MODERN ECOLOGY REQUIRES A STATE-OF-THE-ART DATA ACQUISITION AND MANAGEMENT METHODOLOGY

Submitted by Swiss Federal Research Institute WSL, Switzerland[1]

Contributed paper

ABSTRACT

In today's modern ecology research, the volume of the acquired data is comparable with the volume of data in modern businesses. Likewise, financial resources in ecological research are at least as scarce as in other enterprises. Therefore, it is an absolute necessity to implement efficient workflow for everyday data acquisition and data management. Analysing, planning and implementing this workflow affects IT and the organisational structure of the project and must be done before collecting data.

In this paper, a few activities within one of the LWF (Long-term Forest Ecosystem Research) sub-projects are analysed. The mission of LWF is to improve our understanding of how natural and anthropogenic stresses affect forests in the long term, and what risks for humans are involved. An optimised and a non-optimised version of these activities is described using Activity Diagrams out of the UML (Unified Modelling Language) methodology. This technique is used to analyse and optimise workflow within the LWF project. The Activity Diagrams serve as a basis for information and encourage discussions among project members.

Because data is collected in the field it is an additional challenge to implement workflow to transfer consistent and verified data from the field to the research site. Especially harmful in this context are so-called "media breaks". They occur, for example, when written field data on paper has to be transferred to digital media, i.e. a personal computer. Integrated data acquisition systems help to avoid media breaks. It is important that data is verified in situ, because the information necessary to correct the data will not be available later. The tight integration of the GIS and database in the form of the SDE (Spatial Database Engine) (ESRI 1998) also allows the creation of a more efficient workflow and leads to more consistent data. The comparison between an optimised and a non-optimised workflow clearly shows that integrated systems not only improve data consistency but also reduce necessary job steps and hours of work.

I.　　INTRODUCTION

1.　　The volume of acquired data in today's ecological research is comparable with the volume of data in modern businesses. Likewise, financial resources in ecological research are at least as scarce as in such enterprises. Therefore it is an absolute necessity to implement efficient workflow for everyday data acquisition and data management. As an example of a complex ecological long-term study, the LWF-

---

[1]　　Prepared by Peter Jakob, Christian Ginzler, Norbert Kräuchi and Andri Baltensweiler.

project was chosen, existing workflow was analysed and compared to optimised workflow. LWF stands for long-term (>30yrs) ecological research. Its mission is to improve our understanding of how natural and anthropogenic stresses affect forest ecosystems in the long term, and what risks for humans are involved.

II. METHODOLOGY

2. We used the so-called UML (Unified Modelling Language) methodology (Fowler et al., 1997) to compare activities differing in their degree of optimisation by drawing Activity Diagrams. These Activity Diagrams structure workflow information and encourage discussion among project members. The analysis, planning and implementation of workflow affects the IT and organisational structure of an entire project and should therefore be carried out before any data collection occurs. As most of the data is collected in the field, it is an additional challenge to implement a workflow which can transfer consistent and validated data from the field to the laboratory. Data must be verified in situ, because it will not be possible to make any corrections at a later stage since the necessary information will no longer be available.

III. ACTIVITY DESCRIPTIONS

III.1 Survey of Monitored Objects

3. A lot of research topics within the framework of the LWF Project are space-related. That is the reason why all the monitored objects (e.g. trees, weather stations, catchments) have to be surveyed. LWF is a long-term project so that, periodically, new objects (e.g. growing trees, new catchments) have to be surveyed. In the initial phase of the LWF Project, surveying was a very time-consuming activity whereas now, with all of the 15 research plots (with up to 3,000 objects each) well-established and surveyed, it is an activity of minor importance.

III.1.1 Non-Optimised Workflow

Data Capture in the Field (Activity 1 and 2)

4. The coordinates of the monitored objects on each research plot are measured with a Tachymat. A numerical identifier of the surveyed object has to be entered and is stored together with the coordinates in the memory module of the Tachymat. The numerical identifier serves not only as a unique identifier for each object but also helps to classify the object into different categories (e.g. trees, weather stations, precipitation catchments). For this purpose each category has a specified numerical range for those identifiers. The type of Tachymat (Model TC 1000) that was used for the survey is not capable of storing any alphanumeric or additional information (e.g. species of tree). Additional properties of the surveyed objects have to be collected using a different field-computing device and are therefore not part of this activity. Remarks concerning a specific surveyed object are entered by hand into a field journal.

