ENDURE

European Network for Durable Exploitation of crop protection strategies

Project number: 031499

Network of Excellence
Sixth Framework Programme

Thematic Priority 5
FOOD and Quality and Safety

Deliverable DI4.3

Feasibility study for the development and implementation of a dynamic online database for optimal integration and management of relevant knowledge.

Due date of deliverable: M18

Actual submission date: M18

Start date of the project: January 1st, 2007   **Duration**: 48 months

Organisation name of lead contractor: JKI

Revision: V1

Project co-funded by the European Commission within the Sixth Framework Programme (2002-2006)	
Dissemination Level	
PU Public	
PP Restricted to other programme participants (including the Commission Services)	**X**
RE Restricted to a group specified by the consortium (including the Commission Services)	
CO Confidential, only for members of the consortium (including the Commission Services)	

Table of contents

Glossary

Actor	Defines a role to contain a user or group of users. Note that the user does not have to be a real person. An application may also be an actor.
Application server	Runtime environment implementation for web-based applications
Category	A synonym for one entity of the meta data.
Database	Facility to persist structured data.
EPPO	European and Mediterranean Plant Protection Organization
Java	Object-oriented programming language
Java 5 EE	Runtime specification for web-based applications
JSF	Java Server Faces
NUTS	Nomenclature des unités territoriales statistiques
Online database	In the document used as synonym for Endure ALPS.
OR-Mapper	Object-relational Mapper
ENDURE applications	In detail web-applications of the participants (TwoTracks, VirtualLab, Endure IC) is meant.
RDMS	Relational Database Management System
Report	Summarizes all the data uploaded by an expert to describe a certain application of measure.
Use case	Defines an interaction between an actor and a system

Summary

The technical part introduces the specification of the data for the online database. Thereby the data is divided into two parts. The first part acts as meta data describing and categorizing the managed reports. The second part consists of the reports themselves. This part also contains information about the expert allowed the insert reports. Especially restrictions concerning experts exceed the data structure of JKI ALPS and therefore a migration of JKI ALPS' content is complicated due to the architecture and the intention to integrate ENDURE applications on a common platform.

The technical part also explains the requirement of a dynamic architecture for the new online database. Due to a lack of specific definitions a later common platform (internal back office) is hard to consider. Implementing the single functionalities as separate components provides a high reusability later. This will ensure the use of Endure ALPS as bases for a common platform combining all ENDURE applications currently known.

The technical part introduces the architecture and implementation of Endure ALPS prototype in a dynamic way, allowing reuse of almost each part of work.

It is essential for the goal of integrating of the data bases to adopt the identified entities and the technical structure by all participants and their ENDURE applications. The neglect of this fact will cause severe failures of the system for general integration of all knowledge generated by the network.

The content part has been defined and harmonised procedures have been discussed by the group, improvements will be implemented with the upload and an internal revision process.

The reports representing expert knowledge based on various single sources. Sources will be collected in regard to specific crop/pest or disease/measure/region combinations. A guidance document has been developed ensuring harmonised procedure for the collection of the information and general criteria to be fulfilled. With regard to the description for the practicability a more detailed description of the measure indicating benefits, advantages compared to common practice and bottlenecks is used as well. At the initial phase each partner in IA4 will name an expert who will be authorised to enter the knowledge about non-chemical measures in potato and the close collaboration with the CS Maize group will cover knowledge in that respect.

The technical background for the ENDURE ALPS is defined and first steps in programming have started. A presentation of the prototype will take place in October where first uploaded knowledge will be represented.

ENDURE ALPS will deliver a valuable source raising the awareness and application of non-chemical measure in various crops in such way contributing to a big extend to the reduced reliance on chemical pesticides.

Definitions

All *italic marked identifiers* within a text block generally refer to identifiers from conceptual figures.

1. Technical Implementation

The intension of the chapter is to point out the further planning concerning the technical implementation of the dynamic online database for optimal integration and management of relevant knowledge, further called online database. Incidentally the term online database seems misleadingly from technical point of view. A database is always a defined structure of the arrangement of several kinds of data. A database always provides mechanisms to insert new data into structure or delete data from the structure. Of course the query of data is also provided. However these interactions are always performed on a very low level. The most common interaction is supplied by text-based protocols not comparable with a complex user interface. Web-based applications as Endure ALPS are not databases from technical point of view but may use databases to save their content in a structured manner. Work with this structured data requires processing before representing it to the user. In the document the term online database should be kept if the whole application is meant. For real database in technical context behind the scenes the document will use the term RDMS[1].

1.1. Data Definition

The most important requirement for a knowledge management system is the design of a suitable data model. The definition not only describes a logical structure of data but also the structure used within the RDMS. Thereby the data model has to include two groups of data. The first group contains the meta-data used to describe the context of any content managed by the online database. The meta-data acts as informational base for the real content.
This meta-data must also provide a common base for all ENDURE applications currently in development to realize a common platform offering all participants' services.
The second group of data contains the content itself. Associating the content with the meta-data generates the knowledge, the online database must provide.

1.1.1. Information Base

To realize a common information base for all ENDURE applications currently in development a common data model was already adopted during the meeting in Pisa. The common data model intends to serve as common knowledge pool extended by each application with his own data. As a preliminary condition for the speciation of the relevant knowledge of the online database it seems useful to summarize the most import aspects of the commons data model. A complete overview of the whole commons data model is displayed in Figure 1 Information Base Overview" of the annex.
Later in the document the single entities introduced in the following are summarized as categories.

