Computers and Electronics in Agriculture 132 (2017) 97–107

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

The data storage and analysis system of the Swiss National Forest Inventory

Berthold Traub *, Rolf Meile, Simon Speich, Erik Rösler

Swiss Federal Research Institute WSL, Zürcherstrasse 111, CH-8903 Birmensdorf, Switzerland

ARTICLE INFO

Article history:
Received 19 September 2016
Received in revised form 18 November 2016
Accepted 19 November 2016
Available online 18 December 2016

Keywords:
Forest inventory
Data management
Forest population estimates
Web-based forest information system
Data warehousing

ABSTRACT

Many countries conduct national forest inventories and collect data from different sources such as field sample plots or aerial imagery. The sustainable management of this data requires appropriate systems for the storage and dissemination of this data to a variety of stakeholders. The Swiss national forest inventory started with surveys in 1982 and developed the NAFIDAS system which facilitates highly metadata controlled management and analysis of sample based forest inventory data. Public users are provided with guided access to a comprehensive set of tables and interactive maps on forest statistics. A very user-friendly search and filtering system supports the querying of results. NAFIDAS has proven to be a performant solution which may serve as a blueprint for the complete data processing chain of sample based forest inventories starting with the upload of data and ending with the provision of a huge variety of inventory tables and maps on the internet.

1. Introduction

Many countries have established forest monitoring programs and conduct sample based multiresource forest inventories on national and also on regional level (Kangas and Maltamo, 2006; Köhl et al., 2006; Tomppo et al., 2010). In some countries they have started already more than 100 years ago (Fridman et al., 2014). Collecting data for these inventories is most often based on resource intensive field work and the data collected are an asset of increasing value which requires secure systems for management and storage.

The institutions conducting the national forest inventory (NFI) are usually obliged to regularly report forest statistics to national bodies and also to international protocols and projects such as the Forest Resources Assessment (FRA) (FAO, 2015; Keenan et al., 2015; MacDicken, 2015), the UN Framework Convention on Climate Change (UNFCCC) and its Kyoto Protocol or the Ministerial Conference on the Protection of Forests in Europe (FOREST EUROPE, 2015). Many efforts have been made to harmonize this data in order to ease the integration of country statistics to international reports on forest resources (Vidal et al., 2008; Gschwantner et al., 2009; Gabler et al., 2012).

NFI institutions may also be obliged to provide public users with appropriately prepared NFI data. This user group is heterogeneous in terms of the reasons of their interest in this data and in their knowledge of methods applied in sample based forest inventories. To reach this group at a broad level, easily accessible and user friendly interfaces are needed. Websites have evolved as a standard tool for this purpose; meanwhile many countries provide access to inventory tables and maps or even raw data through web based systems. Examples from North America are the Canadian NFI which publishes standard reports on the internet (https://nfi.nfis.org/en/standardreports) and the Forest Inventory and Analysis (FIA) Program of the U.S Forest Service which offers several online tools and entry points to forest inventory data (http://www.fia.fs.fed.us/tools-data/). A comprehensive overview on European NFIs is provided by the European Forest Information Portal (EUROFOREST, http://forestportal.efi.int/), maintained by the European Forest Institute (EFI, http://www.efi.int/portal/). This service aims to provide an entry point to current information on European forests and the forest sector including links to the national forest inventory websites such as the website of the Swiss NFI which offers various types of data on Swiss forests and information about methods on http://www.lfi.ch/index-en.php. On European level plans for a European Forest Information System (EFIS) existed (Schuck et al., 2005) which has moved to an information system on forest fires (EFFIS) (San-Miguel-Ayanz et al., 2013).

The correct utilization and interpretation of NFI data requires a thorough understanding of the underlying methods, the inventory design, data sources and data acquisition techniques. Harmonized information on forest inventory methods applied is available from

* Corresponding author.
E-mail address: berthold.traub@wsl.ch (B. Traub).

http://dx.doi.org/10.1016/j.compag.2016.11.016
0168-1699/

Tomppo et al., 2010, which also allows comparisons between countries. Further country specific information is available e.g. for the United States (O'Connell et al., 2016), Canada (Gillis et al., 2005; Wulder et al., 2008), Switzerland (Brassel and Lischke, 2001; Lanz et al., 2010), Finland (Tomppo, 2006) and Sweden (Fridman et al., 2014).

