Project no. 018340

Project Acronym: EDIT

Project Title: Toward the European Distributed Institute of Taxonomy

Instrument: Network of Excellence

Thematic priority: Sub-priority 1.1.6.3. "Global Change and Ecosystems"

WP 5 D10. Operational functional model for inventory

Due date of deliverable: Month 16.

Actual submission date: Month 16.

Start date of project: 01/03/2006. Duration 5 years.

Organisation name of lead contractor for this deliverable: HNHM

András Gubányi - Imre Kilián – Veronika Kiss

Hungarian Natural History Museum

Contents

Introduction

Workpackage 5 of EDIT project aims to discover and describe the existing work processes in the taxonomic scientific community. This could later enable the creation of a new software platform for cybertaxonomy that enables researchers to do taxonomic work online, and to access the most important informational resources also online. The current deliverable captures all details of existing software processes, that are practiced for inventorying purposes under the taxonomic researcher community.

To this end the individual members of the community were interviewed, their activities were mapped to UML activity diagrams. The present document describes the modeling details of these activities. The description comprises two packages: the Targets package describes the structure of research target descriptions; the Inventory package describes the inventorying activities, first in a general viewpoint, then also for different branches (Algology, Bryophytology, Colepterology, Lepidopterology, Lichenology, Malacology).

The software being applied was: the SparxSystems: Enterprise Architect 6.5, a leadung UML modeling and CASE software.

The interviews and their metodology

The main purpose of the interviews was to discover the activities of key taxonomists in the scientific community, and to map it into UML activity diagrams. By choosing the actual individual a key question was to cover the most of inventorying work. From the areal distribution most of them were Hungarian scientists, because the approved budget did not allow to make extensive travels across Europe.

In the past year 13 interviews were completed, covering 2 institutes from 2 countries, namely from Hungary and Germany. The interviews were created according to a common questionnaire, the results of questionnaires were used to derive models. Per individual one model was created. The details of the created models can be found at the EDIT Wiki bwebsite at the following URL: http://dev.e-taxonomy.eu/trac/wiki/Inventory

Interviewees and their most important projects

Name	Institute	Branch
Dr. Buczkó, Krisztina	HNHM	Algologist
Dr. Fehér, Zoltán	HNHM	Malacologist - marine invertebrates
Dr. Forró, László	HNHM	Limnologist - fresh water invertebrates
Dr. Földváry, Mihály	HNHM	Dipterologist
Dr. Lőkös, László	HNHM	Lichenology - insects, butterflies
Dr. Merkl, Ottó	HNHM	Coleopterology
Dr. Papp, Beáta	HNHM	Bryologist
Dr. Ronkay, László	HNHM	Lepidopterology
Dr. Arno Wörz	SMNS	Vascular Plants

Dr. Dietger Hausenblas	SMNS	Lepidopterology
Dr. Christoph Häuser	SMNS	Lepidopterology
Dr. Klaus Riede	SMNS	Ortopterology
Dr. Jürgen Schmiedl	UEN	Entomology

Model Description

Managing model variations

The clue of the modeling problem is to get in touch the different model variations. Variations occur along two dimensions:

- abstractional level: regarding that the purpose and the utilization of models is not well defined, modelers must keep an eye on each end of the granularity scale. On one hand models must be simple enough so that even laymen, i.e. biologists, ecologists, taxonomists understand it. On the other hand, when speaking about Web based cyberplatforms, we cant forget, that models turn to be software sooner or later. To cope with the challenge of abstractional nondeterminacy the “single model-several views” has been applied. This means: there is a single and only model, that is fine grained enough even to comply the demands of later software generation, but this model may have several views. These views may already loose certain details of the model so, that they are simplified, they serve other particular purposes. They either focus on certain aspects of the models, or they build a simplified view, that is easily understandable by non IT experts of the project. The abstractional multiplicity of the matter is modeled by “embedded activities”. This means, the taxonomic work is modeled by general activities, which build a sequential execution chain. The most general level of such activities build the simplest model of the matter. Going into deeper details we are revealing the fine grained activity structure of the general activities, which build thereby a more sophisticated, and harder understandable abstractional level. At the current model four abstractional levels are distinguished:
    - abstract process: comprising the most abstract executional sequence, and the most basic object flows
    - general process: comprising a generic executional sequence, that is abstract enough, but still demonstrates basic features of the model
    - inventory process: a full fledged model consisting of a hierarchical structure of activities, and a sophisticated object flow system. It is designed especially for future software design purposes
    - simplified process: a strongly simplified view of the full fledged model, comprising only activities, but no objects flows. It is designed especially for the easier understanding of domain experts of the project
- multiple variations according to multiple disciplines: To cope with the disciplinar dimension of the matter the package structure itself was used. That is, a main package

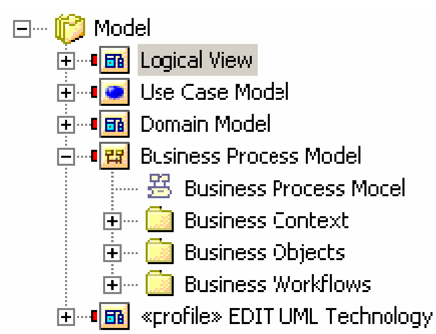
was used to model generic features of the matter, while its subpackages and its further subpackages were used to describe the individual variations, i.e. deviations from the common features. To model the variations the "restricted least common multiplier" approach has been applied. The "least common multiplier" approach means: the generic model is abstract; it can eventually never be realized. It comprises however each feature of individual models so that they can be derived from the common model simply by shortcutting certain features. The restricted least common multiplier approach means: in the generic level only those features are modeled, which occur at least in two individual models. That is individual models either shortcut certain features, or they extend the generic models by features that are used only in the individual model.

Decoration of individual activities

In the compiled diagrams we used a decoration to show, how much individual is an activity. For this end we used the following decorations:

- we painted blue those operations, that were used only in a single individual discipline (model/diagram)
- we painted orange those operations, that were used in a few of disciplinar models/diagrams.
- we left the original colour those, that were used in most of the disciplinar models/diagrams

Package Structure

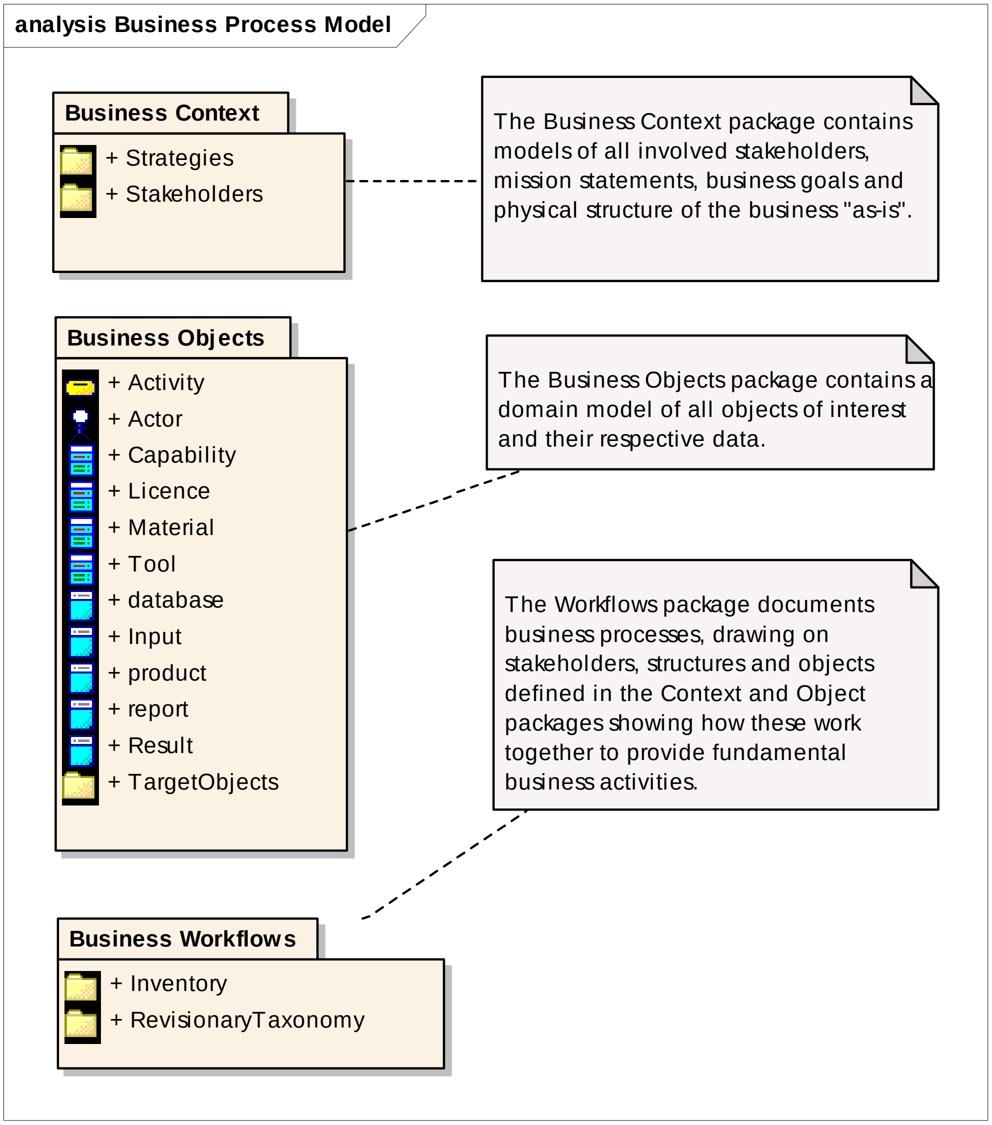


The requirements described above, can be met by the following package structure.

1. The model consists of the following three main packages (see the figure):
   a. Model/Logical View. The package is reserved for future analysis and design activities
   b. Model/Use Case Model. The package contains the elements of a future use case model
   c. Model/Domain Model. The package describes the domain model. That is it describes the domain concepts (taxonomy, biology, common object definitions) and their relationships.
   d. Model/Business Process Model. This package contains the basic workflow details of taxonomic work. The details are acquired in the course of personal interviews with key domain experts (taxonomists). The Business Process Model itself captures the significant events, inputs, resources, processing phases and steps and outputs of relevant taxonomic work. The detailed structure of the package is shown below.

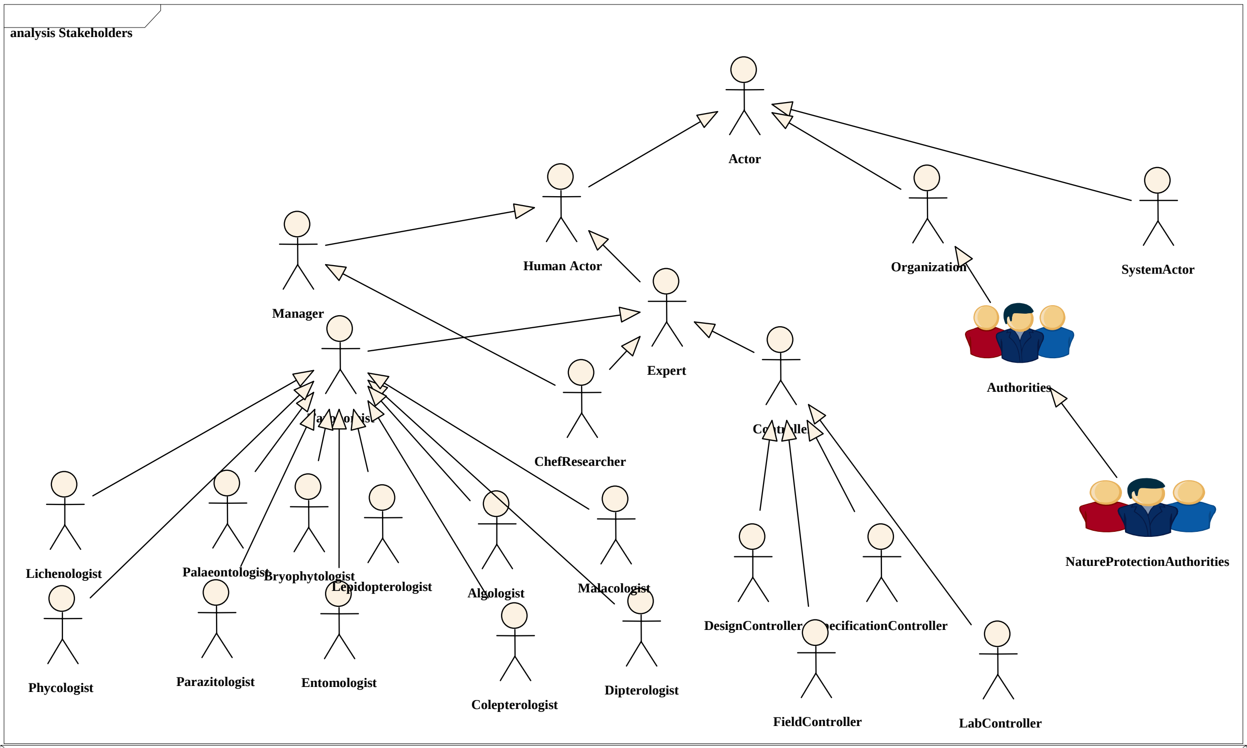
Business Process Model

The structure of the package "Business Process Model" is shown below along with an explanation of its subpackages.



Business Context/Stakeholders

The subpackage Strategies in Business Context is not used at the moment. Stakeholders however contain an analysis of the players and actors in the system to be modelled.



From the actors depicted above, the structure of Actor, Human Actor, Organization, System Actor can be regarded as fixed, the set of distinct taxonomic experts can be freely modified, and other actors can be included in the description.

The list of fixed actors, defined in this diagram:

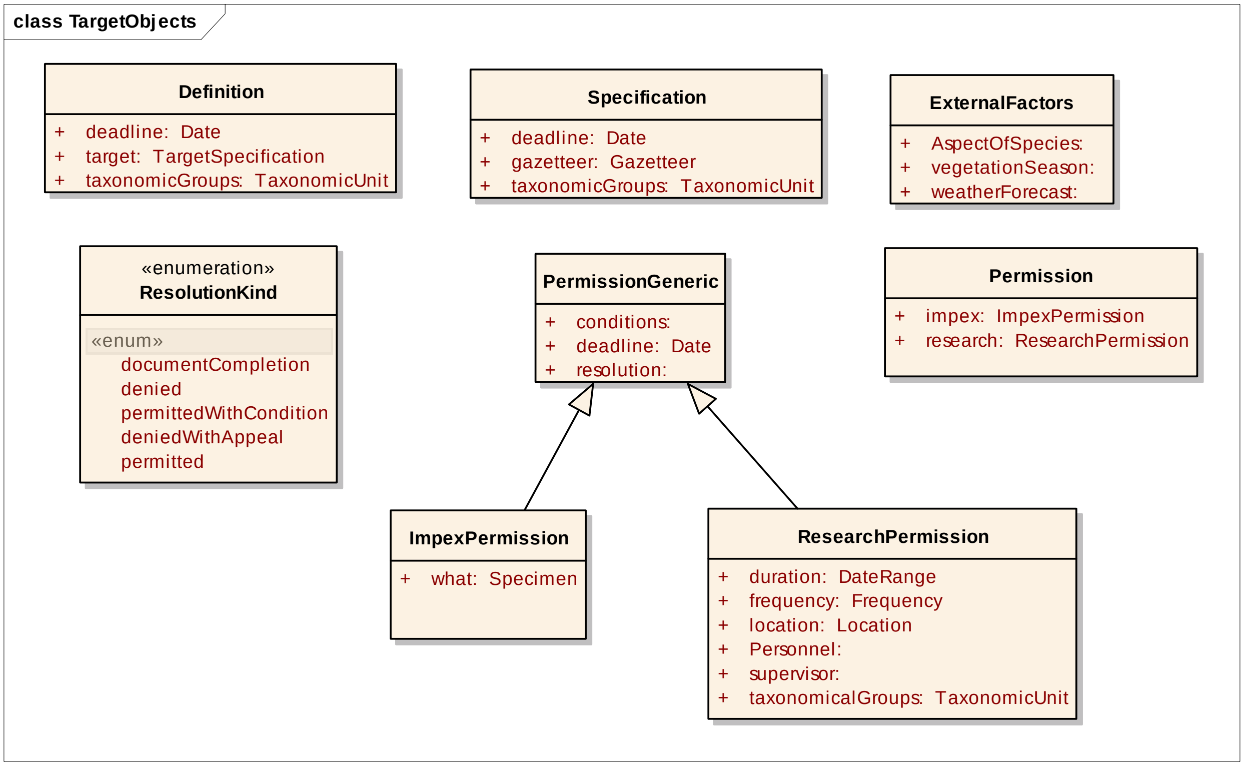
Actor	general actors in the system
Actor/Human Actor	Human actors, i.e. people, who have a specified role in the system (eg: User, Manager, Administrator etc.)
Actor/System Actor	Software/hardware systems, who behave, like external actors in the system (e.g.: Internet, Web-Server, etc.)
Actor/Organization	Organizations, corporate actors.
Human Actor/Expert	People with domain knowledge, but having no other responsibilities
Human Actor/Manager	People having organizational responsibilities
Manager/ChefResearcher	An expert having organizational responsibility in a field project (the leader of the project)
Expert/Taxonomist	Taxonomist expert doing research job in a field project
Expert/Controller	Experts playing the role of controlling the quality of certain activities

Remark that the subclassing may use multiple inheritance, i.e. a role may inherit also from several parents.

Business workflows

The core of the problem is the description of the workflows and their data flow. The distinct workflows are the result of the corresponding interviews with key experts of the topic, and they are stored in a distinct subpackage each. Besides workflow descriptions we also describe at least the raw object structures.

Object structures



The following object structures have been defined as object tags for the general workflow activity boxes:

Definition: The basic information package that triggers the research project. The information being contained in this structure, is not necessary sound, not coherent, not feasible, not justified by experts. It has the following three components (attributes):

- deadline: obvious
- target: target specification structure as detailed below
- taxonomicGroups: TaxonomicUnit. A list of TaxonomicUnits, the target of the research project

Specification: More or less the copy of Definition, with somewhat modified information pieces. Its main feature is, that it is scientifically coherent, as it is justified by experts. Its components:

- deadline: obvious
- taxonomicGroups: TaxonomicUnit. A list of TaxonomicUnits, the justified target of the research project
- gazetteer: a database of geographical localities where the research should take place

ExternalFactors: the set of external factors that cannot be influenced by project members…

- AspectOfSpecies: the description of the aspect of species
- vegetationSeason: the vegetation properties at the chosen locality
- weatherForecast: obvious

ImpexPermission:

- what: Specimen: a description of what kind of specimen is allowed to be exported or imported

ResearchPermission:

- duration: a date range the research permission is valid
- frequency: in case of repeated research activities the frequency must be described accurately
- location: a specification, where is the permission valid
- personnel: who are entitled to carry out the research project
- supervisor: who is the responsible supervisor for the research
- taxonomicalGroups: TaxonomicUnit: the taxonomic units the permission is valid for

PermissionGeneric:

- conditions: extra conditions under what the permission is valid
- deadline: deadline until the permission is valid
- resolution: the resolution, as described below

Permission:

- research: the research permission component of the permission structure
- impex: the impex permission component of the permission structure

Targets

The package describes the scientific target specification structure. The clue of the problem is to model a two dimensional enumeration (a type-subtype construction) in which the actual enumerations of subtypes depend on the actual value of the main type enumeration. The first "type" enumeration is also a basis of a value-narrowed subclassing.

