African and Asian Elephant Database

Functional Specification

April 20, 2009

This Functional Specification is accompanied by a detailed Terms of Reference.

1. Introduction

A. Summary

This document outlines a functional specification for the next generation African and Asian Elephant Database (AAED), a server-based spatially enabled database to house and analyze population, range and other data on African and Asian elephants. The AAED will be hosted on a remote server, allowing various users to access and/or edit the data using various client packages, particularly GIS software and RDBMS as well as a front-end browser. Access to the data will be regulated through an access control system. Existing data in the African Elephant Database (AED) will be migrated to the new database server. The AAED will be designed in such a way as to be able to accommodate additional species in the future.

The project to develop the AAED is a joint project of the IUCN/SSC's African Elephant Specialist Group and Asian Elephant Specialist Group. The database will be managed by the Specialist Groups and used by a number of internal and external users. Only the Specialist Groups will enter and maintain the database, while a wider range of users will use the database to extract data for reporting and analytical purposes, or simply out of interest.

B. Requirements

This functional specification describes the functionality which the AAED should provide and aims to provide the information required for a database developer to prepare a database development plan, as well as the standard against which the database should be tested in the testing phase. Specific requirements are elaborated in the Existing System and Use Case sections.

C. Numbers

Full-access users: These users will have access to the back-end of the database and be responsible for inputting and validating data and performing complex analyses. This is likely to be between 1-5 individuals per species, as well as an overall database administrator.

Front-end users: These users will have access to a web-based interface through which they can submit new data for validation, generate queries and view data. It's unlikely that there will be more than 100 simultaneous users, but this would grow as the number of species included in the database grows.

D. Existing System

The current system is a single-user MS Access database which contains both spatial and non-spatial (attribute) data. The database is managed using ArcGIS software as well as the MS Access front end.

Spatial data layers are currently maintained in ArcGIS 9.1 Geodatabase format and are stored as points, lines or polygons depending on the geographic feature type. These data sets are combined with base map data derived from the Digital Chart of the World. The two primary elephant layers are Range and Survey data, containing information on elephant distribution and numbers respectively. Elephant population survey reports are obtained from wildlife management agencies and other organizations, and questionnaires and maps are distributed to AfESG members and other individuals with possible access to reliable information on elephant distribution and abundance. Data are received in a variety of formats, including paper maps, reports, personal communications and geo-referenced digital data. Information from paper maps is digitized and geo-referenced, while attribute data from reports, communications and questionnaire replies are entered through the keyboard.

Once entered and validated, a number of routines built into the MS Access database (in a mixture of MS Access Macros and VBasic code) are run to categorize data according to quality and to calculate pooled elephant estimates at the national, regional and continental levels. These are stored as attributes in the relevant GIS layers. Estimates exist at the site, national, regional and continental levels.

All information entered into the database is fully referenced to the source, and an ancillary bibliographical database is maintained as part of the AED (currently this is done using Biblioscape, which runs on a DBIsam backend). Crystal reports is used to produce publication-quality tables (in PDF) in an automated fashion. Maps for each country, region and the entire continent are similarly generated automatically in PDF using the MapBook extension for ArcGIS.

Maps and tables are brought together with text in Adobe FrameMaker, and the result is published in the African Elephant Status Report (most recently in 2007, scheduled to be published every 4-5 years).

What problems does the current system have?

1. The current system is strictly geared to the production of African Elephant Status Reports (AESR), and is not designed to maintain a historical repository of elephant population estimates. Currently separate Access databases exist for each AESR that has been published (ie AED1995, AED1998, AED2002 and AED2007). Population estimates that fall between these publication years and are superseded by more recent estimates are often not entered in the database at all, (even though they ought to be). In other words, the current system represents a series of artificial snapshots in time, rather than a continuum of population estimates.

2. The current AED schema is in need of rationalization and normalization. Pooled estimates are stored in tables within the database (rather than being calculated in real time as queries). The proposed system should be able to calculate on demand pooled estimates at any point in the past.

3. The above design makes it difficult to compare how estimates change in time - for those populations for which comparable estimates are available, as no facility exists to link an estimate with previous estimates of the same area. It is not straightforward for the untrained user to generate a time series of estimates for a given site, as this requires querying multiple Access databases. It is worth noting, however, that survey zones can change in size and shape from one survey to the next, and it is therefore never straightforward to generate such a time series.

