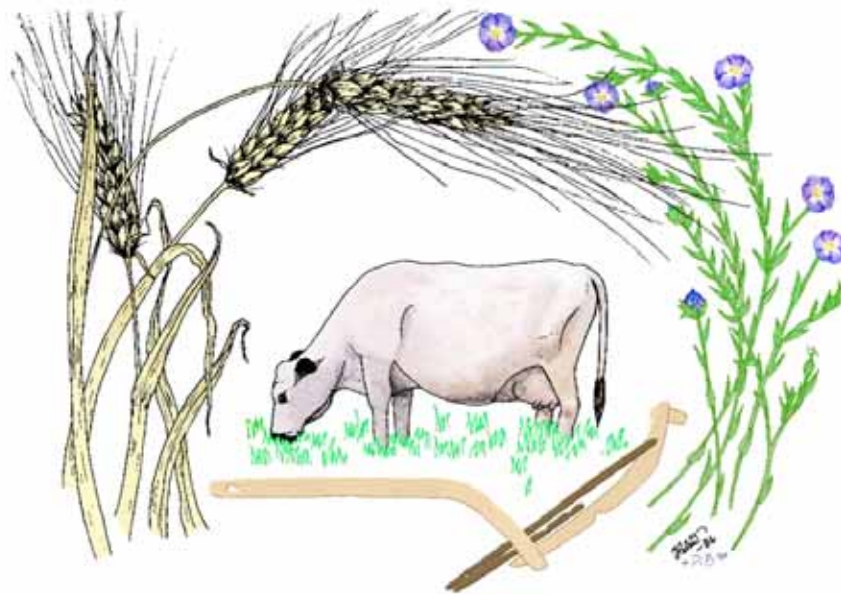# ENVIRONMENTAL ARCHAEOLOGY LAB.

INTERNAL REPORT nr. 2006-031

Planning Report: SEAD
Strategic Environmental Archaeology Database

by

Phil Buckland, Johan Olofsson & Roger Engelmark

DEPARTMENT OF ARCHAEOLOGY & SÁMI STUDIES

Addendum 2007-06-10

This report provides a framework for the construction of a database for environmental archaeology based around the current data holdings of the Environmental Archaeology Lab (MAL) at Umeå University. Whilst internet (ASP) and external partner access are considered, the report focuses on the staff needs and data requirements of MAL, upon which all estimates of development time are based. A functional MAL database, with local client-server functions and simple internet access for data retrieval *may* be possible in the timescale proposed in this report, but a more advanced, fully online system would require considerably more time and a larger developmental team.

With hindsight, and 8 month after completion of the report, the design team feel that the estimates presented here seriously underestimate the investment required to develop a database, online software interface and GIS component that would be of use to a wide group of users outside of MAL's core activities. It should also be made clear that the design below includes only limited modern reference data (ecology, distribution, experiment results), and that this is now considered essential if the system is to be of use to those working with a historical perspective on contemporary problems of biodiversity and landscape management, and the development of new scientific methods.

However, the assessment of MAL's current data, the database design, interface sketches and overviews presented here are still considered valid, and form an important stage in the evolution of a larger, more holistic system. The database structure presented is expandable and adaptable, and will fulfil the needs of a larger system if properly adjusted and implemented by a specialist.

Philip Buckland, Johan Olofsson

Planning Report: SEAD
Strategic Environmental Archaeology Database

by   **Phil Buckland**
           Johan Olofsson
           Roger Engelmark

undertaken at   **The Environmental Archaeology Lab.**
           Dept. Archaeology & Sámi Studies
           Umeå University

Contact details for authors:
Phil Buckland   phil.buckland@arke.umu.se
Johan Olofsson   johan.olofsson@arke.umu.se
Roger Engelmark   roger.engelmark@arke.umu.se

Contents

Appendix 1. Full SEAD database structure diagram (details shown in Section 5.1.5)

List of Figures

List of Tables

List of abbreviations

GIS	– Geographical Information Systems
(G)UI	– (Graphical) User Interface
MAL	– Environmental Archaeology Lab., Umeå University (Miljöarkologiska Laboratoriet)
MS	– Microsoft
SEAD	– Strategic Environmental Archaeology Database
SQL	– Structured Query Language

All trademarks and brand names are copyright of their respective owners.

Database terminology

Database	A collection of related data items, and the information describing them, in this case computer stored.
Table	A subset of the database, containing records and fields (columns) of more closely related data.
Record	An individual row of related data in a table.
Field	A column in a table, where each record/field cell contains the value of the field for that record.
Index	Generally a sorted field in a table, allowing records to be arranged according to some defined sort order such as ascending/descending.
Query	The database term for a request for data that matches specified criteria. SQL is a language for constructing these requests.
Frontend	The software (client) used by the user to interface the data (backend), often stored and used on the computers of individual users. Rarely contains any permanent data itself.
Backend	The actual data part of the system, stored on a server, with which the frontend interacts.

Flow chart symbols
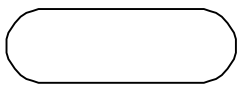
Terminator – start/end of process.

Preparation – steps necessary before undertaking the subsequent tasks.

Process – process or activity.
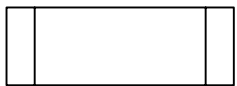
Predefined process – a process that has actions defined by either a method description, or programmatically by a subroutine/function.
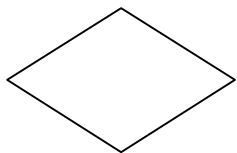
Question/choice – process flow will diverge depending on user choice or data properties (e.g. Yes/No, x>y).
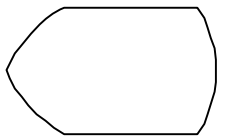
Display – information output to screen or general screen operations.
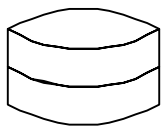
Store – save or load data to/from database.
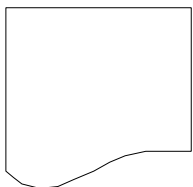
Report/external data – hand written data/information, files produced by external software, or files produced by SEAD for external use.
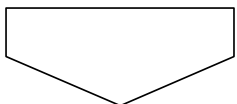
Connector – shows link between multiple diagram parts or figures.

Flow charts were produced with Diagram Designer v.1.15 (© 2006 Michael Vinter) and finished in CorelDraw X3 (© 2005 Corel Corporation).
Database structure diagrams were created in Microsoft Visual Studio 2005 (© 2005 Microsoft Corporation) and finished in CorelDraw X3.
All other diagrams were created in CorelDraw X3.

1 Abstract

This document lays out a strategy for the development of SEAD – A Strategic Environmental Archaeology Database, which will facilitate the digitisation and accessibility augmentation of MAL's existing data from nearly thirty years of work in the fields of archaeology and environmental science. SEAD will also provide a framework for the entry of data from all future research and consultancy work at MAL, and allow guest researchers and external partners to contribute to, and work with the same data. The planned system will be implemented at both local and internet levels, and be designed with an aim towards broadening its scope with external partners in the future.

SEAD will be made available online in order to increase the ease of access to environmental archaeology data and encourage an expansion of both the discipline and Sweden's role in it. This is inline with current EU strategies on enhancing research infrastructure, and providing a greater insight into human-environment interactions for long term planning.

2 Introduction & Background

2.1 The Environmental Archaeology Lab's Part in SEAD

The Environmental Archaeology Lab in Umeå (MAL) has worked in the fields of archaeology and environmental science for nearly thirty years, over which time an enormous amount of raw data on the long term interactions of people and their environments has been amassed. These data consist mainly of proxy variables – that is to say measurements of the changes in physical or chemical properties of sediments, and quantities of macro fossils found on or around sites of archaeological and Quaternary scientific interest. The vast majority of this data is not in digital form, and only available by extraction from published reports: another large portion is stored digitally but in a variety of formats. On their own these individual measures and variables can provide important interpretive information for the reconstruction of prehistoric human activity and environmental change. The combination of several proxies (multi-proxy analyses) greatly increases their analytical power – the relative advantages of each variable complement each other in terms of their ability to reflect environmental changes. Multi-proxy analyses are difficult to undertake without a framework which allows for variable sequences to be overlain and analysed within a common timescale.

One of the primary aims of SEAD is to make this data available in both its raw and interpreted forms. The latter will be facilitated by making all published reports available online, and linked to the relevant data and meta data.

A relational database would greatly improve the accessibility of these data, and allow for the comparison of different variables over the entire geographical and chronological scope of the dataset. Development of user-friendly interfaces for data entry, retrieval, reporting and exporting, and advanced querying will provide a powerful tool for the analysis of environmental change and human interaction with the environment. It will allow complex research questions to be formulated and investigated with a power and efficiency not currently possible due to the disparate nature of the data.

The development of a flexible and expandable system will facilitate the relatively simple addition of further analysis methods, and their related data, as the lab's analysis and cooperation base grows. It will also allow for easy modification of SEAD in response to the desires of external partners such as other labs and research/teaching bodies.

2.2   Need for SEAD

Environmental archaeology now stands at a critical point where the amount of accumulated data permits more complex research questions than ever before to be investigated. This data, however, is currently not in a sufficiently organised form to permit these questions to be realistically analysed. SEAD would provide a framework for this organization. In addition, it would greatly empower and simplify the investigative process for archaeologists and social scientists, especially in tasks where the interpretation of environmental data are crucial to explaining past societies and their interaction with their environments. The data are naturally of fundamental importance for research into sustainable development in that they hold a number of long term proxy records of human impact on the environment.

There is a real demand from Quaternary geologists and environmental scientists for cooperation with archaeologists in the investigation of past climate and environmental change. In particular, there is a general understanding that questions on the nature and extent of past human impacts cannot be answered without investigations into the civilisations that instigated them. At the moment the greater part of this archaeological knowledge and data is not in a form readily accessible or useable by external research groups, even in collaborative projects. There is also a marked polarity between research areas resulting from the traditional disciplinary and faculty boundaries.

Many areas of environmental science examine only the recent past, but with methods and study objects similar to those of environmental archaeology. SEAD would bridge the gap in timescales between the disciplines, and allow for a more holistic approach to modern ecology with an emphasis on human interactions with the natural environment - past, present and future.

SEAD would go a long way towards realising this objective of multidisciplinary, longitudinal research, by both facilitating data exchange *and* providing a common medium for research initiatives. It would, in addition, permit the investigation of single proxies with a greater spatial and chronological perspective than previously feasible.

2.3   Summary of Project Scope

The planned project encompasses the entire scope of MAL's data. The database development project proposed in this report will take approximately 54 person-months to complete. This represents a financial investment of around 4.8 Million Swedish Kronor. Approximate annual maintenance costs are also provided in this report.

3 Development Strategy

3.1 Outline

This document represents the completion of Phase 1 of the SEAD developmental cycle – Planning. The full developmental cycle is outlined in Table 3.1, and described in detail in section 3.2.

Phase	Title
1.	Planning
2.	Database Construction
3.	Client Software Development
4.	Implementation, Training & Parallel Running
5.	Importing of Existing Data
6.	MAL Activation & Fine Tuning
7.	Internet Activation
8.	Evaluation
9.	Use & Maintenance

Table 3.1. Proposed SEAD Developmental Cycle

The phases outlined in Table 3.1 are not intended to be chronologically static or exclusive, and will overlap each other in terms of their content and schedule as deemed necessary under development, and as required by MAL staff. Implementation and training, for example, may necessitate improvements to the interfaces that require more programming; and fine tuning of the system during activation may require minor adjustments to the structure of the database. However, the chronological framework for the full developmental cycle will be effectively limited by the logical flow of database systems development, as suggested in Figure 3.1.

From the third month of development departmental staff and research partners will be encouraged to start applying for the funding of research projects where the use of SEAD is an integrated feature.

3.2 Proposed Developmental Cycle

1. Planning

Analysis of existing system, and comprehensive logical design of replacement system (SEAD). This report represents the results of this phase.

2. Database Construction

Essentially the physical implementation of the logical database design, its immediately associated operational interfaces, and the tools necessary for the client software to interrogate the data.

- Construction of physical database, including a entry of a test subset of MAL's data
- Construction of basic database management interfaces (beyond those provided in the developer tools)
- Construction of preliminary stored procedures/queries

3. Client Software Development

Custom built software for the operation of SEAD, from data input and output to maintenance, mapping and internet access.

- Construction of interfaces, query systems, report templates
- Inline and online documentation
- Mapping interface (GIS)
- Development of internet interfaces
- Debugging

4. Implementation, Training & Parallel Running

Instruction of MAL staff on use of SEAD, and testing in lab routines. A full test and training strategy will be devised in cooperation with MAL staff during the first half year of the project.

- Research orientated interrogation of test datasets (simulated research)
- Testing of:
  - import routines (existing data)
  - data entry systems
  - reporting, export & mapping systems
  - internet access
- Design of training manuals

5. Import of Existing Data

Comprehensive import of all existing MAL analogue and digital data, including translation of meta data when appropriate.

- Identification meta data for compatibility with other archives
- digitization of MAL reports
- Translation of Swedish material

6. Documentation, MAL Activation & Fine Tuning

Full implementation of SEAD as data management system for MAL, and devolution of existing system.

- Minor adjustments to database structure if necessary
- Production of user guide and advanced manuals
- Integration into existing and proposed research & consultancy projects
- Application to teaching

7. Internet Activation

Full implementation of the SEAD web interface, with online access to published data within database. (The domain sead.se has already been acquired for this goal).

8. Evaluation

Comprehensive evaluation (with report) on the successes and inadequacies of the SEAD project's developmental phases.

- Usability & functionality (data input & output)
- Input, transaction & database validation
- Fulfilment of initial and later requirements

9. Use & Maintenance

Operational phase, with real time data entry, use in research & teaching, routine maintenance and improvement.

		Year 1												Year 2												Year 3
	Month from Project Start	1	2	3	4	5	6	7	8	9	10	11	12	1	2	3	4	5	6	7	8	9	10	11	12	...
1	Planning (completed June 2006)																									
2	Database Construction	█	█	█	█																					
3	Program Development (Combined)					—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	
	Local Client					█	█	█	█	█	█	█	█	█	█	█										
	GIS (Local & Web)																█	█	█	█						
	Internet Interface																				█	█	█	█		
4	Implementation, Training & Parallel Running													█	█	█	█	█	█	█	█	█	█	█	█	█
5	Import of Existing Data			█	█				█	█	█	█	█	█	█	█	█	█	█	█						
6	Documentation, MAL Activation, Fine Tuning																				█	█	█	█		
7	Web Activation																								►	
8	Evaluation																									█
9	Use & Maintenance																									►

Figure 3.1. SEAD Developmental Schedule. See section 10.2 for time budget estimates. (Note that these phases do not represent employment units or personnel, but rather the estimated time required to undertake each of the developmental phases, rounded to the nearest month, and excluding vacation time).

4 Analysis of Existing System

4.1 *Data Entities*

The data held by MAL is the result of over 30 years of research and consultancy, undertaken by a large number of employees and external partners. This is reflected in the variety of and limited consistency visible in all data types. As a result it is most convenient to classify the data elements encompassed by MAL's current activities into three zones of acquisition/creation: Field, Laboratory and Office. These zones have remained more or less constant throughout the history of the lab., and this classification reflects the *modus operandi* of MAL, rather than a logical division of data based on type and purpose more typical of systems analysis.

An assessment of the current digitalisation status and the anticipated effort required to convert and import the existing data into SEAD is given for each identified data entity group. Table 4.1 provides a key to the assessment scale. Uncertainty due to difficulty of assessing the material is indicated by a question mark. This may either indicate high variability in the data as described in the associated comments, or that the task showed signs of being more complex than the available data at first suggested.

Current level of digitalisation:	0-5
Volume of data:	1-5
Uniformity of data:	1-5
Complexity of task:	1-5

Table 4.1. Assessment scales for digitalisation of existing data
(5 is highest on all scales, uncertainty is marked with a question mark)

A clear Project/Site/Sample Group/Sample structure can be seen in the data handled by MAL, as shown schematically in Figure 4.1. This structure is illustrated with examples, and shown in relation to the proposed database structure for SEAD in Figure 5.2.
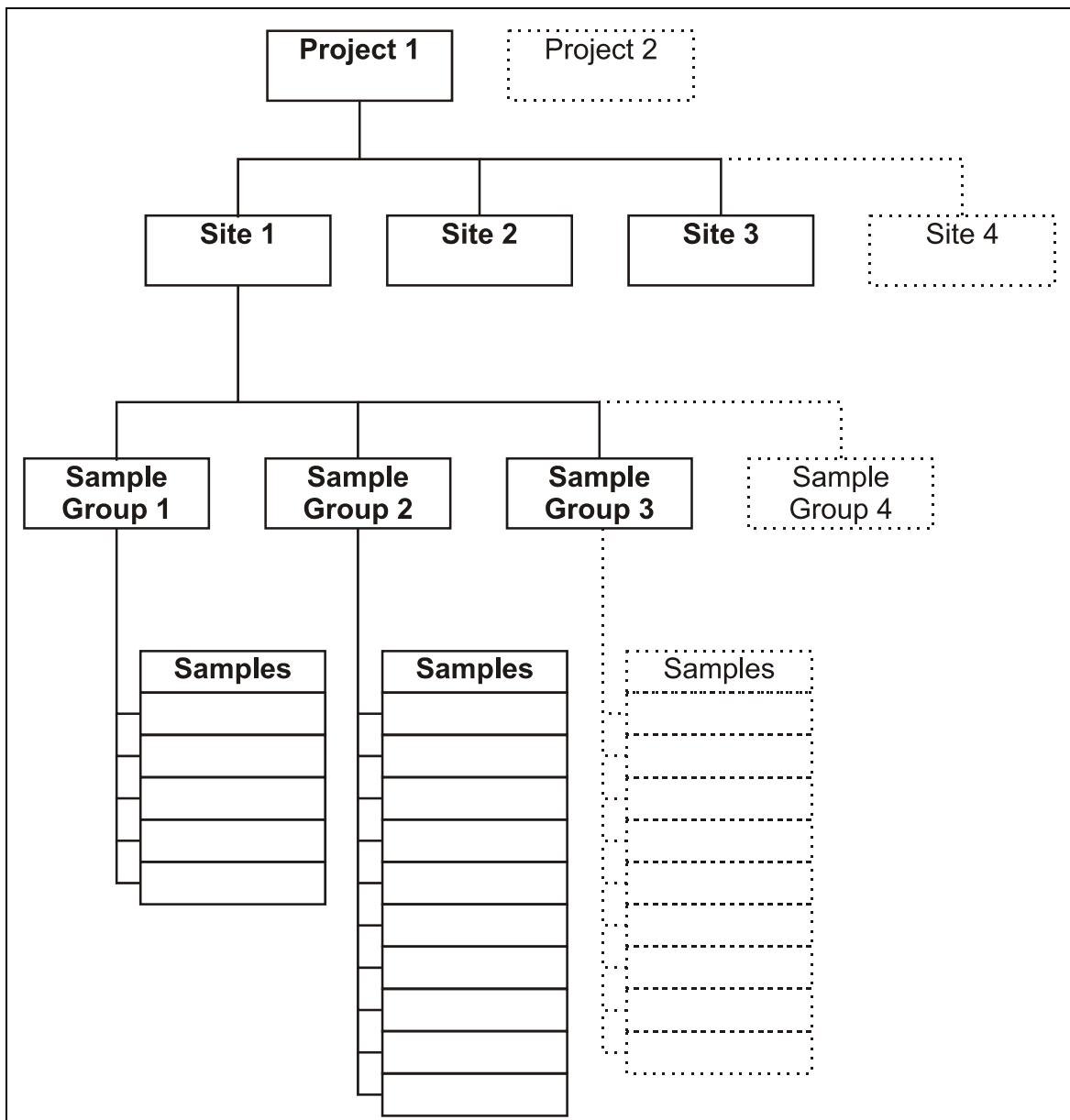
Figure 4.1. General data structure observed at MAL, showing Project/Site/Sample Group/Sample hierarchy.

4.1.1   Field data (internal & external)

4.1.1.1   Summary

A variety of data is collected in the field, from information about individual soil samples, general site information to information about the landscape surrounding the investigated area. Field maps are used to record the location of sample points, archaeological features, variation in soil properties and other data which might be of use interpreting the results of analyses. Additional information is also recorded on the field maps or on the actual sample bags.

Coordinate data (X,Y,Z) is most often produced by a 3rd party (contractor) using a total station. On some occasions these measurements are performed by MAL personnel using a total station or GPS.

Soil samples are marked individually in field using site location, sample number and soil horizon. Field notes are used to describe texture, colour and anomalies such as cultural disturbances in terms of charcoal, burnt clay and similar objects. The marking

of samples collected by external parties (not trained by MAL personnel) can be marked with considerable variation.

4.1.1.2 Component details

Data:	Sample metadata
Comments:	Highly variable and disparate. Many external projects have had questionable sample labelling policies, and ineffective structures. A system must be devised which is both efficient *and* can cater for these problem samples. The effectiveness of the SEAD must not however, be sacrificed in order to cater for badly designed legacy data. Much of this data could be in paper form in structured files, but some is entered into MS Excel files during logging and processing.
Status:	*Current level of digitalisation: 2?* *Volume of data: 5* *Uniformity of data: 1?* *Complexity of task: 5*

4.1.2 Laboratory data

4.1.2.1 Summary

The majority of laboratory data is subject to routines which ensure the creation of appropriate metadata and data elements. There are, however inconsistencies in the storage location and quality of this data, particularly where plant macrofossil analyses are concerned.

An 'Incoming folder' holds a master list of the information necessary for identifying sample batches. All samples handled in the laboratory at MAL are given a unique identification number, referred to as its "MALnummer", constructed by the concatenation of a sample batch number followed by a sample number. This minimizes the possibility of mixing up samples from different investigations. These unique identifiers follow the samples through every step of the process, from pre-processing to reporting and are used for all types of samples processed in the laboratory.

4.1.2.2 Component details

Data related to arrival of samples in lab

Data:	Sample batch information: type, size, date received, external contact and project information. (Incoming folder)
Comments:	System was implemented in 1999 and is used as a temporary store for project meta data. This is digitised for 1999-2003. Can be used as central node for project information (if used consistently), and may be useful when creating project structure in SEAD.
Status:	*Current level of digitalisation: 2* *Volume of data: 2* *Uniformity of data: pre-1999: 2?; post-1999: 4* *Complexity of task: 3*

Proxy data

Data:	Soil chemistry/physical properties
Comments:	pre 1996 data – probably must be manually typed into new template post 1996 data - each MS Excel sheet must be opened, checked, and converted to new template This is a considerable job due to the massive volume of data once converted to standard MS Excel format these will be importable into SEAD relatively easily.
Status:	*Current level of digitalisation: 4* *Volume of data: 5* *Uniformity of data: 4* *Complexity of task: 3*

Data:	Plant macros
Comments:	Degree of digitalisation largely unknown, probably much relatively recent data that must be extracted from tables in Word, earlier data must be scanned of hand typed. Creation of a central species list creation system, with templates for raw data would significantly enhance the speed and accuracy of this work. Systematisation of qualitative descriptors (few, many etc.) is also needed to an extent
Status:	*Current level of digitalisation: 2?* *Volume of data: 4* *Uniformity of data: unknown* *Complexity of task: 5*

Data:	Pollen
Comments:	All data must exist digitally either on server or staff machines. Raw data from the pollen analysis program Tilia can easily be exported to dbf, which can be imported to other database systems or analysis software such as C2. Questions exist as to whether the species dictionary information is easily extractable and whether it could be used as a lookup/central index for plant/pollen species. Older pollen diagrams may have to be scanned and their data reverse engineered if considered of great importance (a very time consuming task).
Status:	*Current level of digitalisation: 5?* *Volume of data: 4?* *Uniformity of data: 5* *Complexity of task:4*

Data:	C14 (Radiocarbon dates)
Comments:	There is a central list of all dates submitted by MAL since 1999, which is digitised up to 2002. Details and metadata (sample type etc.) for dates are however, not digitised. Although how samples are referenced in the analogue records varies to some extent, it is always possible to relate dates to samples by their descriptions. Missing dates may be found in printed reports, although there will inevitably be some loss of data.
Status:	*Current level of digitalisation: 2?* *Volume of data: 3?* *Uniformity of data: 2?* *Complexity of task: 4?*

Data:	Insects
Comments:	Fully digitised since inception in 1996. In addition, the Bugs database contains records of the majority of fossil insect analyses undertaken in Europe (excluding chironomids).
Status:	*Current level of digitalisation: 5* *Volume of data: 1* *Uniformity of data: 5* *Complexity of task: 1*

Data:	Mollusca
Comments:	Raw data primarily analogue, in Word files where it exists digitally. Similar situation to plant macrofossils.
tatus:	*Current level of digitalisation: 2?* *Volume of data: 1?* *Uniformity of data: 3?* *Complexity of task: 5?*

Data:	Wood
Comments:	Raw data primarily analogue, in Word files where it exists digitally. Similar situation to plant macrofossils.
Status:	*Current level of digitalisation: 1?* *Volume of data: 1?* *Uniformity of data: 1?* *Complexity of task: 5?*

Data:	Other
Comments:	Minimal osteological analyses are undertaken at MAL, and only a limited amount of data stored outside of reports.
Status:	Cooperation with other national labs (Stockholm), institutions (Lund) and consultancies where osteological analyses are undertaken is routine. These links could be explored with an aim towards database cooperation (see section 8)

Analysis methods

All methods are systematically described and stored on the central server as MS Word files. The is part of MAL's quality assurance policy.

4.1.3   Office data

4.1.3.1   Summary

Office data creation most often occurs before and after the above data types. Much of this data may be considered metadata and helps to describe the analysis data and put it in context. It is therefore often hierarchically above field and lab data.

Feature data produced by contractors or collaboration partners contains important information on topography, geography and structures found when excavating archaeological sites. Archaeological information such as shape, volume, construction, finds and preliminary interpretation and topographic information are very useful when working with environmental archaeological samples.

Site and sample coordinate data is often provided by collaborative partners, and must sometimes be converted from their local investigation grid to Swedish national grid or latitude/longitude.