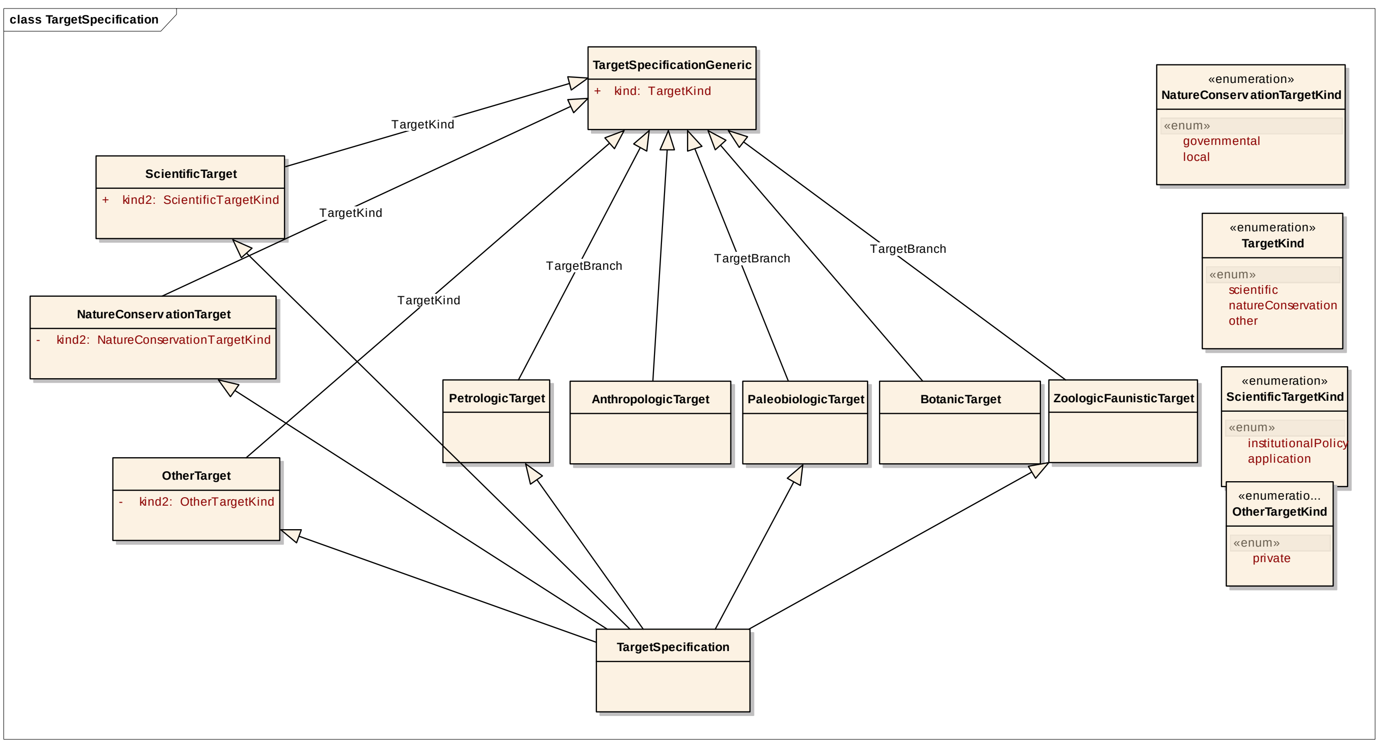
The solution of the problem is to model a disjunctive type construction by UML classes. In this model the subclasses in the TargetKind cluster are complete and disjoint, that is the TargetKind enumeration completely partitions it, the TargetSpecification may inherit from exactly one of them.

The other generalization cluster (TargetBranch) is complete, but may be overlapping, i.e. when TargetSpecification inherits from each of them, there might be actual instances, which belong to several of them.

The generalization structure below ensures, in end effect, the following:

- a Target Specification may cover several of the subclasses of TargetSpecificationGeneric, in the TargetBranch cluster.
- a Target Specification has an enumeration attribute of type:TargetKind
- depending on the actual realization of TargetKind, a TargetSpecification belongs exactly one class of the TargetKind cluster.
- Depending on the actual realization of TargetKind, the “type2” attribute of TargetSpecification may be of different type, according the actual value of the type:TargetKind attribute

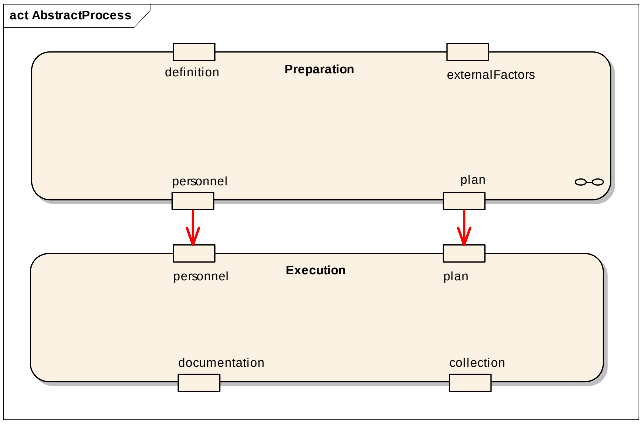
The meaning of the certain subclasses of TargetSpecificationGeneric is straightforward.



Abstract Inventorying Process

The inventory process is described using activity diagrams. The root package contains the generic model of inventorying, while it also contains subpackages for the different concrete branches of inventorying activities. These are: Algology, Bryophytology, Colepterology, Lepidopterology, Lichenology, Malacology.

On the most abstract level of description, Inventorying is a sequence of two actions: Execution comes after Preparation. Preparation needs two object structures as an input:



- definition, that is the statement from state authorities or state financed programs, scientific boards or private corporations, wishing to carry out a certain inventorying project.
- externalFactors: a list of conditions, that lay definitely outside of the scope. External factors can be regarded thus constant, that cannot be influenced by internal forces.

As a result, Preparation prepares two object structures:

- personnel, i.e. the list of people cooperating in the project
- plan, i.e. a pretty detailed and exact description of what, when and how to be done.

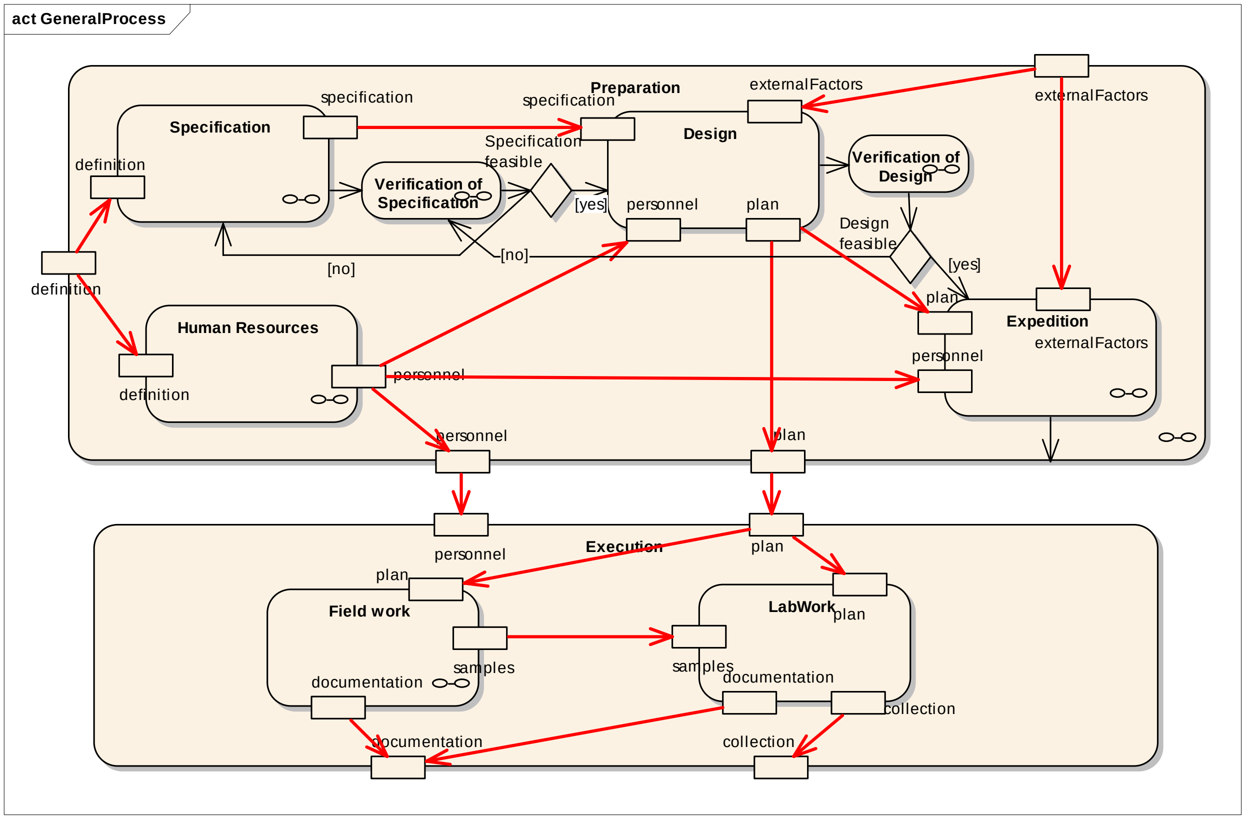
Execution takes the results of preparation, as an input, and produces the following results:

- collection: the physical set of collected and/or prepared and/or mounted samples, together with their all supplementary equipment: cupboards, boxes, plates, flasks etc.
- documentation: all kind of written information, either electronical or paper based, that is produced as a result of the projects. This may include: reports, articles, books, catalogues, databases, etc.

Remark, that to distinguish between object and control flows we color the former to red.

General Inventorying Process

The oversimplified flow of abstract inventorying process is grained in a normal extent on the view of General Inventorying Process. General Inventorying says the following:



The Preparation is broken up to the following four main activities:

a. Specification, that takes the definition as an input, and prepares a refined specification as a result. While the definition is not necessarily feasible, in terms of the general taxonomical and biological concepts, the specification must be scientifically sound, exact and accurately described. Anyhow it mustn't yet contain any details regarding the execution of the project.

b. Verification of specification: a compound sequence of quality-management actions to check the quality and the feasibility of the specification. Design can be done only if the specification is feasible, otherwise control falls back to the specification again.

c. Human Resources, i.e. activities to collect the research team. The input is the definition, and the result is the list of experts to carry out the project.

d. Design, that takes the specification, the personnel list, and the external factors as an input, and produces the research plan, as an output.

e. Verification of design: a compound sequence of quality-management actions to check the quality and the feasibility of the design. Any further implementation activities can be done only if design was feasible. In any other case control falls back to verification of specification, just to allow an eventual respecification again.

f. Expedition, i.e. the preparation of all physical equipment and written material necessary to start the project.

2. Execution takes the results of Preparation as an input, and is broken up to the following subactivities:

a. Field work: a compound activity, comprising all the activities done on the field.
b. Laboratory work: a compound activity comprising all the activities done in laboratory and/or office.

Quality Assurance

The quality of work is, as usually, controlled through loop-backs. The model in its today's outlook consists of three loopbacks, being composed of two.

1. the Verification of Specification loop that tests, if the specification is feasible, and calculates the outcome of the project in terms of numbers of samples.
2. the Verification of Plan that tests, if the design is feasible, and compares the needed resources with the available ones.
3. The third circle contains the previous two. That means: if the design is not feasible, a respecification of the project may also be necessary.

Inventory Process

The special activities are described according to the branch of the subpackage. This is achieved by inheriting activities from the root package, if it is unchanged in the actual subpackage. If there is any change, then the corresponding activities are repeated in the subpackage. Though in the activity diagrams all the activities are depicted, along with the generic, inherited activities, subpackages practically contain only specific activities, they contain only the difference to the generic case.

Note, that using both control-flow and object-flow facilities of UML2, the diagrams shown below demonstrate both the workflow model, and the dataflow model of the described activity.

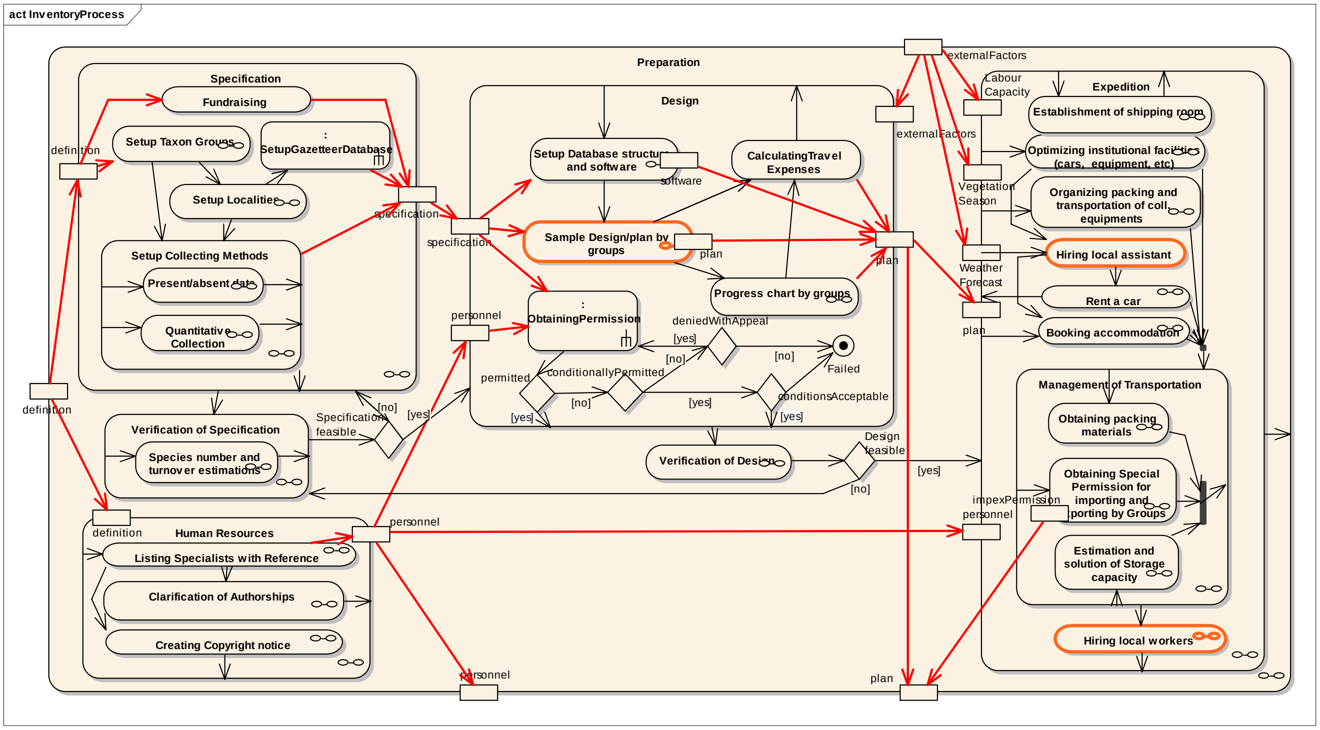
1. The general flow of inventorying has already been detailed in the previous chapter. In the following we are giving the exact details of the workflow.

Preparation

Preparation is the composite activity that comprises all sub activities, which are made still prior to field work. They usually refer to management questions, and they are executed mostly by managers.

The composite activity has two input object parameters, definition and the set of external factors, as it has already been explained above.

The output object parameters that are prepared by the composite activity are the personnel list, and the detailed plan of the field research.



The function of is subordinate activities is explained below.

Specification

The composite activity of specification determines taxonomical groups, localities and methods, to be done in the preparational, management phase of the activities. It requires an object parameter that describes the definition. The activity consists of basically five subactivities.

- First the targeted taxonomical groups are to be determined. This must correspond greatly to the specification of research targets.
- Then the locality must be determined.
- In the following steps, a gazetteer database is prepared
- Finally, the collecting methods are determined

The composite activity has one output parameter, and this is the gazetteer database.

Setup Taxon Groups

Setting up groups means that based on the terms of target specification the topics of interest, i.e. the targeted exact taxonomical groups are to be determined.

Setup Localities

Beyond the rough setup of locality, that is habitat, or national park, a more exact locality plan must also be compiled. This plan should describe the exact localities where the collecting activity should be carried out.

SetupGazetteerDatabase

The action is a reference to the compound activity with the same name. It is described in details below in a following chapter.

Setup Collecting Methods

The applicable collection methods depend greatly on the taxonomical groups of interest, as determined in the previous activity. This step determines the basic collection methods.

Present/absent data

The aim of the activity is only to show, if the species exists or not. In such case a numerical frequency value is irrelevant.

Quantitative Collection

The aim of the activity is to provide numerical data set for establishing the dynamic parameters of the population.

Verification of Specification

As a verification of the collecting method, an estimation of expected results must be done. Verification in this case means that based on literature data and previous experiences, the yield of methods can be previously estimated. One result of this verification is the estimation of species number and the total turnover, but such estimation must be done regarding all collected data, electronically recorded information etc. For the verification the literature can be researched, and the description of previous similar projects must be reviewed.

Species number and turnover estimations

One output of the verification of collecting methods is an estimation regarding the number of collected species, and the number of specimen per species, that is the total turnover of the research activity.

Specification is feasible

A decision, if the specification is scientifically sound, and it conforms to internationally accepted guidelines.

Human Resources

The Human Resources activity covers practically the workflow of setting up a research team. It consists of several steps, and in end effect it results a personnel list.

- In the first step a list of experts with relevant competence is compiled, and they are contacted.
- Then in the next steps the actual labour is divided, the authorships are clarified.
- Finally copyright questions are clarified.

The output of this activity is the list of personnel.

Listing Specialists with Reference

As a first step a list of specialists are compiled. We must always expect that the list is never complete, and a research team, that could cover the specified topics, can never be set up.

Clarification of Authorships

The clarification of authorships establishes a common agreement among the project members, about the authorship relation between documentations, chapters and the members.

Creating Copyright notice

The copyright questions are clarified, agreed and recorded. That is the final copyright holders are named, with special respect of themes, topics, personnel and media.