4. A rudimentary stop-gap system to track changes in estimates was implemented in the AED 2007. This system ought to be made more robust and operationalised. It should also be extended to tracking changes in other layers - namely range and threats (see below).

5. The database is currently geared to maintain data on a single species. The possibility that there

may be more than one species of African elephant, and the desire to use the AED system to maintain Asian elephant data - and perhaps other species in the future - make the accommodation of additional species, and the ability to calculate pooled estimates for any number of species, a necessity.

6. The ways in which data are categorized and estimates calculated are hard coded into the VBcode. Rules for categorizing data change as survey techniques and needs evolve, and it would therefore be necessary to develop a system whereby categorization rules can be specified in an editable table, and these are interpreted by program code to calculate estimates correctly.

7. Decisions as to which surveys to include in a given AESR are made manually. While it is desirable to maintain a manual override, it would be desirable to have a system that automatically selects the most recent, highest quality and most spatially extensive survey for a given area.

8. Currently it is only possible to calculate pooled estimates at the national, regional and continental levels. It is often desirable to calculate such estimates at other spatial levels - populations, survey blocks, ecosystems, districts, provinces, transboundary conservation areas, etc.

9. In the present system, population estimates and maps can only be displayed and tabulated at the spatial level at which they are entered in the database. Hence if the user wants the AESR to display a single estimate to the Tsavo Ecosystem he/she has no option but to enter the survey estimates at that level. There is no option to enter the estimates at the survey block level and have the block/stratum estimates pooled at the survey zone level. It would be desirable to have the ability to enter survey data at the stratum level and yet display it in the AESR at the park or higher level.

10. The AED in its current form can only report pooled estimates in terms of DEFINITE, PROBABLE, POSSIBLE and SPECULATIVE elephants (as per the rules shown on page 12 of the AESR 2007). It would be desirable to enable another way of reporting data - namely as estimates from surveys (expressed as pooled estimates with confidence intervals) and estimates from guesses (expressed as a range min-max)

11. The AED in its present form is a single-user desktop database. If it is to be expanded to accommodate other species, it should be possible for a number of administrators (one or more per species) to have write access to the database). Similarly, all requests for data views and maps not present in the AESR currently have to be manually processed by the AED manager. It would be desirable for other users to have read access to be able to generate maps and tables.

12. The current AED lacks any means of mapping threats to populations, and the new system should provide for this, in the form of threats layers, bearing in mind that a given area can have multiple threats, and that the geographical shape of different threats may or may not coincide. As mentioned above, the system should allow managers to track spatial and temporal changes in threats.

The AAED should incorporate all features of the existing system as well as all the features described above. All existing data from the AED should be ported to the AAED as part of this exercise. The current system is dependent on ArcGIS software and MS Access. The AAED should run on a PostgreSQL/PostGIS backend which enables users to connect and perform full editing and query functions from a variety of computing platforms and GIS clients, such as ArcGIS, GRASS, QGIS, etc.

E. Terminology

- Range - Distribution area of a species
- Survey - A systematic exercise to estimate the elephant population in a given area, using standard methods.
- Survey zone or input zone - The area in which a survey is conducted - ie a study area
- Definite - see AESR 2007 for definition
- Probable - see AESR 2007 for definition
- Possible - see AESR 2007 for definition
- Speculative - see AESR 2007 for definition
- Individual Registration - see AESR 2007 for definition
- Aerial Total Count - see AESR 2007 for definition
- Aerial Sample Count - see AESR 2007 for definition
- Dung Count - see AESR 2007 for definition
- Genetic Dung Count - see AESR 2007 for definition
- Informed Guess - see AESR 2007 for definition
- Other guess - see AESR 2007 for definition
- Known Range - see AESR 2007 for definition
- Possible Range - see AESR 2007 for definition
- Doubtful Range - see AESR 2007 for definition
- Non-Range - see AESR 2007 for definition

F. References

1. *JJ Blanc, RFW Barnes, GC Craig, HT Dublin, CR Thouless, I. Douglas-Hamilton, JA Hart (2007) African Elephant Status Report 2007: An update from the African Elephant Database. Gland: IUCN. Available from* http://african-elephant.org/aed/aesr2007.html

2. *JJ Blanc. AED Documentation.* http://pgfoundry.org/docman/?group_id=1000156

2. Functional Description

The functionality of the AAED will largely mirror that of the existing AED, with the differences outlined above (namely, it will be a PostgreSQL/PostGIS- based client-server system able to accommodate multiple species, and which will be able to maintain a comprehensive repository of all population estimates, as well as to calculate estimates at different spatial scales). It will also incorporate modules to request digital (georeferenced) data from the AED, review and approve/reject requests, and to submit data for inclusion in the database.