4.1.3.2   Component details

Pre-project stage

Data:	Preliminary: project description, preliminary contract specifications
	Background material
	Negotiation notes, phone call logs
	Cost estimation
	Management Decisions
Status:	Currently ad hoc - template and routines exist but are not in use

Contractors & external contacts

Data:	Lists of contractors and previous collaboration partners
Status:	No central list exists, information is mainly held by those individuals responsible for the related projects

Project/Site info

Data:	Project/subproject/site hierarchy
	Project descriptions and definitions (e.g. research & contract)
	Project completion stages
Status:	Structure of major projects usually reflected in file structure on server and individual computers
	No central storage of project descriptions, info held (if at all) on PC of individual responsible for project
	There is no current system for uniquely identifying Projects or Sites. MALnummer is sometimes used as a Project number, but see notes above.

Scientific publications & MAL reports

Comments:	The majority of MAL reports are stored centrally on the server, although not always as single documents. There is a central digital list of all reports after 1996. Reports from before this date will be more difficult to track down.
	Scientific articles are not centralized and are located on the computers of individual researchers
	Centralising these as Adobe pdf files on the server is recommended, so that they can found and distributed more easily.
Status:	*Current level of digitalisation: 3*
	Volume of data: 2
	Uniformity of data: 4
	Complexity of task: 2

Report templates

A standard MAL report template exists, but its use is adapted to individual preferences.

There are only templates for the presentation and publication of insect and pollen proxy data. There is a reasonable amount of systematization in the reporting of soil chemistry data which could be used to produce templates. No templates exist for the other proxies.

4.2 Process (Data Flow) Models

Process models describe the movement of data through the operational structure of the lab., from creation/acquisition through analysis to reporting.

The routines used to handle the enormous amounts of data produced at MAL are a result of many years of method development and laboratory research. In the early years of MAL the vast majority of data were analogue (in creation and storage), but the use of PC:s in daily activities has gradually become indispensable. Software, including MS Office and specialised utilities for controlling analytical instruments assists the creation, storage and manipulation of data. All analytical methods are thoroughly documented and stored on a server to ensure the high standard of quality required by scientific research.

As a result most aspects of the lab's routines are well organised, even if the implementation of routines may vary. The following sections describe the current state of each process, paying special attention to the degree of computerisation. Relevant suggestions for improvement are given where deemed appropriate, and incorporated into the design Section 5 later. Section 5 also includes flow diagrams describing the optimised versions of the routines examined here in a form suggested for implementation through SEAD. The descriptions below are a summary of an ongoing and comprehensive analysis of MAL's operations.

4.2.1 Data acquisition/creation

4.2.1.1 Summary

Two groups of data acquisition/creation can be identified: manual created and digital acquired.

Manually created data

Macro fossil data, pollen data and pH analyses are recorded on paper as analyses proceed and are later entered into non-standardized MS Excel sheets or MS Word

documents. Less common proxies such as wood type and molluscs are processed in the same way. Insect data are recorded on pre-printed recording sheets during analyses and subsequently entered into the Bugs2000([9]) database for storage and interpretation.

Digitally acquired data

Magnetic susceptibility (measured with a Bartington MS-2 meter) and Loss on ignition (using a analytical laboratory scale connected to a PC) analyses create data directly in template based MS Excel sheets. Phosphate content is measured using an Ultrospec 2000 UV/Visible Spectrophotometer which, through specialised software (Swift II) produces results files in a format that is easily converted to MS Excel format.

4.2.1.2   Component details

Data related to arrival of samples in lab (see Figure 5.15, Figure 5.16 & Figure 5.17)

Data:	Sample batch information: type, size, date received, contact information.
Process Status:	Logged in 'Incoming folder' (*inkommande pärm*) per batch, individual samples are not logged until processed.
	Sample batches are assigned a unique identifier which is the first part of the MALnummer (see above). There is some inconsistency in the use of the MALnummer in that they are often used as Project numbers. Whilst this works for small projects with only one batch of samples, it is not workable for multi-batch projects and is liable to cause significant confusion.
	Almost all samples are recorded in this way, although a few single sample projects may slip through the system and be only recorded on the computer of the individual responsible for the analyses.

Proxy data (see Figure 5.20, Figure 5.21 & Figure 5.19)

Data:	Soil chemistry
Process Status:	Automated data acquisition to MS Excel file from Spectrophotometer (Phosphates) & Bartington (Magnetic Susceptibility).
	Standard/calibration values for P are recorded digitally.
	Manual input of LOI, Ph, sample weight data into MS Excel sheets
	Currently no recording of useful qualitative (Grinding method, Calc. lumps, iron concretions, charcoal, artefacts, etc) & semi-quantitative properties (grain-size fractions). There is a need for a flexible logging system during sample processing.
	Currently no facility to log changes/corrections to data (for back-tracking and quality assurance).

Data:	Plant macros
Process Status:	Currently no standard computerised logging of data, individual researchers work in their own routine. Quality of data is ensured by professional expertise, but there is little coordination of metadata to enable raw data, results and project status to be found by other staff.
	Null results (samples without any macrofossils) are often not recorded.
	Suggest the use of standard forms, generated by SEAD for recording.
	Database input of species lists would ensure conformity of styles and taxonomy.

Data:	Pollen
Process Status:	Majority of analyses performed off-site. Samples sent for analyses come back as results in Tilia Graph format, but Excel compatible raw data files are available. Test analyses are most often returned in paper form, or as emails. Storage of raw pollen data in SEAD would be of great research importance, but not necessarily improve the data creation phase

Data:	C14 (Radiocarbon dates)
Process Status:	Metadata on samples sent for dating are recorded on standard paper forms, the index to which has been digitised to an extent. Results are returned in either paper or email form, and stored in physical folders. There is no systematic routine for the digital storage of the data.

Data:	Insects
Process Status:	All raw data are by routine digitally stored in the Bugs2000 database.

Data:	Less common proxies (wood, molluscs, osteology)
Process Status:	No routines are implemented for the digital storage of the data. Routines are similar to those described for plant macrofossils above. Osteological raw data is handled by external partners.

4.2.2   Data manipulation

Data are manually moved between data acquisition and analysis and report writing machines. Automated routines for backup exist, but are only sporadically implemented due to lack of resources.

Calculations, statistics and graphing are chiefly undertaken in MS Excel and to some extent in other software including Golden Software's *Grapher* and Statview Software's *Statview*.

Spatial data is analysed in Golden software's *Surfer* or ESRI's *ArcView*, with which soil chemical and topographic maps are produced. GIS problems are solved in *ArcView*.

Data handling is standardized to some extent. Forms exist for all the current methods but they are not always used and they can be changed when there is a need for it. This has resulted in a system where data may exist in a variety of forms, and there is limited conformity between data areas.

All soil chemical data post 1996 exists digitally in MS Excel format. The layout and form of the underlying template document has, however, changed over the years. In particular, columns have been added, moved, renamed or deleted as required, and the placement of metadata within the files varies. This is not an ideal situation for database import, and requires that *every* Excel sheet be examined before import.

Field maps are digitized using Golden software's *Surfer* or ESRI's *ArcView*. Sample points, features and soil properties are either used directly in these programs or exported to dbf format for later use in MS Excel.

4.2.3   Data storage

4.2.3.1   Manually created data

Data is usually stored on the computer on which the researcher is working, in most cases located in the office of the individual member of staff. Guest researchers and students are often allowed to store data centrally on the lab's server, but data can often remain locally stored.

4.2.3.2   Digitally acquired data

Data are initially stored locally on the PC that the particular analytical machine is connected to. When the analytical procedures for a sequence of samples are completed data are compiled into a single MS Excel file and moved to a specific computer (or the central server). This file is named following a naming convention based on the project number (MALnummer) mentioned in section 4.1.2. The file is later distributed to the person(s) who will write the report.

4.2.3.3   Common data & backups

Photos from sites showing terrain, vegetation, soil profiles, stratigraphy and features (both geological and archaeological) are stored on a PC dedicated to image processing.

A centralized backup system secures the data. The current system is DAT tape based and is suitable for small scale laboratory analyses. It is not suitable for large scale photo documentation and field map digitalization. Backup of the latter kind of data are managed by moving data regularly to other hard drives. The existing backup routine is not currently enforced by the MAL staff, and there appears to be a serious deficiency in this aspect of data security.

4.3   User & Data Requirements

The necessity of a database system has become more and more evident during the past years. MAL has produced about 150 000 (non-normalized) data records and these are stored in about 2000 Excel documents and a number of word documents. A database would greatly help solve problems of data integrity and process uniformity, and accelerate data retrieval, analysis and interpretation.

What follows is summary of requests, ideas and constraints on the form of SEAD as dictated by both the perceived needs and requests of the prospective users and the data themselves. The former were obtained by a combination of personal experience, questionnaires, meetings and interviews with the clients. The authors have also included projected user and data requirements as predicted from the users that were not encompassed by the latter, and the assessed potential user base of SEAD.

The immediate potential user base for SEAD can be described as poorly computer literate, and although there is a general understanding of the benefits of database storage there is limited understanding of the potentials of such as a research and teaching tool. The authors stress that effective implementation of Phase 4: 'Implementation, Training & Parallel Running' (see section 3.1) along with close cooperation between MAL staff and the developers is *essential* in order to rectify this problem. There is however, a thorough understanding of the benefits of making the data held at MAL easily available to the wider research community.

Comments considered of particular importance are summarised below, with developer's remarks where the user's comments differed significantly from the developer's understanding of the project. All comments were considered during the design of the proposed database structure and user interface (section 5).

4.3.1 Data/database scope

"SEAD should encompass the entire scope of the lab's analyses."

"Descriptive sample data must be included (size, volume etc.)."

"Links to electronically stored reports would be useful."

> "Ecological and descriptive data for plant species is *not* necessary due to the ease of access to the literature in the lab and the expertise of MAL staff."
> Developer's comment: *Whilst this may be valid for researchers with a high level of expertise, this data would potentially be of great importance for students and researchers with less knowledge of plant ecology and ethnobotany. The data would also increase the advanced research power of the database by making it possible to query by combinations of vegetation groups, environment and ethnographic information. This principle has been developed in the Bugs Coleopteran Ecology Package[1].*

"Descriptions of archaeological artefacts are outside the scope of SEAD."

"Detailed description of sites and sampling localities is not necessary."

4.3.2 Data entry

"MAL should act as a clearing house for all SEAD data entry. This would ensure that data quality rules are uniformly enforced, and simplify administration. Online interfaces could be developed to enable off-site data submission."

"Routines should be put in place to automate the movement of data from analysis machines to the central database."

"It must be easy to create and edit sites, and project structures should be transparent and understandable. Simple project management facilities would be useful"

"A flexible import interface should be designed to help import the variety of existing soil chemistry/property data. For example, a guided wizard type system."

"Plant macrofossil data should be input directly into SEAD from paper recording sheets. A spreadsheet like interface with samples as columns and species as rows is essential, and species should preferably be selected from a master list to avoid mistakes."

"It must be possible to import pollen data from files exported by the pollen software *Tilia* (Eric Grimm), and possibly *C2* (Steve Juggins)."

"Insect data could be entered into Bugs, and then imported into SEAD."

"Uncommon proxies (wood, molluscs) should be input directly into SEAD at a later date."

4.3.3 General interface design

"Must be easy to use."

"Should be flexible/customisable to an extent so that users with different levels of computer literacy can comfortably use the system."

"Users should be able to save settings so that their preferred (or most commonly used) interface parts are always readily available on their client machine."

4.3.4 Data retrieval

"It should be easy to pose simple data retrieval questions to SEAD, such as:
- Are there proxy data of Type T (e.g. phosphates, plant macrofossils) from Site X?
- What kind of proxy data exist for a specified region/site/context?

- Where and when is a specified species known from?
- What is the preservation status of material from a specified place?
- What is the dating evidence for Site X?"

"Summary data must be readily accessible – e.g. calculated information such as the number of samples from a specified location. It would be advantageous to be able to obtain multi-site summary data in a form that can easily be graphed."

"The investigation of complex research questions should not require advanced computer/database skills."

4.4   Previous Database Related Work at MAL

The following projects and tasks have been under either partly or entirely by MAL, and can contribute to the development of SEAD.

4.4.1   Bugs Coleopteran Ecology Package

Database and analysis software for entomology and palaeoentomology.

Latest version developed by Phil Buckland as part of his PhD project at MAL and in cooperation with partners in the UK, see http://www.bugs2000.org/ and Section 7.5 for details. PhD project funded by The Bank of Sweden Tercentenary Foundation within the Northern Crossroads project, along with external research partners in the UK (Paul Buckland, University of Bournemouth).

4.4.2   Plant macrofossil database preparation

Initial steps towards the construction of a database for MAL's palaeobotanical work have been taken, including: rough database design; identification of essential needs and data components; cursory identification of existing similar systems and taxonomic codes.

Work undertaken by Tina Nilsson, supervised by Phil Buckland, and with earlier involvement of Roger Engelmark and Karin Viklund.

4.4.3   ProvBase

ProvBase was created to store status information for consultancy projects (Figure 4.2). The system is fully functional at the Project/Site/Sample level, but contains only a small data set as limited resources prevented its full implementation. It has the ability to store sample metadata, record the status of analyses and store links to raw data files along with project contact details. It currently contains no facility for the internal storage of raw data or reporting/exporting facilities. ProvBase was designed primarily to retrieve sample and status metadata for existing projects.

Developed by Phil Buckland with the assistance of Roger Engelmark, Johan Linderholm and Johan Olofsson, and funded internally by MAL.
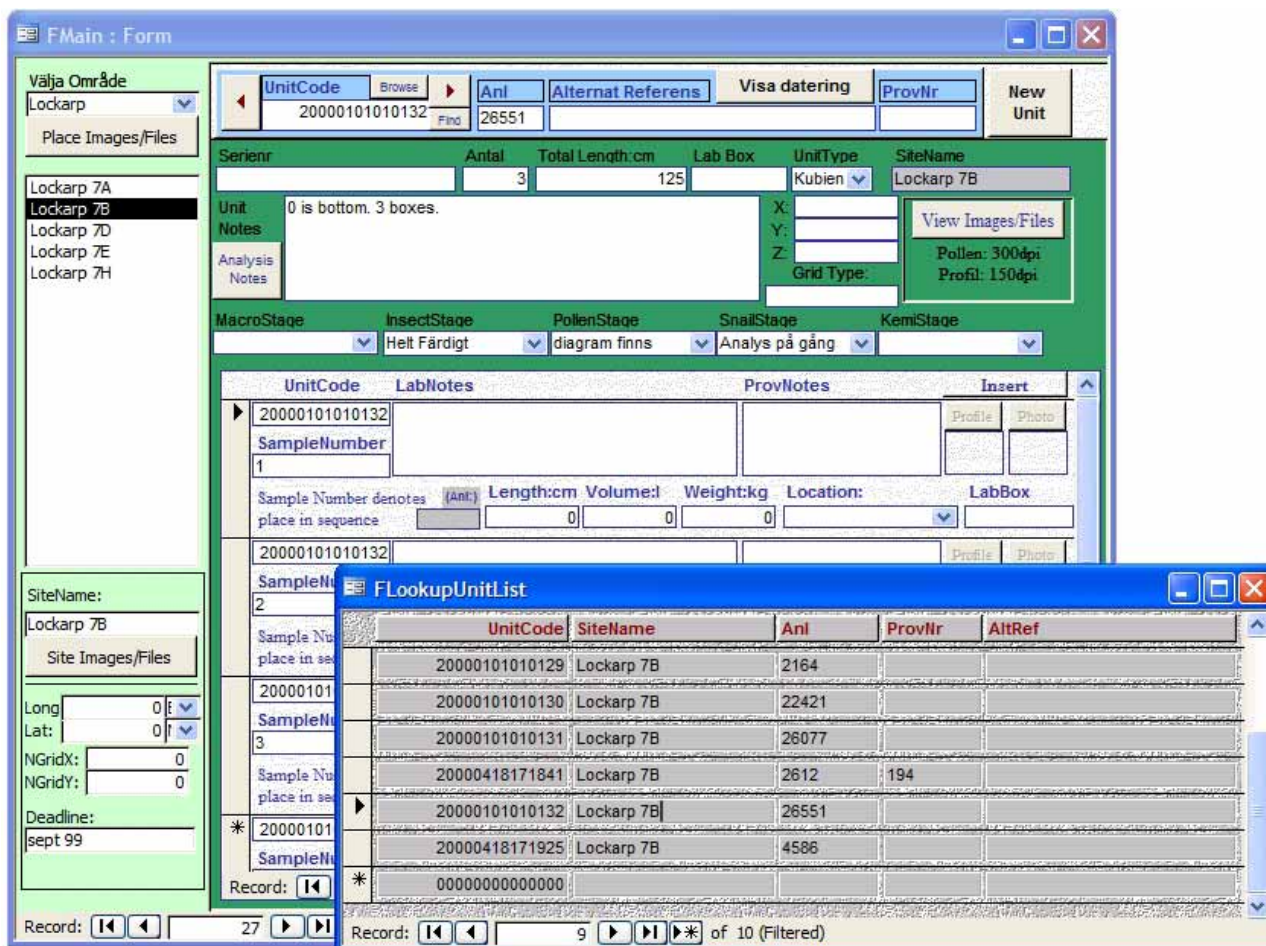
Figure 4.2. Screenshot from ProvBase, showing the main Area/Site browsing interface and the Unit (c.f. sample group) browser popup.

4.4.4 MALdb/MALBase

A prototype project and site management system was created, building on the work undertaken for ProvBase (above 4.4.3), and designed to form the basis for the project management aspect of a system such as SEAD. The 'Browse by Project' form shown in Figure 4.3 shows a list of projects and summary data on the left. Selecting a project shows a list of sites within that project, from which details of each site can be reached by clicking them. The natural browsing progression in SEAD would then be to access sample group and proxy data for the site by the click of a button.

Initial development funded internally by MAL, with subsequent prototype testing under the SEAD planning project. Designed by Phil Buckland. MALdb is evaluated in section 6.2.1.
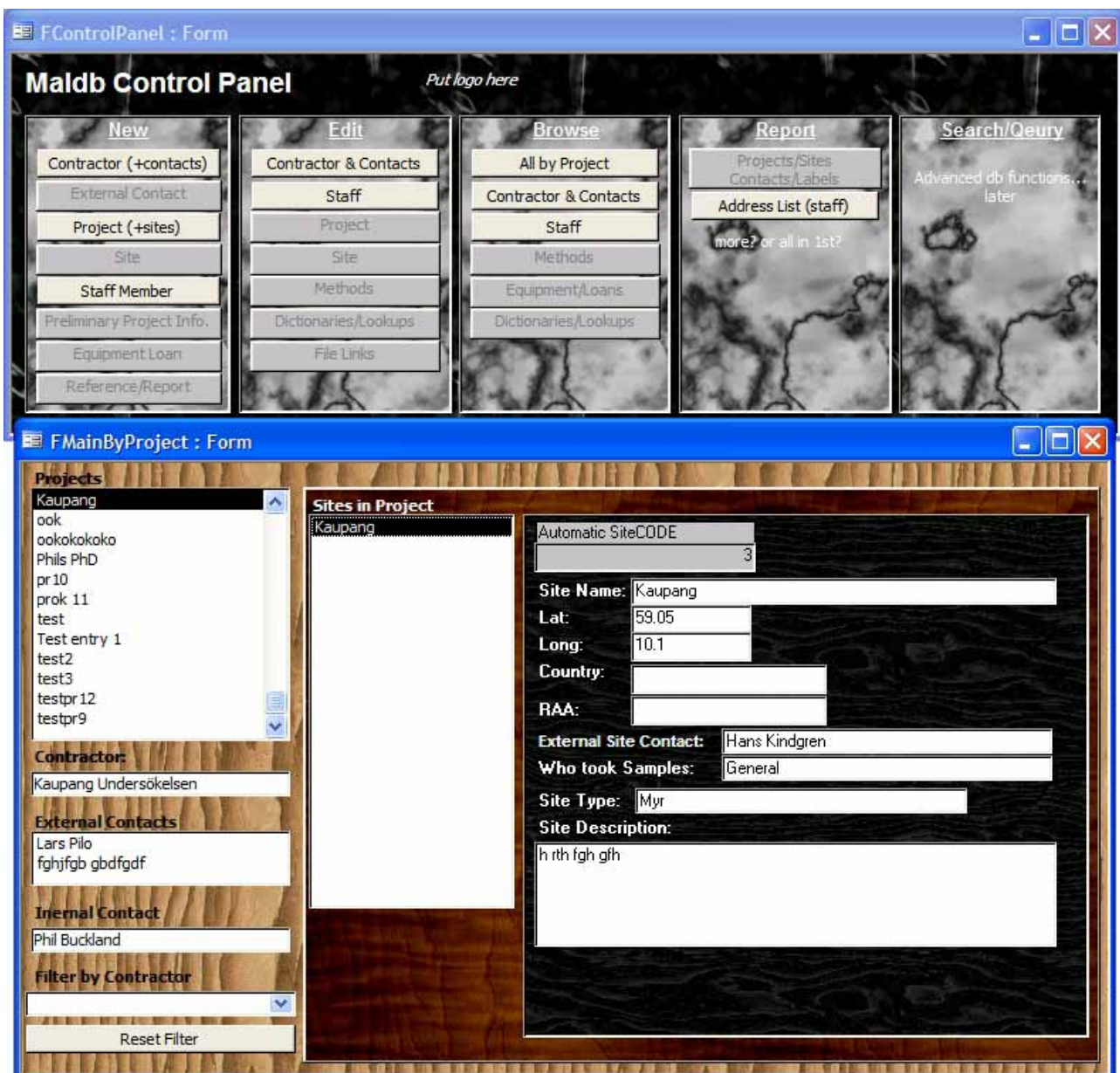
Figure 4.3. Screenshot of the MALdb prototype project management database system, showing the Control Panel used to access different interface areas (top) and the 'Browse by Project' form (bottom). Note that the somewhat unconventional backgrounds are suggestions only!

4.4.5 Report archive

The in-house digital archiving of all MAL reports has been initiated, but resources are limited for its continuation. These reports will be made publicly available through the lab's website. See 4.1.3.2. – Scientific publications & MAL reports.

4.4.6 Image database

MAL has acquired software for the manipulation of a large scale image database (Nikon EclipseNet), connected to a digital microscope camera (Nikon DS-U1). An estimated 20,000 digital images are currently stored at the lab, but have yet to be either linked to the database or assigned metadata. This database system could potentially be linked to SEAD, although the actual implementation of the image database is considered to be outside the initial scope of the SEAD project. *This exclusion should be reappraised during funding application preparation.*

The data consists of a large number of photographs relating to projects (sites, excavations, soil profiles), general reference (landscape types, settlements & structures, landuse) and reference collections (insects, seeds). The systematic photographing of the insect reference collection has been assessed and a pilot study undertaken by Tina Nilsson and Phil Buckland.

The imaging and image database resources are a considerable and valuable investment that should be more fully exploited.

4.5 Potential Users

Several discrete but related initial user groups have been identified for SEAD:

4.5.1   MAL staff

Employees of MAL will be responsible for the bulk of data entry, both during the development of the project and during the operational phase. Researchers at MAL will conduct research, and be involved in teaching using SEAD.

4.5.2   Students on location at MAL

SEAD will be integrated into environmental archaeology taught courses at the basic and advanced levels. Postgraduate students will also be encouraged to use the facilities for both storage of their own data and as a research tool.

4.5.3   Guest researchers on location at MAL

Guest researchers will be encouraged and invited to conduct SEAD based research.

4.5.4   Local interdisciplinary cooperation

SEAD has great potential for use in disciplines from environmental science to cultural history. Multidisciplinary projects will be encouraged primarily, but usage by users from single disciplines will also be most welcome. Researchers from Quaternary sciences have repeatedly expressed an interest in a database of environmental archaeology, and in collaborative research with archaeologists, especially in the areas of environmental change and human impact.

4.5.5   Remote researchers and students

The web interface will enable SEAD based research to be conducted remotely over the internet. Remote users will be registered in order to keep track of the extent of remote usage and provide better support.