Design

Upon receiving a feasible specification, and the list of personnel, the research project can be designed. The final result of this stage is the research plan itself. The performed steps of this activity are the following:

Obtaining Permission

Obtaining Permission is an invocation of the relevant procedure. For its definition see the chapter below.

permitted

Determining, if the permission was given unconditionally.

conditionally permitted

Determining, if the permission for field research was given only under certain conditions.

conditions Acceptable

A decision, if, in case of conditional permission, the stated conditions are acceptable or not.

denied with Appeal

Determining, if appealing is possible, in case of denied permissions.

Establishment of shipping room

The shipping room is the place, where all the equipment is to be collected before a field work expedition. As a preparatory activity this shipping room must also be set up.

Setup Database structure and software

The actually collected data is mostly electronically recorded and stored. Thus before starting activities a database management software must be set up, along with relevant database structure.

Sample Design/plan by groups

Upon summarizing all the detailed information, a concrete and accurate design and/or plan of collecting activities must be compiled. The output of this activity is the design plan itself.

Progress chart by groups

For formalized reporting and evaluation purposes progress charts must be set up already in the beginning of projects. As the working process is different by taxonomic groups, these progress charts are compiled also accordingly. The input of this activity is the design plan.

CalculatingTravel Expenses

One step of inventory management is the calculation of travel expenses. This should summarize everything: pure travel costs, per diem costs, lodging, renting transportation, etc. The input of this activity is the design plan.

Verification of Design

Beyond general verification steps, the activity checks, if the proposed steps are feasible: i.e. they don't exceed financial, material and time limits.

Design Feasible

A decision, if the design fits in the limits of the expedition, and it conforms to internationally accepted guidelines.

Expedition

Expedition management has five information components (object tags) that influence the activities deeply:

- the labour capacity available in the field work
- the aspect of species to be collected
- the vegetation season, and its uncertainty
- weather forecast
- the plan

The activity consists of the following sub-activities: (1) optimizing institutionally available resources, like cars, equipment etc. (2) Organizing packing and transportation of collected equipment (3) Eventually renting a car (4) Eventually booking an accommodation (5) Management of transportation that includes further concrete steps regarding transportation.

Optimizing institutional facilities (cars, equipment, etc)

Optimizing institutional facilities means surveying the actually available facilities (cars, equipment), reserving them, and eventually resolving any over-booking.

Organizing packing and transportation of coll. equipments

The equipment collected in the shipping room must be properly packed, and transported. Transporting can be organized by cars, or through external logistic firms (train, etc.).

Rent a car

Renting a car is necessary only for special projects which need special cars with special abilities, like 4WD, load carrying capability etc.

Booking accommodation

Booking an accommodation is necessary only, when researchers overnight at the field, and many researchers, depending on the actual season or weather conditions, prefer overnighting on the site in tent.

Management of Transportation

An important factor of field work is the transportation of collected materials. As it varies greatly in size and weight, the transportation itself is also a self-standing activity, and it can be thus only in certain cases reduced or completely dismissed.

Obtaining Special Permission for importing and exporting by Groups

The collected specimen must be exported through the customs from the target country, and must be imported into the host country. Thus in such cases a special export-import permission must be previously obtained.

Obtaining packing materials

Depending on the size, weight and physical state of the transported materials, the necessary packing and protecting materials must also be obtained.

Estimation and Solution of Storage Capacity

Hiring local assistant workers

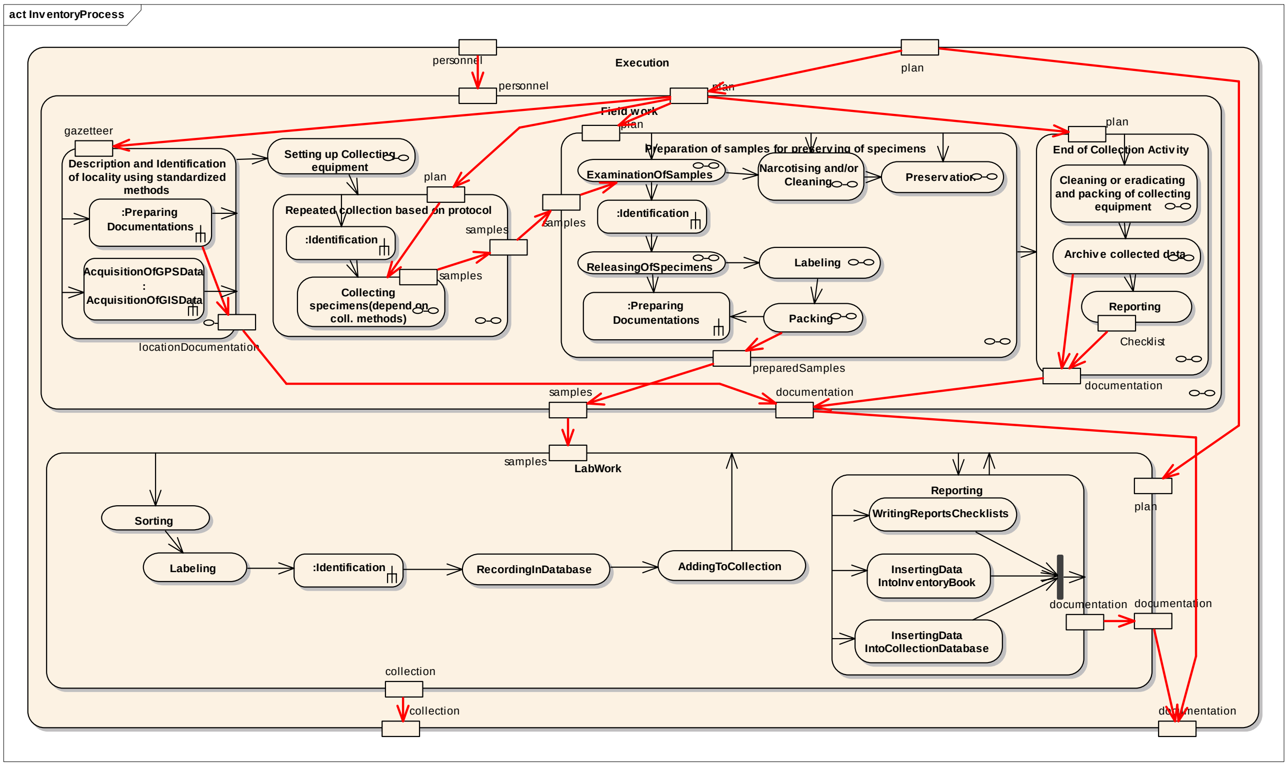
For the collecting activities in certain cases local assistant workers must also be hired.

Execution

Upon receiving the list of personnel and the research plan, actual steps can be started. These include (1) Field work, i.e. doing research on the field (2) Lab work, i.e. collecting, organizing and publishing the results of field work.

Field work

Field work is a generic concept that includes all the necessary phases and steps to be done in or near to the habitat. Field work has four input parameters: 1. personnel list 2. research plan, where the research plan itself includes other important components, like permissions and the gazetteer database.



Given these input objects, the activity is carried out in the field, and it provides the following results: 1. collection of samples 2. documentation

Description and Identification of Locality using Standardized Methods

On arriving at the collection site the locality must be described by standardized methods. One input of this activity is the gazetteer database that should contain the investigated localities and names. The result of this activity is the prepared documentation of the site.

Acquisition of GIS Data

It is an invocation of an activity with the same name. The activity itself is defined below.

Preparing documentation

For a detailed description of how documentations are prepared, see the section below.

Setting up Collecting Equipment

Upon arriving, the collecting equipment must be unpacked, and set to work, or/and set out to the site.

Repeated collection based on protocol

The collection plan should describe repeated collection activities. The exact cycle, frequency and other circumstances of repetitions must be described also in the collection plan.

Identification

In some cases, at least a preliminary identification of biological subjects is to be done already in the time of collecting specimen.

The activity is the invocation of the corresponding procedure.

Collecting specimens (depend on coll. methods)

The exact way of collecting specimens is done according to the description in the collection plan. This plan must describe, among others, the methods to be applied.

The input of this activity is the research plan that in turn produces the collection of specimen.

Preparation of samples for preserving of specimens

In some cases specimens are prepared for preservation still on the site. The inputs of this activity are: 1. the research plan, 2. the collected specimens. The output of this activity is the collection itself.

Examination of Samples

The collected specimens are in most cases examined still on the site.

Identification

The result of examination of samples is also the identification of them. This is also a step that is usually done on the site.

The activity is the invocation of the corresponding procedure.

Releasing of Specimens

The collection plan and the permissions determine also how many and what kind of specimens are to be released after examination.

Narcotizing and/or Cleaning

As the first steps, specimen must be usually narcotized, and/or cleaned.

Preservation

The collected specimens are in many cases preserved, still on the site.

Sorting

Specimen are sorted according to its taxonomical characteristics

Labeling

As a following step, after sorting, specimens are labeled.

Packing

After preservation the specimen must be packed to prepare for traveling back to the office.

Preparing documentation

The process of field work and the collected samples are to be documented.

This is an invocation of the relevant procedure that is defined below.

End of Collection Activity

Having finished the collecting activities, certain closing-up steps are to be made. The input of this activity is the research plan, and as a result it produces documentation about what have been done.

Archive collected data

At the close-up phase of collecting activities the collected data must be archived. Archiving activity includes the archiving of electronical data, that is informatically packing it, and making security copies, and the eventual archiving of other physical, paper based data, sound or movies.

Cleaning or eradicating and packing of collecting equipment

At the close-up phase of collecting activities care must be taken for the collecting equipment. It must be cleaned and packed, or in certain cases it can be eradicated.

Reporting

The last phase of closing up the collecting activities is to compile reports. The result of this activity is the report itself, which is a part of the documentation to be prepared. The result of reporting is the checklist that becomes also a part of the project documentation.

LabWork

All the working phases are included here, which are completed in the laboratory, after arriving from field work again. The input of this activity is the collection itself. The exact way how laboratory work is to be carried out, varies depending on the actual taxonomic group.

Sorting

The collected specimens are sorted and selected in various ways..

Labeling

The sorted and selected specimens are labeled.

Identifying

The collected and sorted specimens are identified. This is an invocation of the procedure, defined below.

Recording in Database

After identifying the sample it can be recorded in the database.

Adding to Collection

The identified sample is added to the collection

Reporting

Reporting is the last step of work after laboratory activities have been closed down. Reporting has three more or less independent phases: 1. writing the reports' checklists, i.e. filling up some checklists, that were prepared still in the planning phase, but haven't been filled in the reporting activities at the end of the field work. 2. inserting the data into the inventory book 3. inserting data into collection database. The three sub activities must synchronize themselves before closing the reporting activity.

Writing Reports Checklists

The purpose of the activity is to fill up checklist items that were still compiled in the preparatory phase of the project. Although there was already a similar, reporting step at the end of the field work, which could have filled up some checklist items, there might well remained plenty not yet filled items.

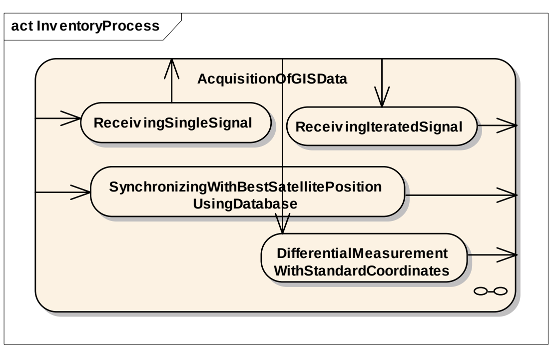
Inserting Data into Inventory Book

The Inventory Book is the documentation that contains the final result of the project, i.e. the actual statistical data concerning the collected specimens.

Inserting Data into Collection Database

The collection database is the electronically stored pool of information, having been collected during the project. This database could have been filled also in the field, but the final data can be expected to arrive first after the laboratory work.

Acquisition of GIS Data



On describing the site the researcher controls and fixes the spatial coordinates of the objects and localities mentioned in the gazetteer database.

The subdiagram describes the different measurement approaches depending on the accuracy requirements, and the quality of GPS, and the received signal.

Receiving Single Signal

The simplest case of determining geographical coordinates is, when we have an accurate GPS navigation system.

Receiving Iterated Signal

When the GPS device is not accurate enough, it might be necessary to perform iterated measurements.

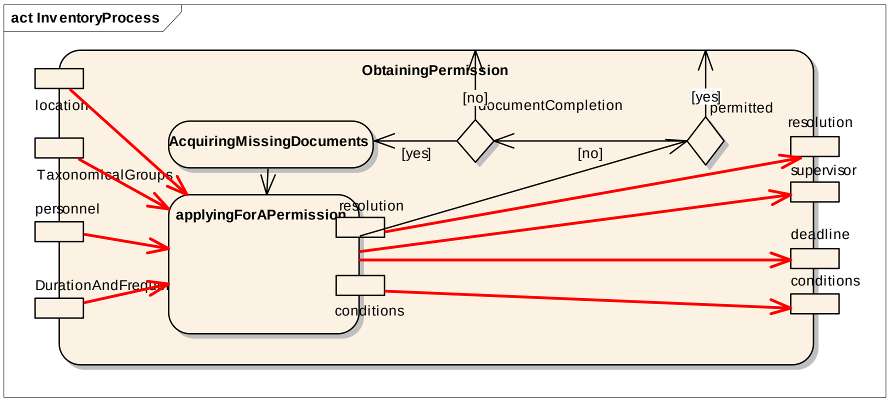
Synchronizing With Best Satellite Position UsingDatabase

To reach the necessary accuracy it is also possible to make corrections using database data of repeated earlier GPS measurements.

Differential Measurement With Standard Coordinates

To eliminate the inaccuracies of measurements, sometimes differential measurements must be used.

Obtaining Permission



Obtaining a permission for field work is in many cases necessary. For obtaining permission, usually, national or local authorities must be contacted. The activity has four input and four output parameters:

As an input the following things must be stated:

- the locality where the field work is to be completed
- taxonomical groups the research is targeted
- personnel (the list of persons) to complete the research
- duration and frequency, in case of repeated activities

In turn an answer is expected, that has the following components:

- resolution: a decision and the circumstances and conditions, if the application is supported or not
- a name of a supervisor, who can and will contact and supervise the collecting activities
- a deadline, until the permission is valid
- any other conditions

Applying for a Permission

As a first step, a permission must be applied for, the relevant documents must be compiled, and must be sent to the authorities.

Permitted

A decision, if the research is permitted under any conditions, or not

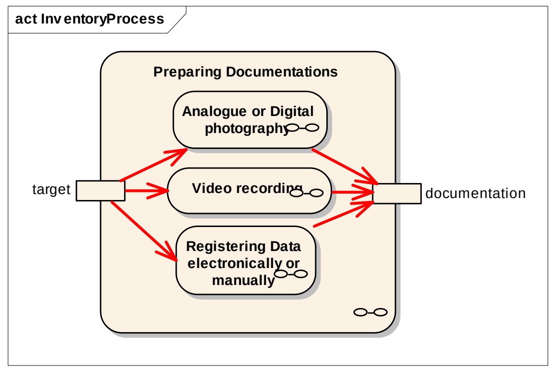
Document Completion

A decision, if the permission was denied only because of the lack of appropriate documentation.

Acquiring Missing Documents

In case of missing documentation these must be acquired, and resent to the deciding authorities.

Preparing Documentations



Preparing documentation is a complex chain of activities, which can be invoked from several points of the activity network. The input of this activity is an object describing the target of the documentations. The output is the documentation itself.

Analogue or Digital photography

Straightforward.

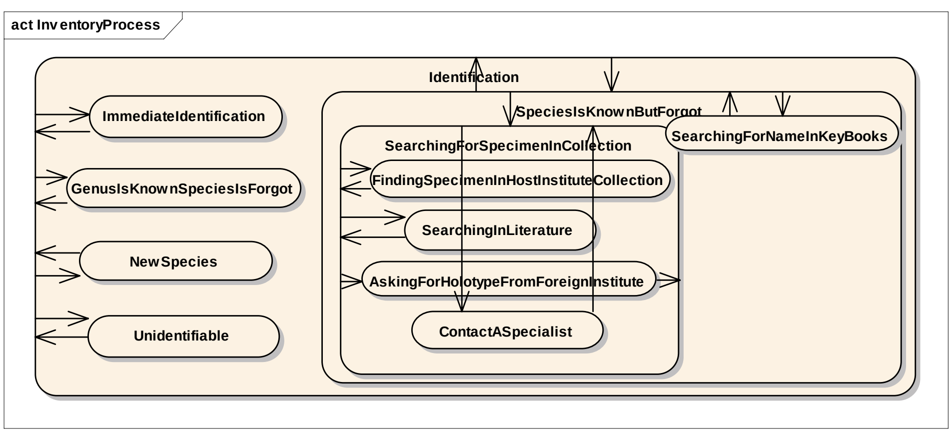
Registering Data electronically or manually

Straightforward.

Video recording

Straightforward.

Identification



The identification if species is done usually pretty often, even if the identification itself is not completed, In fact, in many cases (e.g. in field work) only partial identification is done, because a full identification is either unnecessary or not feasible. Identification consists of the following steps, as described below:

Immediate identification

Immediate identification happens in cases, when the researcher is clever enough to recognize specimen on the fly. It is especially important on the field, when eventually several kinds of specimen may fall into the traps, and the not relevant species are immediately released.

Genus is Known But Species is Forgot

In some cases the genus can be determined, but the actual species is not known.

New Species

This happens, when, with a fair likelihood, a new species has bee, found.

Unidentifiable

This happens, when the researcher is unable to identify the specimen.

Species is known but forgot

When the species is known, but forgot, there is a couple of possibilities to proceed. These are as follows.

1. Searching for specimen in Collection. 2. Searching for specimen in Key Books.

Searching for Specimen in Collection

Searching for specimen means, eventually also with involving other institutions or people, information about the species is gathered. This consists of the following steps:

Finding Specimen in Host Institute Collecion

As a first step, the collection in the host institute is consulted.

Searching in Literature

Searching in written or electronical literature may, in many cases, also help a lot.

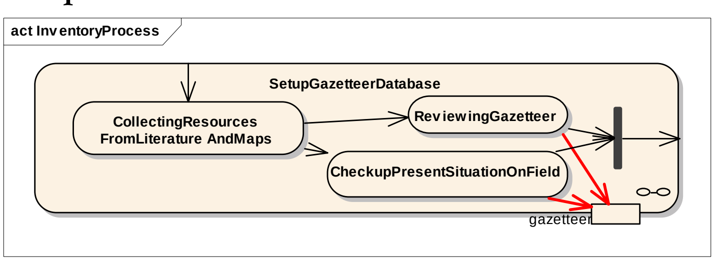
Asking for Holotype From Foreign Institute

For an unknown specimen the holotype of the species also can be required from its host institute.

Contact a Specialist

In case of any doubt specialists can be consulted so, that some specimens can be sent to them.

Setup Gazetteer Database



After describing the locality, a precise gazetteer database must be compiled. This means, maps and literature descriptions must be searched for, and a list of names describing localities must be collected. This list must include also other attributes, like geographical coordinates, elevation and information about any change that might have happened to the geographical object (like: a stream dried out), alternative naming variations etc.