1.1.1.1. Crop

A *crop* entity is used to provide a context for any cultivated plant known for the online database. Beyond a common and scientific name the *crop* entity also provides taxonomical information. The taxonomical information as well as common and scientific names are based on the information given by EPPO Plant Protection Thesaurus (EPPT)[2]. Concerning the *crop* entity identifies a single cultivated plant by the corresponding EPPO Code. Therefore the EPPO Code is used as synonym for any cultivated plant within the whole online database.

---

[1] Relational Database Management System
[2] http://eppt.eppo.org/

1.1.1.2. Pest

The *pest* entity provides a context for all pests, diseases and other harmful effects on plants health. As well as *crop* entity the *pest* entity is based on common and scientific names as well as a taxonomy provided by EPPO Plant Protection Thesaurus. Additionally the *pest* entity also provides specialisations for weeds and diseases.

1.1.1.3. Region

The *region* entity provides spatial information to associate a set of specialized knowledge with a certain location. As common base for regional division the separation specified by Nomenclature of territorial units for statistics (NUTS)[3] is used. Concerning NUTS the *region* entity provides the countries and the NUTS1 regions at least. The sub regions defined by NUTS are possible to integrate but it is not planned at the moment.

1.1.1.4. Topic

The *topic* entity intends to provide a common thesaurus for the online database. The common thesaurus shall serve as a base for a later knowledge integration of the single applications. Therefore the *topic* entity acts as a summary of all problems, measures and pertinent solutions the contained knowledge is based upon. To ease the usage integration of the topics, the *topic* entity provides a scope beyond other features. That scope can be used to group several topics within the same context.

1.1.1.5. Contact

The *contact* entity is defined to allow an access management of the common data model . The *contact* entity describes all users working with the online database. Especially providing information about the author of published knowledge is one of the main reasons of the *contact* entity.

1.1.2. Knowledge Base

Built on the information base the online database has to provide reports related to the common data. A deviation is the definition of an additional *measure* entity. The *measure* entity is a specialisation of the *topic* entity. This specialisation has been introduced for performance reasons only and therefore not discussed in detail.

Any report is generally composed by an expert. Because not every user is an expert and not every expert is an expert for any use case, the online database has to manage experts and their restrictions beyond the reports. Reports, experts and restrictions are discussed in the following. An overview of the complete knowledge base data model can be found in "Figure 2 Knowledge Base Overview".

1.1.2.1. Experts

Authorised users of the online database are already provided by the *contact* entity of the common data model. Because not every user is an expert, a further division must be undertaken. Therefore the *expert* entity serves as a specialisation of the *contact* entity. While all ENDURE partners are authorised to view all knowledge provided by the online database, experts are authorised to upload new knowledge into the online database.

1.1.2.2. Assignments

To increase the quality of published content of the online database, it seems useful to not explicitly allow any expert to publish reports to any kind of category. The online database intends to deny the upload of new content for any expert except for those categories allowed

---

[3] http://ec.europa.eu/comm/eurostat/ramon/nuts/basicnuts_regions_en.html

by the *expert-assignment* entity. Therefore the *expert-assignment* entity always allows the extension of a category defined by a *crop* entity and a *pest* entity by a certain expert.

1.1.2.3. Reports

The core of knowledge provided by the online database are the reports composed by experts. A *report* entity always defines a practicability and a status. The *status* entity is intended to summarize reports to groups so handling the groups in a different manner is generally possible. The main intention of the status is to distinguish between reports for internal use inside the Network and reports accessible and for transfer to the public without further restrictions.

A report is divided into three parts. Beyond the summary, which is a textual description, a report also provides information about the relevant type of measure such as preventive measure, direct measure. Additionally any report also indicates the document it is summarizing.

1.1.2.3.1 Use Case

The *use case* entity always defines a specific area of application. Therefore a use case is always related to a crop, a pest, a measure (topic) and a region. Main intension of the online database is to associate such a measure type with reports already done by experts. Any use case may be described by several summaries including the indicated documents, but a summary may also describe several types of measures.

1.1.2.3.2 Summary

The *summary* entity provides the textual description. Generally any report only includes one description. A *summary* entity also specifies the language it is composed in. Even if only English is intended as online database's language at the moment, specifying the languages for a summary may ease a later internationalization of the database.

1.1.2.3.3 Documents

As already mentioned the textual description of a report is a summary of attached documents. Such documents may either be publicly accessible documents or internal documents not accessible. The latter are considered not available for the public in general but not necessarily meaning restricted access. The online database also has to provide a mechanism to attach such kind of documents while inserting on one side and allow the user to view the document on the other side.

1.2. Data migration

The previously represented data structure already bases upon the data structure of the JKI ALPS (formerly BBA ALPS). However there are also several but important differences between the data.

1.2.1. Structure

Most of the represented aspects are similar to those of the JKI ALPS, but there are still tiny important differences. Because the JKI ALPS only was intended to provide data for Germany, it does not contain a region entity as part of its meta data. The JKI ALPS also just provides a single name for its *pest* and crop *entity*. Several names for the same entity are not intended. Even if Endure ALPS only have to provide the scientific name beyond the common name, it is not possible with the structure given by JKI ALPS.

The most important difference within the report data is the feature of expert assignment. Because JKI ALPS only was intended for a manageable group of internal scientists, a more complex restriction management was not necessary.

1.2.2. Data

The data of JKI ALPS only is available in German language. In order to migrate the data the whole content primary has to be translated, followed by adapting the translations to the new structure. The changed structure involves that migrating the translated content in an automatic way is quite difficult. The most realistic way is to add the translated reports of JKI ALPS using the insert function of the user interface.

1.3. Functionality Definition

The main functionalities of Endure ALPS are quite similar to the functionalities of JKI ALPS. Endure ALPS extends the scale of functionality by interacting with other members of the Endure network. However a minute definition of the interactions is still not decided yet. Especially the integration of all applications of the currently developed applications to a common Endure platform is also still unsolved and not discussed here.

