the raster data (see FUTURE DEVELOPMENTS).

USER INPUT - POINT LOCATION TOOL:



The user may also draw a point on the map, apply coordinate transformations and query the reference maps. The coordinates of a site can be typed directly into a pair of textboxes located on the right panel of the web interface. Several map projections and geodetic datums are provided to choose from a selection list: Geographic coordinates in decimal degrees, two Portuguese map projections, common Spanish map projections (both, official EPSG and customised version) and the new European map projections. The user may then request (by pressing a button) to draw the coordinate location on the map and query its characteristics.



The point location entry tool is valuable for clarifying missing or uncertain projections and geodetic datums. This tool was initially placed on the data input mask, intended to allow the staff to introduce and check new data right from the field site. However, the localization facility which is now found on the web page does not write to the database and spatial attributes are shown only on request. It is thus available to all users and not only to the Seed Bank staff.

The tool also allows the staff to find path and travel distance from any location to the closest alternative sampling sites in case the expected species is no longer present at the site, is not reproducing yet or any longer. In this case it is only necessary to toggle the topographic map on in order to display the reference map with roads, rivers and settlements. The user may then use the distance measuring tool (toolbar: ruler-button) and trace a path from one point to another. If instead the vegetation cover map (CORINE) is toggled on, the user may as well identify other potential habitats where the species could be present. The climate maps –annual mean temperature and annual mean precipitation- may also give hints on likely sample sites.

DATA SHARING AND INTEROPERABILITY:

By making the Seed Bank available on Internet, we hope to contribute to nature conservation. Here the sampling sites are revealed but the seed collection staff may decide to hide single samples or sample sites by marking these records as 'not public'. This may become necessary in case of species or individual populations which are especially vulnerable and where the sample site should not be published. Marked records are automatically filtered each time a map is drawn. Such an exclusion may also be applied to any doubtful records which could not be verified or corrected. Both types of records are not eliminated from the data set and can be recovered anytime later.

Our objective is to share our data with others. We therefore prepare for the implementation of data models, profiles and standards as proposed by relevant initiatives such as the Global Biodiversity Information Facility (GBIF www.gbif.org) and the Infrastructure for Spatial Information in Europe (INSPIRE www.inspire.org).

As web service, PlantFinder implements the OpenGeospatial Consortium (OGC www.opengeospatial.org) specifications. We already make use of our client capability by including ("consuming") third party cartography for spatial reference such as the topographic maps made available as Web Map Services (WMS) by the Spanish Spatial Data Infrastructure (IDEE www.idee.es) and hope to add soon maps from it's Portuguese equivalent. The climatic themes provided as WMS (http://opengis.uab.es/wms/Iberia/) by the Autonomous University of Barcelona give us visual information about climatic conditions.

The usefulness of PlantFinder's data goes probably beyond its original purpose and may be of interest to users outside the seed collection and herbarium world. We are conscious of the importance of metadata which allow potential users to determine if PlantFinder data suits their purposes, but we do not provide standardized metadata at this early stage. So far. each record is documented in detail applying the DarwinCoreV2 metadata profile (http://digir.net/). This profile is widely accepted for collection data and recommended by GBIF. Although not shown in the query output, all of the required and almost all of the optional DarwinCore metadata are registered. However, few additional data fields had to be added, but most of the adaptations are just changes of data formats (i.e. separate the collection date into three fields), automatic transformations (geographic coordinates) and registrations (date-last-modified, coordinate precision etc.). The format changes are done on-the-fly into a separate database table and do not interfere in the normal database processes. The rest of the parameters are inserted or updated by database table functions and do not require any additional time or knowledge from the staff.

FUTURE DEVELOPMENTS:

MAP DISPLAY: Apart from those improvements already discussed before, there may also be a more simplified, multi-lingual version for other botanists and the general public. We propose to replace some of the "working tools" i.e. by selection lists which allow to select the plant family, select whole genders and/or select the species by it's common name. Links to photos of the specimens could also be considered.

Orthophotos are not yet nationwide available neither for Portugal nor for Spain but may be incorporated as online service in the future, as well as the Portuguese topographic map.

DATA ENTRY: After most data from past years have been introduced, data verification and correction methods have been proved efficient and any necessary improvements have been identified, we will return to the development of a data entry mask for individual records. Such an entry tool for individual samples would allow the staff to introduce new data right from the field site, via PDA or laptop. Adaptation to barcode codification is planned as well as additional controls to avoid plant name spelling and coordinate transcription errors.

SPATIAL DATA ASSIGNMENT: The two raster maps, Digital Elevation Model (DEM) and vegetation cover map (CORINE LandCover) are likely to become available at various spatial resolutions and years. The Spanish Spatial Data Infrastructure has recently incorporated a DEM of 200m resolution as WCS and also provides the CORINE LandCover maps for year 1990 and year 2000, which are currently only provided as non-queryable WMS. With some logical modifications of our assignment routines we could then query alternatively either our local 1000m-DEM or the 200m-WCS-DEM. This would allow us to match the most adequate DEM and CORINE map to individual data records according to each samples' localization precision and collection date. We also hope to assign other (online) environmental data into later versions of PlantFinder. Among the WMS maps only those obtained from the Climate Atlas are presently enabled for simple point queries (a GetFeatureInfo-Request would return a not yet standardized text string containing the query results). Hence, we hope that these maps will be soon available also as Web Coverage Service (WCS).

INTEROPERABILITY: PlantFinder is not yet truly interoperable in the sense of being enabled as OGC *server* to other services. Given the structure and nature of it's data, PlantFinder may most likely become a Web Feature Service (WFS). By being interoperable as WFS, PlantFinder would not only allow users to display and query PlantFinder data in their Internet browsers, but also allow other web services to integrate PlantFinder's data. If these services are seed banks or herbariums themselves, data hosted at different locations can be filtered, displayed and queried simultaneously, even though the databases are allocated and maintained at remote places, on different computer platforms and use different software.

We hope to be able soon to actively contribute to GBIF as well as provide ISO19115 metadata as recommended by INSPIRE.