For the sake of this, relevant geographical information must be reviewed, including maps and old reports, and the relevant locality names must be listed. Eventually a preliminary field trip must be also organized, and the old descriptions must be checked and eventually also be corrected.

The output of this activity is the gazetteer database itself.

Collecting Resources from Literature and Maps

For compiling the gazetteer database first the relevant resources must be collected. These include maps, also historical ones, geographical descriptions and descriptions of projects being done formerly on the site.

Reviewing Gazetteer

This activity involves an eventual checkup of the gazetteer database, being completed in the office.

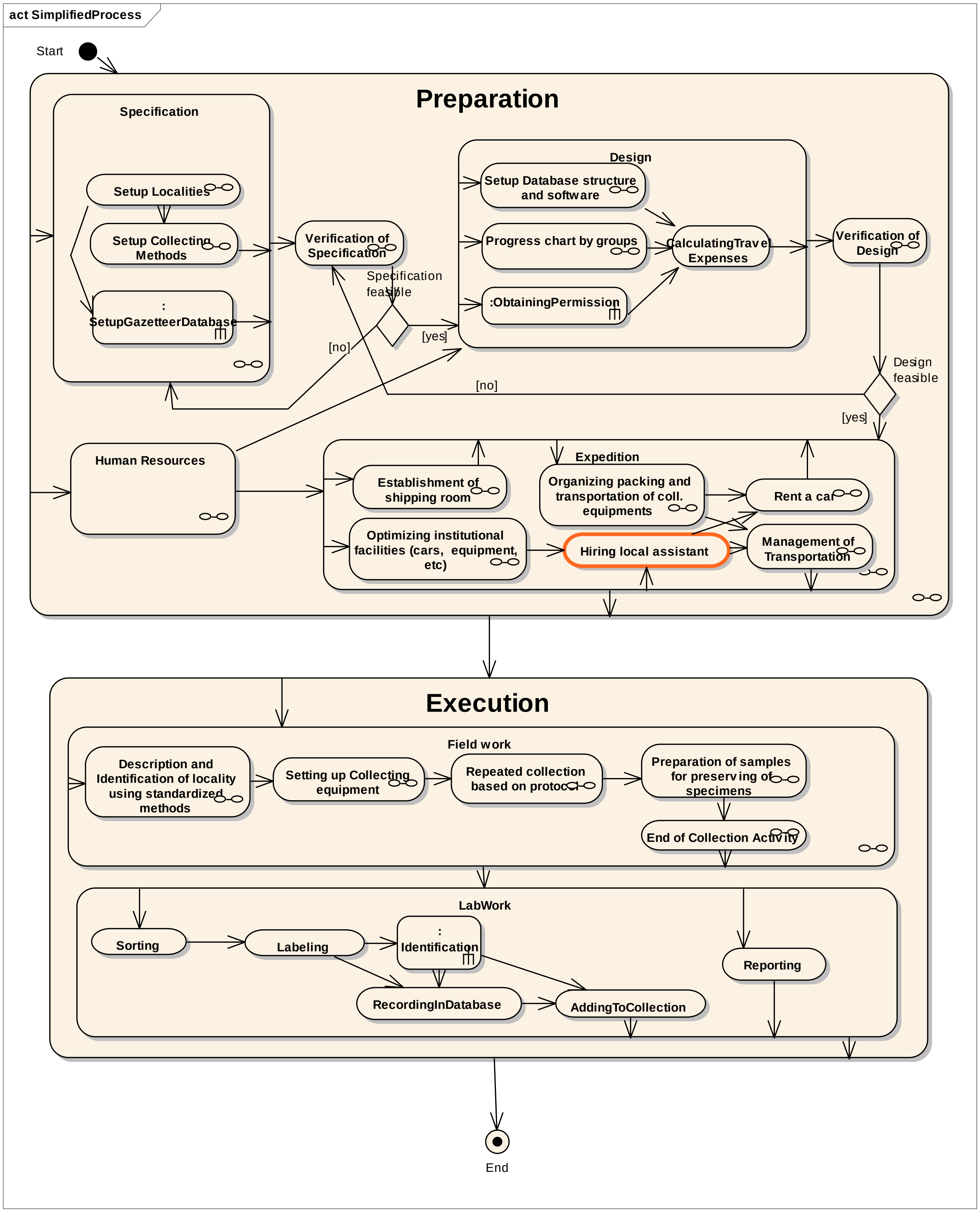
Checkup Present Situation on Field

A checkup of the gazetteer database on the field. The primary purpose of the step is to check, if there were any changes regarding the descriptions.

Simplified process

We introduce hereby the simplified view of the entire inventory process. The most important differences to the full fledged model are the following:

- there are no object tags and no object flows. In places, where the sequence is determined by the object flow itself, an explicite control flow sequence is used
- the deepest details of activities are omitted
- certain – less important – activities are omitted


act SimplifiedProcess
Start
Preparation
Specification
Setup Localities
Setup Collecting Methods
: SetupGazetteerDatabase
Verification of Specification
Specification feasible
[yes]
[no]
Design
Setup Database structure and software
Progress chart by groups
:ObtainingPermission
CalculatingTravel Expenses
Verification of Design
Design feasible
[no]
[yes]
Human Resources
Expedition
Establishment of shipping room
Organizing packing and transportation of coll. equipments
Rent a car
Optimizing institutional facilities (cars, equipment, etc)
Hiring local assistant
Management of Transportation
Execution
Field work
Description and Identification of locality using standardized methods
Setting up Collecting equipment
Repeated collection based on protocol
Preparation of samples for preserving of specimens
End of Collection Activity
LabWork
Sorting
Labeling
: Identification
RecordingInDatabase
AddingToCollection
Reporting
End

Discussion

Bottleneck analysis - Interviews inside the walls of the Hungarian Natural History Museum

Three types of interviews were made during the past 7 months. Interviewees were taxonomists from the Botanical and from the Zoological Department of the HNHM.

Interviews of Revisional Taxonomy

The Revisional Taxonomy interviews took place in autumn 2006. An Algologist (Krisztina Buczkó), Bryophytist (Beáta Papp), Coleopterologist (Ottó Merkl), Dipterologist (Mihály Földváry), Lepidopterist (László Ronkay), Lichenologist (László Lőkös), Limnologist (László Forró), Malacologist (Zoltán Fehér) were asked.

The interviews took approximately four hours per one occasion, and in the course of each interview every taxonomist was twice interviewed. On the first occasion, general questions were asked about their work and workflow. By the second time, they already received a complied workflow diagram of their work with the wish to find out, which parts of the workflow were still incorrect and/or imprecise. Finally, they were kept in contact in case to reveal still unclear details.

Interviews of Inventory, ATBI+M

These interviews took place in spring 2007. An Algologist (Krisztina Buczkó), Bryophytist (Beáta Papp), Coleopterologist (Ottó Merkl), Dipterologist (Mihály Földváry), Lepidopterist (László Ronkay), Lichenologist (László Lőkös), Limnologist (László Forró), Malacologist (Zoltán Fehér) were asked.

The interviews took approximately one hour, per specialist. On this occasion they were showed a rough diagram about their workflow during inventory management. This guideline was compiled by András Gubányi, and helped to understand, what kind of information were to find out. With the aim of this guideline, a questionnaire was compiled and showed to the scientists.

With a little help to fill the questionnaire, we went through it step by step and tried to reveal the unclear details of it. Also a comparison was made between our general model and their work and the differences were emphasized. Finally, they were kept in contact to reveal still unclear details. The fulfilled questionnaires were later compiled to a workflow diagram for each taxonomist.

Interviews of the timesheet in Inventory, ATBI+M

These interviews took place in spring 2007. An Algologist (Krisztina Buczkó), Coleopterologist (Ottó Merkl), Lepidopterist (László Ronkay), Lichenologist (László Lőkös), Limnologist (László Forró), Malacologist (Zoltán Fehér). The interviews took approximately one hour per specialists were asked.

For these interviews only the activities from those models were selected, which were compiled after the second row of interviews (inventory). A timesheet of Inventory was compiled, using some variables. We used Effort to denote the amount of person-days, which were spent with direct work and required to complete an adequate activity. Lead-time Demand denoted that time, which last until an activity was completed.

For instance when a taxonomist applies for a collecting permission, he has to fill out the documents, gain the missing papers, etc. However, these activities do not require so much time, but after having handled his application in, he has to wait from couple of weeks to couple months, until his application is accepted, what requires usually much more time, than the effective work.

In a second question a list of software was compiled that they use in cases of some activities.

With a little help to fulfil the timesheet, we went through it step by step and tried to reveal the unclear details of it.

Using the following equations some results were interfered from the data of the fulfilled timesheets:

1. equation1=IF(Average Effort>0; (Max. Effort / Min. Effort) / Average Effort*100)
2. equation2=IF(Average Lead-time Demand>0; (Max. Lead-time Demand / Min. Lead-time Demand) / Average Lead-time Demand *100)
3. equation3=IF(Average Effort>0; (Max. Lead-time Demand / Min. Effort) / Average Lead-time Demand *100)

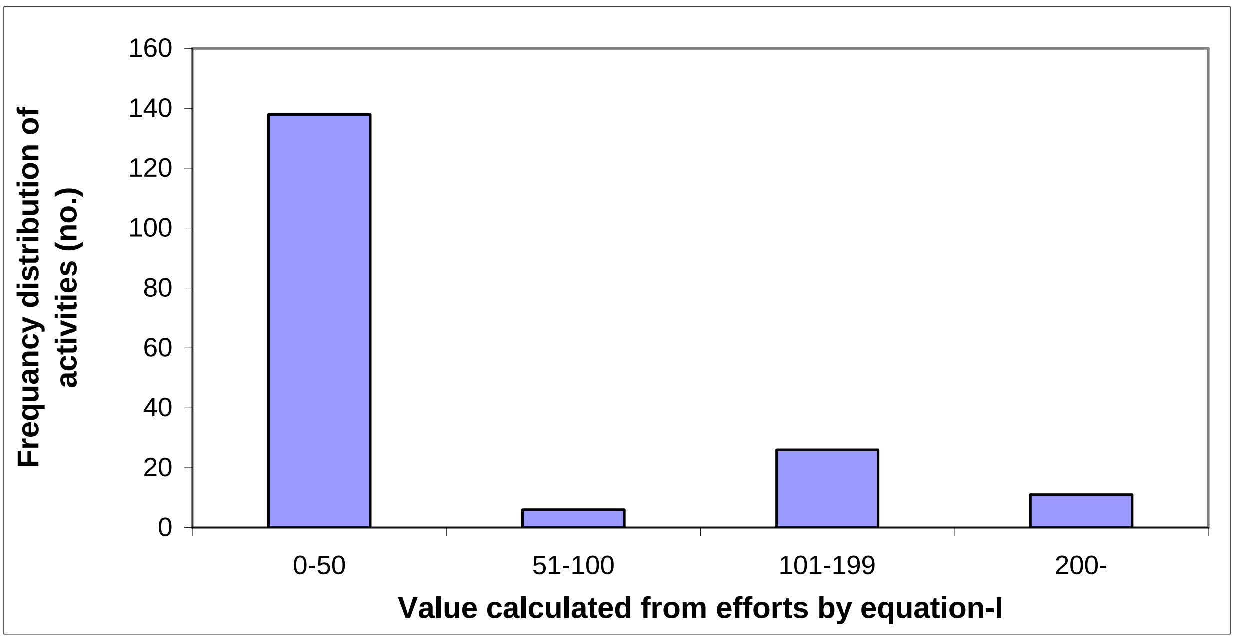


Figure 1. Frequency distribution of activities by equation-I

With the aim of the abovementioned equations, we calculated values. On Figure1. it can be seen, in cases of which activities the calculated value from Effort are higher.

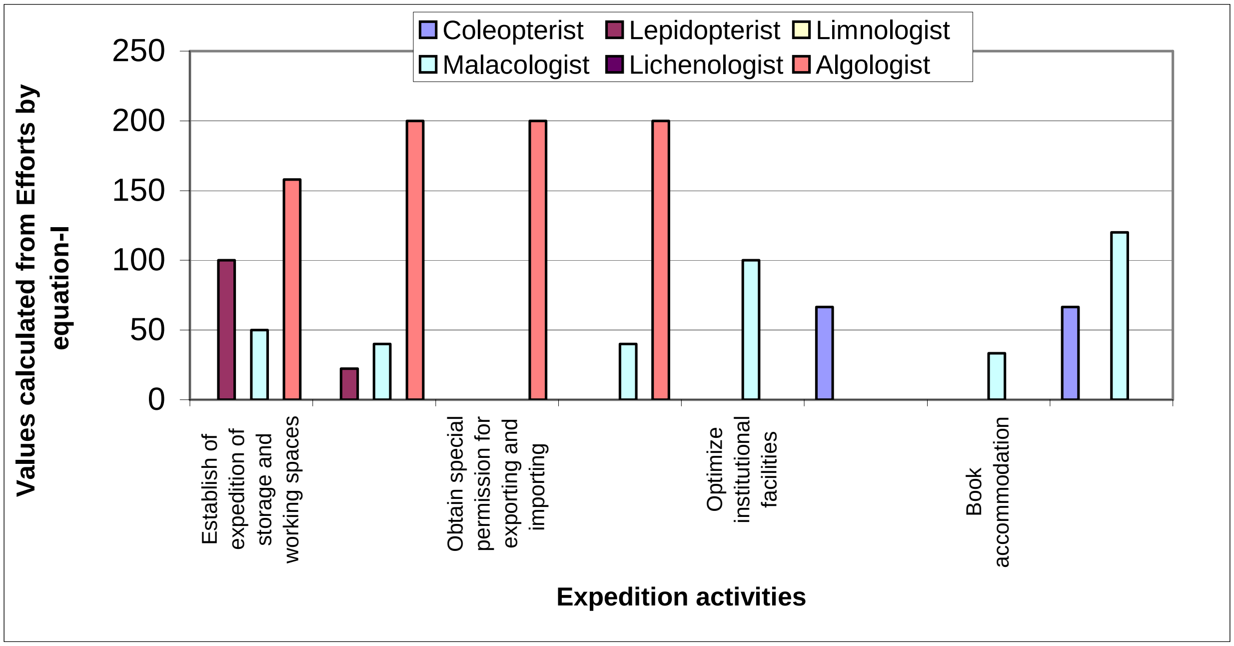


Figure 2. Min-max Effort differences in case of Expedition activities (Activities with zero value are not showed: *estimate of storage necessitiy, optimise collecting activity against financial and human resources*)

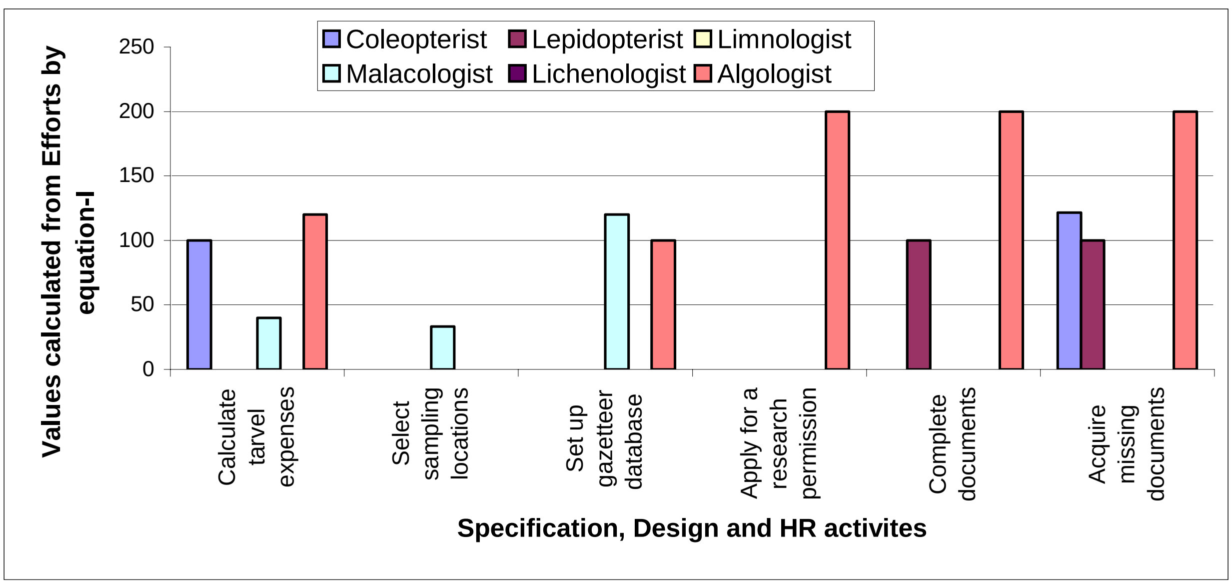


Figure 3. Min-max Effort differences in case of Specification, Design and HR activities (Activities with zero value are not showed: *set up taxon groups, select collecting methods, determine species number and turnover estimators, progress chart by taxon groups, Sample design/plan by groups, create copyright notice, clarification of authorship, clarification of material ownership, list specialist with reference, set up database structure and software*)

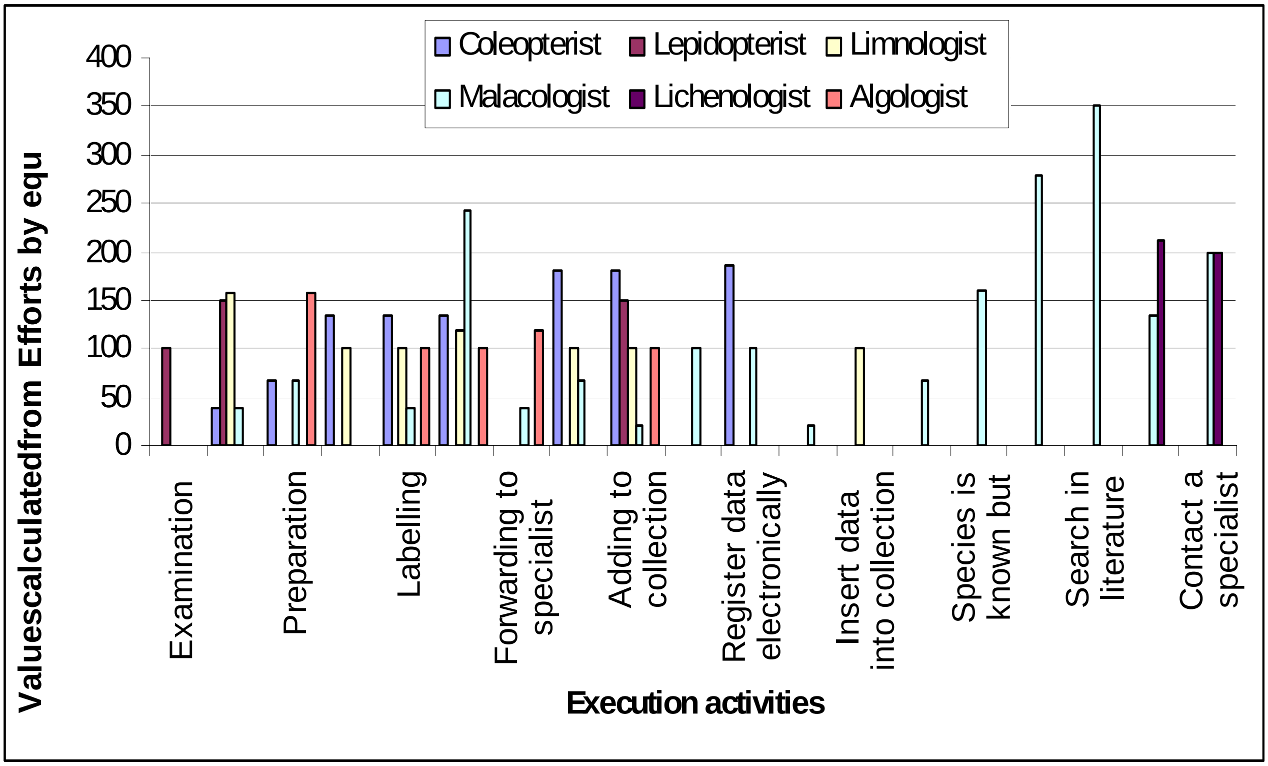


Figure 4. Min-max Effort differences in case of Execution activities (Activities with zero value are not showed: *video or sound recording*)