5 Database Design (Logical)

The proposed structure of SEAD consists of a Microsoft SQL-Server relational database located on a central server PC. Lab users will interact with this backend (see Figure 5.1) through locally installed client software custom built in Microsoft Access, Visual Studio Tools for Office and VBA (or alternative as deemed appropriate at time of construction). Internet users will query and retrieve data through a combination of HTML, Flash and ASP.NET based web pages served from a replicated copy of the backend (see restrictions 5.4.3) hosted by a third party service provider (Figure 5.1).
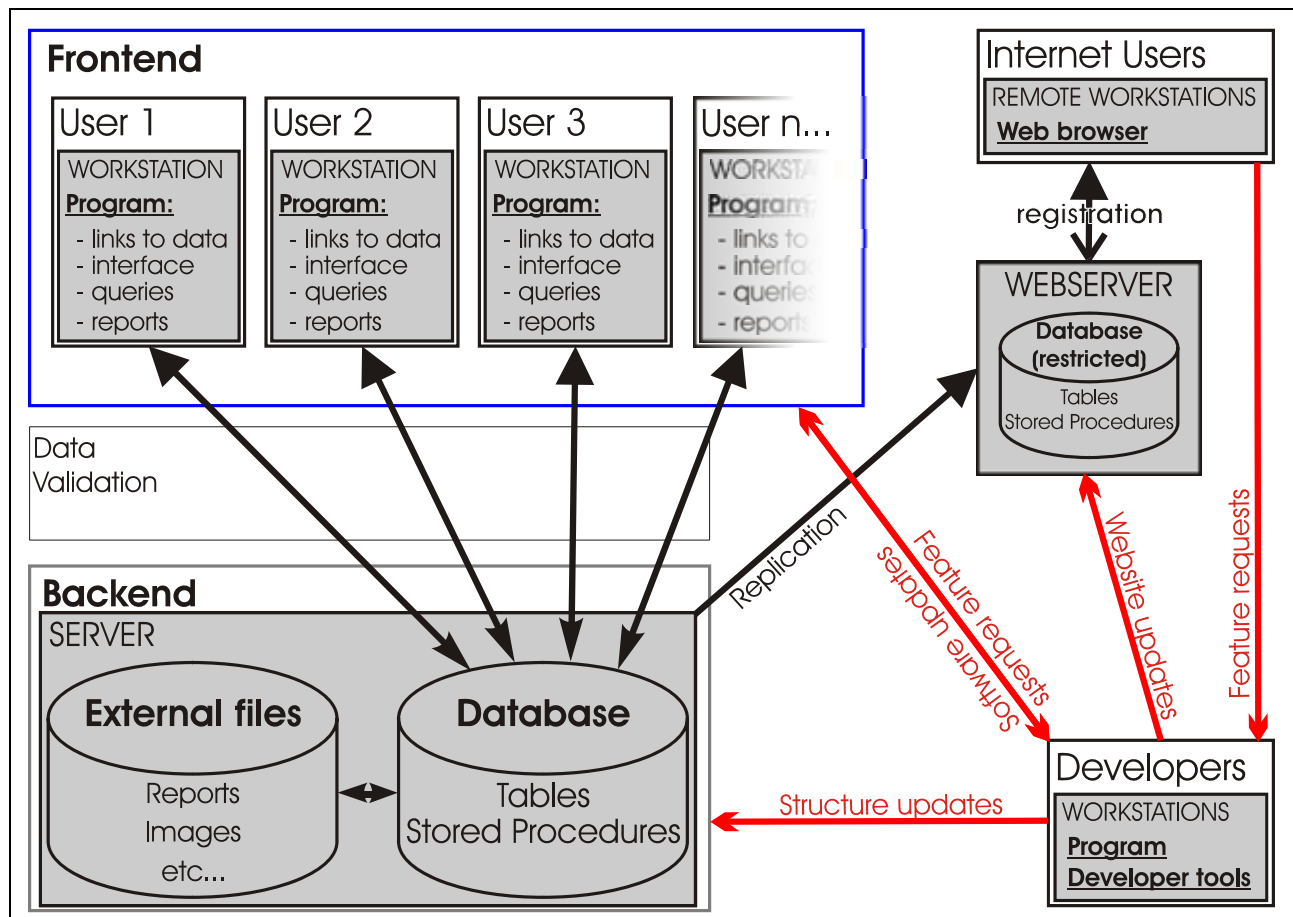
Figure 5.1. Suggested Client-Server-Web structure of the SEAD database system, including developer interaction. The latter will be minimal during the use and maintenance phase.

5.1 Data Structure & Relationships

5.1.1 Introduction

A relational database is, in simple terms, a database designed so that there is no ambiguity over access to data items, no duplication or redundancy, and where data items are linked through a logical system of primary keys and indices which also enforce integrity rules on the data. (Redundancy is the repetition of identical data in a table – in other words, the inclusion of more data than is strictly necessary). By following these guidelines a database can be constructed that is robust and secure in terms of read/write/delete access, and where relatively simple query expressions can be

used to retrieve any combination of linked (related) data items. The process of reorganising data structures into relational database form is known as *normalisation* ([2]).

Environmental archaeology data is by its nature multi-proxy - it consists of several measurable quantities that can give a greater insight into climate and environmental change and human impact together than they can independently. It is also chronologically and spatial extant, and thus forms a multi-dimensional data domain which must be normalised into an efficient data structure if it is to be accessed efficiently. Although the domain is complex, it is logically structured, and thus lends itself relatively easily to normalisation.

MAL's existing data was examined (Section 4.1) and structured according to principles of data normalisation. A small amount of redundancy (denormalisation) is suggested in some peripheral tables in order to improve data retrieval efficiency.

5.1.2 Data area definitions

The proposed structure consists of over 120 related tables, broken down here into manageable data areas. Each area is centred on a nodal table, and reflects the hierarchy of data within MAL projects. The data areas are shown schematically in Figure 3.1 and outlined in Table 5.1.

Note that although SEAD is not designed as a project/laboratory administration system it incorporates all the features of the existing MAL project administration database (MALdb, see 4.4.4). This enforces a data hierarchy which facilitates both efficient data retrieval and administration in keeping with relational database design, and meshes with existing lab practice.

A somewhat compressed image of the full structural diagram is included in Appendix I for orientation, and readers should refer to Section 5.1.5 and the prototype files stored at MAL for full details.
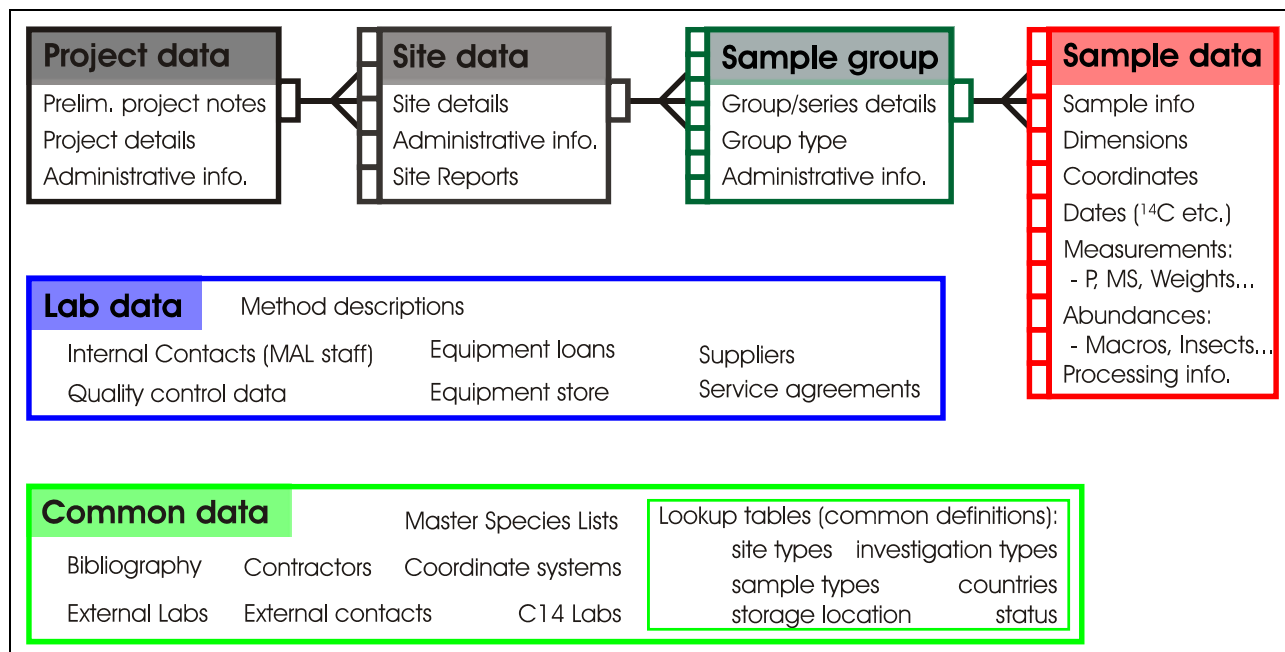
Figure 5.2. Schematic of SEAD data areas, showing the direction of one-to-many relationships as defined by the data hierarchy. (Component lists are non-exhaustive, refer to Section 5.1.5, Appendix I and the prototype files stored at MAL for more details)

Level	Area	Definition – data centred around...
1	Projects	The highest level of administrative organisation, and the first to be defined on project initiation.
2	Sites	Sublevel of Project, may be >1 Site per Project. Data centred on an investigation location, a spatially defined investigative area within a project.
3	Sample Groups	Sublevel of Site, may be >1 Sample Group per Site. Data centred on a group of samples from a site with a common relationship and purpose, for example: a collection of samples from a house foundation taken for plant macrofossil analyses; a collection of spatial samples for soil chemical prospection. Includes (meta)data specific to a cohesive group of samples. (Note that feature/context descriptions and interpretations may not necessarily apply to these same sample group units. It is a feature of variable archaeological excavation systems that sample groups may intersect several features/contexts.)
4	Samples	Elements of a Sample Group. Samples belong to only one Sample Group, and a Sample Group is described exhaustively by the samples that comprise it. An individual sediment unit which is treated with analyses, and the raw and meta data associated with it. *This is the bulk of SEAD's data elements.*
-	Lab Data	Data specific to the Environmental Archaeology Lab.
-	Common Data	Data not specific to the lab, but utilised in various parts of the database.

Table 5.1. Outline of SEAD Data areas schematised in Figure 5.2 (and Appendix I). 'Level' shows the hierarchic position of the data 'Area', where 1 is the highest level. See Section 5.1.5 for structural details.

5.1.3   Bilingual interface and searching

The SEAD local client program will be implemented with a bilingual Swedish and English interface. This will allow for a wider spectrum of users, from Swedish undergraduate students and local researchers to international guest researchers. As the majority of raw data is numerical values/abundances, the main concern is with user interfaces, search terms and metadata. This will be solved by the language duplication of relevant names and descriptions, as shown in the examples in Table 5.2, and the creation of parallel English and Swedish forms within the frontend. Users will be able to choose the interface language on start-up, and switch from within the program.

In addition, the facility to store common names of plants and animals (in the biological proxy master tables – see Figure 5.9) in any number of languages will greatly ease international work, and interaction with the general public .

The web interface will initially be in English only, as it is believed that this will be the common language for the predicted user base.

Field	Example 1	Example 2
AnalysisMethodID	1	2
TechTermEng	Geochemical prospection	Palaeoentomology
MethodNameEng	Soil Chemical Prospection	Fossil Insect Analysis
MethodDescriptionEng	The use of soil chemical methods for the identification of spatial variation with an aim towards identifying possible sites of archaeological interest prior to excavation.	The use of (sub)fossil insect remains as proxies for past environmental and climatic conditions and changes.
Units	Various	Abundance
UnitsAbbreviation		
TechTermSwe	Geokemiskaprospektering	Paleoentomologi
MethodNameSwe	Prospektering med markkemi	Insektsanalys
MethodDescriptionSwe	Användning av variationer i markkemiska spår för att identifiera möjliga områden av arkeologisk interesse innan en utgrävning.	Användning av (sub)fossila insektsrester som proxikällor för förra miljö och klimat förehållander och förändringar.
InternContactID	5	6

Table 5.2. Table structure and data examples for enabling bilingual interface and searching. The example shows the structure of the table tblAnalysisMethods, which holds details of the scientific methods used in environmental archaeology analyses.

5.1.4 Data security

As SEAD will inevitably contain unpublished data from ongoing contracts and research, a security system must be put in place to prevent unauthorised access to this data. The proposed structure allows data to be assigned access levels at the Sites level by users with administrative or developer access rights. The facility of SQL Server to enforce security levels for virtually any object (table, form, query etc.) will enable this policy, and others, to be rigorously enforced.

5.1.5 Data structure

Tables and relationships for the MAL project database (see Sections 4.4.3 & 6) were prototyped using MS Access 2000. This structure was then further normalised and expanded into the full SEAD database using MS Visual Studio's SQL-Server database project interface. The resulting design has over 120 interrelated tables fulfilling the large part of SEAD's anticipated data requirements, and representing all areas of the lab's activities (see Appendix I). The large number of tables and complex linked structure is a feature of a highly normalised relational database, which improves data storage and retrieval efficiency.

The data areas defined in section 4 and outlined above are shown in more detail in this section. Refer to Figure 5.3 for an explanation of the database structure diagrams, and Appendix I to see how the areas are related.

Database tables are linked by one to many relationships where the data held in the *primary key field* is unique for every record in that table (the 'one' side of the relationship). The field with the same name on the 'many' side of the relationship may contain several records where this value is identical, and is referred to as the *foreign key*. For the example shown in Figure 5.3: in tblTaxaInsectGenera the beetle Genus 'Carabus' has the unique GenusID=3, and in tblTaxaMasterInsects there are several records where GenusID=3 which represent the species within the Genus.

Specific attributes, such as field types, sizes and restrictions will be not be presented in this document, and should be considered in the Database Construction phase (see section 3).
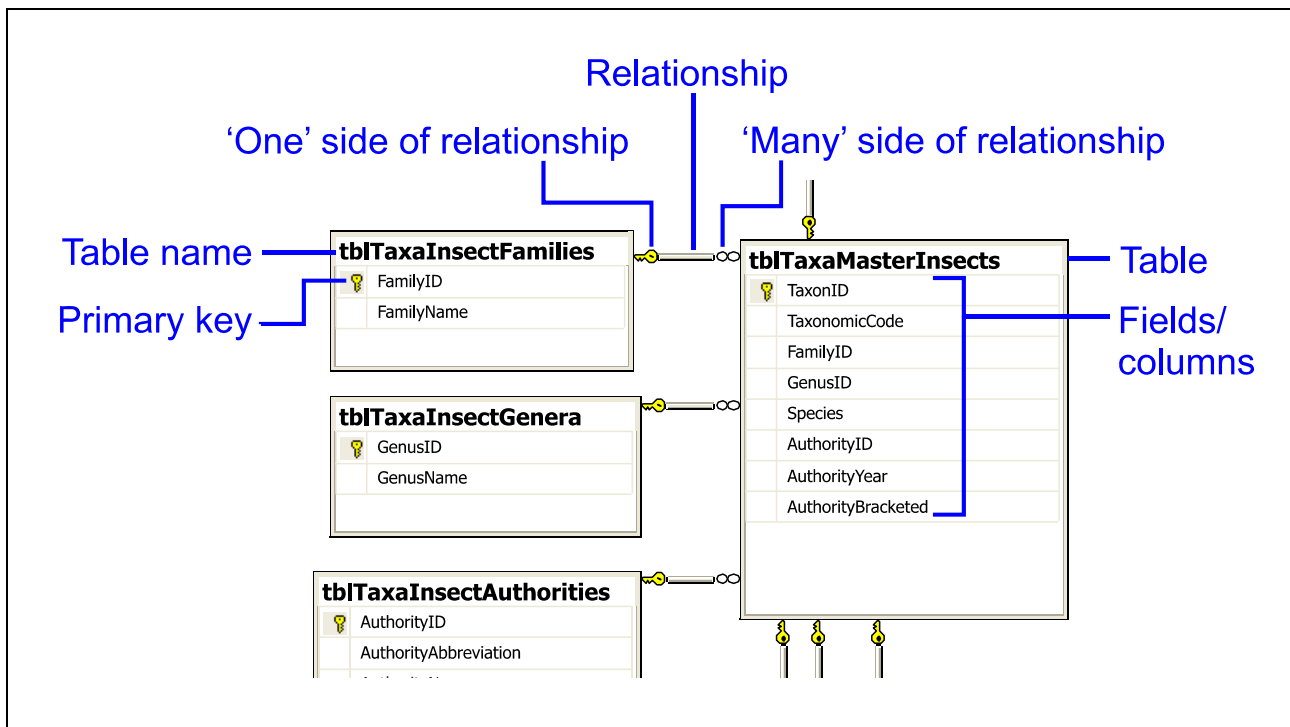
Figure 5.3. Explanation of database structure diagrams.

5.1.5.1 Project data

The project data area is essentially the same as that prototyped in MALdb (see sections 4.4.4 & 6.2). The structure in SEAD has been refined to enable better management of project level metadata, and enforce the data hierarchy as defined in section 5.1.2.
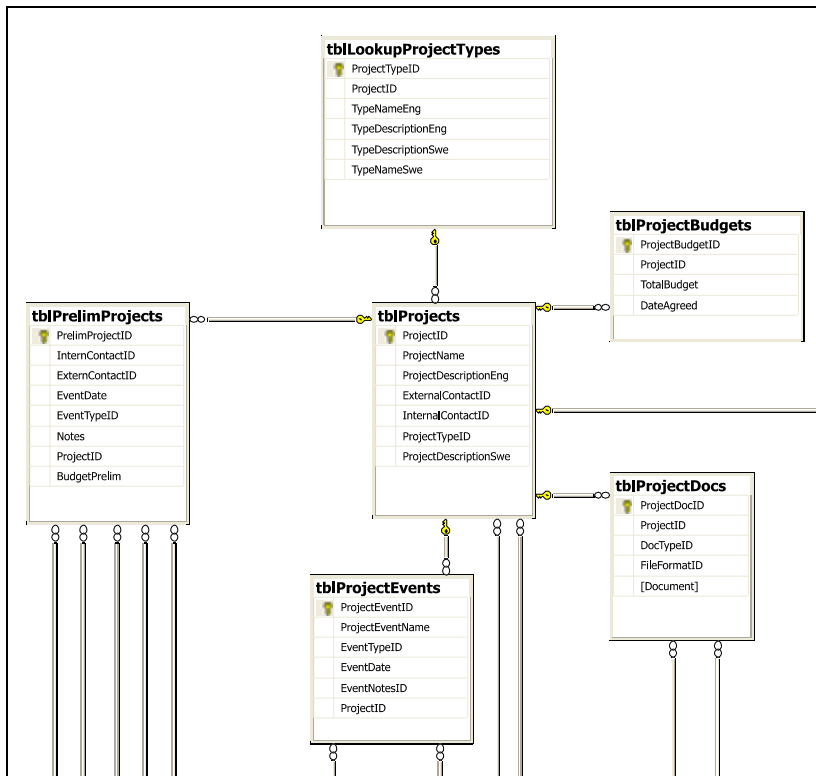
Figure 5.4. Structure of Project data area

5.1.5.2   Site data
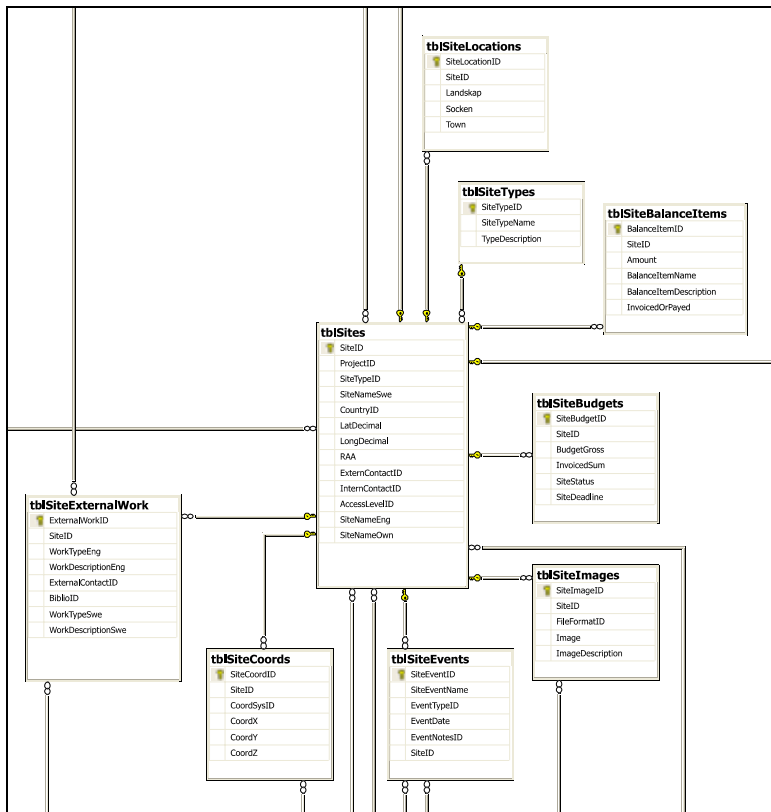
Figure 5.5. Structure of Site data area

5.1.5.3   Sample Group data
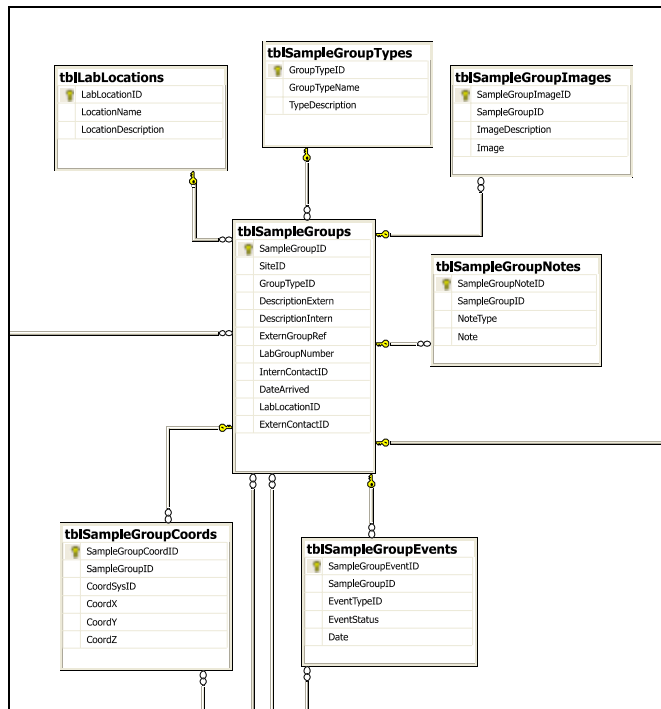
Figure 5.6. Structure of Sample Group data area

5.1.5.4   Sample data (including dating evidence)

Variations in the excavation and sampling practices of external bodies (generally archaeological contractor) is catered for by the use of both external and internal references being held in separate tables.

Tables for molluscan proxy data (tblAbundancesMolluscs; tblAbundIdentLevelsMolluscs) are omitted for clarity.
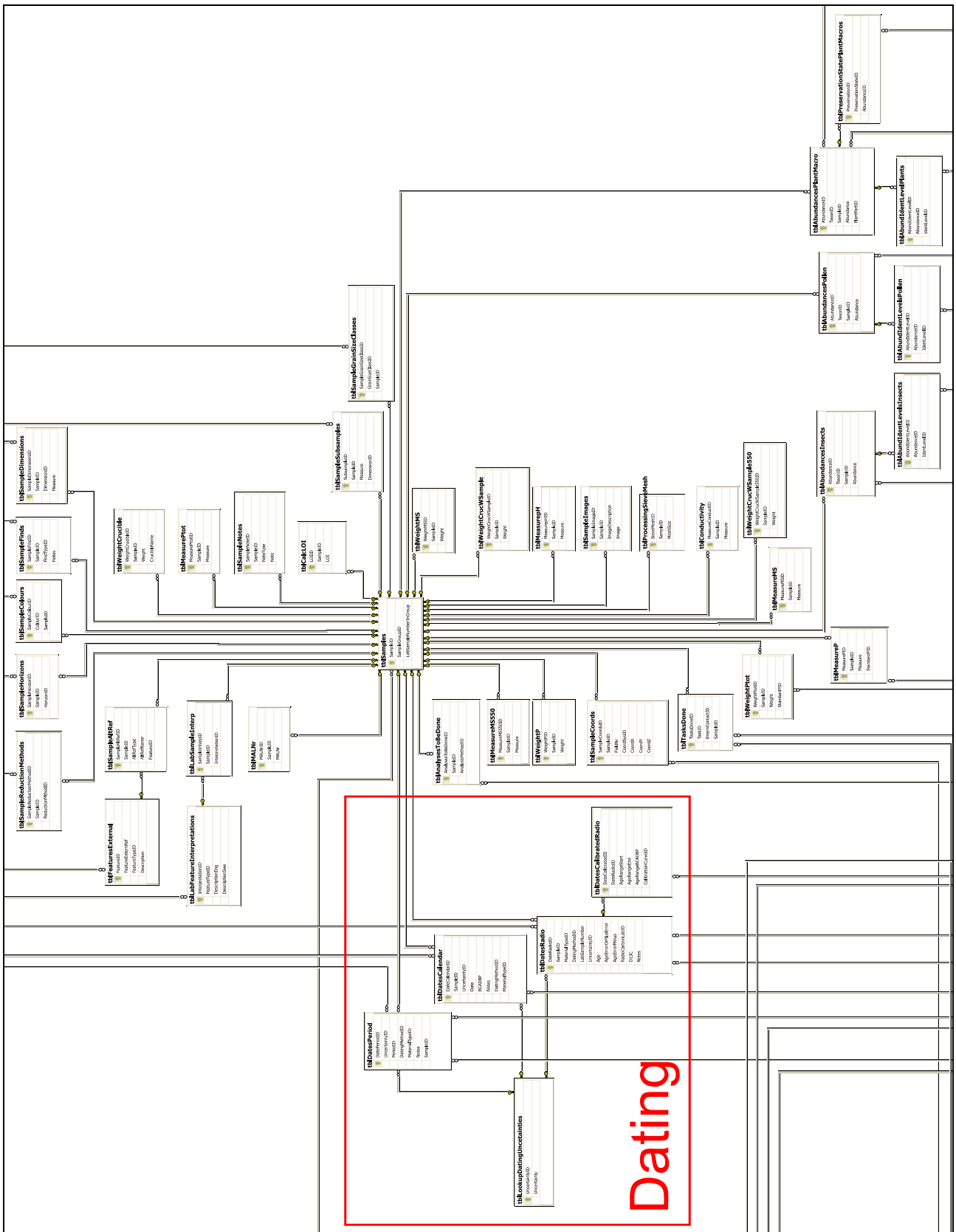
Figure 5.7. Structure of Sample data area

5.1.5.5 *Lab & Common data*

5.1.5.5.1 Common (sample related) data
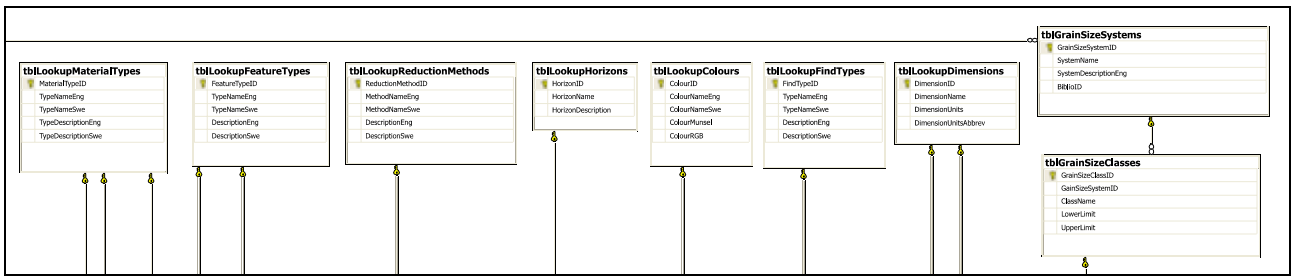
Figure 5.8. Structure of Common data area for samples

5.1.5.5.2 Biological proxy modern data
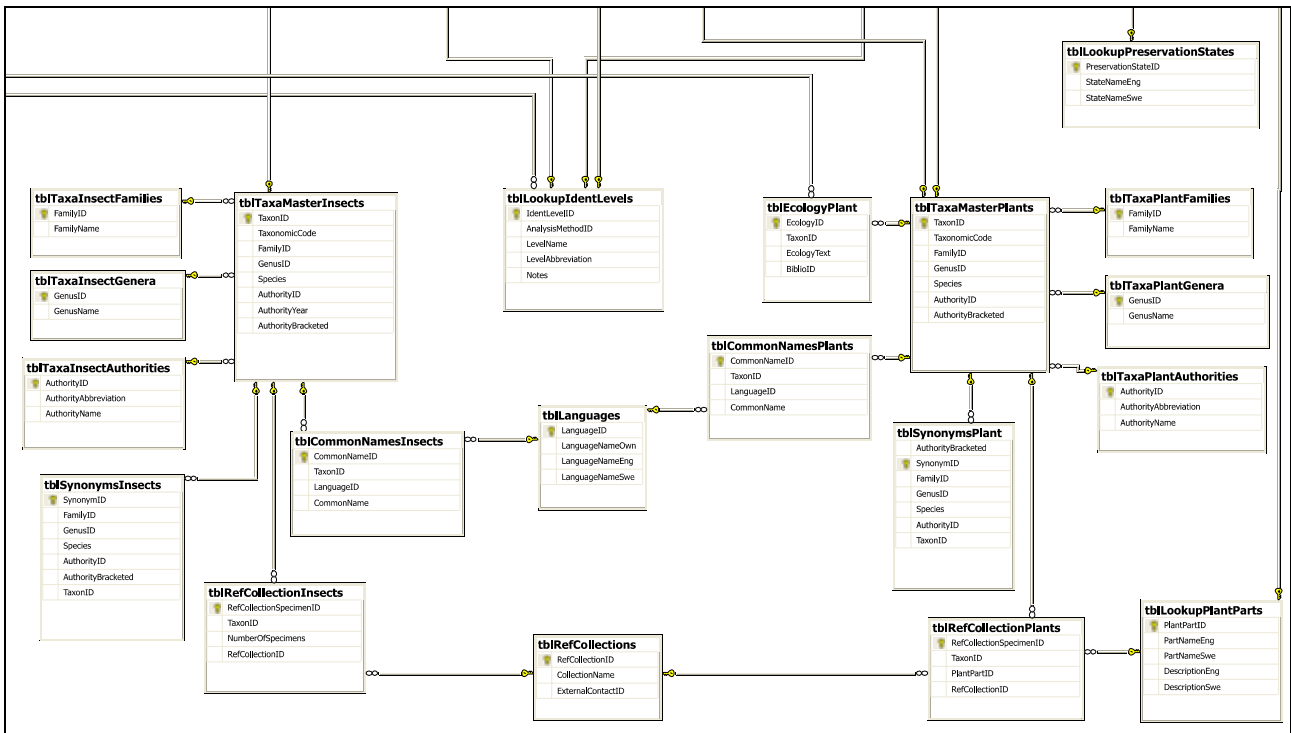
Figure 5.9. Structure of biological proxy master data area

Master tables for molluscs are omitted for clarity, but follow the form of the plant tables. Master tables include the facility for storage of taxonomic codes which will enable lists to be generated in taxonomic order. It is suggested that European taxonomy be followed (e.g. Flora Europea).