Data Editing in the Office (Activity 3 and 4)

5. The captured data has to be transformed into the Swiss Coordinate-System. The data is transferred to a PC and the software "STRATIS" performs this transformation. The additional information from the field journal is added (partly manually) to the resulting ASCII-file on the PC.

Data Storage in the Database (Activity 5 and 6)

6. The ASCII-file is read into an intermediate table in the Oracle Database. A second program performs plausibility tests (all the objects have to be located within a certain range from the centre of the research plot) and puts the data in the final tables within the database. Some monitored objects (mostly

trees) may already exist in the database. In this case the correct co-ordinates are automatically assigned to them.

<u>III.1.2 Optimised Workflow</u>

Data Capture in the Field (Activity 1)

7.   The coordinates of the monitored objects on each research plot are measured with an enhanced model of a Tachymat. Using a mapping software and a field-computer all objects are surveyed in the Swiss Coordinate System in the field. An identifier (numerical or alphanumerical) of the surveyed object has to be entered and is stored together with the coordinates in a field computer. To classify the object into a specific category (e.g. tree, weather station, precipitation catchment), an additional identifier is used. Additional properties (e.g. tree species) or remarks concerning a specific surveyed object may also be stored. In most cases the use of a GPS is problematic because trees and less frequently the landscape (e.g. narrow valleys) make it difficult to receive proper satellite signals.

Data Editing in the Office (Activity 2)

8.   No data editing in the office is required.

Data Storage in the Database (Activity 3 and 4)

9.   The data (ESRI shape-files) from the mapping software can be imported directly into the database. Using the SDE (Spatial Database Engine) it is guaranteed that every attribute has its geographic feature and vice versa.

III.2   Create Point and Polygon Cover

10.   The coordinates of all surveyed objects are stored as simple records in the database. There is no GIS functionality associated with this data. To perform analyses, coordinates need to be extracted from the database to create a GIS topology.

<u>3.2.1   Non-Optimised Workflow</u>

Building Topology in the GIS (Activities 1,2,3)

11.   The coordinates that represent single point features and those that represent polygons are known from the database. Coordinates and primary-key items are exported from the database and then the topology is built.

*Assigning "Business" – Data (Activities 4,5,6,7)*

12.   The expression "business"-data has its origins in the business world and simply means data that has been stored in databases for years (e.g. sales-data for different products in different areas). In our case tree data is treated as "business"-data. Feature-related data is assigned from the database to the point and polygon cover using the primary-key. Consequently, "business-data" is stored twice: once in the database and once again in the GIS. The potential for inconsistency is obvious.

*Plausibility Testing (Activity 8 and 9)*

13.   Plausibility tests are performed in the GIS. The point features of two trees cannot be too close together. Another possible error occurs when a sub-area overshoots the main area. Such questionable points are extracted and have to be surveyed and checked in the field.

Storing of Topology in the GIS (Activity 10 and 11)