In case of the equation-I, the values reveal the differences between the maximum and minimum Effort. Where the difference is huge, the Effort in person-days that the adequate activity requires is quite variable. We observed huge differences in case of the following activities: *hiring local assistant workers, managing transportation and collecting materials, obtaining permission, identifying new species.*

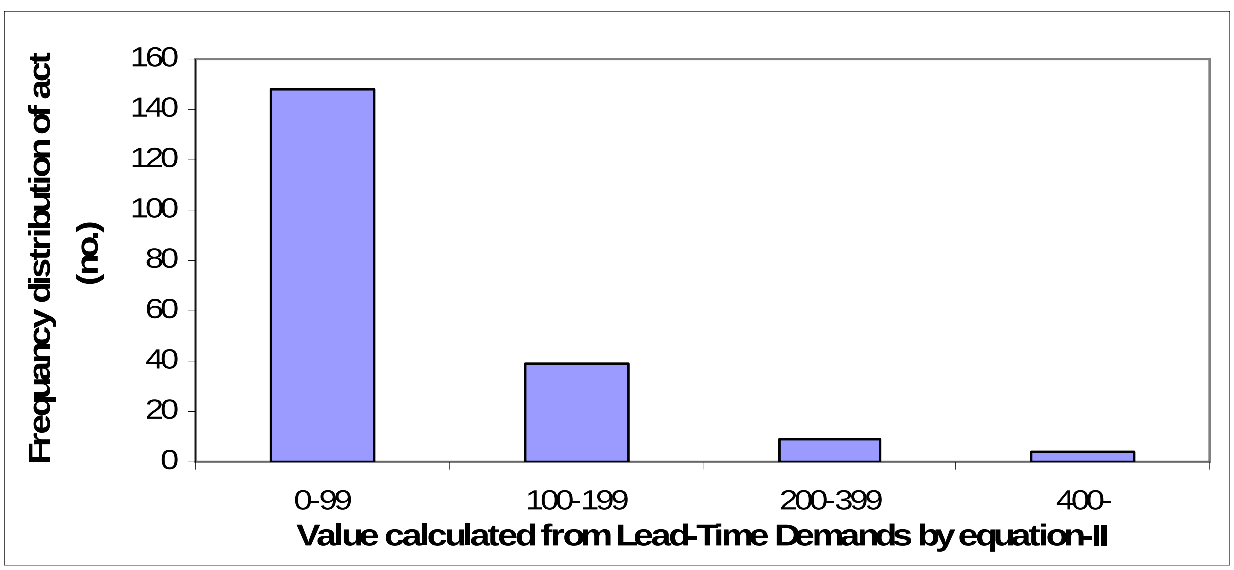


Figure 5. Frequency distribution of activities by equation-II

On Figure 5. it can be seen, in cases of which activities the calculated value from Lead-Time Demands are higher.

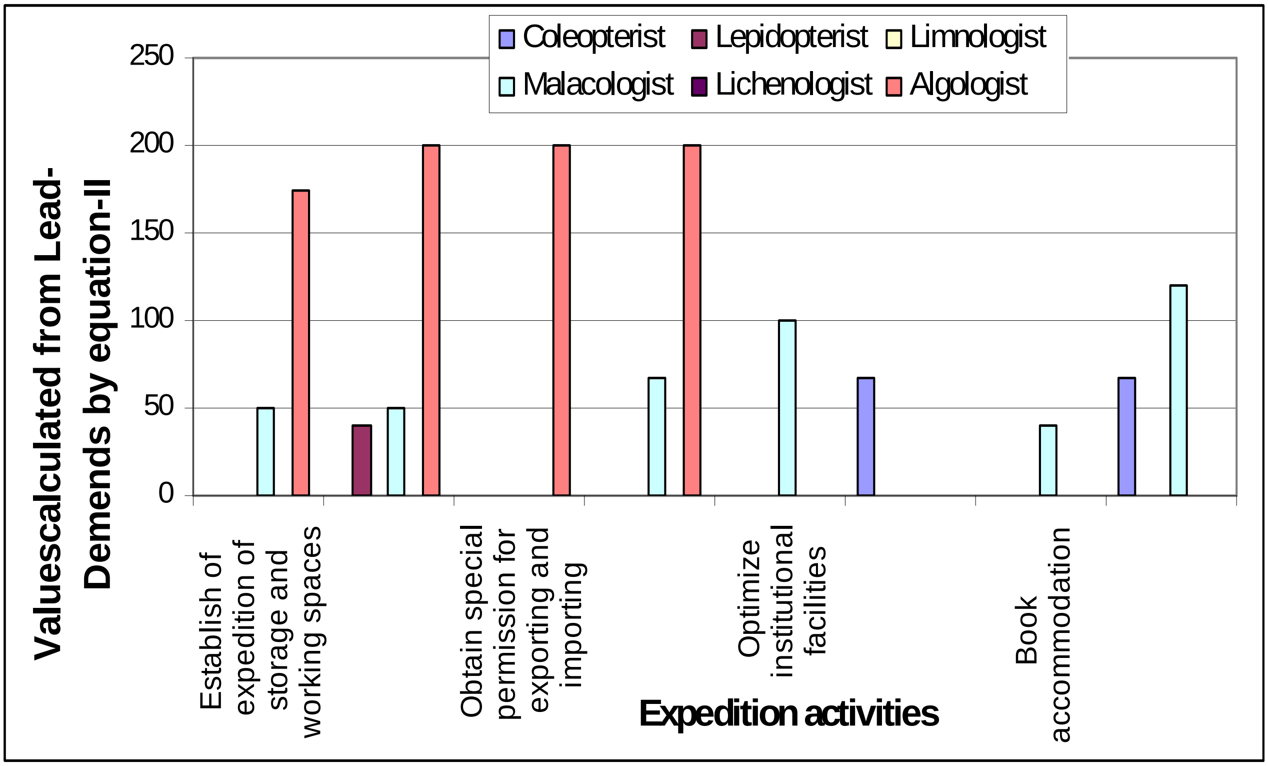


Figure 6. Min-max Lead-Time Demand differences in case of Expedition activities (Activities with zero value are not showed: *estimate of storage necessity, optimise collecting activity against financial and human resources, set up research team*)

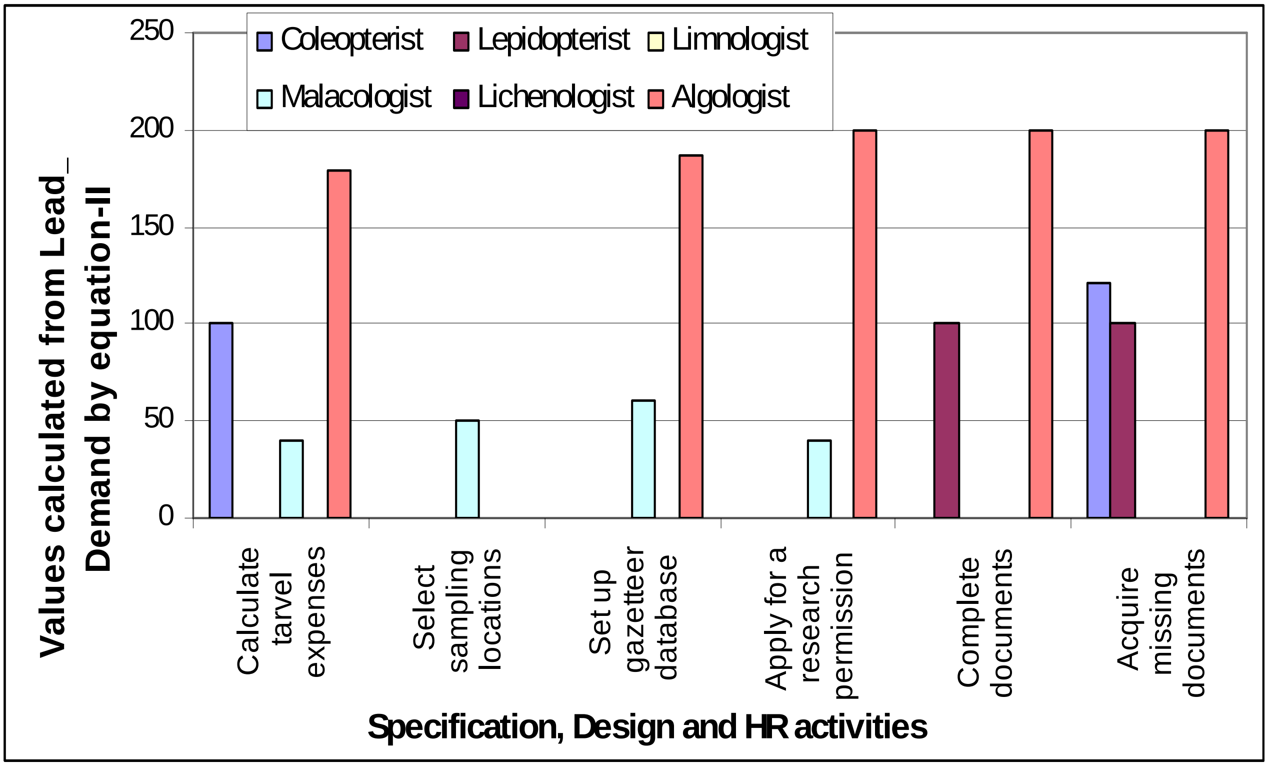


Figure 7. Min-max Lead-Time Demand differences in case of Specification, Design and HR activities (Activities with zero value are not showed: *set up taxon groups, set up research team, select collecting methods, determine species number and turnover estimators, Progress chart by taxon groups, sample design/plan by groups, create copyright notice, clarification of authorship, clarification of material ownership, list specialist with reference, set up database structure and software*)

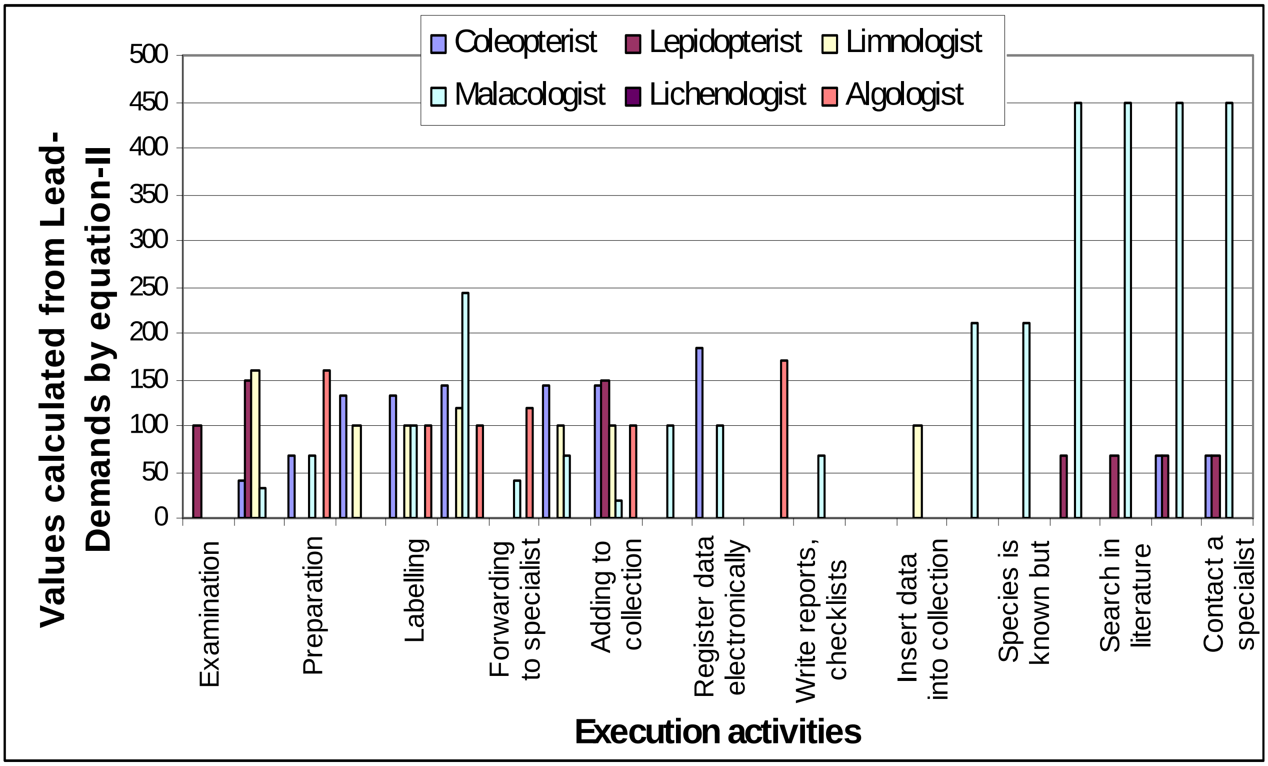


Figure 8. Min-max Lead-Time Demand differences in case of Execution activities (Activities with zero value are not showed: *video or sound recording*)

In case of the equation2, the values reveal the differences between the maximum and minimum Lead-Time Demand. Where the difference is huge, the Lead-Time Demand in person-days that the adequate activity requires is quite variable. Huge differences were observed in case of the following activities: *Lab work, Identifying new species.*

If there are huge differences between maximum and minimum Efforts and maximum and minimum Lead-Time Demands, there differences might be eliminated through some kind of electronic aim. Huge differences were observed in case of the following activities:: *hiring local assistant workers, managing transportation and collecting materials, obtaining permission, identifying new species.*

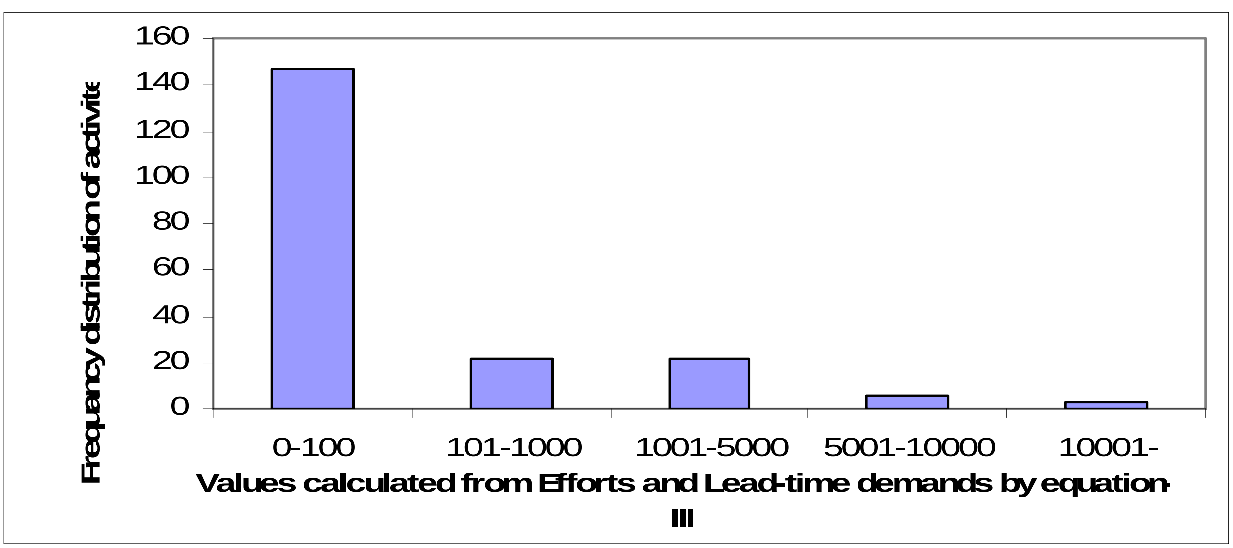


Figure 9. Frequency distribution of activities by equation-III

On Figure 9. it can be seen, in cases of which activities the calculated value from the differences between Efforts and Lead-Time Demands are higher.