A simple modification would enable multiple taxonomic ordering systems to be implemented for increased compatibility between research groups.

5.1.5.5.3 Bibliographic data
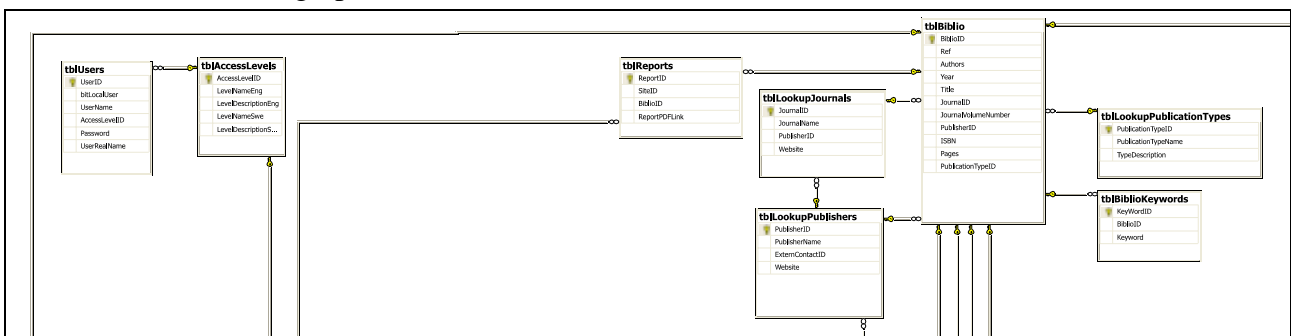
Figure 5.10. Structure of Bibliographic and Users data area

5.1.5.5.4 Users data

The Users data is used to retain information on registered users and limit access to data from the site level. Further data security and access rights will be handled by SQL-Servers security model. The need for six initial user groups have been identified as shown in Table 5.3.

Group	Description
System Administrator	Developer level access, full data entry and retrieval plus system modificaion/programming.
MAL Staff Onsite	Full data access, entry and retrieval.
External Onsite	Restricted to only published data retrieval and specific project data entry.
MAL Staff Online	Full data retrieval access to the extent permitted by the online database.
External Online	Restricted to only published data retrieval.
Student	Restricted access, customisable, but initially as External Onsite.

Table 5.3. Initial user groups for SEAD access.

5.1.5.5.5 Other Lab & Common data
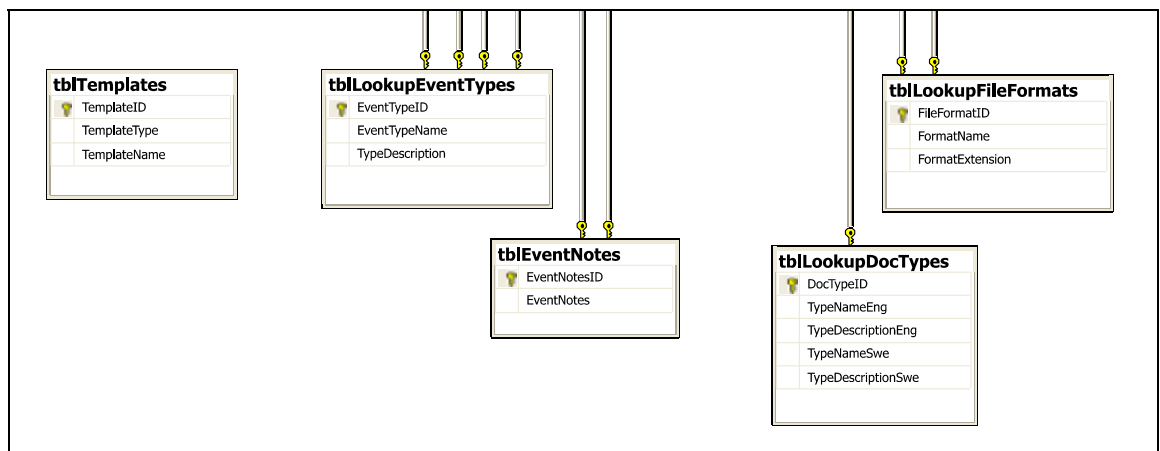
Figure 5.11. Structure of Lab & Common data area (continues in Figure 3.1)
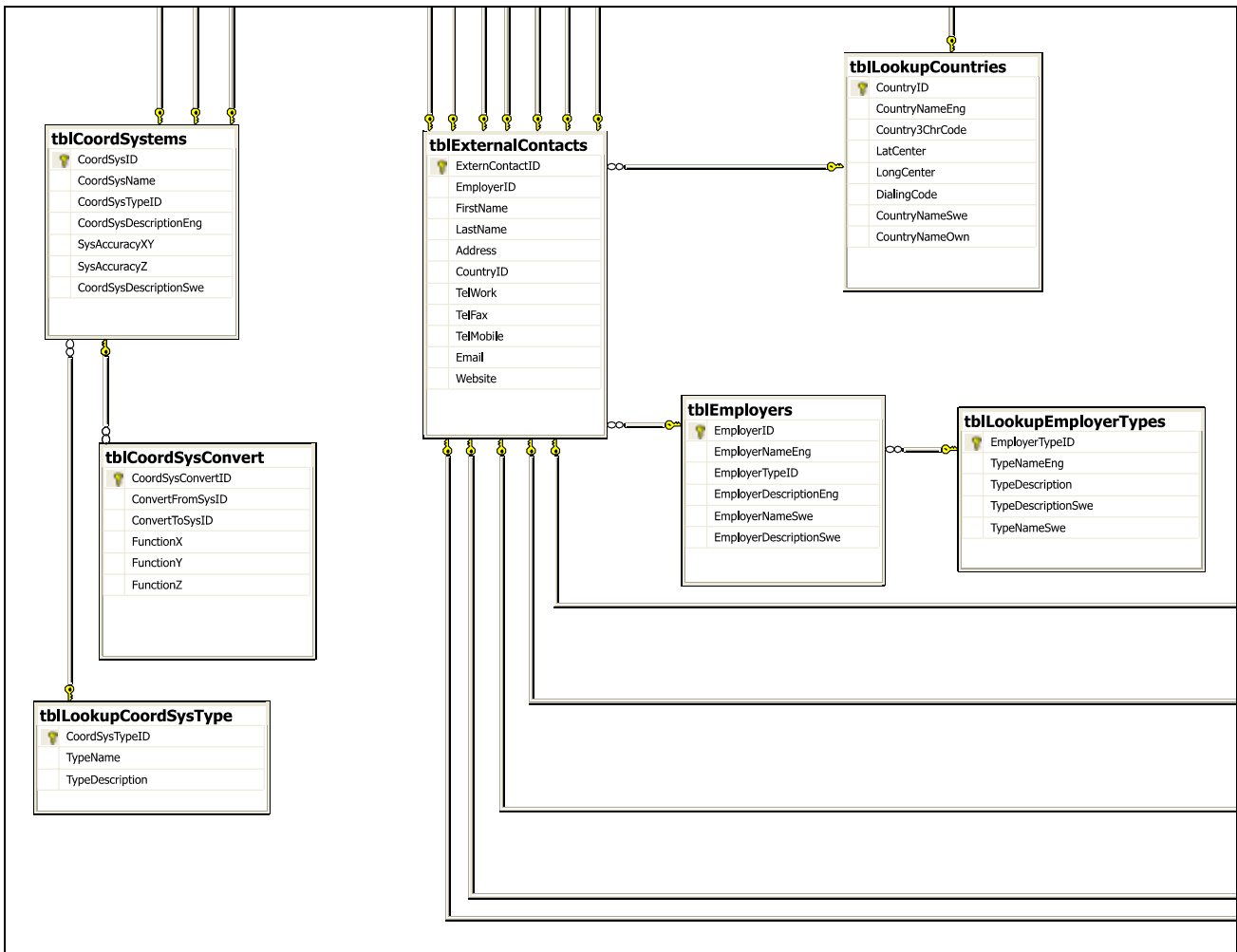
Figure 5.12. Structure of Lab & Common data area (continued from Figure 5.11, continues in Figure 5.13)
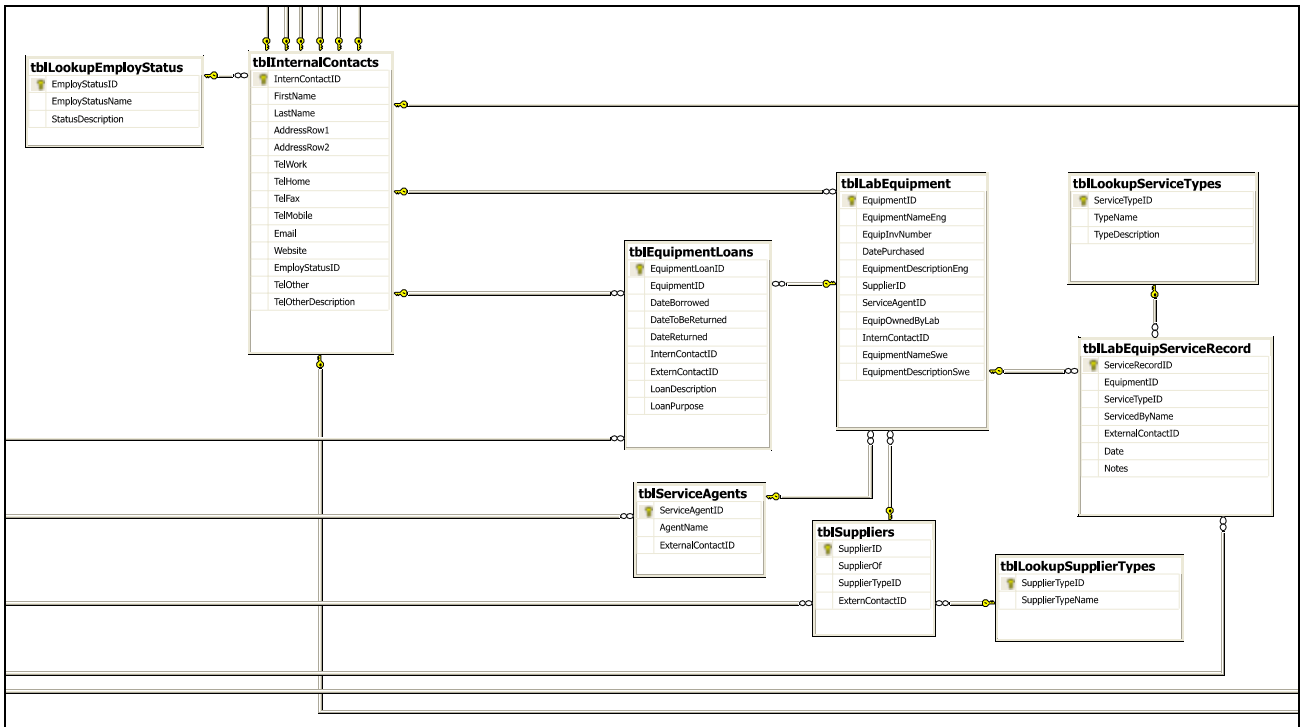
Figure 5.13. Structure of Lab & Common data area (continued from Figure 5.13, continues in Figure 5.14)
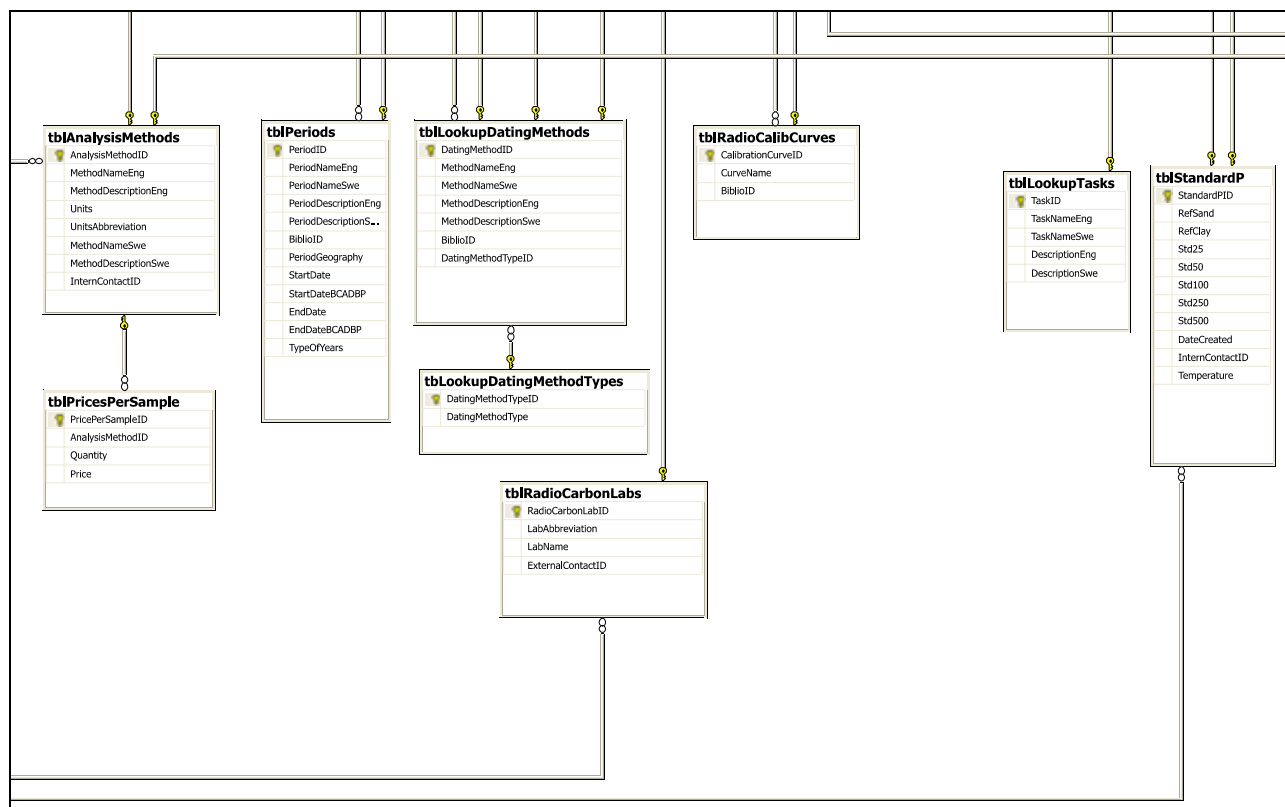
Figure 5.14. Structure of Lab & Common data area (continued from Figure 5.13)

5.2 Potential Queries and Analysis Routes (Processes)

Generalised SQL statements are given below for a few of the (predicted) most common data retrieval queries. Note that the syntax has been simplified in some cases for the sake of clarity.

More details of how data retrieval is facilitated through the user interface is given in Section 5.3.1. The queries given below are based on the examples given in Figure 5.22, and represent part of the evaluation (see 6.1) of the data structure outlined above.

The queries given below were not tested with data, and are not saved as stored procedures within the SEAD (Visual Studio) design files. They are only intended to give an indication of how the proposed structure permits data retrieval, and to validate the structural logic of relationships.

5.2.1 Standard Queries

The majority of analysis routes will concern the retrieval of sample based data for particular projects, sites or groups of samples.

Retrieve phosphate measurements for all samples at Site 'Lockarp7B

```
SELECT     tblMeasureP.Measure, tblSites.SiteNameSwe
FROM       tblSites INNER JOIN
               tblSampleGroups ON tblSites.SiteID =
           tblSampleGroups.SiteID INNER JOIN
               tblSamples ON tblSampleGroups.SampleGroupID =
           tblSamples.SampleGroupID INNER JOIN
               tblMeasureP ON tblSamples.SampleID =
           tblMeasureP.SampleID
WHERE      (tblSites.SiteNameSwe = 'Lockarp7B')
```

Retrieve 14C dates for Sample Group with External Reference 'Context147'

```
SELECT      tblSampleGroups.ExternGroupRef, tblDatesRadio.*
FROM        tblSampleGroups INNER JOIN
                tblSamples ON tblSampleGroups.SampleGroupID =
            tblSamples.SampleGroupID INNER JOIN
                tblDatesRadio ON tblSamples.SampleID =
            tblDatesRadio.SampleID
WHERE       (tblSampleGroups.ExternGroupRef = 'Context147')
```

Retrieve taxonomically ordered plant macro species list, with abundance data, for Site 'Lockarp7B' (note that a crosstab variant of this query could be used to retrieve this data in the spreadsheet like form commonly used with biological proxy data)

```
SELECT      tblSites.SiteNameSwe, tblAbundancesPlantMacro.Abundance,
                tblTaxaPlantFamilies.FamilyName,
                tblTaxaPlantGenera.GenusName,
                tblTaxaPlantAuthorities.AuthorityName,
                tblTaxaMasterPlants.TaxonomicCode
FROM        tblSampleGroups INNER JOIN
                tblSamples ON tblSampleGroups.SampleGroupID =
            tblSamples.SampleGroupID INNER JOIN
                tblSites ON tblSampleGroups.SiteID =
            tblSites.SiteID INNER JOIN
                tblAbundancesPlantMacro ON tblSamples.SampleID =
            tblAbundancesPlantMacro.SampleID INNER JOIN
                tblTaxaMasterPlants ON
            tblAbundancesPlantMacro.TaxonID =
            tblTaxaMasterPlants.TaxonID INNER JOIN
                tblTaxaPlantAuthorities ON
            tblTaxaMasterPlants.AuthorityID =
            tblTaxaPlantAuthorities.AuthorityID INNER JOIN
                tblTaxaPlantFamilies ON
            tblTaxaMasterPlants.FamilyID =
            tblTaxaPlantFamilies.FamilyID INNER JOIN
                tblTaxaPlantGenera ON
            tblTaxaMasterPlants.GenusID = tblTaxaPlantGenera.GenusID
WHERE     (tblSites.SiteNameSwe = 'Lockarp7B')
ORDER BY tblTaxaMasterPlants.TaxonomicCode
```

Retrieve bibliographic details and PDF files for all reports published by MAL during 1999

```
SELECT      tblBiblio.*, tblLookupPublishers.PublisherName,
                tblReports.ReportPDFLink
FROM        tblBiblio INNER JOIN
                tblLookupPublishers ON tblBiblio.PublisherID =
            tblLookupPublishers.PublisherID INNER JOIN
                tblReports ON tblBiblio.BiblioID =
            tblReports.BiblioID
WHERE     (tblLookupPublishers.PublisherName = 'MAL') AND
                (tblBiblio.Year = 1999)
```

Retrieve contact details for all current MAL staff sorted by surname

```
SELECT      tblInternalContacts.*,
                tblLookupEmployStatus.EmployStatusName
FROM        tblInternalContacts INNER JOIN
                tblLookupEmployStatus ON
            tblInternalContacts.EmployStatusID =
                tblLookupEmployStatus.EmployStatusID
WHERE     (tblLookupEmployStatus.EmployStatusName = N'Current')
ORDER BY ContactLastName
```

Retrieve bibliographic details for all MAL reports which either include the word 'soil' in the title or have the keyword 'soil'

```
SELECT      tblBiblio.*, tblBiblio.Title AS RefTitle,
            tblBiblioKeywords.Keyword
FROM        tblBiblio INNER JOIN
                tblLookupPublishers ON tblBiblio.PublisherID =
            tblLookupPublishers.PublisherID INNER JOIN
                tblBiblioKeywords ON tblBiblio.BiblioID =
            tblBiblioKeywords.BiblioID
WHERE       (tblBiblio.Title LIKE 'Soil') OR
                (tblBiblioKeywords.Keyword = 'Soil')
```

5.2.2   Advanced Queries

One of the potentially powerful features of SEAD will be the ability to retrieve the values of certain variables where other variables meet particular criteria. This research potential of being able to perform multi-proxy data interrogation is enormous, and has implications well outside of the lab's research base. Some examples of questions, and their usefulness are listed below.

Show plant macrofossil data (in taxonomic order) for all samples where pH<4.
This will essentially retrieve a list of all plant species found fossil in acidic samples – which in itself may be a proxy for past environmental conditions.

```
SELECT      tblTaxaPlantFamilies.FamilyName,
            tblTaxaPlantGenera.GenusName,
            tblTaxaPlantAuthorities.AuthorityName,
            tblTaxaMasterPlants.TaxonomicCode,
                tblMeasurepH.Measure,
            tblAbundancesPlantMacro.Abundance
FROM        tblSamples INNER JOIN
                tblAbundancesPlantMacro ON tblSamples.SampleID =
            tblAbundancesPlantMacro.SampleID INNER JOIN
                tblTaxaMasterPlants ON
            tblAbundancesPlantMacro.TaxonID =
            tblTaxaMasterPlants.TaxonID INNER JOIN
                tblTaxaPlantAuthorities ON
            tblTaxaMasterPlants.AuthorityID =
            tblTaxaPlantAuthorities.AuthorityID INNER JOIN
                tblTaxaPlantFamilies ON
            tblTaxaMasterPlants.FamilyID =
            tblTaxaPlantFamilies.FamilyID INNER JOIN
                tblTaxaPlantGenera ON
            tblTaxaMasterPlants.GenusID = tblTaxaPlantGenera.GenusID
            INNER JOIN
                tblMeasurepH ON tblSamples.SampleID =
            tblMeasurepH.SampleID
WHERE       (tblMeasurepH.Measure < 4)
ORDER BY tblTaxaMasterPlants.TaxonomicCode
```

Retrieve magnetic susceptibility data for all samples from Skåne, including RT90 coordinates for the samples.
This would provide spatial data that could be used to examine the extent of prehistoric occupation of Skåne with GIS. Note that this query uses a custom built function called funcConvertCoords to convert the samples coordinates from whatever system they use into the standard RT90 (or SWEREF99) grid commonly used in Sweden.

```
SELECT      tblMeasureMS.Measure,
                funcConvertCoords(tblSampleCoords.CoordX,
            tblSampleCoords.CoordSysID, 'RT90') AS ConvertedX,
                    funcConvertCoords(tblSampleCoords.CoordY,
            tblSampleCoords.CoordSysID, 'RT90') AS ConvertedY,
            tblSampleCoords.FieldNr
FROM        tblSiteLocations INNER JOIN
                    tblSites ON tblSiteLocations.SiteID =
            tblSites.SiteID INNER JOIN
                    tblSampleGroups ON tblSites.SiteID =
            tblSampleGroups.SiteID INNER JOIN
                    tblSamples ON tblSampleGroups.SampleGroupID =
            tblSamples.SampleGroupID INNER JOIN
                    tblMeasureMS ON tblSamples.SampleID =
            tblMeasureMS.SampleID INNER JOIN
                    tblSampleCoords ON tblSamples.SampleID =
            tblSampleCoords.SampleID INNER JOIN
                    tblCoordSystems ON tblSampleCoords.CoordSysID =
            tblCoordSystems.CoordSysID CROSS JOIN
                    tblCoordSysConvert
WHERE       (tblSiteLocations.Landskap = 'Skåne')
```

Retrieve a taxonomically ordered list of insect species from samples where barley (*Hordeum* sp.) has been found in the plant macrofossil analyses. (A slight variation of the query could be used to retrieve sample and site details etc.)