14.   All features which have passed the plausibility tests are stored in the GIS.
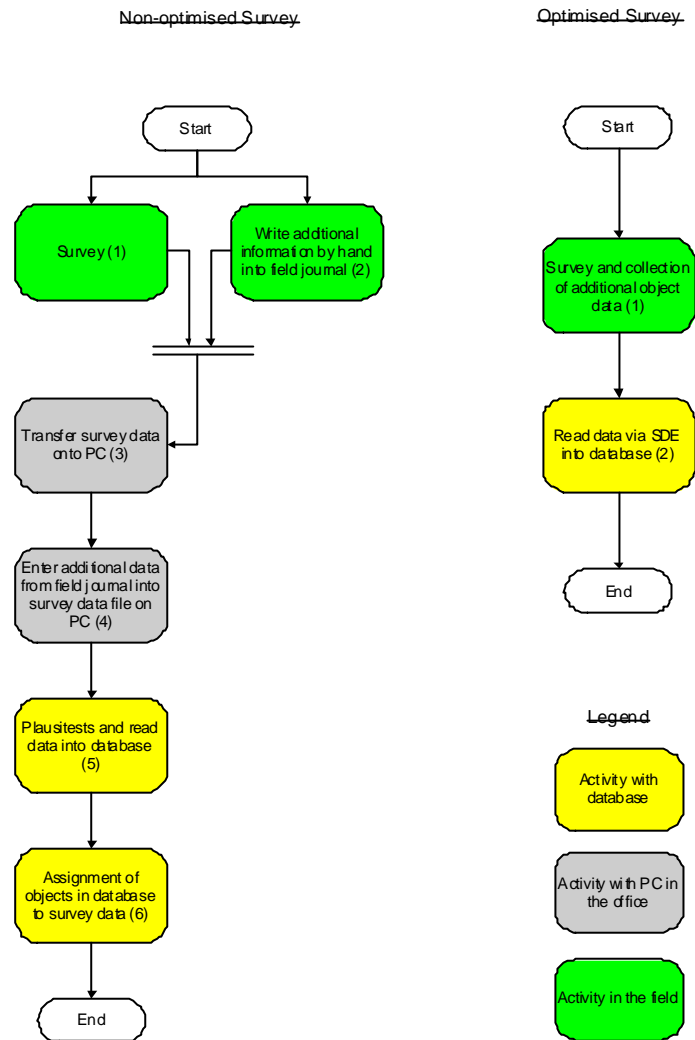
Fig. 1: Activity Diagram "Survey". In the optimised form the number of the activities involved is smaller.

Create Point and Polygon Cover (non optimised)
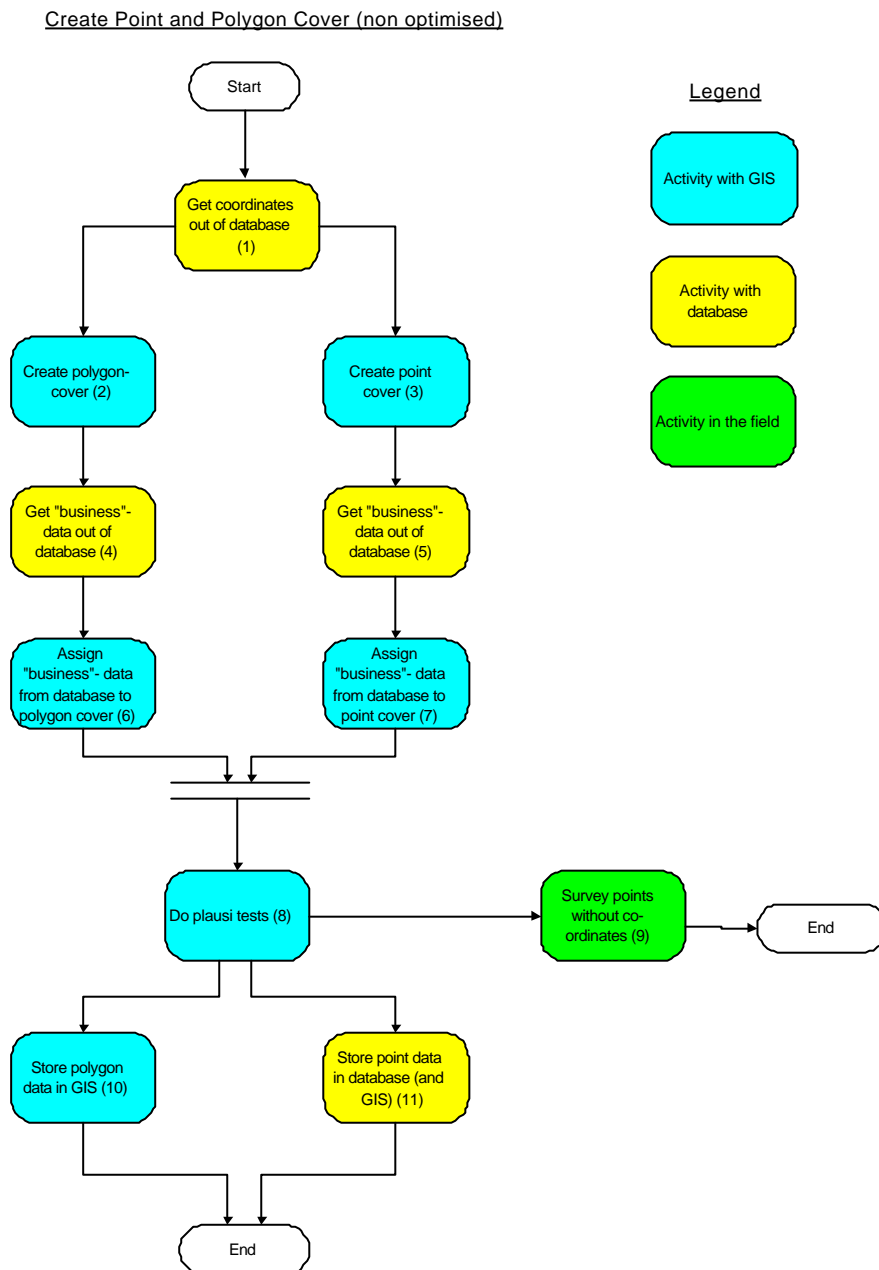
Fig. 2: Activity Diagram "Create Point and Polygon Cover" in the non-optimised form only. By using the SDE the whole activity becomes obsolete because GIS data is integrated in the database.