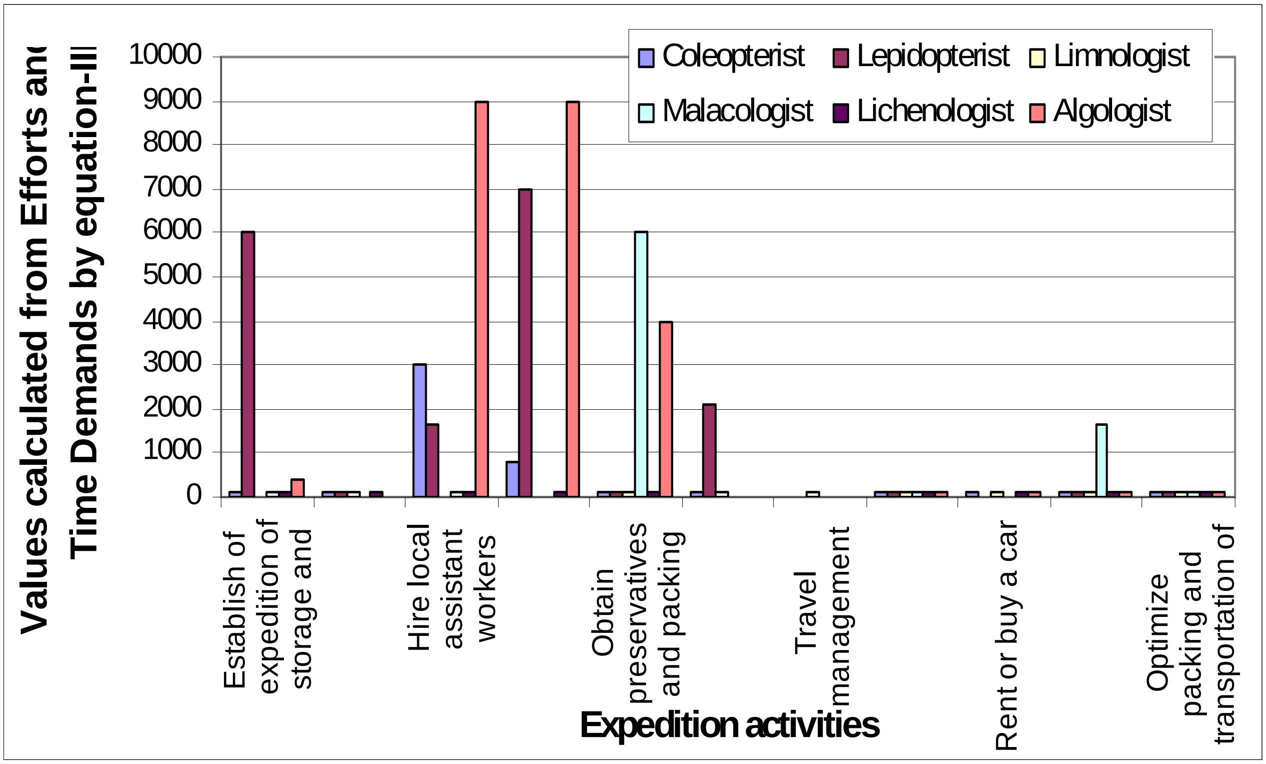


Figure 10. Average Effort and average Lead-Time Demand differences in case of Expedition activities (Activities with zero value are not showed: *manage transportation and collected materials)*

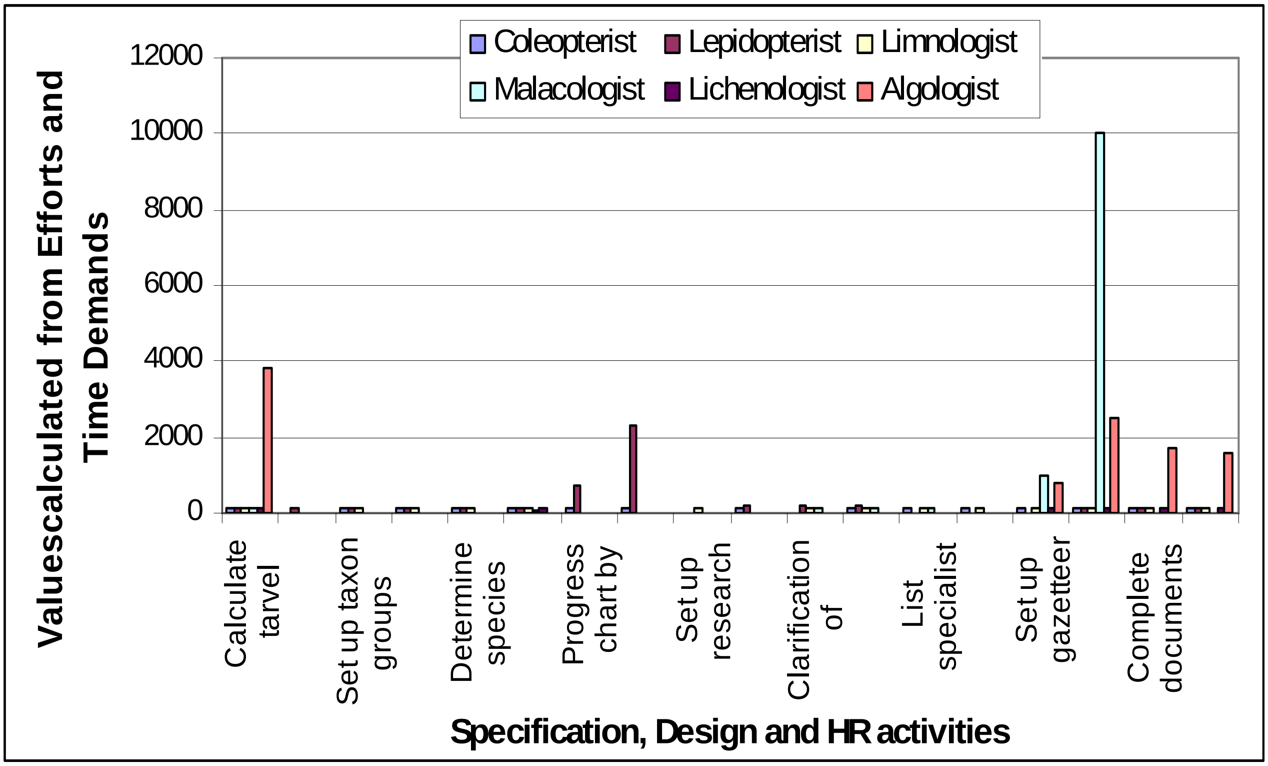


Figure 11. Average Effort and average Lead-Time Demand differences in case of Specification, Design and HR activities

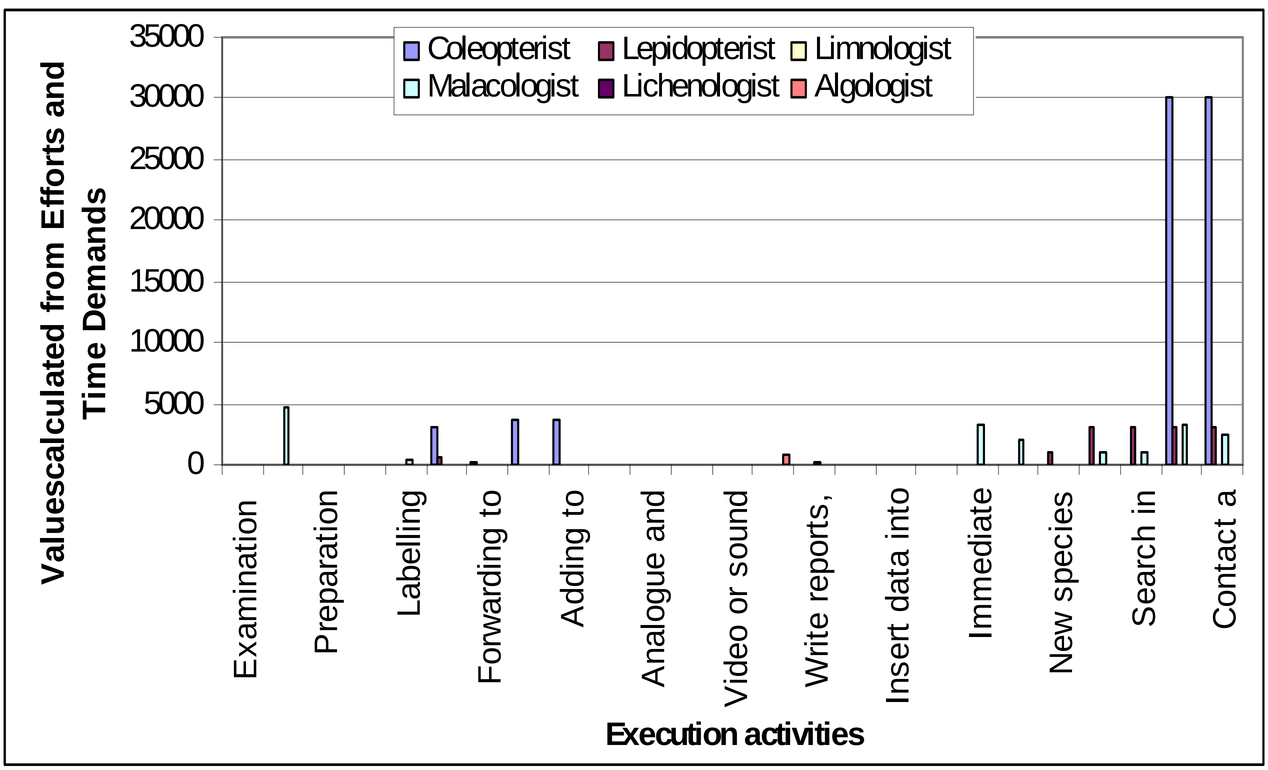


Figure 12. Average Effort and average Lead-Time Demand differences in case of Expedition activities

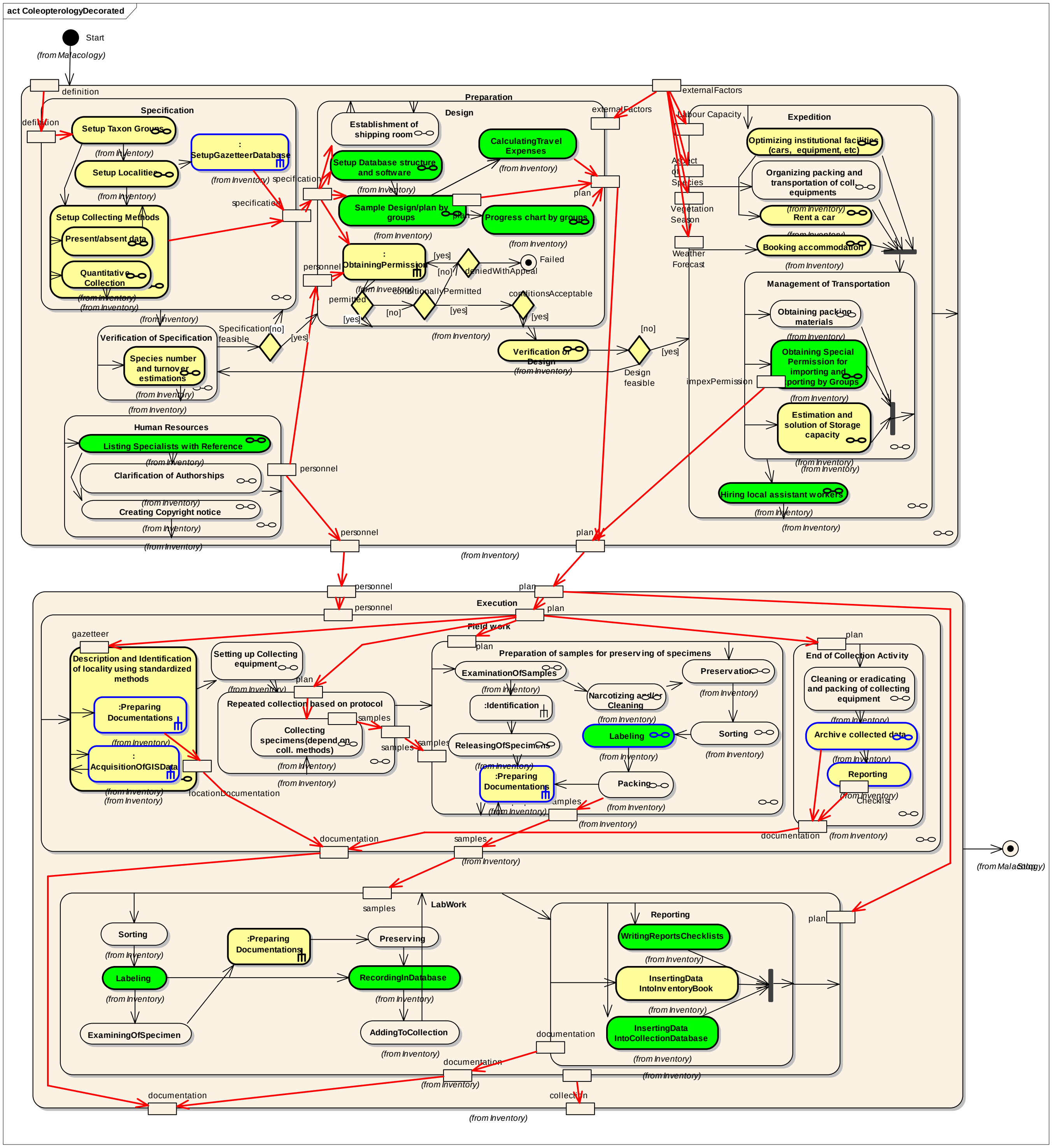
In case of the equation3, the values reveal the differences between the average Lead-Time Demand and average Effort. Where the difference is huge, it means that not much work of activities is done at the same time, but these types of activities have to be done regularly. Thus more activities have to be dealt with constantly, which requires separated attention that proved to be quite hard for anyone.

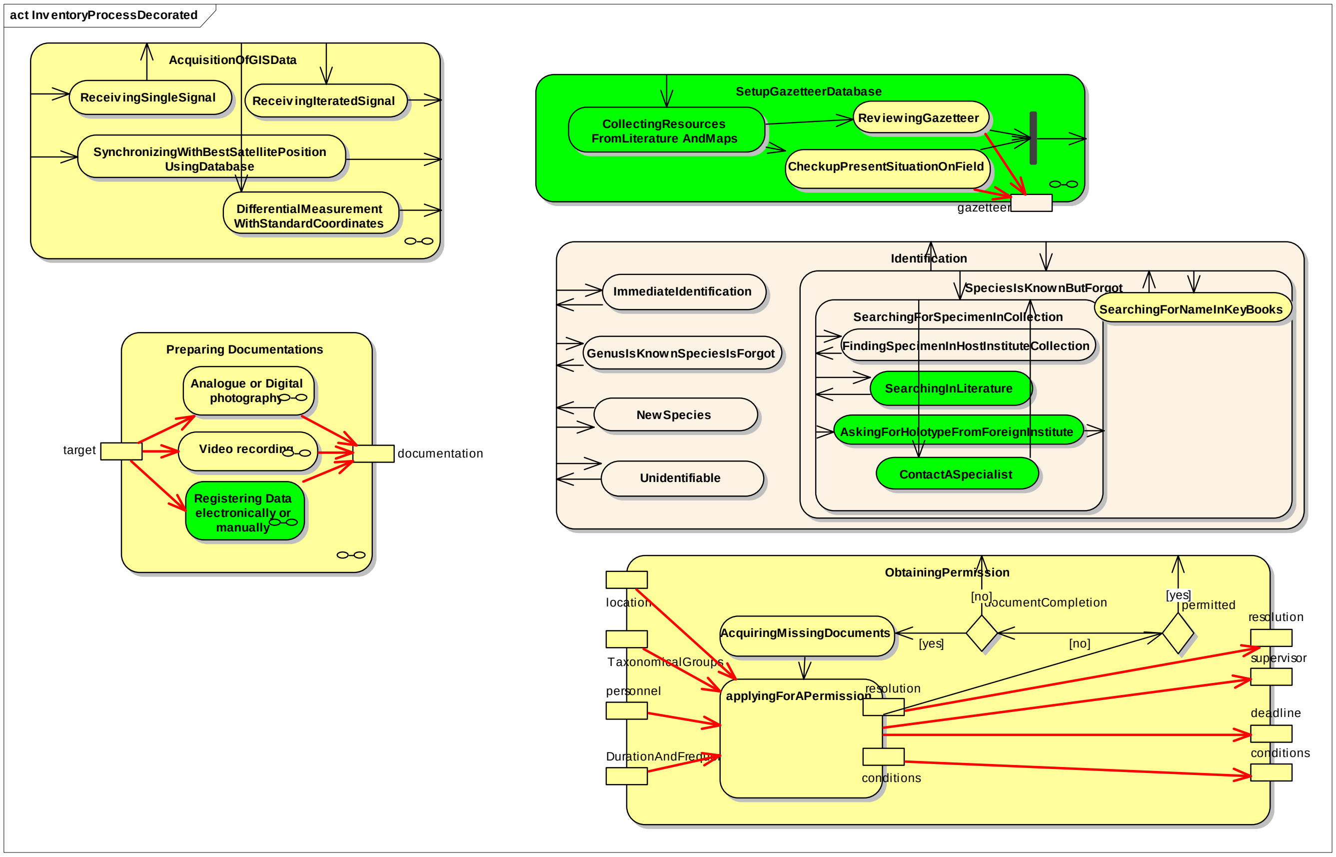
Decorated workflows

As it was already mentioned above, András Gubányi compiled a general model for Inventory. During the last type of interviews, the timesheet of the taxonomist was filled out. Beside the Effort and Lead-time Demand we were interested weather software were used during the activities.

On that general Inventory model, Imre Kilián used different type of colours according to the details of the timesheet and his own knowledge and experience.

Green colour was used in cases of those activities, where the scientist used software. While yellow colour signed those activities, where the scientist did not use any software, but some kind of electronic aim would help to complete the activity quicker. Finally, blue outlined those activities, which are usually fulfilled on field or on such a locality, where the network of his institute or the Internet can’t be reached. Generally, they bring a laptop with them, onto which they export the necessary information, and then they work on field, and after in the institute they import the information gotten on field onto the network.


act ColeopterologyDecorated
Start
(from Malacology)
definition
Specification
Setup Taxon Groups
Setup Localities
:SetupGazetteerDatabase
Setup Collecting Methods
Present/absent data
Quantitative Collection
specification
Verification of Specification
Species number and turnover estimations
Specification feasible
[no]
[yes]
Human Resources
Listing Specialists with Reference
Clarification of Authorships
Creating Copyright notice
personnel
Preparation
Design
Establishment of shipping room
Setup Database structure and software
CalculatingTravel Expenses
Sample Design/plan by groups
Progress chart by groups
:ObtainingPermission
Failed
deniedWithAppeal
conditionallyPermitted
permitted
conditionsAcceptable
Verification of Design
Design feasible
plan
externalFactors
Labour Capacity
Aspect of Species
Vegetation Season
Weather Forecast
Expedition
Optimizing institutional facilities (cars, equipment, etc)
Organizing packing and transportation of coll. equipments
Rent a car
Booking accommodation
Management of Transportation
Obtaining packing materials
Obtaining Special Permission for importing and exporting by Groups
Estimation and solution of Storage capacity
impexPermission
Hiring local assistant workers
(from Inventory)
Execution
Field work
gazetteer
Description and Identification of locality using standardized methods
:Preparing Documentations
:AcquisitionOfGISData
locationDocumentation
Setting up Collecting equipment
Repeated collection based on protocol
Collecting specimens(depend on coll. methods)
samples
Preparation of samples for preserving of specimens
ExaminationOfSamples
:Identification
ReleasingOfSpecimens
Narcotizing and/or Cleaning
Preservation
Labeling
Sorting
Packing
End of Collection Activity
Cleaning or eradicating and packing of collecting equipment
Archive collected data
Reporting
Checklist
documentation
LabWork
Sorting
Labeling
ExaminingOfSpecimen
Preserving
RecordingInDatabase
AddingToCollection
Reporting
WritingReportsChecklists
InsertingData IntoInventoryBook
InsertingData IntoCollectionDatabase
collection
(from Malacology)

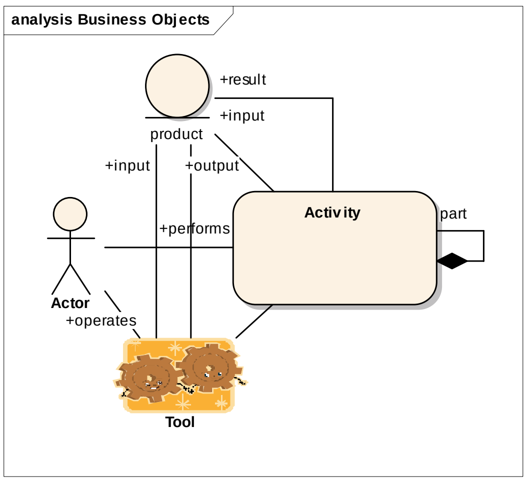

act InventoryProcessDecorated
AcquisitionOfGISData
ReceivingSingleSignal
ReceivingIteratedSignal
SynchronizingWithBestSatellitePosition UsingDatabase
DifferentialMeasurement WithStandardCoordinates
SetupGazetteerDatabase
CollectingResources FromLiterature AndMaps
ReviewingGazetteer
CheckupPresentSituationOnField
gazetteer
Identification
ImmediateIdentification
GenusIsKnownSpeciesIsForgot
NewSpecies
Unidentifiable
SpeciesIsKnownButForgot
SearchingForNameInKeyBooks
SearchingForSpecimenInCollection
FindingSpecimenInHostInstituteCollection
SearchingInLiterature
AskingForHolotypeFromForeignInstitute
ContactASpecialist
Preparing Documentations
Analogue or Digital photography
Video recording
Registering Data electronically or manually
target
documentation
ObtainingPermission
location
TaxonomicalGroups
personnel
DurationAndFrequ
AcquiringMissingDocuments
applyingForAPermission
resolution
conditions
[no]
documentCompletion
[yes]
[no]
[yes]
permitted
resolution
supervisor
deadline
conditions

Proposal of a Workflow Management Software

One natural follow-up of the workflow modeling activities, as described in the previous section, is to construct an integrated workflow-management software. The basic assumptions of such a software, the first steps of use case modeling are summarized in this chapter.