In addition, the AAED will incorporate a new system, code-named MOD (for Mammal Observation Database), which will enable the collection and maintenance of sighting data at various spatial scales.

A full functional specification should be one of the outputs of the development process.

Use Cases

Draft use case diagrams are presented in Figures 1-4 (below).

User Community

The system is primarily aimed at data managers in species Specialist Groups (or other relevant agencies). A second group of users would be registered web users, who can generate queries and map views from

the database. Web users should also be able to place requests for georeferenced data files (eg shapefiles of species distribution). Depending on the data release policies of the various Specialist Groups, such requests may either be served automatically or may necessitate review and approval by the Specialist Group in question.

Administration Functions

Access to the database will be handled through an access control system based on usernames and passwords. There will be an overall database administrator who sets up new species and administers user rights. Registration of web users (who have the ability to submit data and generate simple queries) should be handled automatically by the system.

Error Handling

Errors should be trapped to the extent possible, giving human-readable output and, ideally, also any debugging information that can be submitted to the community of developers. Types of error to be trapped include user input error/validation violations.

Security

Wildlife distribution and abundance data are potentially sensitive, particularly so in the case of endangered species of commercial value. It is of the utmost importance that unauthorized users do not gain access (not even read access to parts of the database) through vulnerabilities in the system.

Help

The system should provide online help for web users. It should also provide help for users of various DMBS and GIS clients to interact with the system - i.e. how to open layers for editing, generate pooled estimates, etc.

Printing

Most printing would be done via GIS Clients and reporting software (see Interfaces below), and mainly to PDF, but it would be desirable for web users to be able to print good quality maps and tables with the results of their queries in hard copy or PDF. Such printouts would automatically include organizational logos, legends, sources, etc.

Interfaces

User interface
In most of the use cases, users will be interacting with the database through third party applications (GIS, reporting software, etc., see below), and hence no interface is required. The exception to this is the web users interface. The detailed functionality of the web interfaces is deferred to the design stage.

Software Interface
The system should be able to interface with the following applications:

- ArcGIS as well as other GIS clients (such as Quantum GIS, uDIG) through a connection to PostgreSQL/PostGIS;
- DBMS front-ends such as pgAdmin and MS Access (or equivalent application that enables visual generation of queries for the SQL-illiterate;
- Crystal reports and other reporting software - through ODBC;
- R (language for statistical computing); and

- Bibliographical management/citation software - ideally multi-platform open source solutions built on a RMDBS, such as Zotero.

Boundary Conditions

It must be established whether it is possible to maintain a PostgreSQL/PostGIS database and to interact with it in a fully functional manner through ArcGIS. A user-friendly GIS and RDBMS interface is seen as crucial components of the system.

Constraints

The funds to complete this project are limited and there is little option for extending them. A functional system should be the final product of this process, but it is hoped that future enhancements can be implemented through an open source community development effort, with additional fundraising as and when required.

The project is financed by the US Fish and Wildlife Service, and must be completed by October 30, 2009.

Platforms

The database will run on a server, preferably under *nix, but ideally would be platform independent. The RDBMS will be PostgreSQL, using one or more of a number of open source, platform independent languages (such as Python, PHP, PL/SQL) to implement various parts of the system. At the very least, the system should support Windows, various flavours of Linux and Mac OS X.

Internationalisation

The database system should be language neutral to enable the production of maps and reports in a number of languages. At the initial phase, the system should at least be able to handle English and French (eg for country and place names, AED-specific terminology, etc).