Main tasks of the online database are adding and searching for the already discussed reports. Around the main tasks there are also less important tasks. In the following all of the tasks provided by the online database are described in detail. "Figure 3 Use Case Diagram" gives an overview of all tasks in form of a use case diagram.

As seen in "Figure 4 Actors" the system intends three kinds of actors. The *contact* actor is a registered user. The user has reading access to any kind of knowledge intended for publication. Another kind of actor is the already mentioned *expert* actor. The experts are intended to provide knowledge for the online database. The least kind of actor is an *administrator* actor. As the name implies this actor is only intended for administrative tasks.

1.3.1. Search for reports

A main task of the system is to allow a *contact* actor to search for reports. After setting the wished search parameters the online database has to provide all relevant reports as result. The relevance depends on the specialized kind of search. Currently only two search facilities are planned, described in the next sub items. A combination of both search facilities is not planned.

Results are generally published as list independently from the search facility. The list has to be set by relevance of the included reports.

Additionally the online database has to provide mechanisms to restrict the result lists by filter criteria. How far the usage of the filter criteria can be chosen by the user is unresolved at the moment. A specific example for a filter may be the exclusion of reports marked as non-practicable.

1.3.1.1. Search by criteria

In this case the search parameters are criteria in form of the single categories. Depending on the user one, several or all categories can be specified. According to the specification of categories the online database displays all relevant reports. In this case relevant reports are those exactly matching the specified search criteria. Additionally the online database also has to display those reports matching after generalising the search criteria. Generalising search criteria means going a level upward within categories' taxonomy. Search results have to be set by relevance. The single search result is most exact match the less search criteria had to be generalised. Because this task is one of the main tasks of the system the complete one can be found in form of an activity diagram under "Figure 6 Search for reports by criteria".

1.3.1.2. Text search

Beyond a search by criteria a user must also be able to search for reports by text modules. Text modules may be a single term, several terms or a phrase. Therefore only the word stem

of given terms is relevant. The search facility must not distinct between singular and plural. If more than one term is given the user must be able to decide whether the results must contain any of the terms (disjunction) or all of the terms (conjunction). *Text search* results must be arranged according to relevance as well as to results of *search by criteria*. Relevance for several results might be the number of matching terms (disjunctions) per result for example. Further relevance criteria might be possible.

1.3.2. View reports

Search results are already represented as report list which is only containing the report title and the cue for the practicability of the described measure. A detailed view on each of the reports must be ensured containing the textual summary and the related documents. A detailed view includes also bibliographical reference for the report given. Within the view the user must be able to navigate between the single reports of the result list.

1.3.3. Insert reports

A report has to refer to a crop, a pest/disease, a non-chemical measure and a region.
An expert is authorised to add reports after registration. The report has to include a summary and a title. The summary may contain simple text formatting (bold, italic, underlined). Bibliographical reference can be inserted in the summary by the expert. Furthermore the expert must specify the documents he relates to. If those documents are public sources (websites) a simple reference to the site is sufficient. For non-public sources the documents have to be attached and should be submitted in line with copyright.
A complete definition as an activity diagram can be found under "Figure 5 Insert reports".

1.3.4. Edit reports

An expert is also able to edit his existing reports. Any detail of information might be changed, as well as the summary, the title or a reference to any related document. The reference to crop, pest, measure or region might also be changed if needed.

1.3.5. Provide common data

Insertion of categories occurs before the start-up of the online database. Insertion will be done by the database administrator on database level; therefore an input mask is not required. Changes on the categories may occur seldom or never; if necessary they will be realised by the administrator on database level.

1.3.6. Invite expert

The expert has to be registered into the online database to submit reports. With registration, the assigned categories the expert is authorised to submit reports are defined. As well as forinserting the common data, the experts are registered by a database administrator; an input mask is not necessary.

1.4. Preliminary consideration

As the name already implies, the online database must act as a web-application accessible through users' web browser. The Endure ALPS online database is based on the already existing ALPS (JKI) online databases. The content and the web-based user-interface are quite similar. Main difference between both is the interaction with other members of the Endure network on a technical level. Currently it is planned to establish an interaction between the Endure ALPS and the other applications such as the Virtual Lab, the Two tracks system. Furthermore an integration of Endure ALPS' content within the Endure IC is required. Several concepts were already presented for that purpose but a definite definition does not exist yet.

1.4.1. Basic approaches

The next target for Endure ALPS is running prototype. Interactions between the prototype and the other applications are not intended, but a final integration of all applications to a common platform is planned. These requirements necessitate a flexible, dynamic architecture model, open for extensions. A fixed implementation alike ALPS (JKI) is not able to comply with the requirements. Therefore two different models are shortly introduced here which are general options.

1.4.1.1. Component-oriented approach

A sufficient basic approach is the implementation of a component-oriented architecture. The basic idea of that policy is to encapsulate the most important core features into closed reusable bricks. Each of those bricks applies well-defined interface for data input and data output. Each of that bricks is a component.

1.4.1.2. Object-oriented approach

Working with different components upon the same data requires the definition of a clear data model. This is already done by the used RDMS. However state of the arts is to work with an object-oriented presentation of the data, called an object model. Working with the object model also requires the implementation of a mapping mechanism to transform RDMS' structured data into the object model and back. Working with the object model is quite easier than working with RDMS' text-based interface and for the purposes sufficient.

1.4.2. Runtime Environment

According to the basic approaches the Java 5 EE environment seems the most matching technology for our purpose. Therewith the Java Servlet technology of JKI ALPS is also included. Beyond the Servlet technology which is already part of Java 5 EE several others are also provided, especially those needed to realize the component-oriented approach.