Workflow modeling basics

When speaking about workflow modeling we must first fix the metamodel, i.e. the basic set of concepts used by describing workflows. There are also a numerous workflow modeling systems, sites and metamodels (OMG's Software Process Engineering Metamodel (SPEM), Business Process Engineering Language (BPEL) and others), so it is most likely not necessary to reinvent them, it is enough only to apply an existing framework.



The most basic question is: what to model. This question had already to be answered in case of interviews and business process modeling. The modeled concepts are depicted on the figure, and can be summarized as follows:

- activities, which are certain packages of tasks, that are performed in a good circumscribed way. Activities may be elementary (steps), or may be compound, in such a case they consist of other activities.
- business actors, mainly humans, who perform the above mentioned activities.
- Products, that are inputs or results of the mentioned activities
- tools the actors operate to complete the activities. Certain tools have direct input or output relationships with products, which are in most cases also the input and output of the corresponding activities.

How workflow management software looks like?

Workflow management software is a corporate/groupware solution, which operates on a network, and connects the actors. Workflow management can be applied only in case of information products, that are processed electronically, and they are useful when the flow of processing is complicated.

Such software maintains the different processing steps, products, and their actors. The activities are ordered to the actors, who can browse them in their task-box, they can open, and complete them. When an activity is completed, it is passed automatically forward to the next fellow, who makes some other processing step.

One activity-step is either a simple form with information bits to be filled out, maybe also with decisions, or wrappers of commercial software (word processors, databases, image/movie processors etc.)

Workflow management software therefore must be a client-server solution, where the basic workflow and the management of much corporate information are done by the server side, while the distinct steps are done by local software.

Demand driven design – concluding a use case model

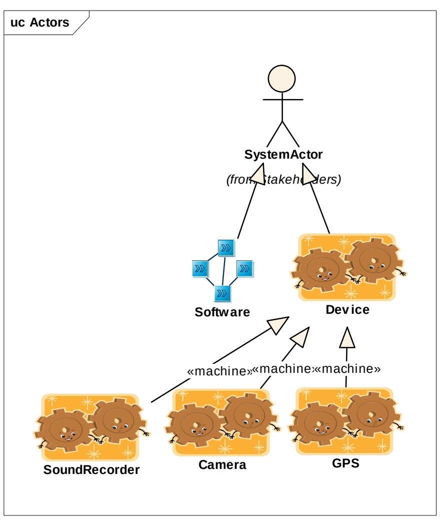
Use Case modeling is the first step when designing new software. The point in use case modeling is to define the borders of the system. Use case analysis uses the following concepts:

- actors: all those who get in touch with the system. This includes people, using the system, gaining any benefit from the system, servicing and/or maintaining the system. Beyond single people also organizations are mentioned, but external devices also belong to actors.
- use cases correspond to functions, the system completes.
- relationships among use cases. Here two kinds of relationships are used. Inclusion: a use-case includes another, if the second is performed every time the first is performed. Extension: a use case extends another, if the extension is also performed in certain cases when the extended use-case is performed.
- relationships between actors and use cases. For modeling such features "normal" UML associations are used. The direction of the association denotes the activity flow of the operation. When the arrow points at the actor, the actor is passive, the operation is initiated by the system. When the arrow points at the use-case, then the actor is active, i.e. she is the one that initiates the operation.

Actor model from stakeholder analysis

As a first step the actors of a system are identified. The actors are basically those, being collected in the stakeholder analysis of the previous chapter. However stakeholders are only human actors, whereas in actor model device (hardware and software) actors, and other technical actors must be identified.

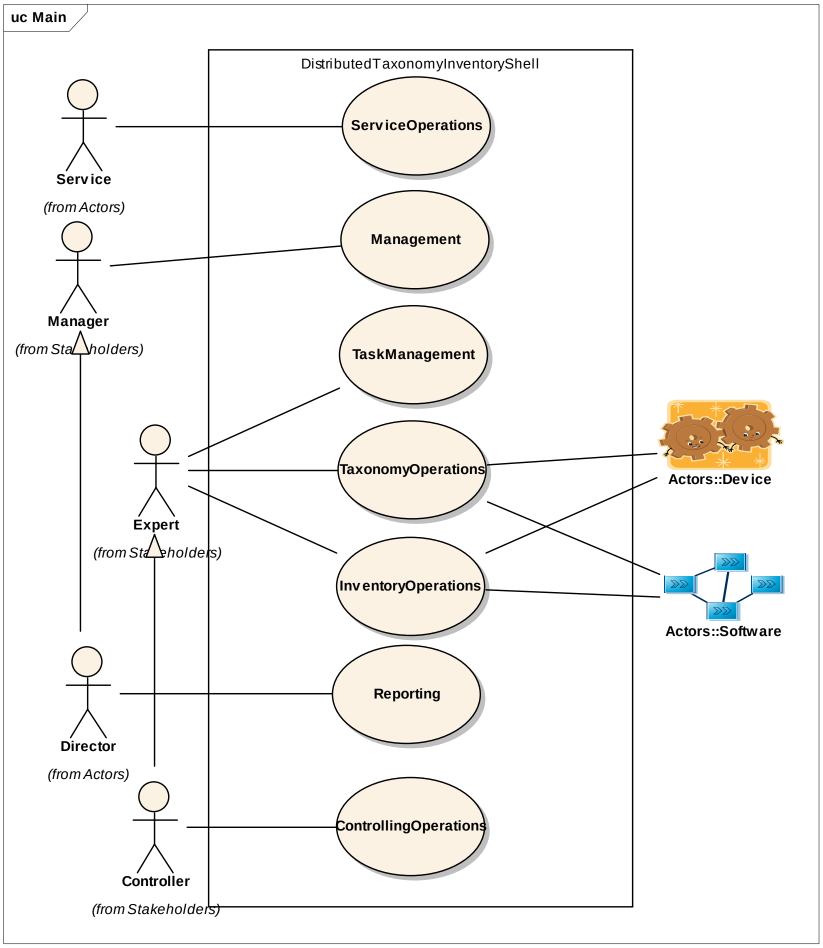
One important bunch of actors is the class of System Actors (see below). System actors are either software, or hardware actors. We don't detail software actors at the moment, whereas from hardware actors (Devices) Sound Recorders, Cameras and GPS devices have been identified.



Another important difference to workflow analysis is the identification of controllers. Controllers are those actors, who perform some quality control operations. At the moment Specification Controllers, Design Controllers, Field Controllers and Lab Controllers have been identified.

Use Case Model – An Overview

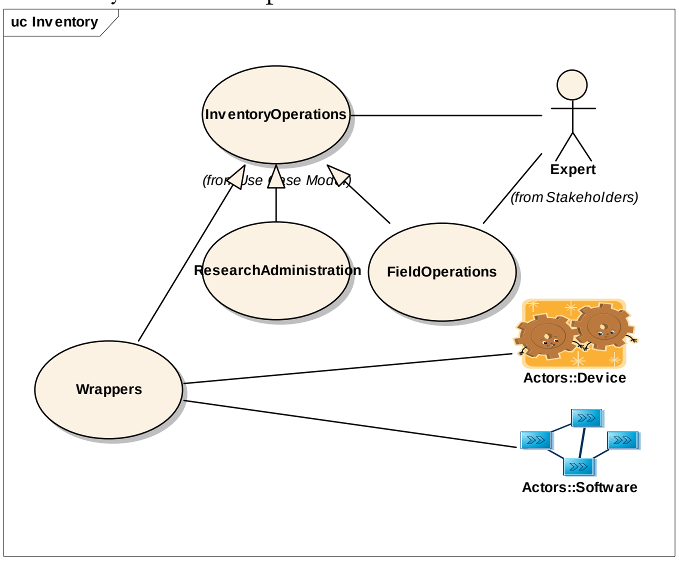
The most important function of the Use Case model is to set the borders of the system. To reach this the most important common user groups, and the most general operations are identified.



- A concrete software system must also include service functions; those are operated by service personnel. Service functions are those, maintaining basic information chunks, users, locations, authorizations etc.
- The workflow-management system includes also Management Operations that are carried out by managers. Those are the operations at the specification and design of a project, and some overview/reporting possibilities that give an impression, to what extent a project is completed at a given moment.
- Task management operations are those, which may influence the normal flow of control, and that is available for everybody. Under task management we may take operations, like: postponing a task, sending back a task on its lifeline if the results are not satisfying, and the like.
- Inventory and Taxonomy operations are those, corresponding to the workflow activities that are executed by taxonomists and biologists. These operations also refer to a special group of actors, namely: special software that the workflow management serves as a wrapper, and special devices the human actors work with.
- There are some special tasks in both of inventory and taxonomy operations, for securing the quality of the work. Those steps are carried out by a special group of actors the so-called Controllers.

Use Cases from Activities

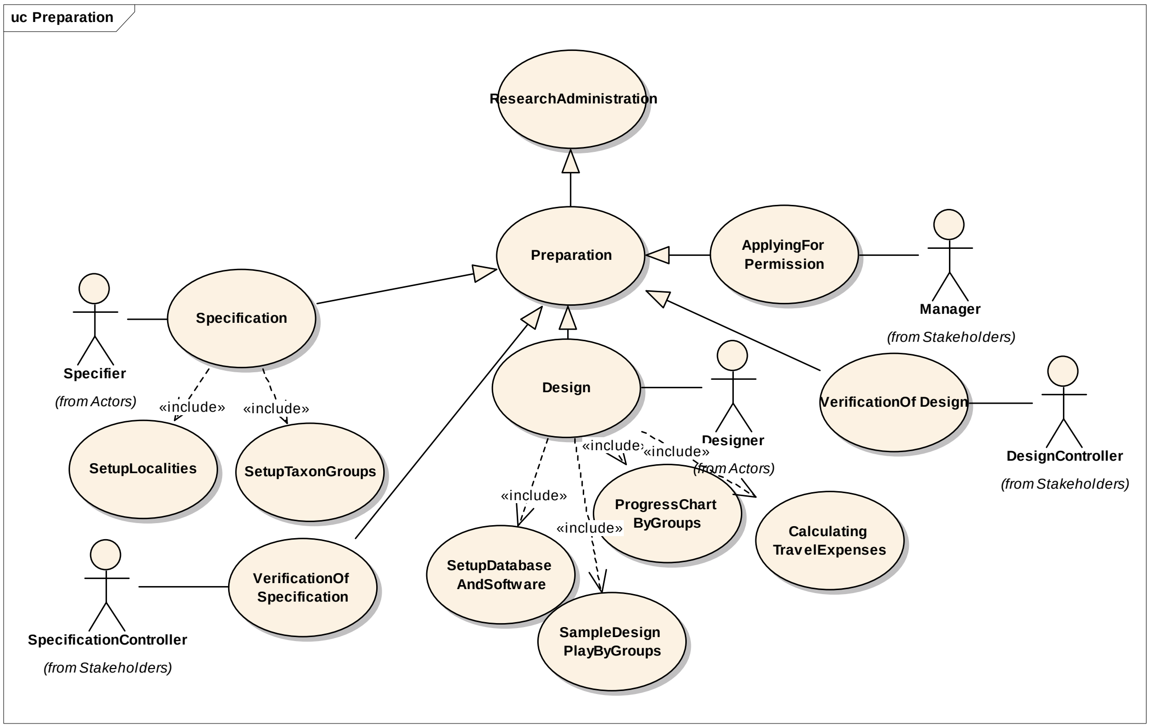
When refining the Use Case Overview model above, the activity model can be a good starting point. The decorated workflow diagrams define the set of activities, that could be carried out by a computer aided way. The modelers task is merely to define the interaction among them, and define eventually common "partial use cases".



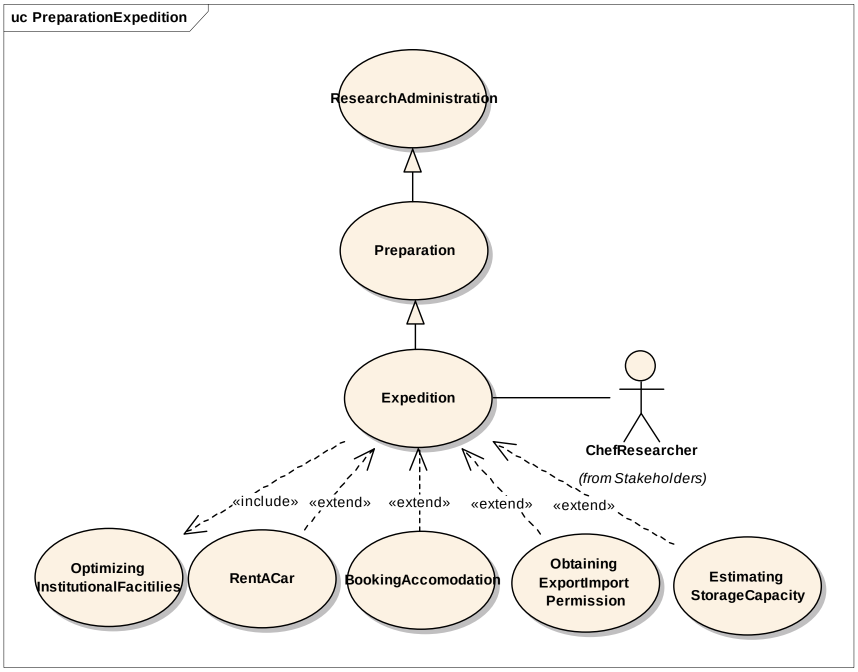
The most general model says: inventory operations are those, carried out by experts, and they consist of three groups:

- wrappers: those wrapping up some common software for word processing, database manipulations etc.
- research administration: those, mainly forms for information input.
- field operations: those, being carried out on the field, offline, with somewhat limited computing capacity

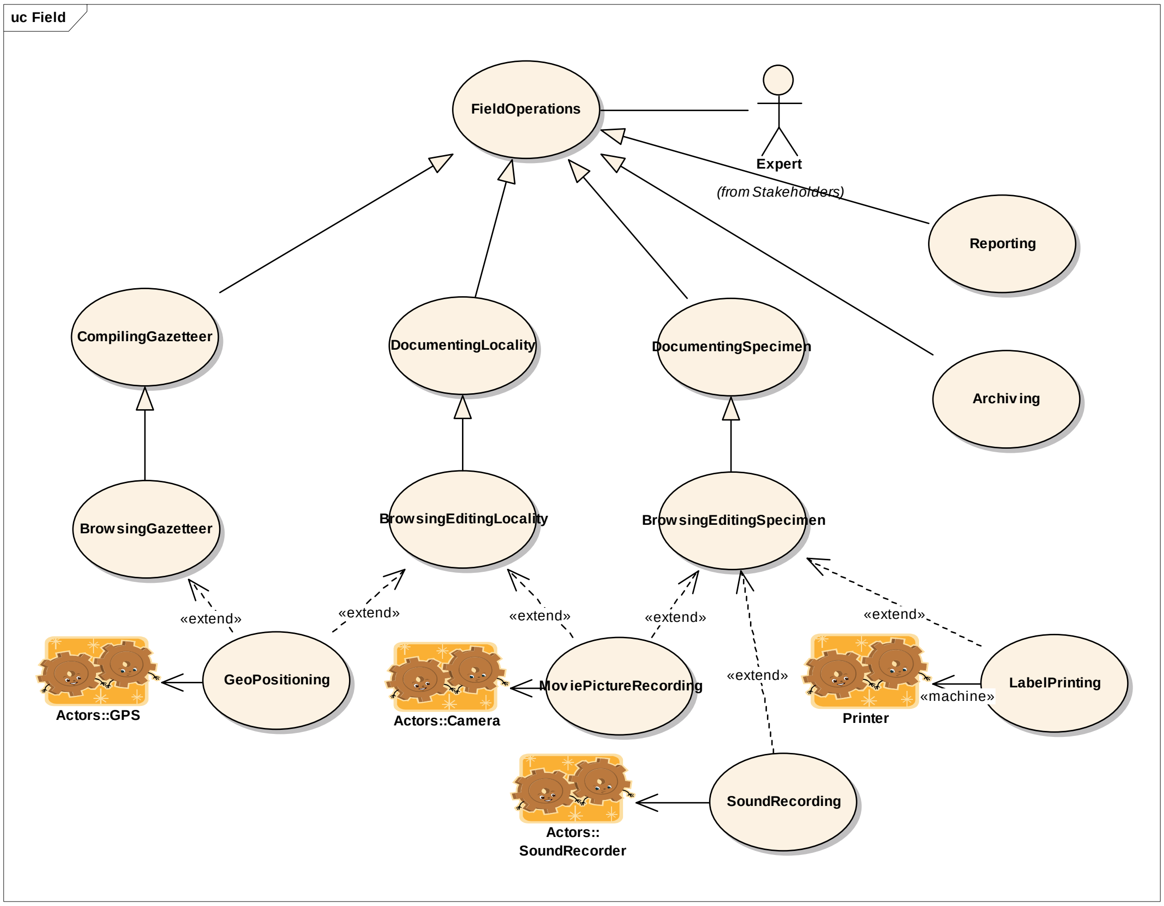
Preparational activites belong to the group of research administration operations, and its structure is depicted by the following picture:



- Specification is the responsibility of Specifiers. By specification we understand two basic thing, that is included compulsorily: 1. Setting up the localities, and 2. Setting up taxon groups
- The specification is controlled by Specification Controller actors.
- Designing is carried out by Designer actors. Its included sub-activities are: 1. Setting up database and software, 2. Compiling a sample design plan by taxon groups, 3. Compiling progress charts by taxon groups, and finally 4. Calculating travel expenses
- Design is controlled by Design Controller actors
- Applying for permissions is the task of Managers