Performance

The system would be expected to handle simultaneous editing of different layers by different users in remote locations. Editing performance would be expected not to differ much from editing GIS vector layers in a local machine. If response times are anticipated to be a problem (particularly bearing in mind the often slow internet connections prevailing in many developing countries, a master-replica system could be implemented, allowing species data managers to keep a local replica of the master database and regular synchronization of the databases (limited to their species of relevance).

Portability

The system should be designed to be platform independent from the start.

Expandability

As mentioned above, it is expected that additional species will be added in the future. These may necessitate additional base map layers, or alternatively basemap layers for the entire planet could be included from the start.

Customisation

It is to be expected that different species will use different means of categorizing the reliability of estimates and range. It would therefore be desirable to establish a flexible framework under which different rules could be set up for different species. This framework should also allow rules for a given species to be changed by Specialist Group data managers.

Support & Maintenance

Provisions should be made for easy upgrades of the database engine and procedures. The software will be maintained as an open source community project. A bug tracker, web fora/mailing lists and other project management tools should be set up as part of the system in a project forge such as pgfoundry.org. Web traffic should be monitored through an analytics system such as urchin.

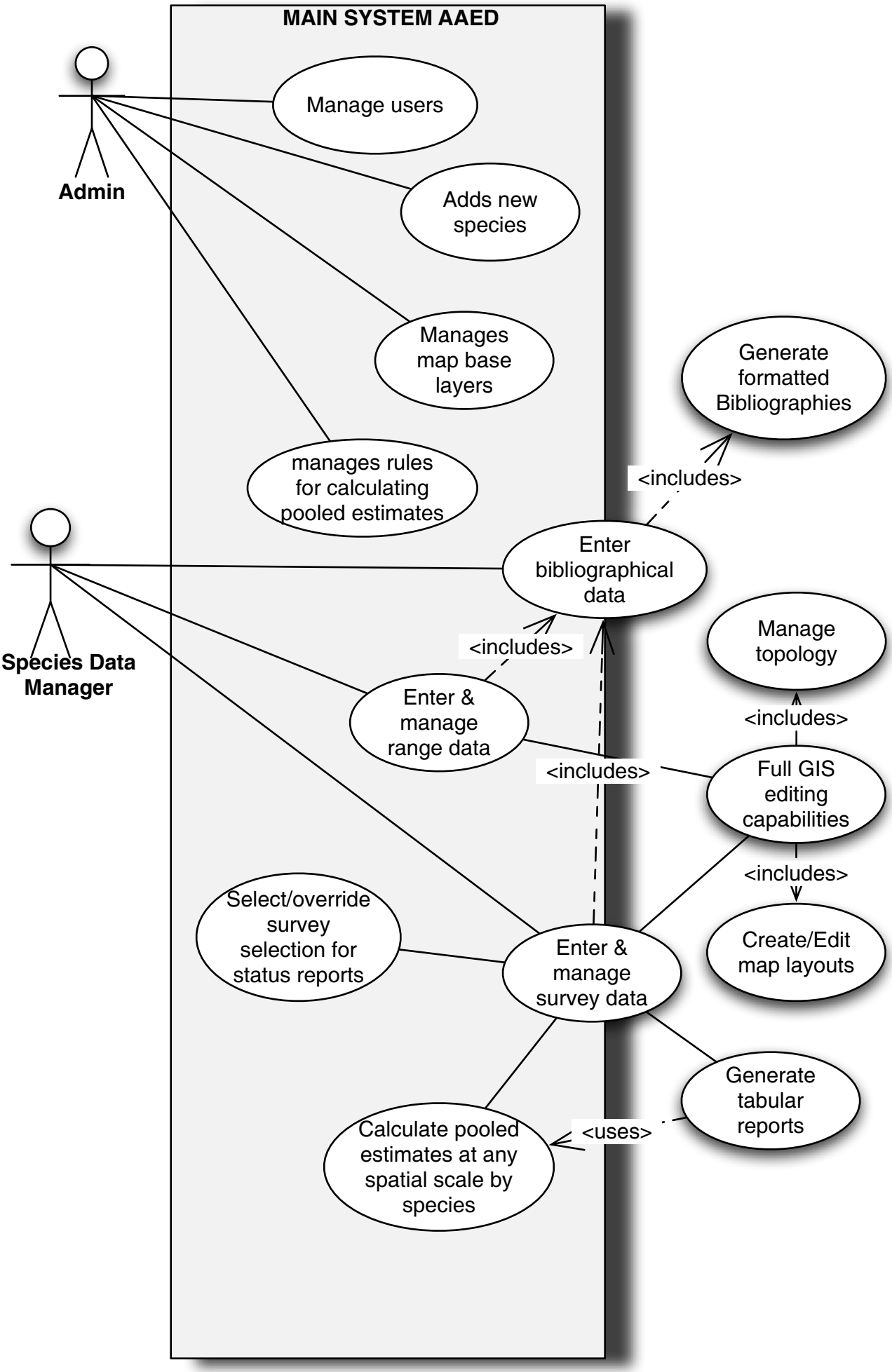
Configuration Management

Software versions should be maintained in an open source version control system, but easy to install "service packs" should also be available.

Documentation

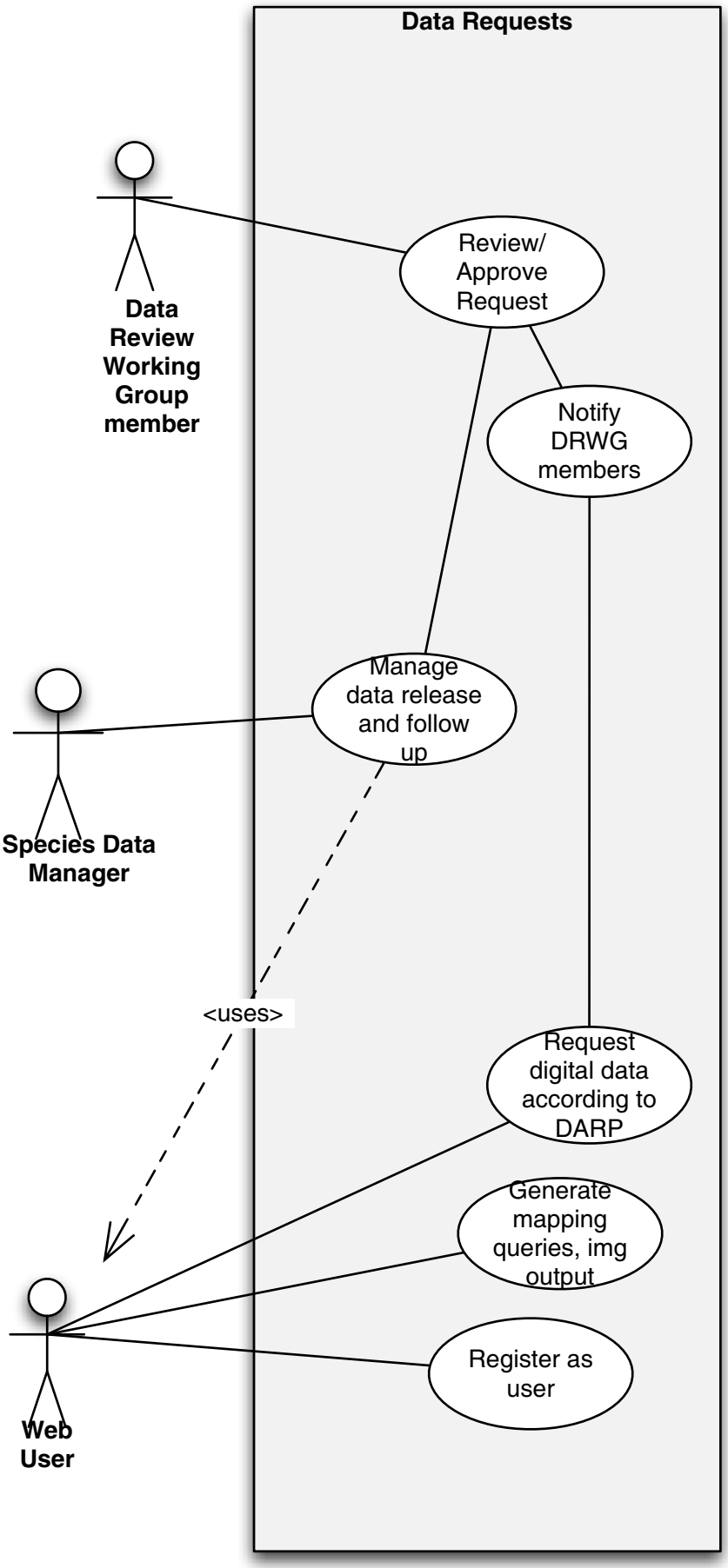
Software code should be fully documented in-code. In addition, a functional description should be provided, including a clear and concise description of the framework, together with a description of the database schema and procedures/triggers. Finally, instructions for connecting to the database from various client configurations should be provided. Web-based training materials should also be provided.