A potentially very informative query that could provide data on the past relationship between crops and pests of stored products. By enhancing the query it would be possible to provide dating evidence for the extent of cereal damage throughout that extent of prehistory represented by the available data.

```
SELECT      tblTaxaPlantGenera.GenusName,
                tblTaxaInsectFamilies.FamilyName,
                tblTaxaInsectGenera.GenusName AS Expr1,
                    tblTaxaInsectAuthorities.AuthorityAbbreviation,
                tblAbundancesInsects.Abundance
FROM        tblSamples INNER JOIN
                    tblAbundancesPlantMacro ON tblSamples.SampleID =
            tblAbundancesPlantMacro.SampleID INNER JOIN
                    tblTaxaMasterPlants ON
            tblAbundancesPlantMacro.TaxonID =
            tblTaxaMasterPlants.TaxonID INNER JOIN
                    tblTaxaPlantGenera ON
            tblTaxaMasterPlants.GenusID = tblTaxaPlantGenera.GenusID
            INNER JOIN
                    tblAbundancesInsects ON tblSamples.SampleID =
            tblAbundancesInsects.SampleID INNER JOIN
                    tblTaxaMasterInsects ON
            tblAbundancesInsects.TaxonID =
            tblTaxaMasterInsects.TaxonID INNER JOIN
                    tblTaxaInsectAuthorities ON
            tblTaxaMasterInsects.AuthorityID =
            tblTaxaInsectAuthorities.AuthorityID INNER JOIN
                    tblTaxaInsectGenera ON
            tblTaxaMasterInsects.GenusID =
            tblTaxaInsectGenera.GenusID INNER JOIN
                    tblTaxaInsectFamilies ON
            tblTaxaMasterInsects.FamilyID =
            tblTaxaInsectFamilies.FamilyID
WHERE       (tblTaxaPlantGenera.GenusName = N'Hordeum')
```

5.3 Local User Interface Design

As SEAD will inevitably contain unpublished data from ongoing contracts and research, a security system must be put in place to prevent unauthorised access to this data. This will be solved in part by restricting the scope of the data that is replicated on the web server, i.e. data classified as restricted will not be uploaded. In addition, user level security will be implemented at the Site data level, enabling onsite users to set the desired level of access to the data they are inputting or importing, including choosing whether to allow it to be available online.

Only those with administrative or developer level access will be able to access financial data – none of which will be made available online.

To conform with personal data integrity laws the names and contact details of individuals will not be made available online without their prior consent.

It is proposed that the local client interface be developed using a combination of MS Access, MS Visual Studio for Office and the VBA programming language.

5.3.1 Data entry and retrieval (process models)

Data entry will be performed by either direct machine input or assisted file import into SEAD, or manual typing of raw data and metadata into purpose built interfaces as outlined in this section. The bulk of data are either soil chemical/property data or biological proxy data, the process models for the entry/import of which are shown below. These have been developed by optimising the present system as analysed in section 4.2, and with special attention being paid to the practicalities of lab operation and the user and data requirements highlighted in section 4.3. Note the processes illustrated here focus on the flow of *data* and may not accurately illustrate physical lab routines.

See the beginning of this document for an explanation of flow chart symbols.

5.3.1.1 Data & database preparation

Project and site structure, as well as sample group and samples must be defined before sample related data can be entered into the database. This enforces the data structure as illustrated in Figure 4.1 & Figure 5.2. Figure 5.15 shows a generalised process model for the handling of samples on arrival at the lab with SEAD implemented. Samples are registered after either assignation to an existing project and site or the creation of a new one. Recording sheets can be created to standardise the recording of particular proxy raw data. Details of the Project and Site selection/setup procedure are illustrated as a process model in Figure 5.16. A similar process model illustrates the creation of samples groups and samples in preparation for proxy data assignation (Figure 5.17).

Notice that the recording sheets may be created during any of these processes, as it is hoped that this will encourage their use and aid in standardising the recording of proxy data. This is of special relevance for plant macrofossil analyses, for which workers exhibit a large degree of variation in recording strategies, and the qualitative information gained during sample processing for all proxies. This facility should be offered to all researchers – onsite or remote – to promote uniformity and ensure data quality.
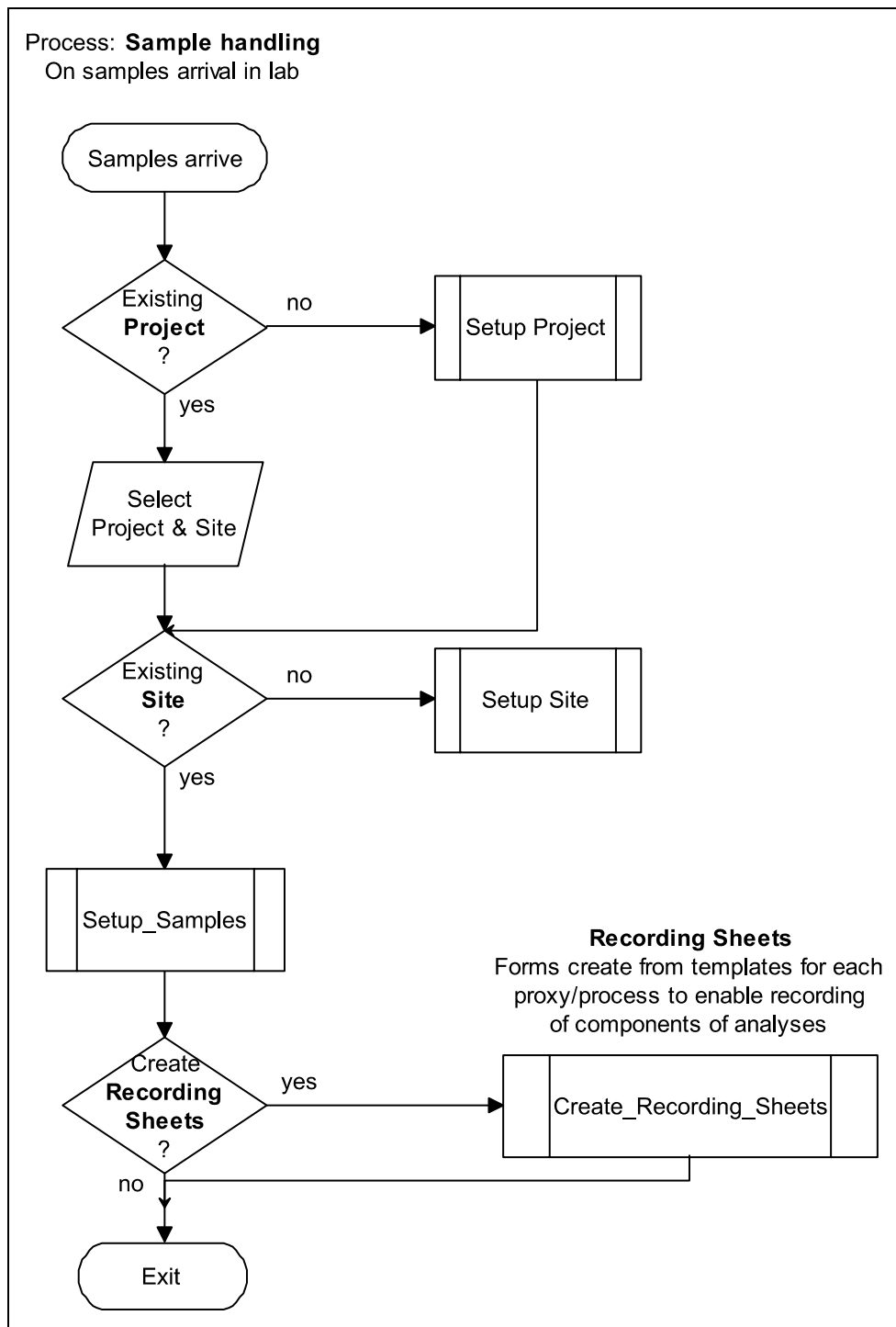
Figure 5.15. Process model for sample handling in SEAD.
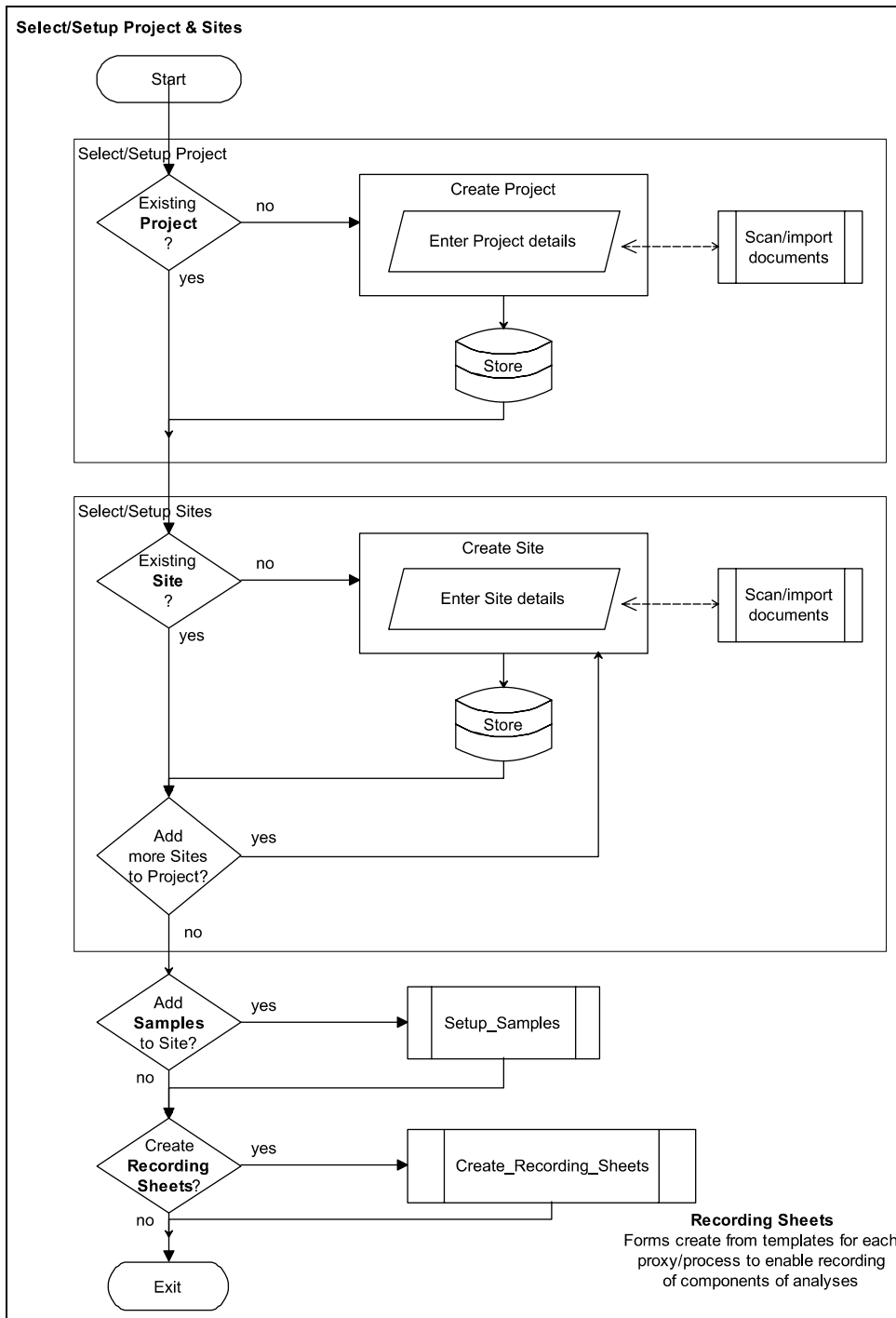
Figure 5.16. Process model for Project and Site selection/setup.
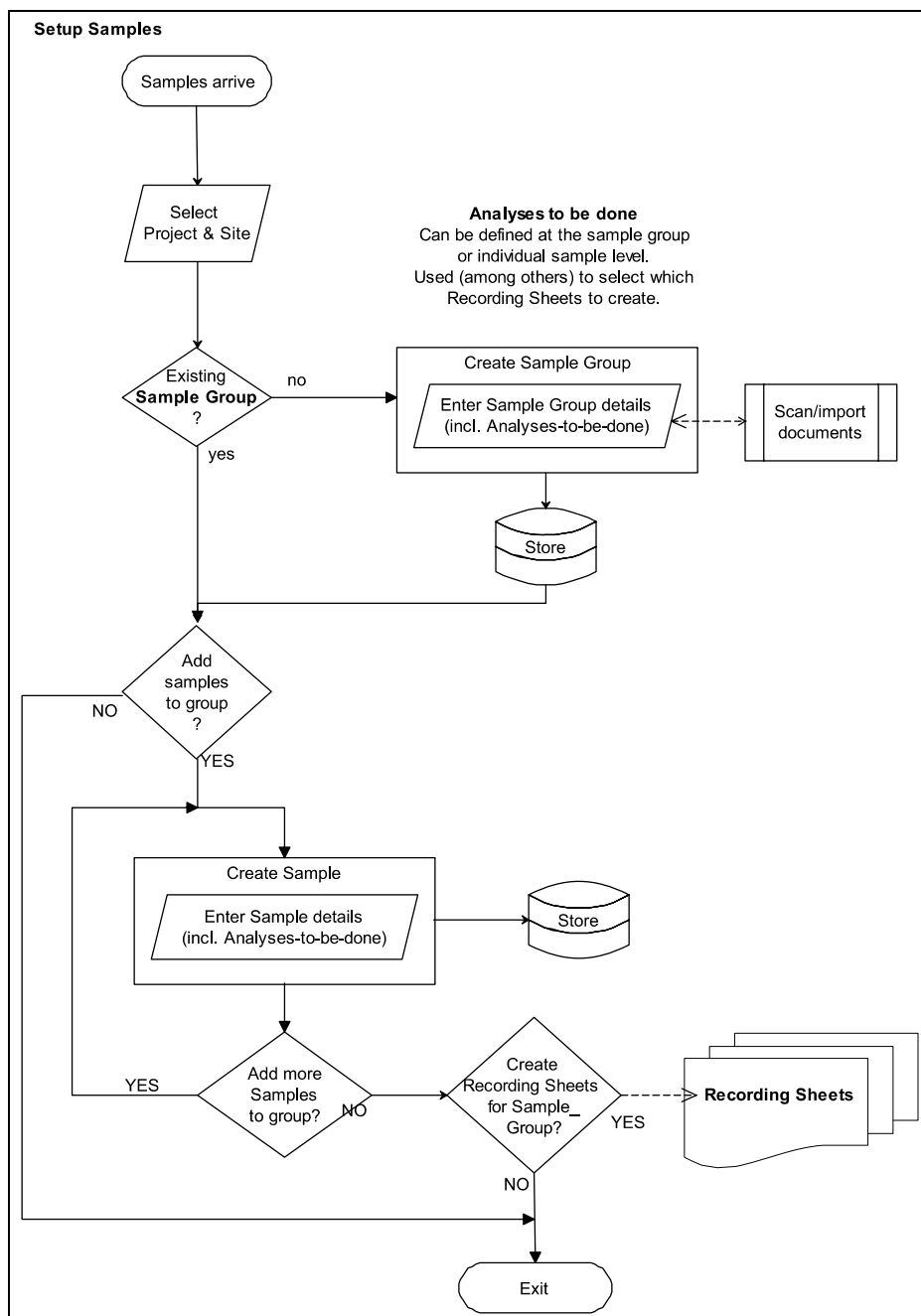
Figure 5.17. Process model for Sample setup showing Sample Group creation and optional export of recording sheets.

5.3.1.2 *Biological proxy data entry*

Printable templates will be created for output from SEAD, as described above, to simplify the recording of quantitative count data (see section 5.5.3). Recording is an activity generally performed beside the microscope and access to a computer cannot be guaranteed. Users will subsequently log on to SEAD and manual enter the recorded data into custom built interfaces, as illustrated in Figure 5.18. Figure 5.19 shows a generalised process model for the input of all biological proxy data. Note that although only Insects, Plant macrofossils and Molluscs (Snails) are shown the model applies to all hand recorded biological proxies.
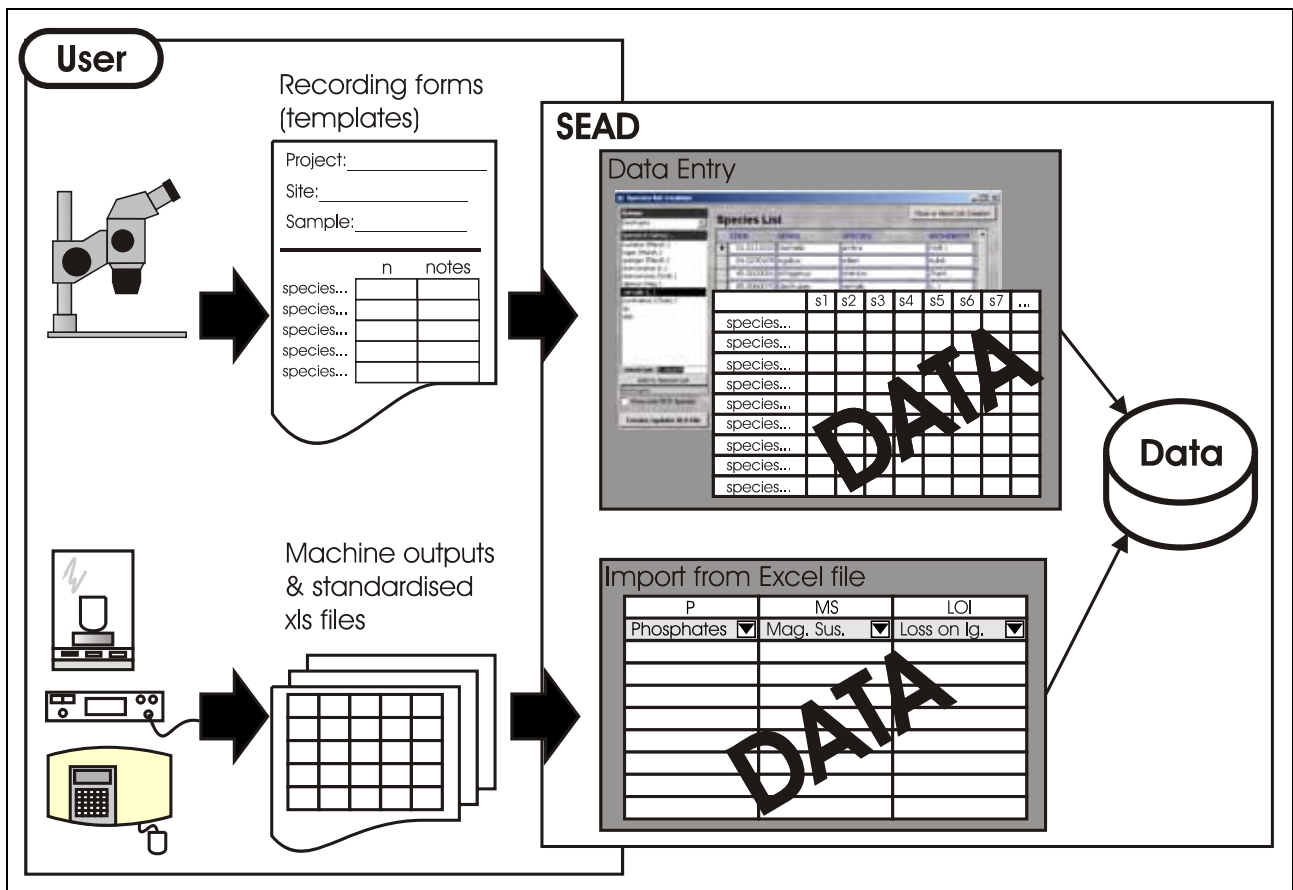
Figure 5.18. Schematic showing input process for measured chemical/physical property data and quantified biological proxy into SEAD.
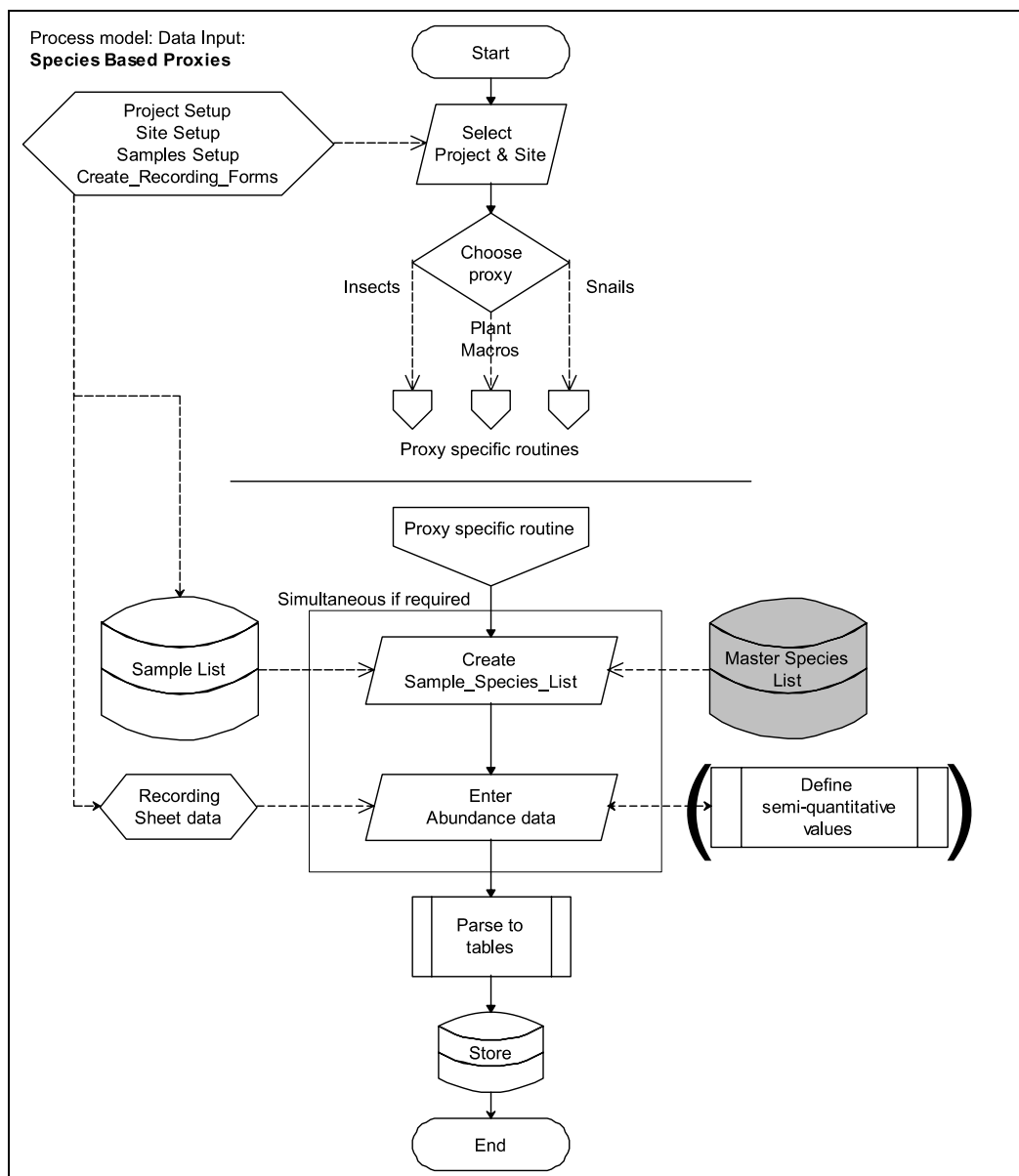
Figure 5.19. Process model for input of biological proxy data.

5.3.1.3   Soil chemistry/property data entry

The majority of soil physical and chemical property data (weights, magnetic susceptibility, phosphates) will be fed directly into SEAD from computer linked analysis equipment. Where this is not possible interfaces will be created to guide the import of files created by analysis equipment (as illustrated in Figure 5.18. This will ensure the correct storage of all data. Import templates, and a custom import interface will also be designed to allow existing data to be semi-automatically imported from MS Excel files.

These interfaces will also facilitate the import of data created by external parties, and thus help to ensure a wider user base.

The model for the current scope of soil chemistry/property analyses is shown in Figure 5.20, and continued in Figure 5.21 to show the import of files and entry of recording sheet data. As Figure 5.20 shows, measurements are undertaken in three stages which are dependent on the analyses required.