Expedition preparation activities also belong to preparation activities, and further to research administration activities. It includes only one activity compulsorily: Optimizing Institutional Facilities. The optionally extensions of Expedition preparation are the following: 1. Renting a car (in many cases not necessary) 2. booking accommodation (in many cases researchers sleep on the field in tents) 3. Obtaining export/import permission (not necessary e.g. in domestic expeditions 4. Estimating storage capacity (not necessary for small volume expeditions)

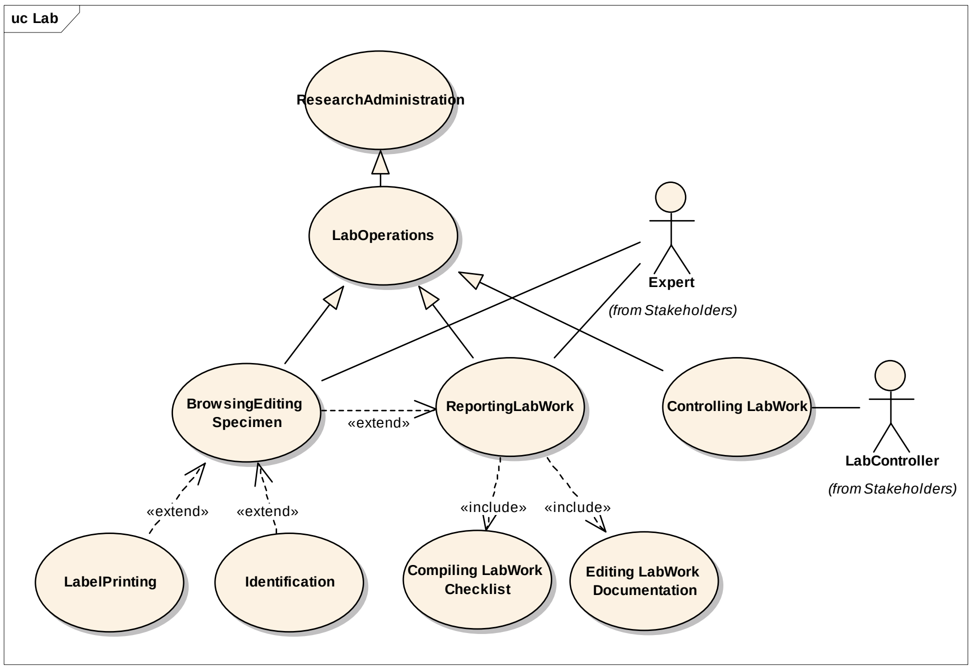


Field operations are those, being carried out off-line. When an expert goes to an expedition, prepares (exports) a local environment of working material, and when she comes back, makes an import to feed everything back into the corporate database again. The concrete operations belonging to this group are:

- Compiling Gazetteer, an operation, implemented practically, as browsing gazetteer. This is extended by the GeoPositioning use case, that interfaces GPS as a passive actor.
- Documenting Locality, an operation, implemented practically, as browsing and editing locality database. This use case is extended firstly by GeoPositioning, secondly by Movie Picture Recording use case. This latter interfaces Camera, as a passive actor.
- Documenting Specimen, an operation, implemented practically, as browsing and editing specimen database. This is optionally extended by Movie Picture Recording, or Sound Recording use case, that interfaces Sound Recorder, as a passive actor. It may also be extended by the Label Printing use cases, that refers to a printer.
- Reporting: a use case that is carried out also on the field, and practically is used only as a wrapper of a common word processor.
- Archiving: a use case that makes an archive of the locally used database.

Lab Operations are those, carried out in the laboratory, after returning home from the field expedition. It has basically the following three operations:

- Browsing Editing Specimen, that is extended firstly by the Label Printing use case, secondly by the Identification use case. Both are carried out by the expert. The use case is also extended a third use case (Reporting Lab Work) that is described in the following paragraph
- Reporting Lab Work can be used independently, or also from the previous use case (Browsing Editing Specimen). This is also carried out by an Expert.
- Controlling Lab Work is completed by a Lab Controller actor.



Conclusions

In the previous chapters we have introduced the modeling activities having been carried out at the Hungarian Natural Museum, Budapest, in the framework of the EDIT project.

The first chapter summarized the basic information and agreements that were necessary to understand and/or interpret the following chapters properly.

In the second chapter the actual results of workflow modeling has been described. The results included a disciplinarily divided, and an abstraction divided dimension of modeling inventorying activities. We have thus achieved to identify activities that were carried out by experts of any actual discipline, and modeled those in the package of common, discipline-independent activities. Under this package we have identified ten different subpackages for Algology, Bryophytology, Coleopterology, Dipterology, Entomology, Lepidopterology, Lichenology, Malacology, Ortopterology and VascularPlants. For the abstraction related dimension we have used the UML tool of encapsulating activities into each other. We have identified therewith four abstraction levels, and gained three projections of different granularity of the one and the same model. These were: *Abstract processes*, *General processes*, and *Detailed processes, Simplified processes*.

The third chapter speaks about the bottleneck identification of the matter. For this end we have extended the activity information by elapsed time and effort information tags, and evaluated those by Excel based equations. The results of this phase were far from being exact, because of the following reasons:

- The Hungarian interviewees were not equally cooperative to invest energy to estimate the mentioned time-tagged data for our investigations
- The mentioned Excel equations don't give any 100% method for finding out the bottlenecks of inventorying activities. Anyhow, they must be regarded as indicators, that can be compared to intuitive guessing
- In spite of all the mentioned insecurities, the end results correspond fairly to the guessing and expectations.

We didn't report in the third chapter, that we made an experimental processing of the time-tagged data by Microsoft Project. Unfortunately we haven't found it bringing us closer to the final result of the matter.

In the fourth chapter a proposal is put forward to understand "EDIT Cyber platform" as a workflow processing software, and an initial use case model is sketched. Remark that it is far from being ready: it should be considered preliminary. Finalizing the use case analysis can be done when the approach has got green light in a later phase of EDIT project.

References

[OMG1] Unified Modelling Language: Infrastructure
Object Management Group 04-Feb-07.

[OMG2] Unified Modelling Language: Superstructure
Object Management Group 03-Feb-07.

Annex 1: Disciplines

Algology

There are numerous differences with respect to the generic inventorying process. The main differences can be summarized as follows:

- there is no need of obtaining permissions
- renting a car is not necessary, but it is necessary sometimes to rent a canoe or a boot
- as algae are quite small, and in the examined case only domestic activities were carried out, there were no need of obtaining export and import permissions, and no need of obtaining packaging material and storage capacity
- only actual collecting is done in the field, all examination activities are carried out in the laboratory
- documentation is also compiled only in the laboratory
- field work is carried out mostly alone, or with family in form of family boat trips. Although for bigger project a more complex research team might be necessary, for one man projects the Human Resources package is fully irrelevant.

LabWork

Laboratory work in case of algology is different from the generic case inasmuch as practically no examination and documentation activities are carried out in the field. The activity has two input parameters, the collection itself, and a research plan.

Packing

Packing means boxing, whereas they are stored in cupboards on slides.

Rent a Boot or Canoe

For experts carrying inventory activities out in water habitats, it is in most times more important to rent a boot or a canoe, than to rent a car. This is also true for algologists.

Bryophytology

The points differing from the generic description are as follows:

- testing collecting methods. Testing is done habitat by habitat, one by one. The bigger the concerned are is, the more information can be gained.
- Renting a car: it happens very rarely because it is expensive
- Repeated collection by groups: it means a repeating over habitats.
- Examining and identifying the samples happens only in laboratory.
- As there is usually only a single taxonomic group collected (bryophytes), no progress chart is prepared.
- There is no releasing

- There is no narcotizing
- Preparing documentations happens also in the laboratory

Colepterology

Colepterology differs from the generic inventorying process only in the following slight details.

- In the examined case for example there is no shipping room, because the researcher uses his/her own room for this purpose.
- In the field only the easy and common specimen are identified, those which can be identified using microscope.
- Common species, and females are released, because males are easier to examine
- Only pre-labeling is made in the field.
- In the laboratory only the yet unknown specimen must be identified

Lepidopterology

In the course of interviews we have interviewed three lepidopterologist. The difference among then is summarized in the following diagrams.

The difference between lepidopterology and the generic case is also very slight, and can be summarized as follows.

1. It is sometimes quite hard to obtain collecting permission because there are different rules in and out of the EU. It is also a problem that you can not state definitely for permission, which kind of specimen to collect.
2. There are always constant financial problems. Therefore trips are usually financed from different resources.
3. Cars are rented sometimes, but very rarely.
4. Local workers are hired also very rarely, mostly for carrying load, or as a local interpreter.
5. Obtaining an import/export permission is not necessary, only in case of getting loaned material from other institutions.
6. Repeating based on protocol: In Hungary Lepidoptera must be checked in every 3 weeks.
7. Examining and identifying the samples happens always under microscope, in laboratory. For examination sometimes genetic test is also applied.
8. Releasing of specimens is rare, because live capturing methods are less effective, and for identification they must be narcotized.
9. Documentation is prepared only in laboratory.

Lichenology

The difference to the generic collecting process can be summarized in the following way:

- collecting methods depend strongly on the group under examination
- they obtain collecting equipment already in the specification phase of the project (Determining Groups, Localities and Methods)
- Permissions must be obtained only for somewhat special localities and/or species, in such case always governmental institutions (ministries) are to be contacted
- In the labour division steps also deadlines, area and species are contractually laid down for the scientists
- Own accommodation room is used for the purpose of shipping room
- Import/export permissions are complicated, because plant transport is much strongly controlled in order to avoid the import of invasive or pathogen plants
- Collecting is carried out repeatedly only for monitoring activities, for inventorying never.
- Specimens are identified always in laboratory.
- No releasing of specimens, and no narcotizing is made
- Under an extra step (Others) the consideration of other conditions (public holidays, local customs and other unforeseeable circumstances is understood.

Dividing Labour

The division of labour includes also a contractual record of deadlines, area and species.

Obtaining Permission for collection by Groups

Obtaining permission is done in case of somewhat special species or special locality. The permission is applied from the relevant ministries.

Obtaining Collection Equipment

Since the actual collecting methods vary very strongly according to taxonomic groups, the actual collection equipment is obtained already in the specification phase of the project.

Malacology

The most important differences to the generic process are the following:

- The eventual locality of malacologist inventorying activities is set out only very roughly at home, the actual locality is determined always on the field.
- Verifying the collecting methods never happens
- Booking accommodation never happens
- Establishing an expedition room is meaningless
- Compilation of a sample design plan happens only very roughly in the preparational phase. It is finished only in the field.
- Compilation of progress charts by groups is meaningless

- Repeating cannot be made by a rule because the occurrence of specimens depends on fluctuations. Protocol is rather in the scientist's head, than written down. The work relies deeply on the experience and intuition of the scientist.
- Samples are never released.

Dividing Labour

The division of labour covers not only experts' tasks, but also all tasks, starting from helping tasks, and also those for which there are no special experts.

Producing Maps

As a subactivity of travel management, maps are also produced or obtained. For this end, literature is searched and GPS coordinates are also used.

Dipterology

The most important differences to the generic process are the following:

- Quantitative collection is never done, because it is quite hard in case of flies
- Setting up a database structure: the database of HNHM is used, no private database is set up
- There is no point of testing the collecting methods, because everything can change very quickly.
- No research team is built up because researchers work alone or at most in twine.
- There is no need to optimize institutional facilities, because of the lack of appropriate equipment, and because colleagues are not enough cooperative.
- Hiring of assistant workers is very rare: local information is collected from colleagues by email.
- There is no need to organize special transport for the collected material; in most cases some regular transportation means is perfect.
- Obtaining packaging materials: Massive steel boxes or glass tubes.
- Obtaining permission for import/export: only exporting permission is necessary.
- Repeated collection based on protocols: scientists return to the place, they were before.
- Examining and identifying the samples is made together, and to main purpose is to decide if they are to be released or not.
- No video recording takes place for documentation purposes.
- Data is recorded manually, except GPS coordinates, which are recorded electronically.

Annex 2: UML Basics

The structure of UML

The Unified Modelling Language (UML) is a general-purpose visual modelling language that is used to specify, visualize, construct and document the artefacts of a software system. It captures decisions and understanding about systems to be constructed. It is used to understand, design, browse, configure, maintain and control information about such systems.

UML captures information about the static structure and dynamic behaviour of a system. The static structure defines the kinds of objects important to a system and to its implementation, as well as the relationships among objects. The dynamic behaviour describes the history and the change of state of objects over the time, and also their communications among objects to accomplish certain goals.

The static model is presented in static views. The application concepts are modelled as classes, each of which describes objects that consists of information pieces, and communicates with other objects. The information pieces are called attributes, their behaviour are modelled as operations. Classes may share similar structure using generalization. A child class adds incremental structure and behaviour to that of being inherited from its parent class. Object-to-object relationships are modelled as associations between/among classes. The static view is notated using class diagrams.

The dynamic behaviour can be modelled in two ways. One is to describe its life history, as it interacts with the surrounding world. The other is the communication patterns among objects, as they interact to implement certain behaviour. To view an object in isolation is a state-machine – an object responds to certain events, based on their current state. They perform certain actions, and transitions to a new state. State machines are displayed in state-diagrams.

To view a system in an interacting way is possible through their collaboration. A collaboration diagram is a context-dependent view of objects and their links to each other. These links also describe the flow of message-passing mechanism among objects. In this way, data structure, control flow and data flow is unified. All of this is presented in collaboration and sequence diagrams. The behaviour of the system and their actors is depicted by a set of use cases.

Extensibility mechanisms. UML provides a limited extensibility mechanism. A stereotype defines a new kind of model element, with similar structure to the basic element, but having additional constraints, different interpretation, and a new iconic representation. Stereotypes are usually having also different code generation patterns and other back-end tools. A tagged value is a pair of strings, that can be attached to any kind of model element that holds arbitrary information, e.g. information for project management, code generator and/or required values for stereotypes. Both tag and value are represented as strings. Constraints are well-formed conditions that are represented as a string, attached to other model elements. UML has a constraint description language, called Object Constraint Language, (OCL).

Activity diagrams and their semantics

To describe the working process of taxonomist experts, the means of UML activity diagrams is used. UML activity diagrams consist of the following features:

- activity objects: (rounded boxes) describing certain activities. Activities may be primitive (steps) that cannot be broken up to parts, or they may be compound, that consists of a series of sub-activities.

- Object tags: activities may be extended by object tags, which are attached to certain activities. Special object tags are activity parameters. The direction of parameters can be determined only by the direction of their incoming and outgoing object flow arcs.
- Control flow arcs are arrows starting in one activity, and ending in another activity. Control flow arcs may attach guard conditions. If they exist, than the condition decides, if the prescribed change in activity is to be completed only if the guard condition evaluates to true.
- Object flow arcs that are heading out from object tags, and arriving to other object tags. As shorthand they may also connect object tags with activities. This is regarded as if there were object tags attached also to the activity side of arcs.
- Events: are practically starting points of activities, which trigger certain chain or subnetwork of the entire activity network.
- Activity calls: compound activities that are detailed elsewhere. The point of compound activities is twofold: 1. to simplify the object flow, while not loosing the details 2. to define reusable activity subnetworks. In general they act very similarly to subroutines or subprograms in traditional programming.

Beyond these there are still a few constructs to mention. These constructs are already to

- Forks that start new activity threads. Remark, that this construction is not new… the same effect can be reached by several outgoing tokens for any activity.
- Joins that synchronize and unify several activity threads. The same effect can be reached by several incoming tokens for any activity
- Decisions, where, depending on the result of a condition, one of two or more outgoing arcs are activated. The same effect can be reached by any activity with two or more outgoing conditional arcs, where the guarding conditions are expressed accordingly.

We say that along the arcs of activity networks, there are tokens traveling. An activity has incoming and outgoing arcs, both of control flows and object flows. If arcs are active, we say, they have tokens. If all the incoming arcs are active, we say the activity is firing. Upon firing the concrete actions connected with the activity are carried out. Having completed the actions of activities, tokens are multiplied, and are placed to the outgoing arcs of the activity. The exact scheduling policy of outgoing tokens is unspecified.

Annex 3: The Rational Unified Process (RUP)

Based upon the Unified Modelling Language, as a tool, different modelling methodologies can be developed. The most widespread such methodology and de facto standard is RUP itself. Although RUP gives proposals for most of the questions UML leaves open, still there is enough room for customizations, and for adaptation of the methodology to actual needs.

RUP has been developed mainly for actual software/application development; still, it is flexible enough to cope with the development such non-conventional software, like the artefacts of EDIT project.

RUP methodology allows also elaborating its steps not strictly sequentially – i.e. certain steps may also begin BEFORE previous steps had been declared 100% ready. In case of EDIT we shall rely on this feature very heavily.

Requirement management

Still before starting any modelling activities a series of requirement management steps are necessary. Requirement management comprises the following activities:

1. Developing a concept catalogue. The concept catalogue is in a first approach nothing else, than a dictionary, with or without explanation part, that must encompass all the concepts, the software must manage. The concepts are exclusively those of the business ….., that is they mustn't contain any implementation-specific, no IT specific concepts, unless the domain of the software to be developed is not IT industry itself. In case of EDIT the concept catalogue can be thought as an ever growing list. That is after new and new interviews the concept list may
2. Developing a requirement catalogue. Requirement catalogue is a list of written requirements, one sentence each, that is attached to modules/packages/use cases. A careful requirement management keeps track of implemented modules and fulfilled requirements in the form of a matrix.

Business Modelling

Business Modelling means in general, that only the externally visible and distinguishable system components, actors and concepts are modelled. The main steps of business modelling are the following:

- Identifying Business Actors: Anything, which lies outside of the system, is an Actor. Business Actors are mainly people, but even external electronical or software services are also actors. Actors mean a mainly role, that is a physical person may be assigned also several different roles to become an actor.
- Identifying and Specifying Business Use Cases.
- Defining a domain model. The domain model contains all necessary information that is captured by the software/model. It is basically an ontology, representing the concepts of the captured problem domain.

By modeling inventorying workflow activities the following questions must have been taken into account:

1. The inventory and revisionary taxonomic activities must be modeled in a uniform way, in a unified model. It is essential, that the two research paradigms are modeled according to the same approach, and they should be described in the same modeling framework, and possibly share common activities or other features.
2. One model – several views approach. The model should be stored in a uniform and unified database, and it should describe everything down to the smallest details of the matter. Views serve to project certain features of the one and only model. Modelers must make as many views, as it is necessary to present all significant features of the model.
3. Modeling must be made, in terms of granularity, in a multiple level way. On one extreme hand a view should present the most distilled, the most general features of the matter, while on the other hand another view should go into the deepest details of the matter.