Figure 1



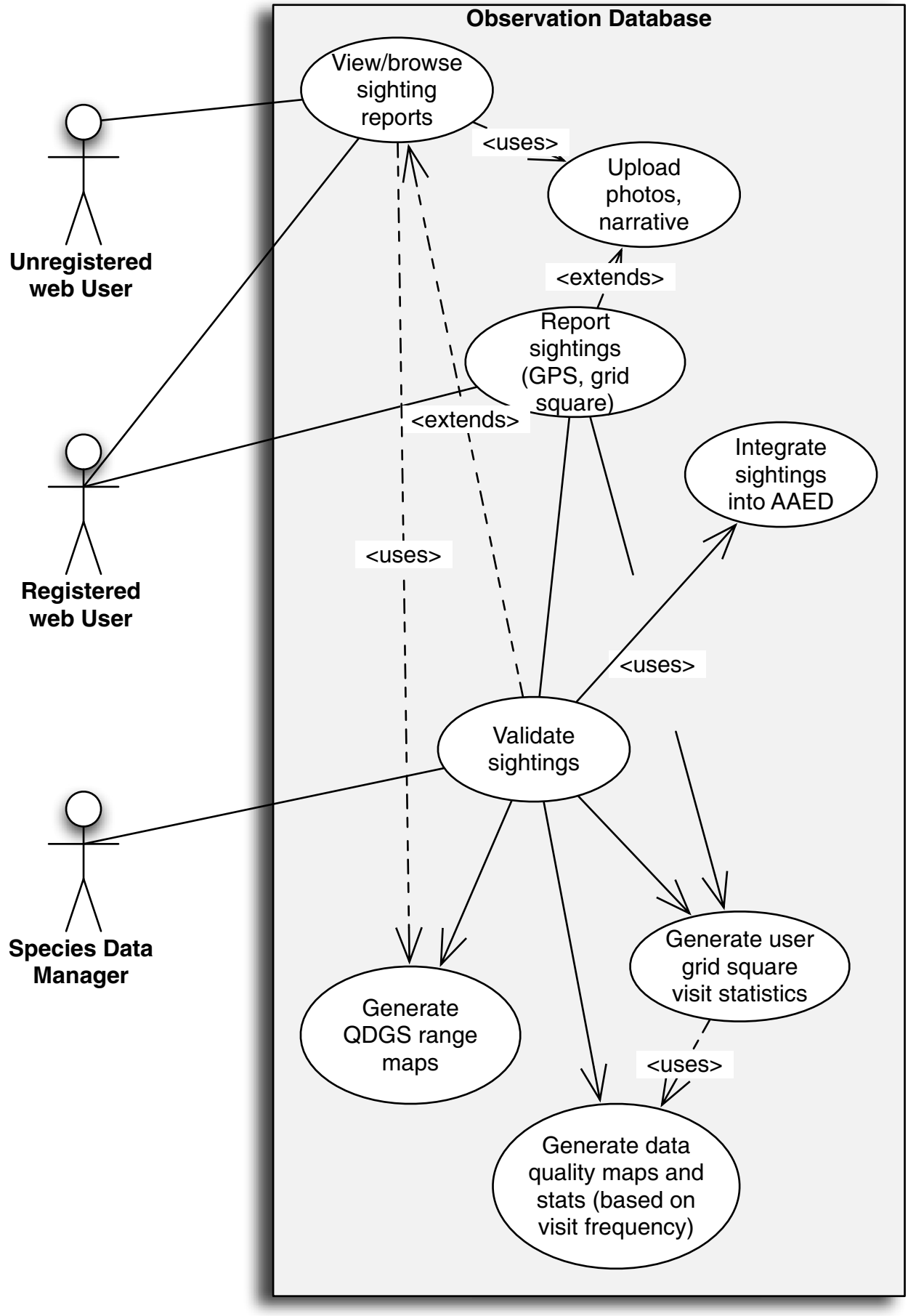
Use Cases: African and Asian Elephant Database April 2009