Process: **Soil chemistry/property workflow**
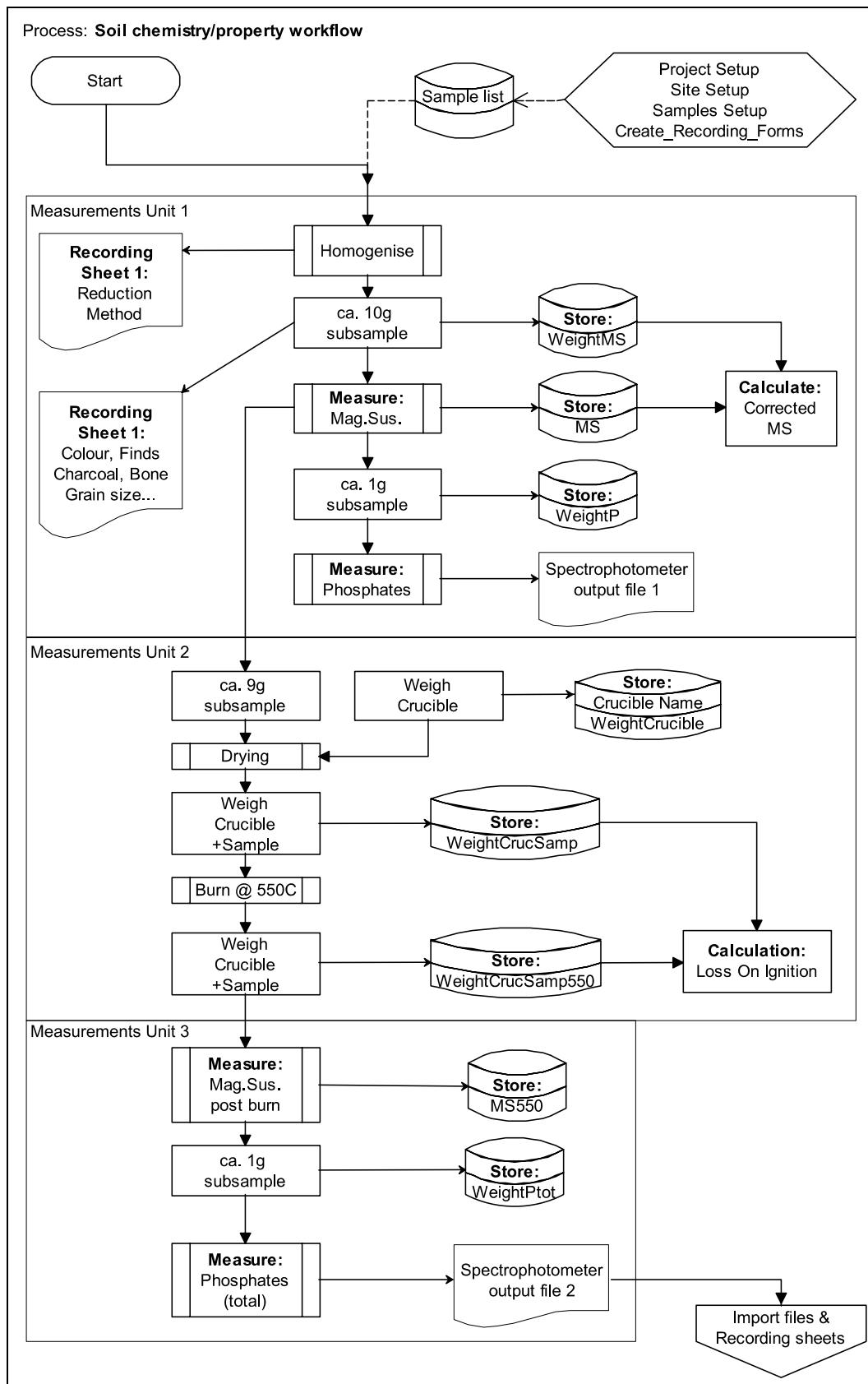
Figure 5.20. Process model for soil chemistry/property sample processing. Units represent complete task groups. Calculated items may be stored to improve efficiency. P = Phosphates; MS = Magnetic Susceptibility; Ptot = Total Phosphate. (Continued on Figure 24).
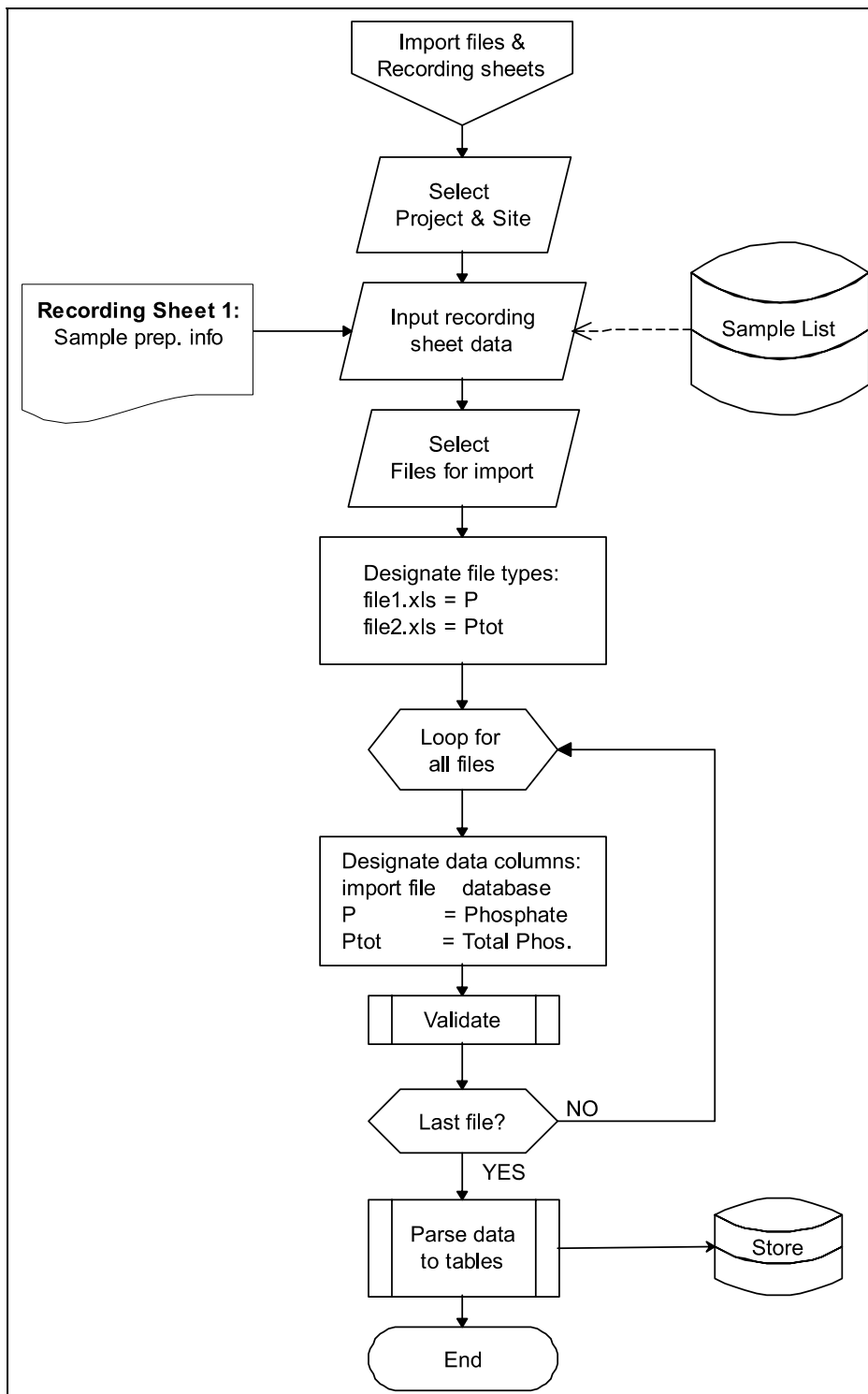
Figure 5.21. Process model for import of analysis machine created files and recording sheets for soil chemistry/property data. (Continued from Figure 5.24).

5.3.1.4 *Other data and metadata entry*

Dating evidence will be manually typed into forms after the creation of samples. The 'sample' is the critical data unit, to which all data items must be related – a site, for example, cannot have a date without having any samples. Figure 5.7 clearly illustrates the nodal nature of the samples table (tblSamples).

Project, Site and Sample Group information and other metadata will be input manually into the appropriate management interface. Method descriptions, equipment descriptions, contacts and other lab and common data will likewise be entered manually. A facility will be created for the import of bibliographic references in standard formats in addition to the facility to enter them manually.

5.3.1.5 *Data retrieval*

Data retrieval will be performed using three interface media:

1) Local Clients – installed on all lab PCs.

2) GIS – Geographical Information Systems map based interface, installed where needed.

3) Web – internet access to SEAD.

Four primary data retrieval paths have been identified for implementation through these media (as illustrated in Figure 5.22):

i) Quickly extract specific Project/Site data

ii) Browse the database for specific data

iii) Extract data which fulfil specific criteria

iv) Retrieve data by geographical location by map selection, and display the results of any search cartographically (described below, 5.3.2)

These paths will be implemented by the creation of interfaces which use stored procedures or queries built on the fly as the bridge between user and data. Figure 5.22 shows a generalised model of the paths described here, and Section 5.2 (Potential Queries and Analysis Routes) describes some of these in detail.
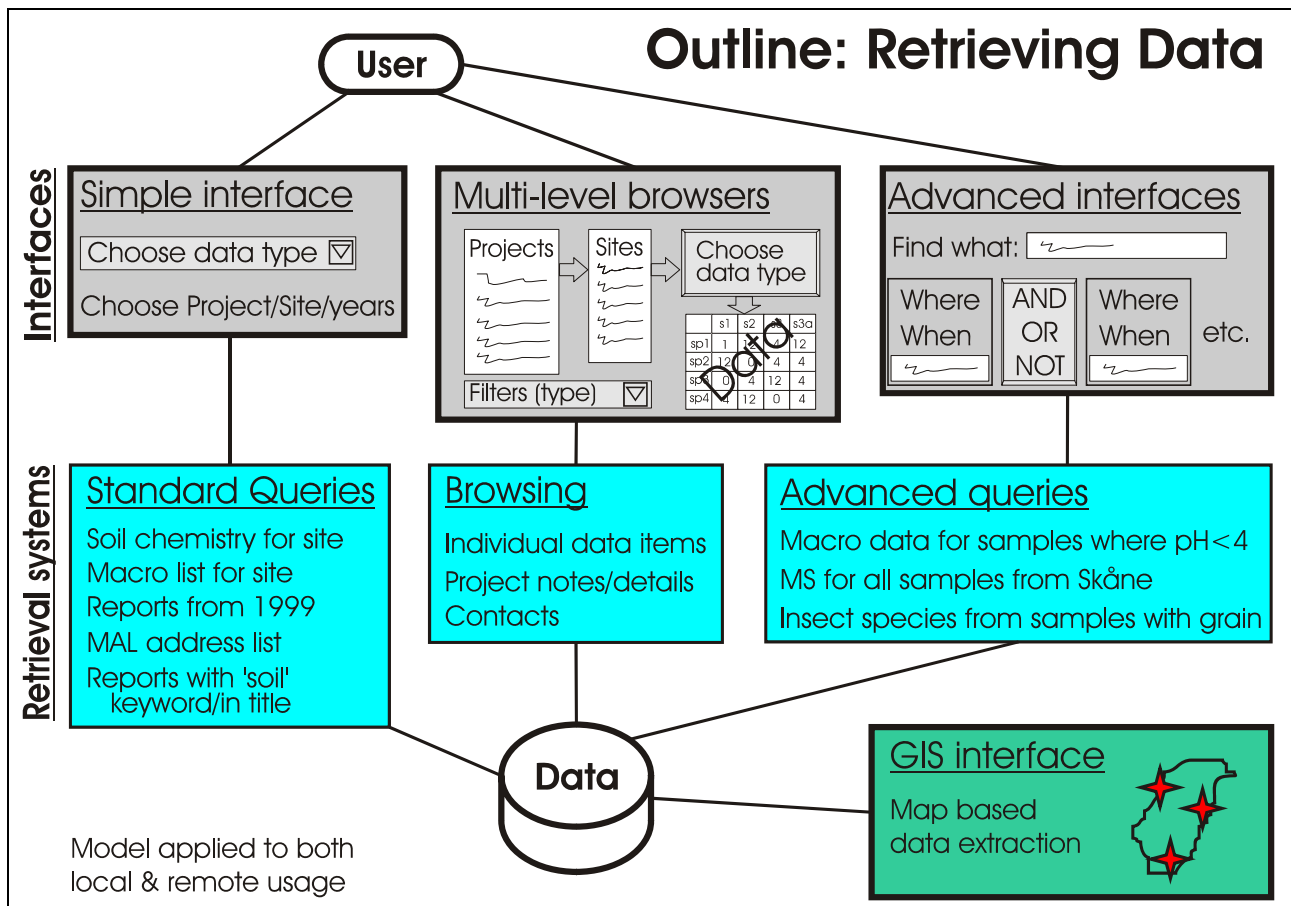
Figure 5.22. Generalised model of SEAD data retrieval (web and local client interfaces are not differentiated here).

5.3.2 GIS

Onsite GIS functionality will be achieved by linking the ArcMAP component of the ESRI ArcGIS package directly to the SQL Server database. A custom interface will be designed using ArcGIS's authoring facilities, which include the MS VBA programming language. This interface will essentially be a set of custom toolbars and menus to simplify access to SEAD, and allow users to perform the most common GIS tasks such as map production, querying by geographical location/property and spatial analyses. Map templates and common query types will be created to increase the efficiency of repeated retrieval of similar data from different projects and sites.

An interface sketch is not provided here, as there will be little visual difference from the standard ArcMAP interface apart from a selection of new wizards and toolbars.

Access to ArcGIS's other features will not be restricted, thus allowing for the full spectrum of GIS analyses should it be required.

The resolution required for the GIS interface will only require freely available base maps.

5.4 Internet Access

Internet access to SEAD will take the form of :

i) a website utilising Flash and ASP to retrieve user data requests from a web server replicated copy of the SQL-Server database (see replication restrictions in 5.4.3)

ii) a Flash based mapping interface using ASP to interact with the SQL-Server database

Note that no internet based *data entry* system is planned at this stage, but may become a viable option in the future if the SEAD project is successful in its establishment of external research partners. A web based *data submission system* will be put in place to encourage remote users to contribute and interact with MAL. MAL staff will then cooperate with the remote user to ensure the correct onsite import of the submitted data. In other words, MAL will act as a clearing house for SEAD data.

5.4.1 Data retrieval

Microsoft Visual Studio provides a developer environment tuned to producing websites for database interaction. It is proposed that this be used to reproduce the data retrieval and querying interfaces outlined above (5.3.1) as much as possible for web browser use (noting the restrictions below 5.4.3).
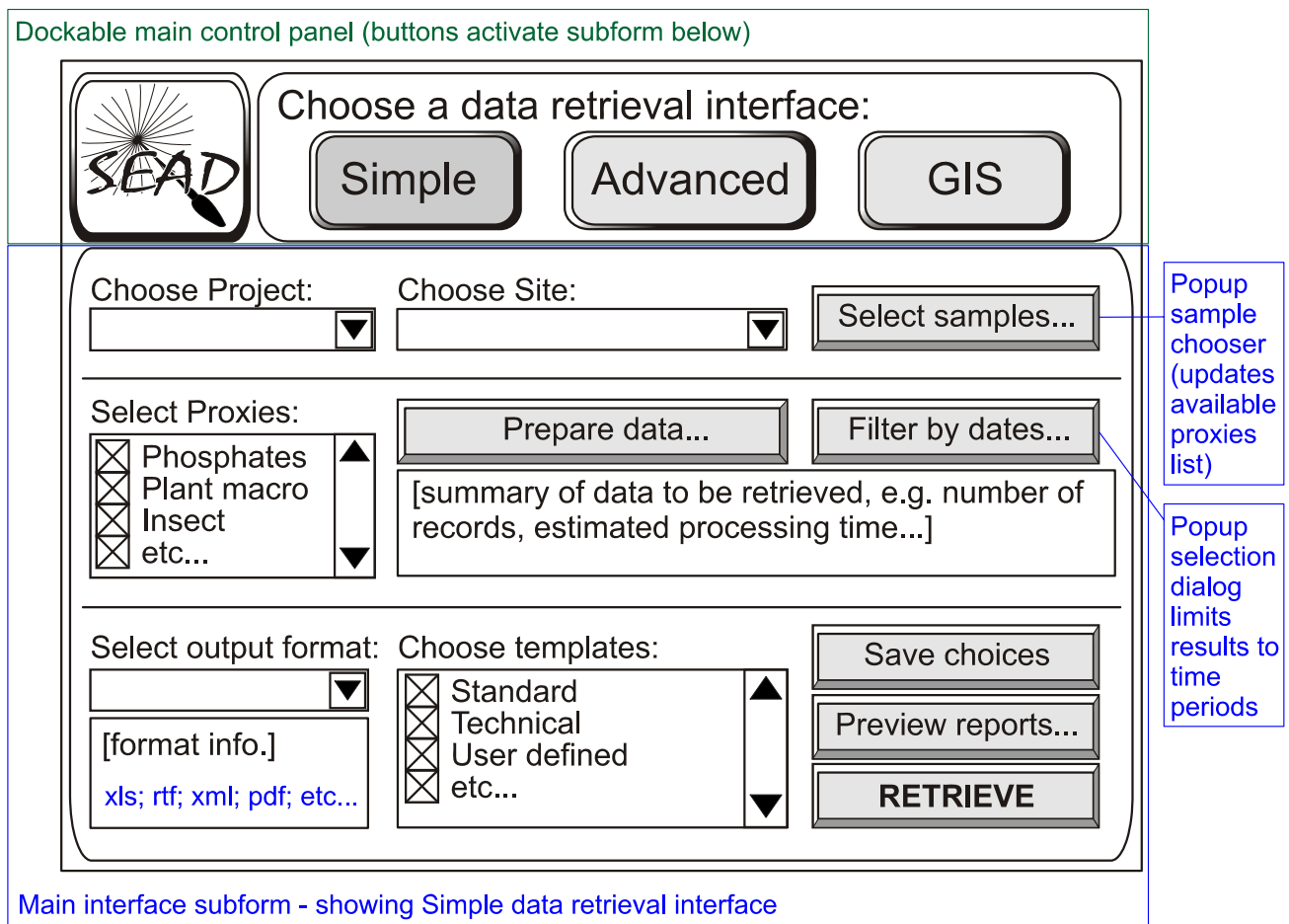
Figure 5.23. Sketch of SEAD web interface, showing control panel for switching between subforms, and the 'Simple' subform. Note that this is a suggestion only, and will be developed in line with user requirements.

5.4.2 Geographical display of data (maps)

To avoid the necessity of expensive map hosting software with a potentially heavy maintenance implication, a custom GIS simulation will be built for web access. This will be constructed with Flash authoring software to create a seamless interface between the user and the SQL Server database. ASP will be used to handle data retrieval. Although this will in no way be as comprehensive as a full blown online GIS interface it will allow users to:

i) Interrogate the data through a clickable map based interface

ii) Display retrieved data on maps

Flash also affords a greater degree of flexibility in terms of user interface design than many purpose built online mapping interfaces.
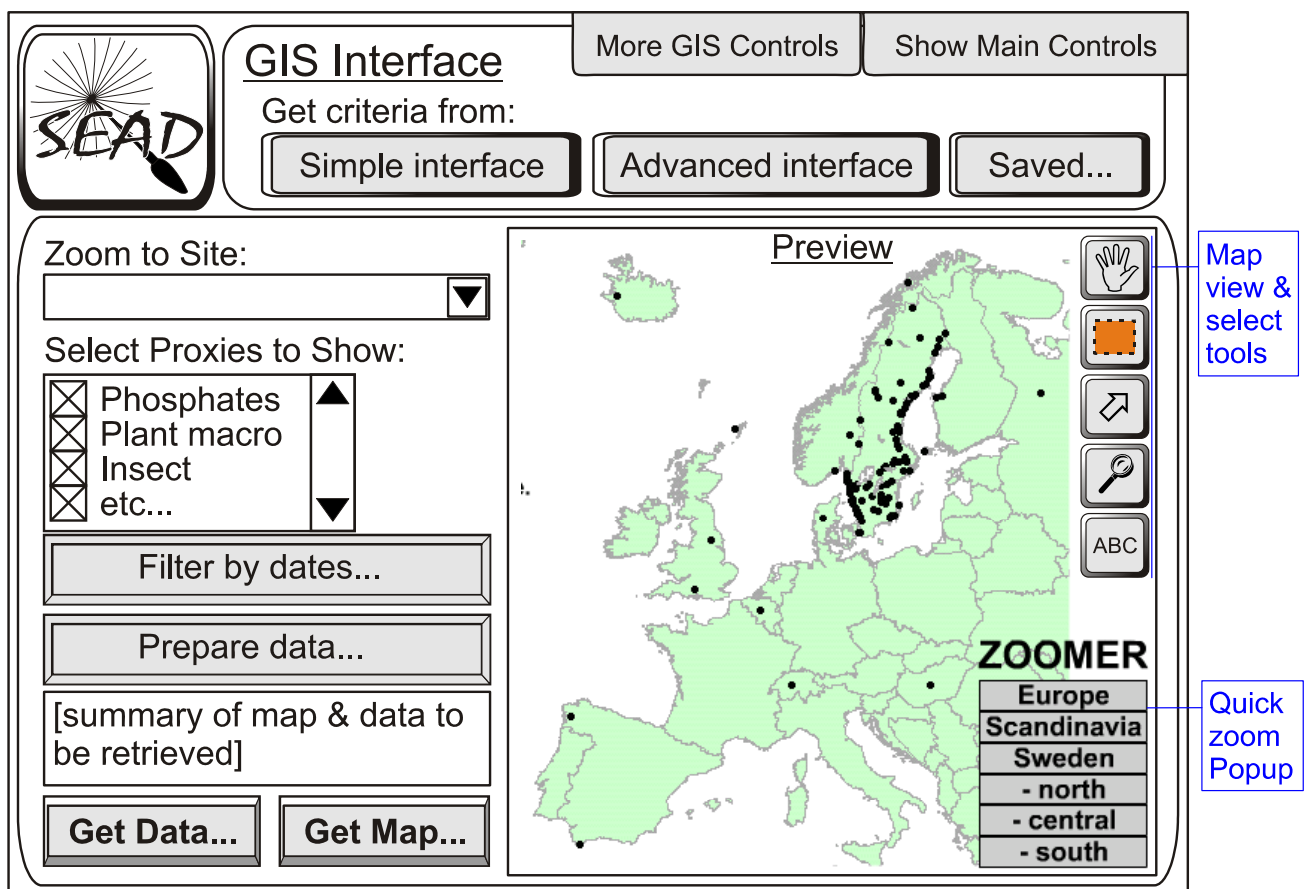
Figure 5.24. Sketch of SEAD web interface, showing GIS interface subform.
Note that the preview map image is taken from the *MAL Flash Project Browser* online at http://www.umu.se/envarchlab/mal/projects/images/projectmap2.html

5.4.3   Security and restrictions

Internet data security issues are highly relevant for online databases, and it is recommended here that a third party hosting solution is used to house the web accessible component of SEAD. This reduces the extent of the need for in-house webserver support, for only a small annual cost (see section 10). The local SEAD database will be replicated (with the restrictions outlined below) to the web space as necessary.

To further reduce data security risks the following data will never be replicated on the web server:

- personal data such as staff or contact addresses – internet users wishing to contact staff will directed to the appropriate departmental website

- unpublished and ongoing research data

- unpublished contract work data

- material limited by other copyright restrictions

Internet users will be required to register to gain full access to the web interface. This will enhance the lab's ability to provide support, and also allow for the easier calculation of (anonymous) usage statistics. Users will be encourage to cooperate more closely with the lab in the form of collaborative research projects.

5.5   Exporting and Reporting Needs

5.5.1   Export types

The ability to output selected data in formats that can be usefully integrated into research reports will be critical to the success of SEAD. It must also be possible to output data in formats that enable further analyses in third party statistical and GIS software.

As Figure 5.25 and Figure 5.26 show, three groups of data export (the optional results of data retrieval) have been identified as follows:

1) Reports – integrated documents combining datasets and summarising several data areas.

E.g. ecological information for all plant species represented at a site, along with bibliographic information for the ecology sources.

2) Exports – raw/calculated data output in tabular (spreadsheet/database) form (including formats compatible with the Tila/Tilia Graph pollen data plotting software)

E.g. MS Excel file with tabulated abundance data for samples and species from a site.

3) Maps – distribution maps of various kinds

E.g. Map showing geographical location of all known Medieval finds of flax from Sweden.