A suitable implementation of the Java 5 EE standard is delivered by Sun Application Server Glassfish. Except for Glassfish all environment requirements are almost the same as those of JKI ALPS. The complete list of environmental requirements is already covered by the adopted runtime environment as follows:.

1.4.2.1. Software

- Client:
  - HTML 4.0 compatible web-browser
  - JavaScript
  - CSS
- Server:
  - Java 5 EE
  - Sun Application Server Glassfish
  - MySQL 5.x
  - SSO Server zur Authentifizierung

1.4.2.2. Hardware

- Client:
  - internet-capable computer
- Server:
  - internet-capable computer
  - suffient hard disk space

1.4.2.3. Orgware

- connection to the internet

- database administrator

1.5. Implementation Concept

A functioning prototype is the first target. Therefore the prototype implementation must be flexible in a way either allowing the reuse of many components or act as a base for a final common platform. Several concepts for reusable components are already provided by the Java application server (Glassfish).

1.5.1. Layer separation

The Java Servlet Technology was used for the JKI ALPS. Such Servlets generally represent reusable parts of an application, but the Servlets of JKI ALPS were built in a way used as core of the whole application. This means they were responsible for data access on one side, but also for website creation on the other side. Therefore these Servlets are not reusable in the proper meaning of the word because they are tailor-made for JKI ALPS especially for its data structure.

1.5.1.1. Presentation layer

Endure ALPS also will use Servlets, but in a more common and reusable way. In detail an extension of the Servlet Technology is used, the Java Server Faces (JSF) framework. The operational area of JSF is the implementation of web-based user interfaces. Therefore the JSF already offers a wide range of components used to build website. Further custom components can be created easily.

Even if the creation of the common platform requires a complete re-creation of a user interface, it is almost possible not only to reuse the standard components provided by JSF, but also those components created for Endure ALPS especially.

Therefore by presentation layer only has to read user input and write output for the user. Any processing aspects are moved to logic layer.

1.5.1.2. Logic Layer

Following the principle of layer separation allows an easy exchange of each of each layer. It is unlikely that such an exchange is required, but a clear separation also allows changes within the functionalities without affecting the presentation layer.

To provide the reuse of the functionality it is also intended to separate the core functions of Endure ALPS into components. Currently the most import core functions are searches for reports and inserting new reports. It is planned to implement the function based on two different Enterprise Java Beans (EJB). EJB is the appropriated specification for functional encapsulation within logic layer.

Whereas the process of inserting reports may not change, it is possible the exchange the Search EJB by a more powerful implementation, not only able to search in Endure ALPS database but also in other sources like those of other applications.

1.5.1.3. Persistence Layer

Persistence layer contains those mechanism used to persist any kind of data this generally an RDMS. In our case we do not access the RDMS by its own text-based protocol, but by an object-relational mapper (OR-mapper). The main function of the OR-mapper is to create an internal object-oriented representation of the data easily to work with.

The Java environment already provides implementations for OR-mappers, as open source products. Using OR-mappers inflicts fewer changes within the data structure without concerning other parts of the application.

2. Content implementation

2.1. The collection of content

This part of the document focuses on procedures to collect the content for the ENDURE ALPS data base. The primary focus is laid on harmonised formats for the content and harmonised understanding about the extraction of content from various sources. The defined guideline allows all ENDURE members to add reports to the ENDURE-ALPS data base fulfilling a number of criteria.

2.1.1. Guidance document

As already stated in the technical part of the document the reports representing the expert knowledge based on various single sources. Sources will be collected in regard to specific crop/pest or disease/measure/region combinations, the relevant information will be extracted by the expert on maximum 1 page.

Sources can be research reports, PhD thesis´, journal and magazine articles, "grey literature" such as trial reports and other.

The following information should be included in the extract:
• General information (region, global climatic conditions)
• Description of symptoms (caused by disease or pest)
• Special conditions for measure (e.g. ontogenesis of disease)
• Description of measure (direct, preventive, integrated)
• Practicability of measure (the evaluation)
• Bibliography of sources (not included in the 1 page max.)

Concerning the evaluation of the practicability of measures additional indications are required. This will enable the user to assess the value of the measure for potential adoption in different regions.

The following criteria shall be used as indicators for the practicability of measures (displayed as traffic light):

GREEN: scientifically sound, tested in the field, practical to adopt and cost-effective.

YELLOW: scientifically sound, only practice at experimental farms, subsidies needed, etc.

RED: strategy is still in a development stage, not practical to adopt, not cost-effective

It was agreed that the general criteria of the SA4.1 activity would be adopted concerning the practicability indicator in ENDURE-ALPS, and that the Dutch criteria for Best Practice will be used to indicate and advantages or benefits.

Being aware of the need to comply with requirements of the Endure IC which eventually will be the window to the public for knowledge of the ENDURE ALPS the core criteria (scientifically sound, tested in the field, practical to adopt, cost-effective) of the ENDURE IC have been adopted. A more detailed description of the measure indicating benefits, advantages compared to common practice and bottlenecks is used as well. The information about bottlenecks and benefits will be included in the abstract. Taking into account criteria such as:

Bottlenecks: - Costs
                   - Labour input
                   - Risks (bad control)
                   - Risk perception (unknown)
                   - No official admission (legal)

Contribution to the decrease of environmental problems:
                   - Reduced dependency of chemical inputs

- Large effect
- Moderate effect
- Small effect
- No effect

Organic: - Applicable in organic farming
            - Not applicable in organic farming

2.1.2. Procedures and cooperation for the collection

Each partner in IA4 will name an expert who will be authorised to enter the knowledge for a certain crop and will be responsible for the quality of the information entered. After the first trial period and improvements after an internal test more expert form other ENDURE partners will be included. For the initial phase a close collaboration with the CS Maize group takes place.