Figure 2



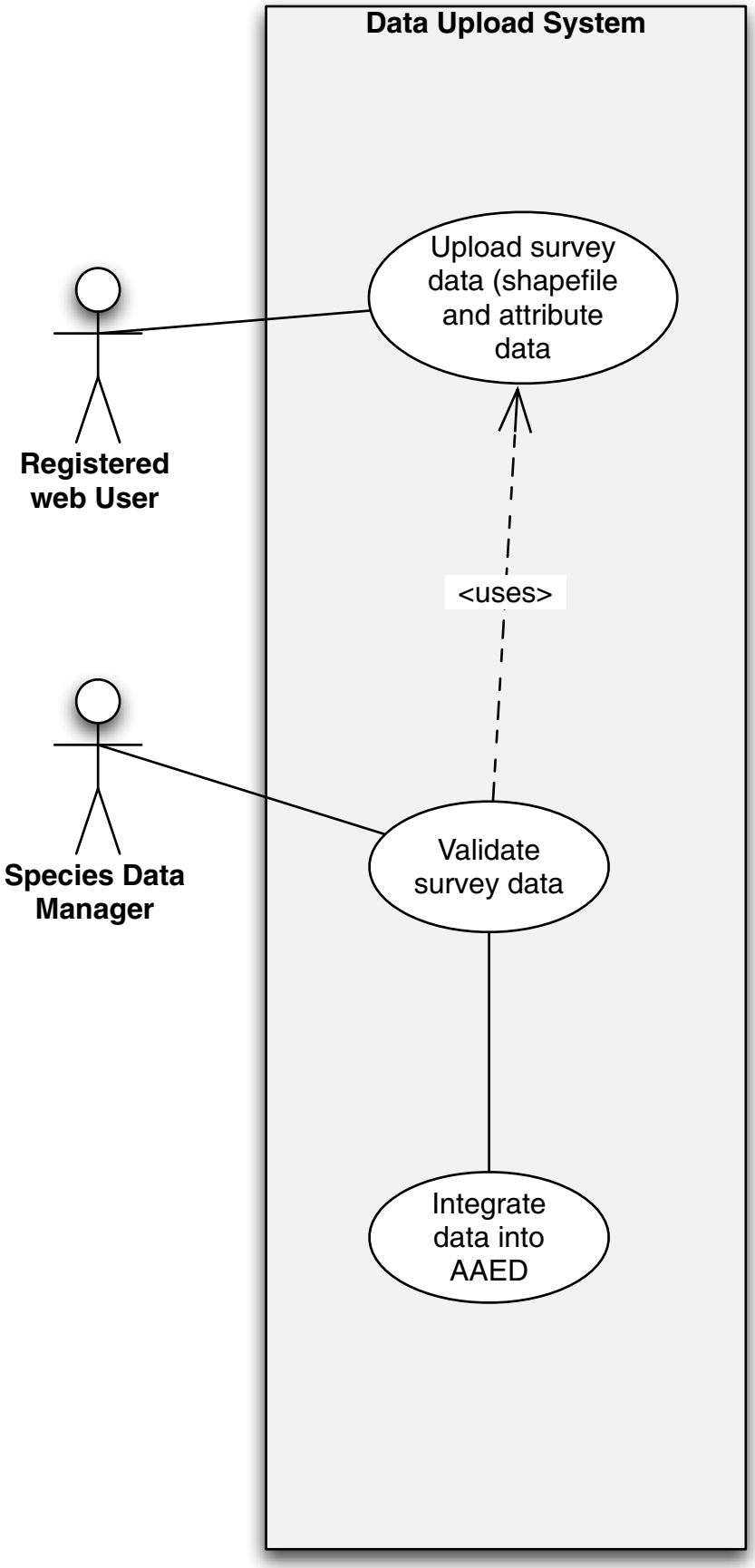
Use Cases: African and Asian Elephant Database April 2009

Figure 3



Use Cases: African and Asian Elephant Database April 2009

Figure 4



Use Cases: African and Asian Elephant Database April 2009