III.2.2 Optimised Workflow

15.   In an optimised workflow there is no difference between GIS data (formerly topology) and "business-data" because all data is stored in the one database using the SDE. Plausibility tests can be performed much more easily, consistency is guaranteed and all data is stored only once.

III.3   Periodical Tree Inventory

16.   One of the main sources of information within the framework of the LWF Project is periodical tree inventories where certain properties of the same trees are re-evaluated within a period of one or more years. The people who evaluate the trees in the field need some tree data from earlier inventories in order to locate the trees. In addition, certain properties of a tree do not change over time (i.e. species of a tree).

Field computing devices have been used for data-collection since the beginning of the inventories in 1985. The use of such devices has the advantages that no media breaks occur and that plausibility checks can be done in the field, when correcting data is still relatively easy.

III.3.1   Non-Optimised Workflow

Select Trees on Sub-Area of Research Plot (Activity 1)

17.   Tree inventories are carried out periodically but not always on the same sub-area of the research plot. In one year all the trees on a research plot are evaluated whereas in the next year only the trees on a sub-area are dealt with. With the help of the GIS the trees within the sub-plot are defined. The result of this process is a file of tree numbers containing all trees located in the relevant areas of each research plot.

Select additional Tree Data in Database (Activity 2)

18.   The file from Activity 1 serves as a basis for adding additional data from the database to the tree data. This additional information (i.e. species, diameter) facilitates the location of trees on the research plot and permits the use of more specific plausibility checks on the data entered in the field. All data is stored in text files on ram cards.

Data Collection in the Field (Activity 3)

19.   All the trees are numbered and the field group locates trees with the help of a map of the research plot. The map contains all the numbered trees and any other research installations. Tree properties must always be evaluated from the same position.

Transfer Data into Database (Activity 4)

20.   The data collected on ram cards is sent or brought to the research institute and then transferred into the database.

III.3.2   Optimised Workflow

Select Trees via the SDE (Spatial Data Engine) on Sub-Area of Research Plot (Activity 1)

21.   The SDE allows the storage and integration of GIS data together with other "business"-data in a relational database (e.g. Oracle). With the SDE it is possible to select the appropriate tree in a specific area of a research plot in a single job-step or activity. This considerably reduces the amount of work involved in preparing the field data. The sources for potential errors in this activity are also reduced and the quality of the data is improved.

Data Collection in the Field (Activity 2)

22.   There is no difference in this activity between the optimised and the non-optimised version of the workflow. The computing device used in the field is an old (ca. 8 years) relatively inexpensive (US$ 600) model of a DOS Palmtop from Hewlett Packard (HP 200LX). The software was produced by the US Forestry service about ten years ago. The highly generic approach in the design of the software allows its easy adaptation to the specific requirements of a tree inventory. An optimised workflow does not necessarily lead to expensive investments in integrated instruments and/or computer hardware. Careful evaluation of the requirements can result in surprisingly cost-efficient and appropriate solutions.