3. Conclusion

Setting up the technical background and implementing all procedures will ensure a smooth upload of knowledge to ENDURE ALPS. The definition of content and its structure builds a good base for the extracts consisting of information from the various sources. The first internal test and the presentation at the Annual Meeting in October 2008 will show the relevance and importance of this collection of expert knowledge. ENDURE ALPS will deliver a valuable source raising the awareness and application of non-chemical measure in various crops in such way contributing to a big extend to the reduced reliance on chemical pesticides.
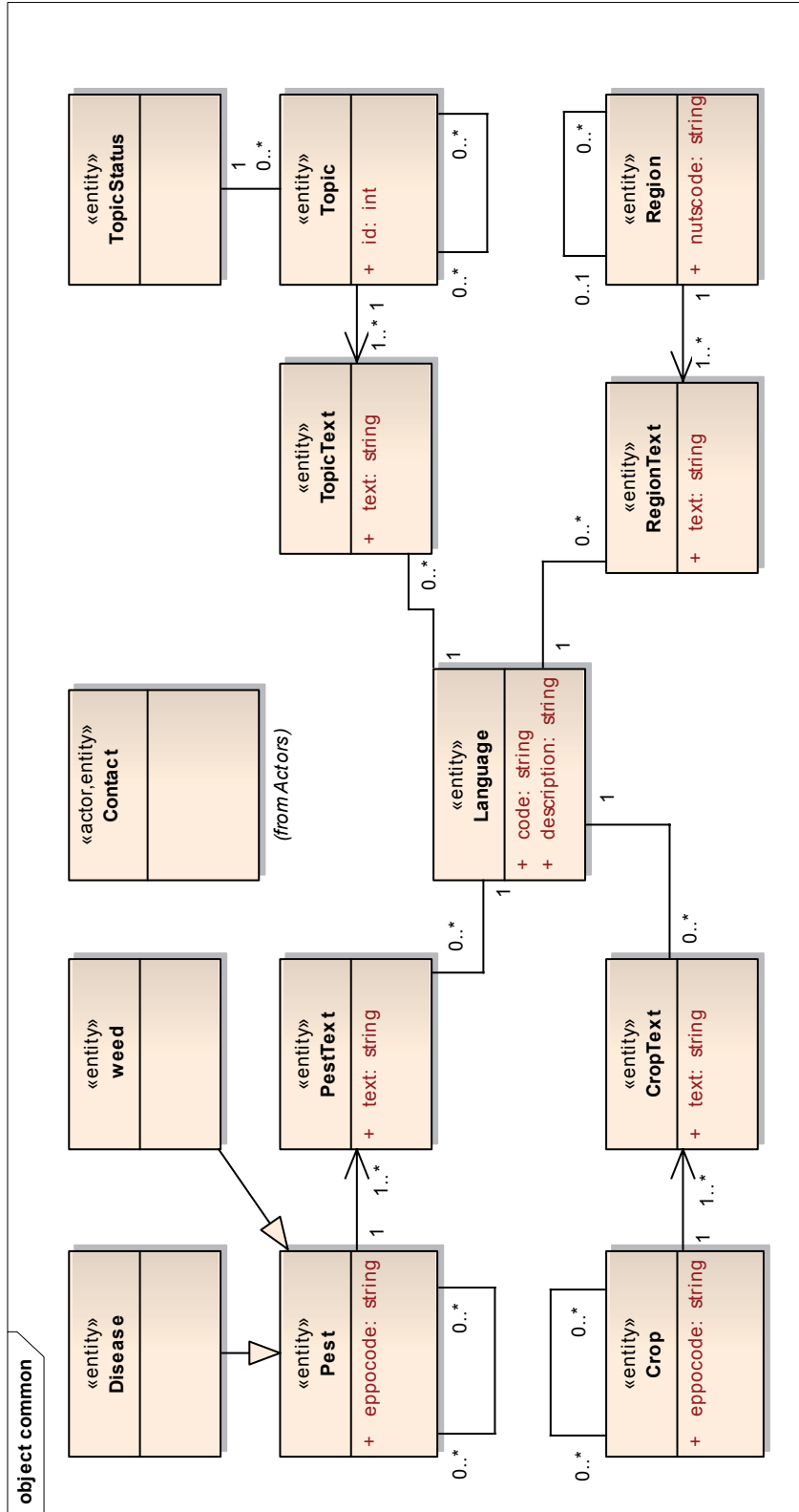
Annex I - Figures
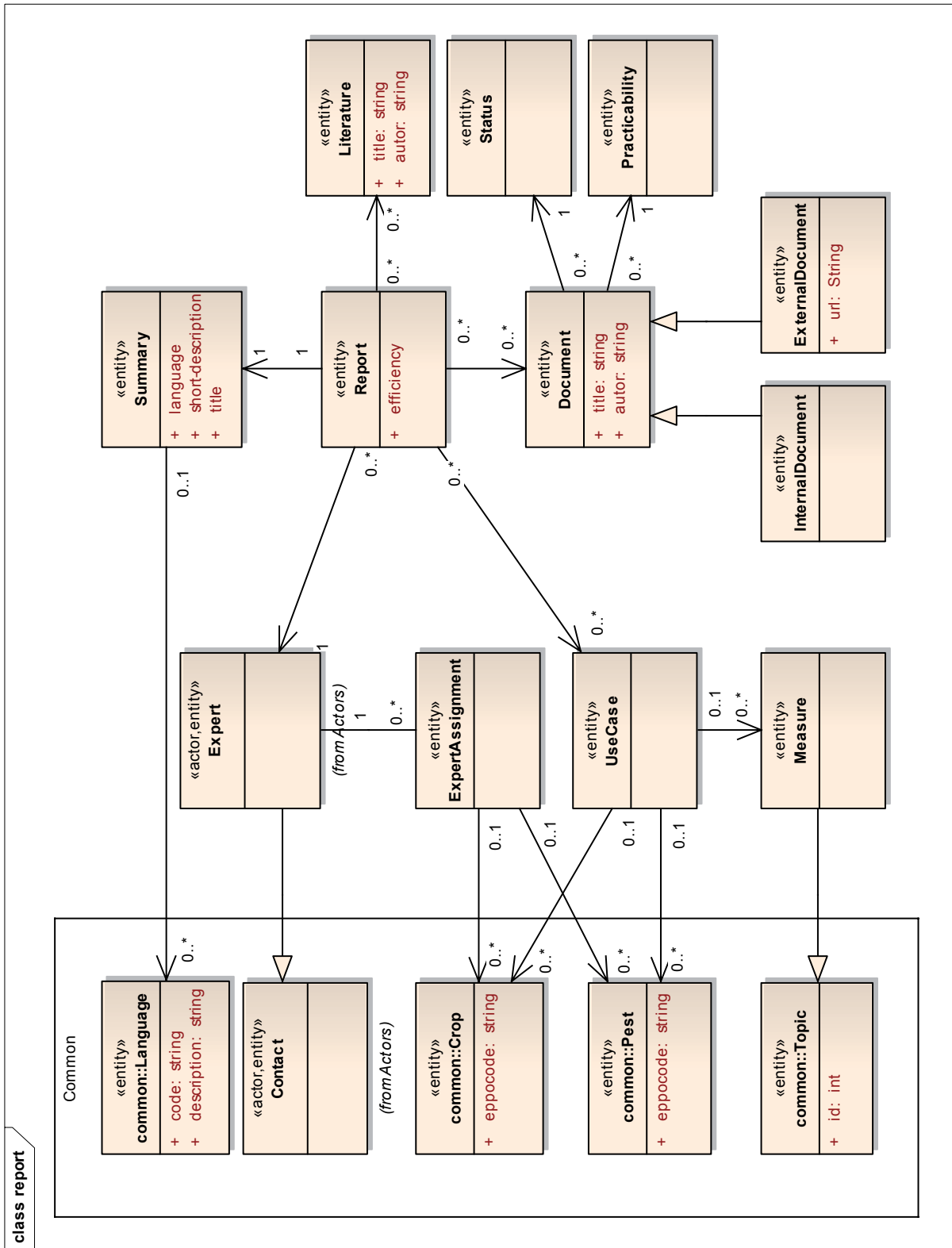