Templates will be made available for common data retrieval tasks and output formats (see Figure 5.26) in addition to the ability to create custom exports of exactly the data required in a specific format. Groups 1 and 2 have a proven track record of usefulness in the Bugs software [1,9].
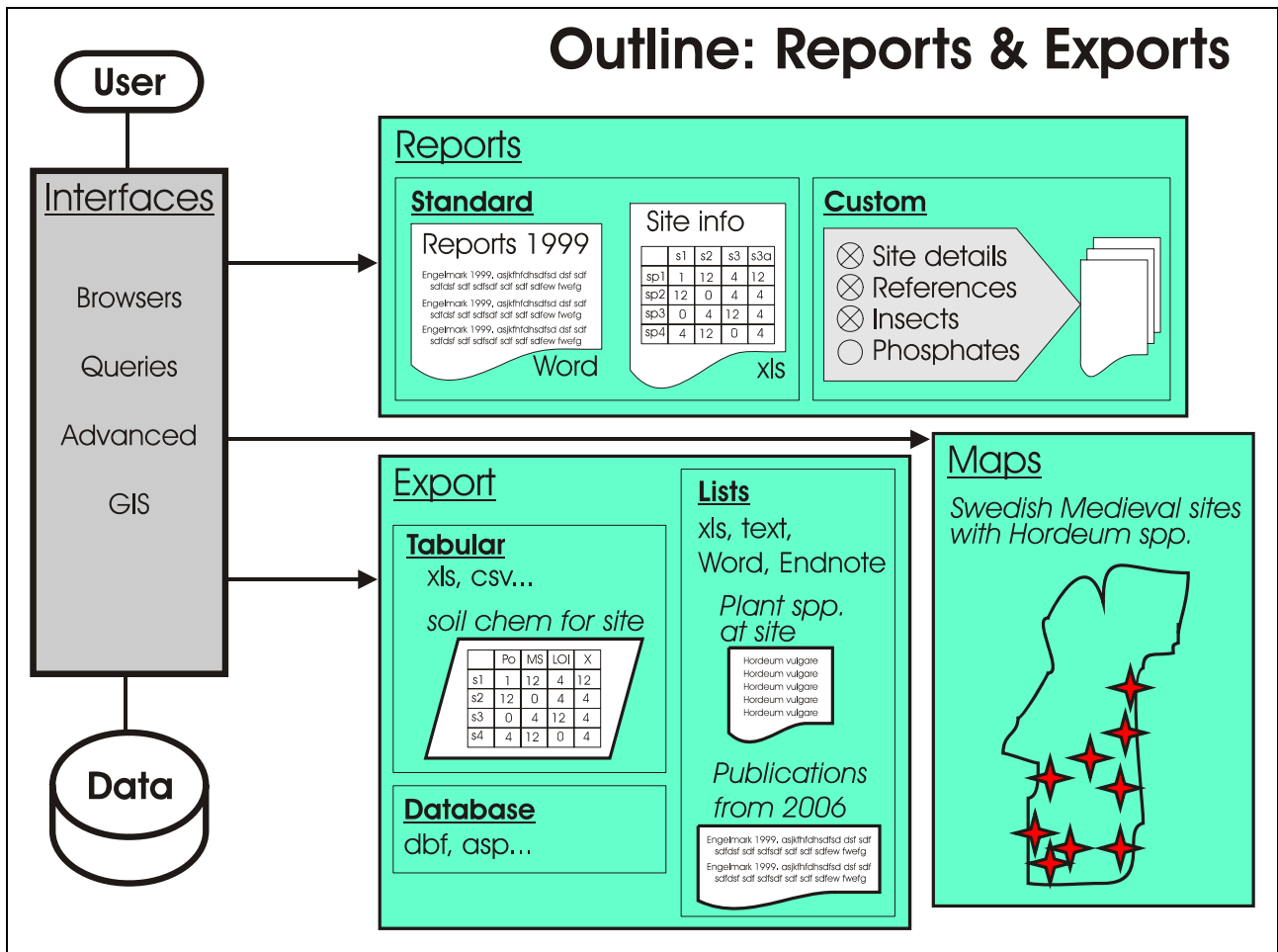
Figure 5.25. Schematic of selected reporting and exporting features of SEAD, including output for external GIS analysis.
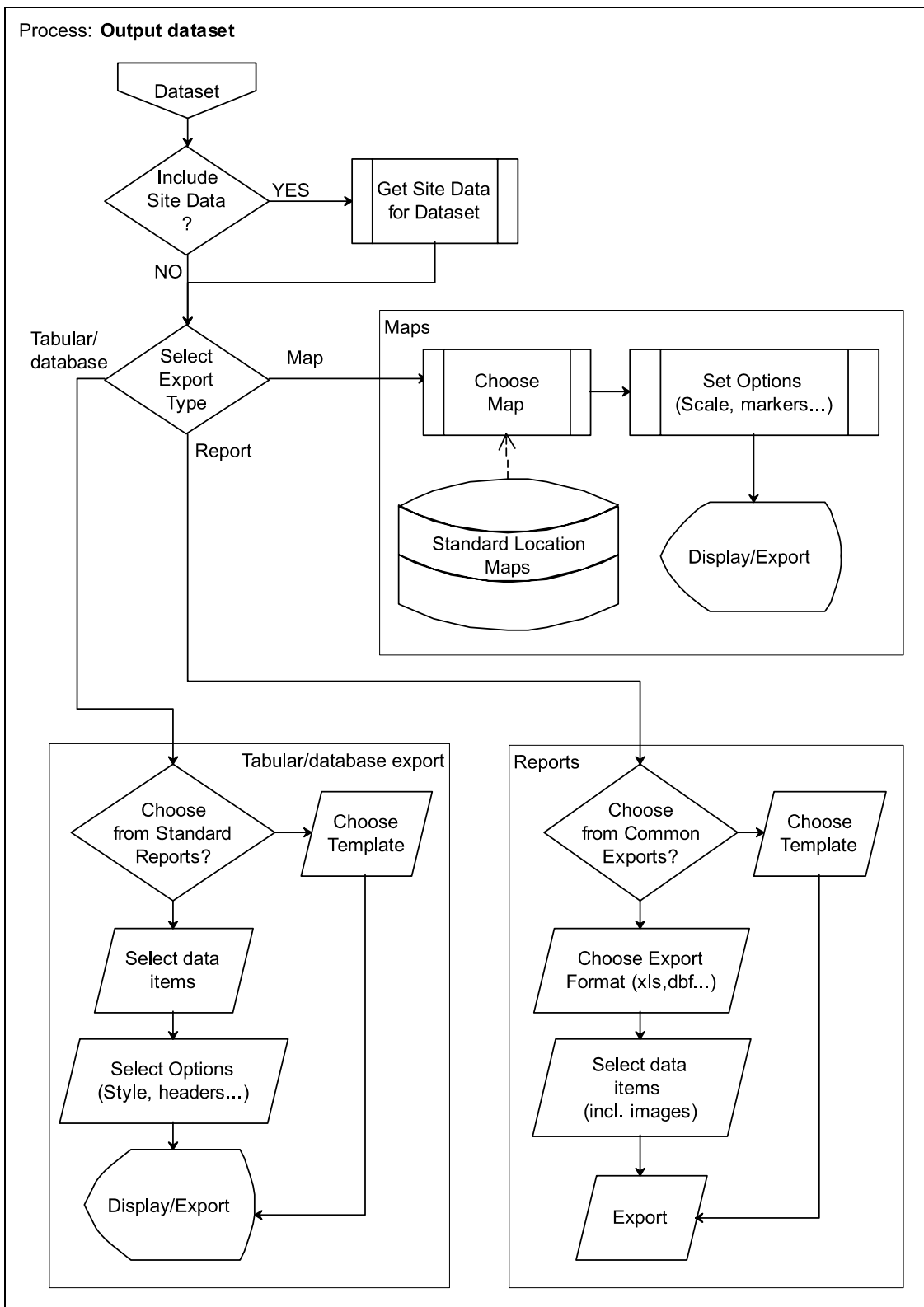
Figure 5.26. Process model for dataset outputs, schematically showing data selection and use of templates.

5.5.2 Export formats

A number of data formats (file types) are commonly used in environmental archaeology and Quaternary science. Table 5.4 shows those identified as essential for the current needs of SEAD. It is further anticipated that the availability of XML formats will eventually be required for all non-graphical output types. Additional format needs will become evident under development and as the scientific community's needs change.

Name	File extensions	Purpose
Excel spreadsheet Database cross-tabulation/tables	xls, dbf, csv	Soil chemistry/property data for samples Biological proxy data & species lists (with abundance values) Data with coordinates for GIS ... plus other tabulated data (e.g. for Tilia).
MS Word Text files	doc, txt	SEAD Reports (site summaries, Publication/reference lists.
Web based data retrieval	asp	Data returned by and displayed on web interface (note that user will be the facility to output this in the above formats from their browser).
Maps	wmf & other ArcGIS supported exports	Maps created with the GIS interface, for inclusion in reports.

Table 5.4. Predicted common export formats for data from SEAD.

5.5.3 Templates & standard recording sheets

To improve the standardisation of the recording of biological proxy counts and sample processing notes, SEAD will store standardized recording sheets which can be printed either on the creation of a project/site/sample group or on demand. This will help to ensure the quality of data entered into SEAD, and their use by all potential user groups will be encouraged as a form of quality assurance. These sheets will be designed in collaboration with MAL staff and external partners during the design phase. Recording sheets for fossil insects are already in use at MAL and by partner universities in the UK and Ireland. External parties will be consulted during the design of those for other proxies.

In addition to the above, a number of documents that are commonly used by SEAD's target groups will be stored to improve the efficiency with which researchers can undertake these tasks. Forms for the submission of samples for carbon dating, for example, are routinely filled in by hand by the majority of researchers. SEAD will replicate these forms as templates, and provide the facility to add/enter all relevant data through the interface and output a completed form, as modelled in Figure 5.27.
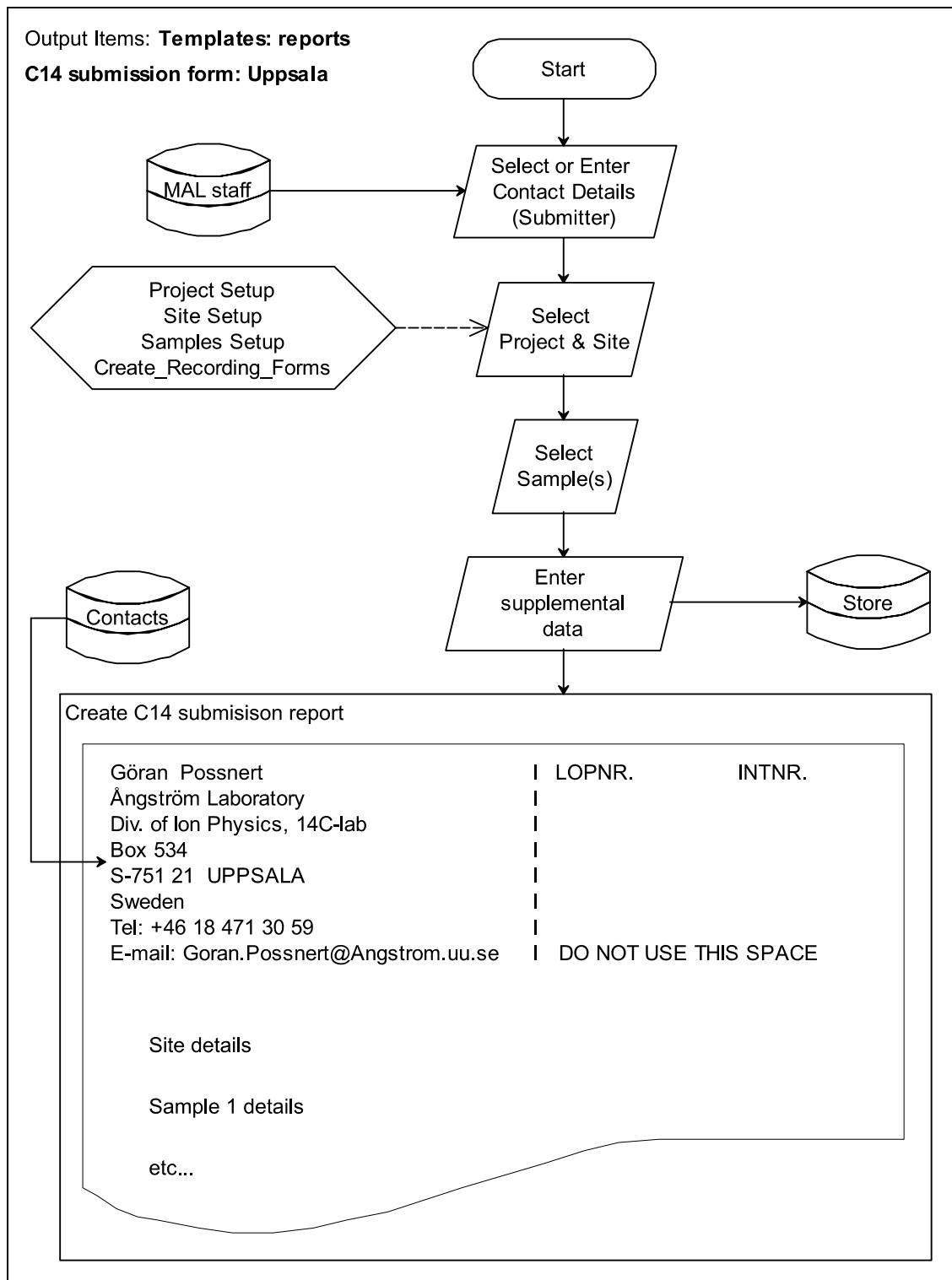
Figure 5.27. Process flow model showing the creation of a standard C14 submission form for Uppsala laboratory, with user input/selected data (including sample and submitter contact details).

5.6 Documentation

Good documentation is essential in both the continued use and future enhancement of a database system. Four levels of project documentation are envisioned:

i) Inline documentation of programming code and database objects – to enable future developers to rapidly gain an understanding of the system. In addition, programming and development logs will be maintained in which all changes will be recorded.

This will improve coordination between members of the development team, and help trace errors between versions.

ii) Application help – built in help files and tool tips accessible from the client software. This will not only provide instructions for users but will help to reduce the amount of support time needed.

iii) User guide and detailed operational and maintenance manuals – basic instructions for general users, and comprehensive instructions for the more advanced database functions, explanation of maintenance routines and customization possibilities (e.g. addition of templates or export formats).

iv) Web based manual and online help. The former will double as online marketing and profiling material - highlighting the features and research potential of SEAD.

In addition to the above, marketing and promotional material will be produced for distribution to institutions and at conferences.

See section 9.1 for details of the project's publication strategy, which is external to the documentation strategy outlined above.

5.7 Expansion Potential

Properly normalised data structures lend themselves well to expansion. The use of primary key fields allows additional tables to be easily related to the existing ones. Good documentation is, of course, a key issue when modifying any system at a later date. SEAD is designed with expansion in mind, and will be thoroughly documented at all levels – from the data structure through to the design and program code of the user interfaces.

Beyond the obvious addition of data, there are two key potential areas for expansion: 1) the addition of more proxy data sources; and 2) adding more functionality, which would enhance the analytical power of SEAD. These are both described below.

5.7.1 Addition of further proxies

Nodal tables such as tblSamples, provide natural hubs for the addition of further proxy data tables, and methodology descriptions can be easily added as new records into the lookup table for analysis methods: tblAnalysisMethods. The addition of the extra tables required to manage the new data would be a relatively simple procedure once the nature of the data is established. The creation of queries and interfaces would require closer consultation with researchers working with those proxies in order to establish their needs in terms of data retrieval and cross-correlation. Naturally, the input of raw data would be the most time consuming aspect of such a venture.

Examples are provided below for both an additional chemical property variable; and a biological proxy variable, which requires considerably more complex support data. The appropriate interfaces and queries for the input, manipulation and extraction of the new proxies could be derived from those already functional for the existing proxies, although there may be a need for additional interfaces to cater for greater variety in new user requirements (e.g. from environmental scientists). The addition of more unusual data (from the SEAD point of view), such as soil micromophology (images and quantifications), ceramics and osteology would require further consultation with the relevant experts before designs can suggested.

Simple presence/absence recording of further proxies used at sites will be afforded by a generic 'Other proxies' table which will allow users to at least see whether data outside of the scope of SEAD exists for their area of interest.

5.7.1.1 Example 1: Expansion to include Lead Isotope data

Lead isotopes are valuable indicators of past industrial activity and recent pollution. Only measured values should be stored – ratios and other derived values would be calculated on demand.

5.7.1.1.1 Additional tables required:

tblMeasure204Pb

tblMeasure206Pb

tblMeasure207Pb

etc... for all isotopes measured. These would be related to tblSamples using the SampleID field (Figure 5.7)

5.7.1.1.2 Records to be added to existing tables:

tblAnalysisMethods (Figure 5.8)	
Field	**Data**
AnalysisMethodID	(automatic)
MethodNameEng	206Pb
MethodDescriptionEng	Lead stable isotope 206 measurement...
Units	micrograms per kilogram
UnitsAbbreviation	ppb
MethodNameSwe	206Pb
MethodDescriptionSwe	Blyisotop 206 mätning...
InternContactID	(foreign key in tblInternalContacts)

...plus similar records for each isotope measured.

tblInternalContacts – Contact details for persons responsible for measurements in– if not already on the list of internal contacts (Figure 5.17).

5.7.1.2 Example 2: Expansion to include Diatom abundance proxy data

Diatoms provide information on palaeotemperatures and past influences on aquatic conditions, as well as being valuable indicators of present day water quality, and are even used in forensic science.

Note that a similar set of tables could be used to add chironomids to SEAD.

5.7.1.2.1 Additional tables required:

tblAbundancesDiatoms – abundance data (foreign key = SampleID in tblSamples)

tblAbundantIdentLevelsDiatoms – identification levels

tblTaxaMasterDiatoms – master species list

tblTaxaDiatomFamilies – family names

tblTaxaDiatomGenera – genera names

tblTaxaDiatomAuthorities – authority names

tblTaxaDiatomSynonyms – synonyms

tblRefCollectionDiatoms – reference collection information

tblEcologyDiatoms – modern ecology data

These tables would be related in a similar fashion to those shown for plant macrofossil, pollen and insect data in Figure 5.7 and Figure 5.9.

5.7.1.2.2   Records to be added to existing tables:

tblAnalysisMethods – description of diatom analyses (Figure 5.14)

tblBiblio – bibliographic information for ecology and methods (Figure 5.10)

tblInternalContacts – contact person details (Figure 5.13)

5.7.2   Expanding the analytical power and target audience of SEAD

The potential for enhancement is practically limitless, and more a question of prioritization and resources than possibilities. Listed below are a few high priority areas that this report has identified.

Biogeographical information

The current SEAD proposal includes the storage of ecological data for plants, which allows researchers to easily obtain information on the ecological and ethnobotanical significance of the plant macrofossils and pollen species found in samples. Note that this data is not included for insects, which are already catered for by the Bugs software package ([9]). The addition of modern distribution data would enhance this analytical power giving researchers access to comparative spatial data – a powerful tool in both climate and environmental research, which could help give insights into issues such as changing resource availability; the impact of climate change on vegetation; and human impact on the landscape. This data could of course be used to enhance the possibilities for GIS based analyses, and increase the user base of SEAD to encompass a wider variety of biodiversity and environmental change studies.

Images

There is currently no plan to include digital images of biological proxies due to the large amount of time involved in assembling them. This is an obvious area for future expansion which would be programmatically very simple, but requires a large investment in time. There are numerous, although generally incomplete, relevant images databases available online, for example, those available through NOAA[3] or the US Department of Agriculture[4], and various others either directly connected to environmental archaeology/Quaternary science, or otherwise related by their content. It is possible that cooperation between SEAD and the managers of these datasets could lead to mutually beneficial solutions. There are real possibilities for joint funding in this area.

Analysis tools

Another potential area for expansion is the enhancement of data interrogation tools, such as the inclusion of statistical methods for data analyses. The Bugs software ([1]) includes facilities for producing graphical environmental reconstructions, calculating correlation coefficients and the ability to perform limited resampling (e.g. Jackknife) on climate reconstructions. These are all features that could easily be ported over to SEAD,

and used as the basis for programming more analytical tools. This would increase the potential user base by providing access to methods that would normally require external statistical software. Internet implementation would be of particular advantage, and provide a practically unique resource (there are few free online statistical packages)

The potential for expansion along these lines is enormous. Cooperation with external partners and an assessment of their needs should be used to prioritise areas for enhancement.

Identification keys

Taxonomic keys to enable the identification of biological specimens would both expand the target audience of SEAD into modern ecology and museum collections work and provide a useful teaching and research aid. The Bugs database[1] already contains over 200 keys by the entomologist Peter Skidmore.

Swedish national archaeological laboratories

Cooperation with the other national archaeological laboratories of Sweden (which cover dendro, ceramics and osteology) could expand the scope of SEAD towards a *national resource for scientific archaeology*. The current design plan covers biological and physical proxies, and not artefactual or osteological data, the inclusion of which would require an additional planning period to ensure that the needs of the laboratories involved are comprehensively covered. The incorporation of these data areas would potentially enhance SEAD into being one of the most powerful databases and analytical tools for archaeology, environmental archaeology and Quaternary science on an internationally. Such a system would also have great teaching potential, especially for distance/internet based courses.

5.8 Backup Strategy

A rigorously enforced backup strategy will ensure the survival of SEAD data through system crashes, user error, and the inevitable teething troubles of a newly developed system. The following components are suggested, to be integrated into a strategy dependent on the volume of data stored and the personnel and physical resources available:

i) Onsite (tape) backup – daily incremental, weekly and monthly full backups

ii) Offsite server – external storage space though web hosting service. Although SEAD will be routinely replicated to the web server for internet access, an additional (compressed) backup of the full onsite database is highly recommended.

Backup implementation is considered as an integrated part of the operational maintenance strategy outlined below.

An additional offsite backup could potentially be provided by ADS (The UK based 'Archaeological Data Service' see section 7.1), who have expressed an interest in cooperation and have several years experience of managing and disseminating archaeological data. This could also increase the public accessibility of SEAD data through their web portal and provide a useful international development partner.

5.9 Maintenance Strategy

Without a coherent maintenance strategy a database's projected lifespan is limited. Both physical (hardware) and software components must be continually updated and repaired

to ensure smooth running, and compatibility with any future updates in the local and internet domains. An integrated maintenance strategy is suggested, where the degree of user interaction is high, and a certain amount of marketing and research orientated development is encompassed. This will ensure that future demand for new data areas and analysis features can be implemented rapidly, and thus providing a better service for current and potential users.

Maintenance will be documented with inline commenting and supplemental manuals. Logs will be maintained of all changes.

General office administrative costs must be added to the items below.

The following areas have been identified as priorities for maintenance (adapted from MS SQL Server 2005 documentation):

5.9.1 Database server and database

Monitor and troubleshoot physical server performance.

Monitor and troubleshoot database performance.

Troubleshoot and maintain query performance.

Monitor data quality and integrity, clean and repair damaged data.

5.9.2 Backup and emergency recovery strategy

Maintain and verify backups, tuning backup strategy when necessary.

Diagnose causes of failures and prevent repeats.

Recover from failures, including restoration of backups and minimizing data loss.

5.9.3 Data security, clients and users

Assess changing user's needs and database capabilities.

Provide support for users.

Identify new potential target groups and research areas (in order to ensure the future of SEAD)

Training of new onsite users.

Monitor and maintain database security policy (hardware and software), and ensure legislative regulations are met.

Maintain existing and process new (local and remote) users.

Install and maintain client software on new computers.

5.9.4 Update management

Ensure Microsoft service packs and security updates are implemented and that SEAD functions correctly with them.

Monitor and maintain replication to webserver.

Ensure metadata is entered and conforms to standards.

Maintain website and webserver components.

6 Limited Physical Prototype

6.1 Database Structure

The data structure for suggested SEAD was modelled using MS Visual Studio with an aim towards testing its ability to store the required data and the viability of the data flow (process) models. The resulting database structure diagrams are shown by data area in section 5.1.5, with an overview included in Appendix I. Note that not all biological proxies were modelled, and that the dual language facility was only implemented in a selected number of tables. This was purely a time saving omission, as the missing items would not differ in structure from those already modelled.

Potential common data retrieval queries were designed first as textual questions (based on the information gathered from potential users), and then constructed using Visual Studio's Query Designer. Examples are given in section 5.2, and schematized in Figure 5.22.

6.1.1 Evaluation

The data structure responded well to the construction of test queries, and the simulation of data flow paths. The highly normalised structure was shown to be efficient, logical and expandable.

6.2 Project Management Database (MALBase/Maldb)

A series of prototype project and sample management databases have previously been developed at MAL (see section 4.4), and these were evaluated for their potential contribution to the project structure and interface for SEAD.

6.2.1 Evaluation

A single file (combined database and interface) database application was developed using MS Access 2000 to test the proposed data structure and demonstrate the project/site management possibilities of a SEAD type system. A limited dataset, including a mix of real and simulated test data, was imported successfully, and variations on simple user interfaces (forms) were tested. The prototype was developed under the working names MALBase and later Maldb, a screenshot of which is shown in Figure 4.3.

It was shown that the proposed data structure adequately catered for the lab's current data domain, and that only minor adjustments to the entities and relationships would be necessary to create a more flexible and upgradeable system. These modifications are included in the design above (section 5.1). The interface concept, although limited in scope by the time available for prototype construction, was deemed to be efficient and intuitive in that users could quickly browse to and extract the level of data required for the interpretation of sites. Other forms of interface for the retrieval of advanced query data were discussed with staff and are described above (5.3.1).

The ability of the existing MAL database (Provbase – section 4.4.3) to provide answers to enquiries as to the status of work on different projects was considered particularly time saving.

7 Overview of Existing Similar or Related Systems

There is currently no publicly available database that covers the proposed scope of SEAD, both in terms of its data or functionality and web accessibility. Below are listed number of existing systems with which cooperation may be of mutual benefit in terms of both project development and user features. This is by no means an exhaustive list, and there are a number of small regional databases that should be further investigated should SEAD be expanded internationally.

7.1 ADS – Archaeological Data Service

http://ads.ahds.ac.uk/ads_summary.html

Maintained by

Department of Archaeology, University of York, UK.
Arts & Humanities Research Council, UK.
Joint Information Systems Committee, UK.

Purpose

Provide a central digital repository for all archaeological data in the United Kingdom, and support "...research, learning and teaching with high quality and dependable digital resources".

Scope of Data

Archaeology, sites, reports, metadata, UK.

Database/System Form

Public.
Online, searchable with a certain amount of downloadable datasets. All metadata is searchable, but not the individual sample and proxy data.
There is currently very little environmental data and no modern ecology data. The environmental dataset available would require significant reworking to be useful in the SEAD context.

Potential for Cooperation/database interaction

To quote from contact person at ADS:
In short ADS would be very interested in the proposed project and I can think of a number of areas in which we could contribute:

(1) As a digital archive, which special expertise in the preservation and curation of digital data.

(2) As a body with expertise in harvesting metadata and the interoperability of distributed datasets, including at European level through ARENA, and most recently through our involvement with cyberinfrastructure projects in the USA.

(3) As a data provider. We maintain several environmental datasets which might be usefully cross-searched with some of the data sets you mention, and have others in the pipeline.

Cooperation status

Contact established, and mutual interest acknowledged.

7.2 AIP – Archaeological Investigations Project

http://csweb.bournemouth.ac.uk/aip/aipintro.htm

Maintained by

English Heritage & Bournemouth University (UK)

Purpose

"English Heritage commissioned The Archaeological Investigations Project to undertake a detailed study of the nature and extent of archaeological fieldwork carried out in England annually"

Scope of Data

England, archaeological fieldwork meta data and summaries.
1990-2003 (current) – present day (intended)

Database/System Form

Online (ASP, SQL-Server), searchable, browsable.

Potential for Cooperation/database interaction

As a comprehensive database of archaeological fieldwork conducted in England, AIP could form an extremely useful starting point for the collation of data necessary for the construction of an English national version of SEAD. AIP does not contain any post excavation data, but includes summary information for reports which include environmental data.

Cooperation status

The AIP team have been contacted and have expressed an interest in getting involved in the SEAD project at an international level.

7.3 Intrasys

http://www.intrasis.com/

Maintained by

Riksantikvarieämbetet - The National Heritage Board of Sweden.

Purpose

Intrasis or *Intra-site Information System* (Riksantikvarieämbetet 2002) is an archaeological information system for on site use. The system is built to store and arrange GIS based data collected at archaeological excavations.

Scope of Data

Swedish, archaeological sites and features. This database allows the recording of "samples" (prover) but is not sufficiently detailed to enable the analyses required by environmental archaeology. The system lacks the ability to handle stratigraphic (chronological) and multi-scale spatial data in a manner that would conducive to detailed inter-site multiproxy interrogation.

Database/System Form

Commercial.
MS SQL Server relational database with ESRI ArcView integration.