Transfer Data into Database (Activity 3)

23.   There is no difference to Activity 4 of the non-optimised workflow: the data collected on ram cards is transferred at the research institute into the database.
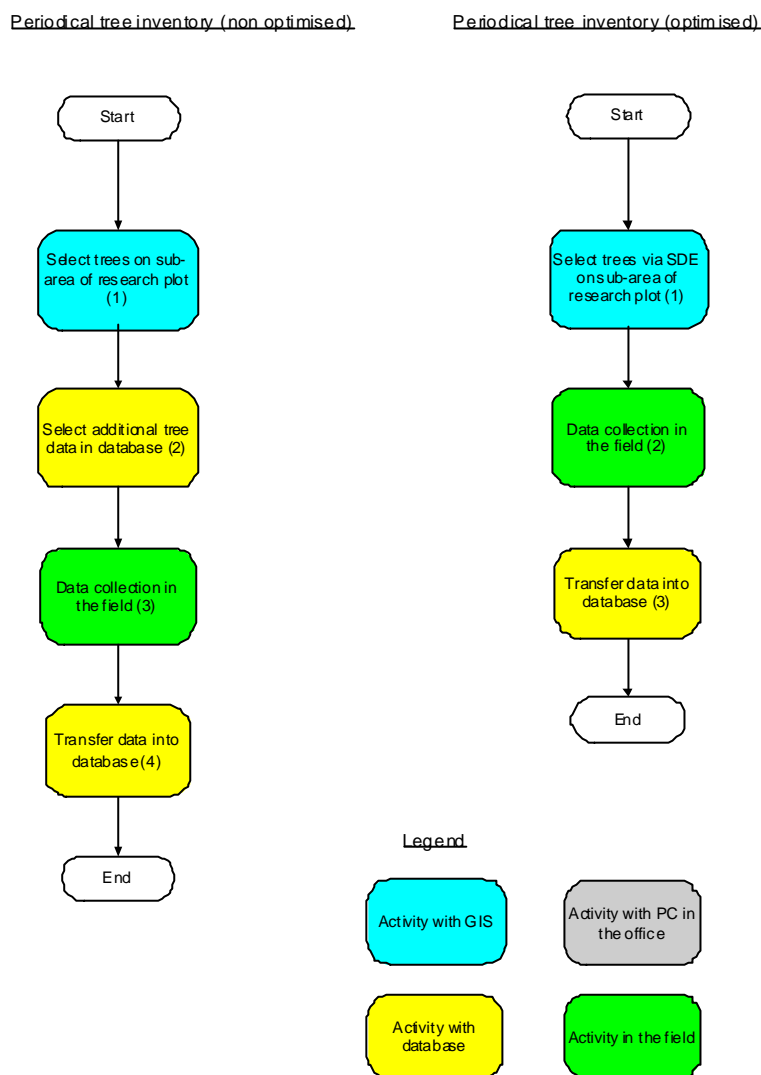

Fig. 3: Activity Diagram "Periodical Tree Inventory". The use of the SDE reduces the number of activities involved and the overall time spent on them.

IV.   DISCUSSION

24.   Especially critical in this context are the so-called "media breaks" which occur, for example, when field data that was recorded on paper has to be transferred to digital media, i.e. a personal computer. One appropriate solution is the use of integrated systems which eliminate media breaks. In the LWF sub-projects, the tight integration of the GIS and the database in the form of the Spatial Data Engine (SDE) allows the creation of a more efficient workflow. The comparison between an optimised and a non-optimised workflow clearly shows that integrated systems not only improve data consistency but also reduce the required number of job steps and hours of work. The implementation of an optimised workflow often goes hand in hand with a replacement of manpower costs by capital investments in integrated systems.

V.   LITERATURE

Fowler, Martin; Kendall, Scott 1997: UML Distilled. Applying the standard object modelling language, Addison Wesley.

ESRI, 1998. Spatial Database Engine. ESRI White Paper, Environmental Systems Research Institute, Inc., Redlands, CA.

Long-term forest ecosystem research LWF, 2000: http://www.wsl.ch/forest/risks/lwf/lwfintro-en.ehtml.