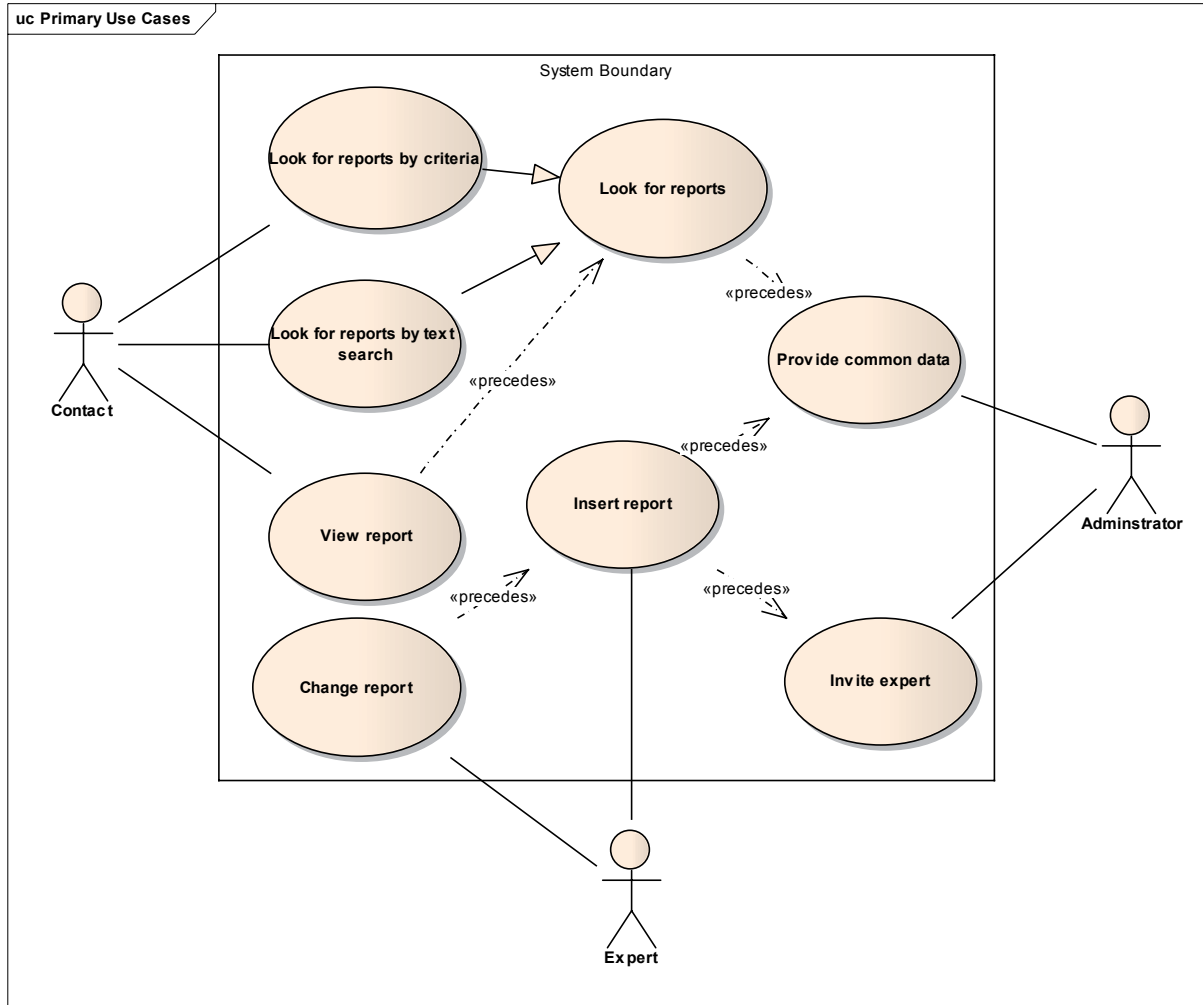**Figure 1 Information Base Overview**

Figure 2 Knowledge Base Overview

Figure 3 Use Case Diagram

uc Actors
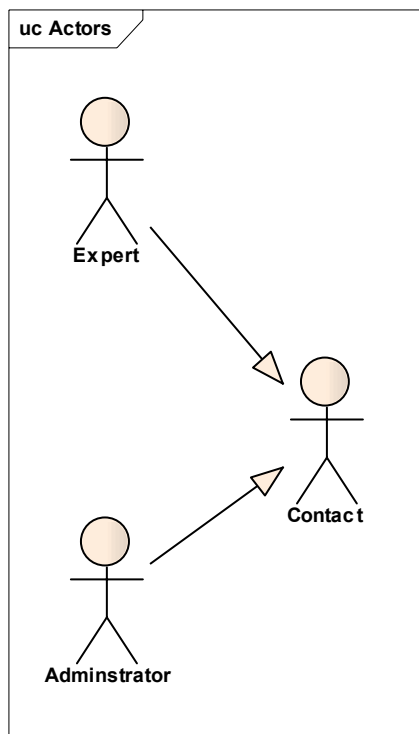
Expert

Contact
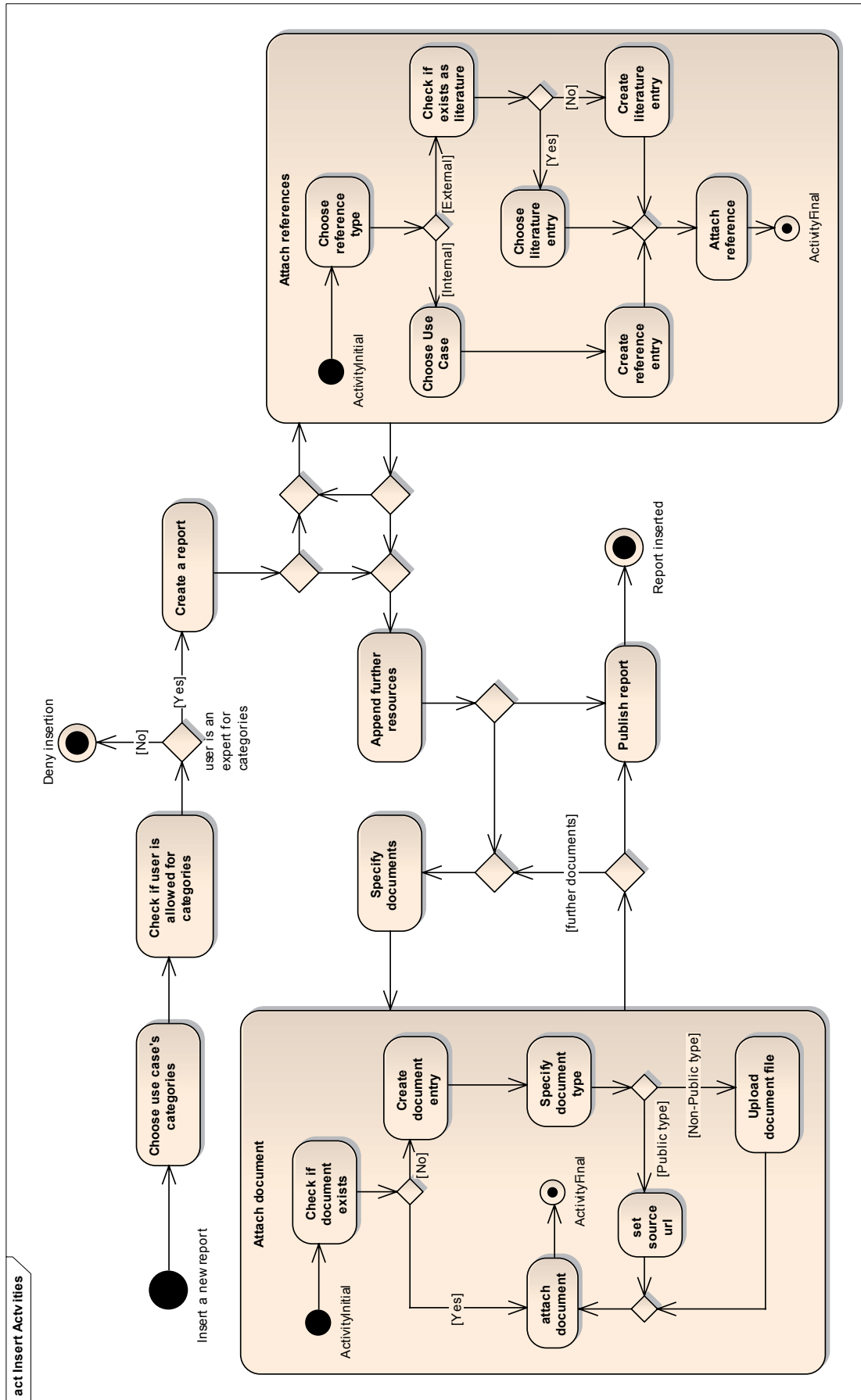
Adminstrator

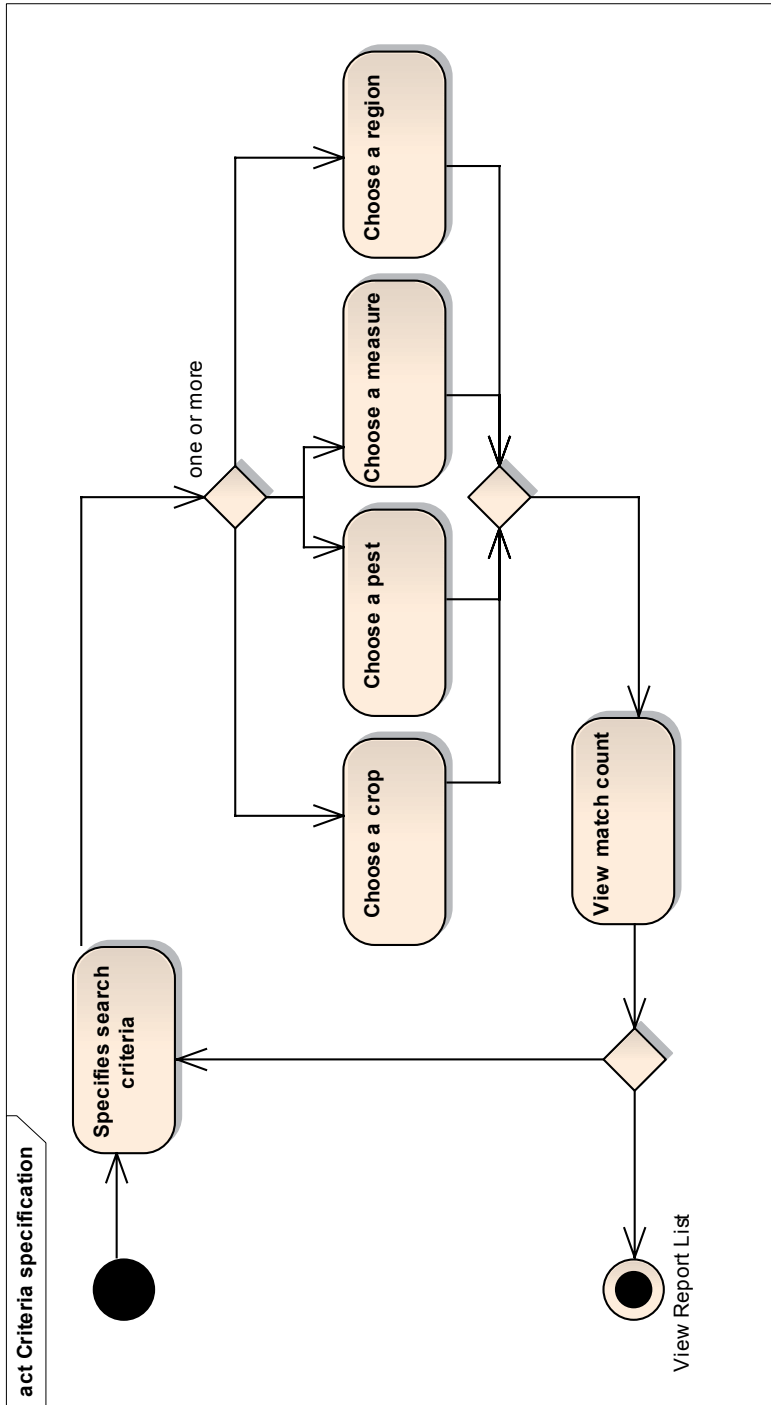**Figure 4 Actors**

Figure 5 Insert reports

Figure 6 Search for reports by criteria

Annex II – Guidance document

Collection and evaluation of information for ENDURE-ALPS (ALPS means: non-chemical preventive and direct alternatives in plant protection)

Aim
The final objective is to define guidelines to allow in a near future all ENDURE members to add information to the DB ENDURE-ALPS on the condition that it fulfils a number of criteria.

Sources