Potential for Cooperation/database interaction

The intention is to examine to what extent Intrasis and the new environmental database SEAD could be integrated. The benefits would be significant if the two systems could be linked.

Cooperation status

Not initiated.

7.4 EPD – European Pollen Database

http://www.ncdc.noaa.gov/paleo/epd/epd_main.html

Maintained by

An Advisory Board and Executive Committee
http://www.ncdc.noaa.gov/paleo/epd/abec.html

Purpose

Make European pollen raw data publicly available.

Scope of Data

European wide storage of raw data from palynological investigations.

Some facility for the storage of other plant macrofossil data is built in.

Related projects: African, Canadian, Global, and North and South American Pollen Databases.

Database/System Form

Public data, with restrictions, downloadable.
SQL interface available on request; some online query facilities at
http://www.ncdc.noaa.gov/paleo/ftp-pollen.html
Online GIS interface for site based data retrieval at
http://www.ncdc.noaa.gov/cgi-bin/paleo/webmapper.cgi
Corel Paradox relational database.
No ecology or modern distribution data (?).
Limited interface – i.e. purely a database.

Related projects include regionally specific climatic, landscape and ecological calibration data. A variety of statistical tools are available for pollen data.

Potential for Cooperation/database interaction

Integrated search system, possibly through web interface could be mutually beneficial.

Cooperation status

Not specifically established, but EPD contacts with the lab exist.

7.5 Bugs Coleopteran Ecology Package

http://www.bugs2000.org/

Maintained by

Paul Buckland, Bournemouth UK.
Phil Buckland, MAL, Umeå.

Purpose

Bugs is a research and teaching tool based around a database of modern and fossil beetle (Insects: Coleoptera) ecology and abundance data. Bugs is used by entomologists, ecologists, archaeologists and Quaternary geologists alike.

Scope of Data

European fossil insect records.
Ecology and distribution data (non-GIS).
Climate and environmental calibration datasets.

Database/System Form

Public, downloadable.
MS Access database.

Potential for Cooperation/database interaction

An author of this report (PIB) is primarily responsible for the development of Bugs and cooperation is thus ensured. Much of the structure and utility of SEAD outlined in this report is a direct result of the development of Bugs.

Cooperation status

Guaranteed. A new version of Bugs will be released with Phil Buckland's PhD thesis ([1]).

7.6   *The Swedish Sites and Monuments Record (FMIS)*

http://www.kms.raa.se/cocoon/fmis-public/index.info

Maintained by

Riksantikvarieämbetet - The National Heritage Board of Sweden.

Purpose

Record geographical location and basic summary data of all archaeological sites and finds in Sweden.

Scope of Data

Limited to object specific metadata and geographical position. Meta data is not well structured and could be difficult to search effectively. Varies with county.

Database/System Form

State owned, can be queried online.

Potential for Cooperation/database interaction

Cross-referencing of SEAD and FMIS could be extremely useful in identifying the state of knowledge for particular places and times. This could be used to both prioritize research areas and acquire background information for projects.

Cooperation status

Not initiated, but cooperation assumed on grounds of state ownership.

8 Research Collaboration

8.1 Sweden

Contact has been made locally at Umeå University with the Department of Ecology and Environmental Science, and in particular members of the Environmental Change Assessment group (Prof. Ingemar Renberg)

Possibilities for collaboration between the other national archaeological resource labs should be investigated (see section 5.7.2). The proposed SEAD data, however, has physically more in common with Quaternary science/environmental change data than ceramics or osteological data, and more extensive additions to the suggested structure would be required to enable their incorporation. If this path is chosen then additional funds should be acquired for design and development at all phases.

Many sustainable development issues investigated by the Swedish University of Agricultural Sciences could benefit from a greater time depth, and SEAD could prove a valuable tool in this respect. MAL has a long history of cooperation with SLU in Umeå, collaboration opportunities are under investigation.

8.2 International

Environmental archaeology related laboratories and institutions, both national and international, will be encouraged to review and use SEAD. Interdisciplinary research is a high priority, and can only be achieved through the establishment and development of extensive contact networks.

Specific contacts have already been established with:

- The School of Conservation Sciences (Prof. P.C. Buckland, Bournemouth University, UK)

- The Institute of Archaeology (Dr R. Macphail, University College London, UK)

- Dept. de Edafologia y Quimica Agricola (Dr A. Martinez Cortizas, Santiago de Compostela, Spain),

- The Institute of Archaeology (G. Guðmundsson, Reykjavik, Iceland)

- The coordinators of other database systems as listed in section 7.

Inter-project cooperation is also essential in the acquisition of further funds and securing the future of research success. Cooperation has been initiated with the following projects and networks:

- Nordic Archaeobotanical Group (NAG) (Dr. Karin Viklund, Umeå)

- The UK based Leverhulme Trust funded project 'Footsteps on the Edge of Thule' (Prof. Kevin Edwards, Aberdeen; Dr. Eva Panagiotakopulu, Dr. Andy Dugmore, Edinbrough; Prof. Ian Simpson, Stirling).

- Sami Prehistory & Early History in the Western Barents Region (SAMARC) (Prof. Roger Engelmark, Prof. Thomas Larsson, Umeå)

- The Hansa-Network Project (Dr. Karin Viklund, Umeå)

9 Research & Teaching

9.1 SEAD Publication Strategy

A two component publication strategy for is envisaged:

1) Database publications: scientific papers on the database system itself, its concept, construction and possibilities.

2) Applied publications: scientific and popular articles relating to work using the database as a research tool.

It has been shown from previous database development projects ([5,6,9]) that a publication strategy based on these two principles produces an effective feedback basis for academic database systems, covering both the development and use of the system. Publications will be directed towards both scientific journals and conference proceedings, as well as to a lesser extent, popular scientific media.

9.2 Potential for Use in Research

The pooling of 30 years of research and consultancy results into a single, user-friendly database system, has immediate and obvious research potential, as discussed above. A number of priority areas are outlined below.

Sustainable development and past land management

- SEAD's concept of environmental data integrated with archaeological information and interpretations, provides a unique tool for analysing the scale and intensity of past landuse practices. It could also be used to provide empirical support/opposition for any existing theories, and provide the footings for new ones – including future predictions based on past evidence.

Biodiversity and cultural impacts

- Intimately connected with the concept of sustainable development when considered in the light of human impact, SEAD could be used to store and provide data on the presence/absence (and to a lesser extent abundance) of particular species at particular times and in particular regions. The GIS interface could easily be used to map the observed changes, and even simulate the potential future effects of modelled changes. The changing pattern of weed distributions with the expansion of agricultural practices, for example, could be effectively mapped using SEAD. Similar work has been undertaken with respect to the insect fauna of Iceland[7].

Environmental and cultural interactions in the North

- With a large part of its initial dataset coming from investigations into the past of Northern Scandinavia, and being based in Umeå, SEAD would be ideally placed for providing empirical support and analysis tools for backing studies into the history of culture-environment relations through time. One possible direction is cross-referencing with the Demographic Data Base[8] to look for possible correlations between past health and mortality and environmental change through proxy analyses. Changes registered in proxy archives (i.e. sediments) will often not be the same as those recorded in historical records, and the two sources strongly complement each other. Such investigations would also identify deficiencies in the proxy and historical records and act as a guide for future research priorities. This theme is both a research profile area at Umeå university and the theme of a major research initiative to be undertaken by colleagues in the UK ('Footsteps on the Edge of Thule', see section 8.2)

Scientific research programmes and research prioritization
- SEAD could be used to undertake background research, providing information on the current amount and type of evidence available for a particular region and/or period. In this way SEAD could be used to help assess the state of knowledge of environmental archaeological/archaeological investigations on a regional basis, and identify critical areas for future research. This would also be useful in designing comprehensive research programs, in that SEAD would be able to quantify the amount of work done in the proposed research area, and thus allow researchers to strategically target funding to achieve the most beneficial results. One could, for example, use SEAD to see exactly how much data exists on the Neolithic period of Skåne, relate this to the more intensively researched Iron Age of the same area, and then devise an investigative strategy for best analysing the development of agriculture in southern Sweden.

Collections and conservation prioritization

- SEAD could easily identify gaps is modern reference collections and preserved fossils, and thus direct researchers to preferentially collect or preserve particular specimens. This concept could be expanded to include digital photography of specimens (as described in section 5.7.2), and would be of benefit to a wider audience including natural history museums and collections coordinators.

Environmental and cultural resource management

- As record of archaeological and environmental investigations SEAD would provide a unique system for highlighting cultural environments in need of preservation due to either their special character or rarity. Alone, this would be extremely useful in the development of conservation programs at a national or local level, and could easily be integrated into community planning projects. In combination with research prioritization exercises this be an extremely powerful tool for shaping the future of

Methodological research and development

- With its ability to hold several analysis datasets from any set of samples, in addition to storing method descriptions and publications, SEAD could be used to compare analysis techniques and assess their relative advantages under differing situations. Even the 'simple' collation of data, and ability to export and summarize it, would provide unique test sets of a scale which permit the application of statistical methods normally avoided due to lack of data. (the environmental reconstruction methods implemented in BugsCEP[1] are an example of statistical techniques which have required database implementation.). SEAD could also be used to highlight deficiencies in certain methods or areas that lack a desired degree of quality or quantification – thus giving potential directions for the development future methods.

Multi-proxy analyses

- The integration of datasets from several proxy sources will be facilitated through SEAD's interface system. This will enable multi-site and multi variable analyses to be undertaken on a larger and more complex scale than is normally feasible. Such multi-proxy investigations could give new insights into the history of the interrelations of environmental and cultural change.

Advanced uni-proxy analyses

- The investigative tools built into SEAD, and its ability to collate and summarize vast amounts of data from a variety of spatial and temporal scales will allow for a deeper analysis of single proxy variables than is practical today.

Towards expert systems and The Grand Unifying Database

- The application of state of the art computational techniques such as neural networks and multi-database data mining could lead to the development of semi-intelligent analysis systems for the interpretation of data from archaeology and other Quaternary sciences. A multi-nodal system, essentially spanning several online databases, would be an interesting alternative to the age old concept of a single database for all Quaternary data.

9.3 Potential for Use in Teaching

Recent years have shown both an increase in student computer literacy and student demand for instruction and teaching involving computer software – especially GIS and information retrieval systems. The Bugs database system[9] has been successfully used as a tool in teaching palaeoentomology modules of lab courses during 2002-2005, proving useful both as a pedagogical medium for ecology and environmental science and increasing the computing competence of students. Bugs components are also used in teaching in palaeoecology and palaeoclimatology in a number of departments throughout Europe. It is anticipated that SEAD will perform equally well in a broader spectrum of teaching areas, and be especially useful as a data source for student theses.

The integration of SEAD into teaching in Umeå would give the department a competitive advantage which could lead to an increase in both undergraduate and postgraduate student numbers. In particular, the incorporation into a Masters programmes would be particularly strategic. An international Masters in Environmental Archaeology is under development at the department, in cooperation with the University of Bournemouth (UK) and potentially others.

10 Detailed Time/Cost Analysis

Hardware and software cost are estimates based on summer 2006 academic prices. Licensing situations and discounts may change and should be reassessed on the preparation of funding applications. Note that this cost assessment does not include any additional charges that could be incurred by the utilization of external data repositories such as ADS (see section 7.1), or any funds to be used by external partners.

10.1 Hardware & Software Requirements

10.1.1 Software

10.1.1.1 Development

It is suggested that SEAD development be primarily undertaken in the integrated developer environment of Microsoft Visual Studio Tools for Office and MS Office Pro. Umeå university has good support for this environment, which has a proven track record and exceptional online support through the Microsoft Developer Network and user community. MS SQL Server Developer Edition will be used to construct the database (backend) and stored procedures (queries). Visual Studio provides a system for developing database linked web applications, and affords enhanced developer functionality with the MS Office Pro Suite. The included Access Developer Extensions facilitate the seamless deployment of the local frontend application to client machines, the Access Runtime Licence allows the software to run on machines that do not have, or do not have the correct version of, MS Access installed.

Additional third party tools, such as Automated Form Resizing utilities and other existing software for improving the efficiency of development will be acquired as necessary. Table 10.1 and Table 10.2 summarise the proposed developer software solution and associated cost. Note that more than one license of some software packages will be required as at least two developers will be active.

Development area	Software
Database structure (Backend)	MS SQL Server - managed through Visual Studio

Onsite interface application (Local frontend)	MS Office Pro (primarily Access & Excel)
	MS Visual Studio Tools For Office
	    - programming in VBA & VB
	    - deployment with Access Runtime License where necessary
	3rd party tools and solutions as required

Web interface (Remote frontend)	MS Visual Studio
Onsite GIS interface	ArcGIS
Website (non-database interface)	Macromedia Dreamweaver (website) SwishMax (Flash authoring)

Table 10.1. Summary of proposed developer software

Item	Price (SKr)	Quantity	Totals
MS SQL Server	200	1	200
MS Office Pro	1000	2	2000
MS Visual Studio Tools For Office	2000	2	4000
ArcGIS (licensing/purchase cost estimate only due to ongoing negotiations – could either be annual license or one-off cost)	6000	2	12000
3rd party tools	4000	1	4000
Manuals & Reference literature	2000	1	2000
Macromedia Dreamweaver	(existing departmental licences)		
SwishMax	300	2	600
Consumables			5000
		Total:	**29800**

Table 10.2. Estimated cost of developer software (based on academic prices from summer 2006)

10.1.1.2 Operation

MS Office support will ensure that the SEAD client will run on the majority of PC's. The creation of runtime versions will ensure that no additional software is required beyond the MS Office Standard package.

ArcMAP will be required to run the onsite GIS interface and licenses should be obtained for every machine where its need is anticipated. See Table 10.2 for pricing details.

10.1.2 Hardware

10.1.2.1 Development

Two developer specification MS Windows compatible PC's, along with access to an onsite server and web hosting (as detailed below), will be required for the proposed development team. Additional PC's will be required should the team size be increased.

10.1.2.2 Operation

As described in section 5, SEAD will be developed both (i) as an onsite client-server system and (ii) for remote access via the internet. The hardware solutions for both of these components are given below:

i) Low end server capabilities will suffice for running SQL Server at the current scope (minimum 1GHz CPU/1Gb RAM recommended). Some redundancy should be obtained with the purchase of a more powerful server.

ii) To avoid the risk of data loss and hacker induced damage, and to reduce maintenance costs, external web hosting will be rented to store a replicated MS SQL Server database and web interface. This space will also be used as an offsite backup of the onsite database, and optimize the use of the rented space. The currently acquired domain name www.sead.com will be used as the online database portal.

iii) A high capacity backup solution should be acquired for onsite use. The use of removable media will provide a fire and burglar proof system, with the storage of the media in existing safes at the lab in addition to an offsite location.

iv) Tablet PC for conference and meeting use. Enables effective sketching, planning and display of ideas in any meeting environment.

Item	Initial cost	Quantity (3 years)	Totals	Annual cost
PC: Developer workstation (Medium spec's, two screens, ergonomic)	15000	2	30000	
PC: Local database/file server	25000	1	25000	
Backup solution	25000	1	25000	
Tablet PC	15000	1	15000	
Web hosting (MS SQL Server capable)	1000			1000
Consumables	5000	1	5000	5000
		Totals:	**91000**	**6000**

Table 10.3. Estimated hardware costs (summer 2006)

10.2 Programming, Development & Implementation time

Note that time estimates indicate the projected time required for one person to complete the given task. Estimates are calculated from an analysis of the data and task volumes and complexity, and from standard guides to database project budgeting. See section 3.2 for proposed scheduling of phases and more details of specific task.

Time estimates are rounded up to the nearest month. Estimates also include the time required of staff outside of the development team – e.g. training requires that those to be trained allocate time for the process.

Table 10.4 summarizes the projected times and costs, which are described in more detail below.

Developmental costs			
Phase*	Item	Months	Cost (SEK)
2	Database (backend)	2	
3	Local client software (local frontend)	11	
3	Onsite GIS interface	4	
3	Web interface including GIS simulation	4	
6	Documentation	4	
5	Data Conversion, Import and Entry	14	
4	Implementation, Training and Parallel Running	12	
8	Evaluation	1	
	Publication, Presentation, Meetings and Marketing	2	(p/a)
	Consultation with experts		10000
	Software		29800
	Hardware		91000
	Sum	**54**	**130800**

Annual costs (p/a)			
	Operational Maintenance (website, data, software & hardware)	2	
	Conferences, meetings, marketing, publications		25000
	Software ¤		12000
	Hardware		6000
	Annual Sum	**2**	**43000**

Table 10.4. Summary of time required and projected cost for SEAD development.
* Developmental phases as shown in Figure 3.1, and described below.
¤ Annual software costs dependent on ESRI licensing situation

10.2.1 Skills required for development team

- A working knowledge of environmental archaeology and laboratory based scientific data. Knowledge of taxonomic systems and/or geoarchaeology would be of considerable advantage.

- Experience of database development (preferable using MS SQL Server or MS Access).

- Knowledge of programming in an MS Windows environment (preferably object orientated and MS VBA/VB).

- Advanced skills in MS Word, Excel and Access would be of considerable advantage.

- GIS knowledge, preferably including application development.

- Advanced web authoring skills, preferably including graphic design, Flash development and database access (ASP skills would be advantageous).

10.2.2 Estimate of programming and database development time

10.2.2.1 Database (backend)

This planning phase document outlines the required database structure, and as such the time required for development includes only the physical construction of the data elements and relationships and a limited set of queries. Fine tuning of this structure is inevitable during the other phases.

Approximate months required: 2

10.2.2.2 Local client software (local frontend)

From past experience it has been seen that, aside from the actual input of data, the development of user interfaces is one of the most time consuming components of database system development. This is primarily due to the complex and variable nature of the data involved, and the unfamiliarity of the target users with database (and in some cases computing) concepts. Development in such cases becomes as much a dialogue between the developers and the users on the possibilities of interface design, as a series predetermined development tasks. This rings particularly true for systems with a wide and varied user base.

It is estimated that a period of at least one week per user form group is adequate, including the design and coding of interfaces, construction of underlying queries, reports and validation checks. Fine tuning of these interfaces will be inevitable during the Implementation, Debugging and Parallel Running phase.

43 initial form groups have been identified, covering the interfaces required for the entry, retrieval and management of the data areas outlined in the database design (section 5), and shown in the structure diagrams (see Appendix I for an overview). This excludes GIS and Web interfaces.

Approximate months required: 11

10.2.2.3 Onsite GIS interface

A customized ArcMAP interface integrated through MS Visual Studio to SQL Server.

Approximate months required: 4

10.2.2.4 Web interface (remote frontend) including GIS simulation

A combination of Dreamweaver, SwishMax Flash authoring and MS Visual Studio database integration will provide the necessary tools for advanced website development.

Approximate months required: 4

10.2.3 Documentation

 i) Inline documentation of database and code – included in programming & database construction time.

ii) Application & web help files

iii) User Guide

iv) Detailed operational and maintenance manual

Approximate months required: 4

10.2.4 Data Conversion, Import and Entry

The most time consuming single component of the SEAD database development. See section 4 for details of the relative volumes data and complexity of tasks involved.

Approximate months required: 14

10.2.5 Implementation, Training and Parallel Running

Close contact with MAL staff and partner researchers will be essential in activating SEAD. The estimate includes time for MAL staff and external cooperation partners to interact with the developers. MAL staff, who will initially be the primary data managers will need to be instructed on the use of SEAD. They will also form the primary test population.

Approximate months required: 12

10.2.6 Evaluation

SEAD will undergo a fully documented evaluation on completion.

Approximate months per year required: 1

10.2.7 Operational Maintenance (data, software and hardware)

A certain degree of maintenance, as outlined in section 5.9, will be required to keep SEAD running. One day a week should be an *absolute minimum* for a database manager and administrator. This will not include data entry or any major redevelopments or enhancements. This figure should be increased should the scope and cooperation base of SEAD be increased.

Approximate months per year required: 2

10.2.8 Publication, Presentation, Meetings and Marketing

It is suggested that time be allocated for the preparation of presentations and marketing material in order to promote awareness of SEAD. This will be a combination of website promotion, brochure distribution and poster presentation.

Meetings, both internal and with external partners will be required during the entire developmental phase.

Developer time should also be allocated to provide close support with those preparing and writing up research which implements SEAD. The developers will also be required to write papers concerning the construction and workings of SEAD as outlined in the publication strategy in section 9.1.

Conference participation, and the associated material production is included in this unit.

Approximate material and logistic costs:

>*1 international conference per year for two people: 20000 Kr/yr*
>*Review meetings at 6 monthly intervals: 5000 Kr/yr (will increase if more external partners involved)*

Approximate months required:

>*Writing publications, brochures & posters: 2 months*

10.3 Commercial Potential

SEAD will most probably be a commercially viable product and service. There are, however, currently no plans to realize its commercial potential as this is regarded as detrimental to the principles of academic freedom and open access.

The physical presence of, and the experience gained during the development of SEAD will greatly enhance MAL's national and international research status and competitiveness and will thus lead greater possibilities for attracting funding and research to Umeå University.

10.4 Assessment of Available Resources at the Institution

10.4.1   Hardware & software

Current fileserver – below minimum specifications for MS SQL Server

Developer PCs - one developer specification PC will most likely be available at the time of project commencement. MS Office Pro is installed, but no developer software.

Client machines - MS Office is installed on all lab and staff machines, which would most probably be able to support the SEAD local client software.

This situation should be reassessed on funding application preparation, and relevant experts should be consulted before purchases are made.

10.4.2   Human resources

Environmental archaeology

The permanent staff of the Environmental Archaeology Lab have considerable knowledge of environmental archaeology and scientific data in general.

Development team

The lab is in contact with individuals with skills covering the majority of those required for the development of SEAD (see section 10.2.1 for an overview). The remaining skills could be acquired under the early developmental stages with a minimum amount of training. Time should be allocated for the development team to acquire training in any areas not covered by their previous training or experience.

Consultation time with database experts should be arranged in order to confirm the plans outlined in this document.

Data entry

Data entry requires a lesser degree of computational proficiency, but a working knowledge of environmental archaeology. Previous experience of work at MAL would be of considerable advantage where the conversion and import of existing data is concerned. Data entry staff should be the opportunity to acquire basic training in software debugging and development so that they are able to better provide feedback to the development team.

11 Conclusions

11.1 Final SEAD Design Criteria

On completion, SEAD should:

- Be easy to use and intuitive for a group of users consisting of environmental archaeologists, Quaternary geologist and other palaeoecologists; students in these fields.

- Encompass the present full data type scope of MAL and be expandable into related areas.

- Contain the entirety of MAL's existing published data.

- Have internet access.

- Have a working data retrieval system with both database form based and GIS access.

- Be able to export selected data (query results) in common formats, as reports and distribution maps.

- Be available for use in teaching and research at the department and among partner institutions.

In addition it should:

- Have a working maintenance, backup and future funding strategy.

- Be involved in current or near future research at the department or among partner institutions.

11.2 Check List for Funding Application Preparation

- Check academic prices for software.

- Confirm ESRI ArcGIS licensing situation

- Check Hardware costs.

- Consult with database & hardware experts on technical issues (although this is more important upon commencement of the project).

- Check staff availability and commitment possibilities.

- Reconfirm any external cooperation.

- Calculate employment time (including holidays, training & staff commitment)

- Check status of report digitization project (may have to incorporate into SEAD if not progressing)

- Cursory examination of existing systems to look for updates

- Re-establish cooperation contacts, and probe for new ones.

- Look for EU partners

11.3 Final Remarks

As mentioned in the introduction, there is a real need for SEAD. There is currently no existing system with similar coverage, both in terms of data type and area, and functionality and access facilities. The construction of SEAD would be a important step forward in the public/research community dissemination of environmental archaeology data and results, and could lead to a significant upswing in research and collaboration on the national and international playing fields.

SEAD will contain a wealth of data with significant importance for analysing the progress of Holocene environmental change and human impact. Whilst this in itself is interesting, it can also be applied to viability studies for future sustainable development, in that the interplay between past farming and other industries with the natural environment is the *only* data set we have for predicting the impact of future practices. The analysis tools proposed in SEAD would greatly enhance the efficiency with which researchers could undertake such studies.

Studies on long term changes in biodiversity could be boosted by building on existing, and initiating new research cooperation with related fields in Quaternary science, environmental change, forestry/agriculture and ecology, and by cross-querying SEAD with databases in those areas. If the comparative data is lacking or not yet collated then SEAD could be adapted to include it.

The human aspect is often missed in ecological and environmental sciences, often due to a lack of access to the data or the skills to interpret it within those fields. SEAD would facilitate closer cooperation between archaeologists and these groups and thus further enhance the frontiers of multidisciplinary studies into biodiversity, sustainable development, environmental change and human impact.

12 References

[1] Buckland, P. I. (in prep.) *The BUGS Coleopteran Ecology Package. The Development and Implementation of Software for Palaeoecology and Palaeoclimatology*. PhD Thesis, Umeå University. (http:/www.bugscep.com)

[2] Gifford et al. (1996). Access 95 Unleashed. Sams Publishing, Indianapolis

[3] National Oceanic & Atmospheric Administration (USA) http://www.ncdc.noaa.gov/paleo/paleo.html

[4] http://www.geocities.com/seedimages/

[5] A. Kreuz/E. Schäfer *A new archaeobotanical database programme* Vegetation History and Archaeobotany 11, 1-2, 2002, 177-179.

[6] Sadler, J. P., Buckland, P.C. & Rains, M. (1992) *BUGS: an entomological database*. Antenna. 16, 158-166.

[7] Buckland, P. C., Dugmore, A. & Sadler, J. P. (1991). *Faunal Change or taphonomic problem? A Comparison of modern and fossil insect faunas from south-east Iceland. Environmental Change in Iceland Past and Present*. Maizels, J. K. & Caseldine, C. Dordrecht, Kluwer: 127-146.

[8] Demographic Data Base (Demografiska databasen) http://www.ddb.umu.se/index_eng.html

[9] Buckland P.I., Buckland P.C, Yuan Zhuo Don & Sadler J. (2002) *Bugs Coleopteran Ecology Package* [CD-ROM].