Sources such as research reports, PhD thesis´, journal and magazine articles and other "grey" literature are emphasised in the context for ENDURE-ALPS and will be collected in regard to specific crop/pest or disease/measure/region-combinations.
The relevant information of many sources will be extracted by the experts on maximum 1 page (2500 characters without space) per pest or disease/measure-combination and made available in the DB.

Structure
The following should be included in the extract:
- General information (region, global climatic conditions)
- Description of symptoms (caused by disease or pest)
- Special conditions for measure (e.g. ontogenesis of disease)
- Description of measure (direct, preventive, integrated)
- Practicability of measure (the evaluation)
- Bibliography of sources (not included in the 1 page max.)

Evaluation
The following criteria shall be used as indicators for the practicability of measures (displayed as traffic light) uploaded in the DB:
GREEN: scientifically sound, tested in the field, practical to adopt and cost-effective.
YELLOW: scientifically sound, only practice at experimental farms, needs subsidies, etc.
RED: strategy is still in a development stage, not practical to adopt, not cost-effective

The following criteria shall be used for the further more detailed description of the measure indicating benefits, advantages compared to common practice and bottlenecks.

Bottlenecks: - Costs
- Labour input
- Risks (bad control)
- Risk perception (unknown)
- No official admission (legal)

Contribution to the decrease of environmental problems:
- Reduced dependency of chemical inputs
- Large effect
- Moderate effect
- Small effect
- No effect

Organic: - Applicable in organic farming
- Not applicable in organic farming