Many of these publications on NFI methods however lack detailed information on the database and data analysis systems utilized. To address this gap we describe the design and architecture of the NAtional Forest Inventory And Analysis System (NAFIDAS) developed for the Swiss NFI. NAFIDAS is a fully automated web based system which has been established to store, analyze and to disseminate Swiss NFI data. Such a system requires

- efficient storage techniques and sustainable organization of long-term and large-scale inventory data,
- reliable, reproducible and transparent data analysis,
- as well as functionalities to provide data to a broad user community in an understandable, reproducible and timely manner.

The crucial issue for the NAFIDAS development was how those requirements could be integrated into one system. Since there are many ways to store data collected from forest inventories, there is no simple conclusion which way is optimal. Efficiency and quality requirements demand a flexible, sustainable and well maintainable system. However, users are primarily interested in having intuitive access to the forest statistics accompanied by additional information allowing for proper scientific interpretation. The current NAFIDAS integrates the complete data processing chain from data upload through data management and data analysis from sample based inventories. Particularly due to its integrated flexible and user-friendly solution to disseminate a huge variety of inventory tables and maps on the internet it could serve as a general blueprint for the development of data management and data processing systems applied in forest inventory.

2. Material and methods

2.1. Development of NAFIDAS

The Swiss NFI spans three decades and almost four inventory cycles, starting as early as 1982 and is close to start its fifth cycle in 2018. In the first Swiss NFI (1983–1985) (Bachofen et al., 1988), data storage and analysis systems were completely separated and any access to data in electronic format was restricted to staff in-house. Experience with computer infrastructure and advanced knowledge of the NFI methodology was a prerequisite to work with the system. The system used to analyze data from the second NFI (1993–1995) (Brassel and Brändli, 1999) already offered a graphical user interface to define statistical analysis. It allowed direct access to raw and derived[1] data from various data sources (Traub and Schnellbächer, 2001). However, computer skills were still an important requirement to work with the system.

The web based access to the database and data analysis system was a qualitative milestone reached at the end of the third NFI (2004–2006) (Brändli, 2010). For the first time, the system could be accessed in the intranet and facilitated metadata controlled data management and data analysis processes (Böhl and Bierer, 2004). A huge set of inventory tables and interactive maps were published on the internet in the four languages German, French, Italian, and English and with unlimited access for public users (Speich and Meile, 2013). The decision to distribute forest statistics over the internet was mainly driven by the ambition to quickly supply several thousand tables and maps to a broad audience based on the most current data at comparably low cost. This approach required a completely reengineered data model.

The current NAFIDAS has reached a stable state after major revisions and improvements have been made to all parts of the system since the beginning of the 4th cycle (2009–2017) (Abegg et al., 2014). Particularly, the web applications were significantly extended and the application used for data analysis was completely restructured. Other revisions aimed at increasing efficiency of data processing flows, quality and security of the system, which mostly involved the comprehensive use of metadata. Currently, about 40 internal WSL users are registered to the system, 10 out of them work permanently with the system.

2.2. System overview

The NAFIDAS design is generally comparable to a data warehouse architecture (Inmon, 2005; Kimball and Ross, 2013), thus according terminology and concepts will be used in the following. Particularly the 'dependent data mart and operational data store architecture' as described in Hoffer et al., 2011 (p. 422) is very close to that of NAFIDAS (cf. Fig. 1).

It consists of an operational data store which has interfaces to source data systems (internal and external data sources), a data and metadata storage area and end user presentation tools. Similar architecture was applied in the Swedish NFI (Fridman et al., 2014). The INARIS data warehouse for the agricultural data of India has similar objectives but is more complex in terms of architecture and ownership of data (Nilakanta et al., 2008). The operational data store and the enterprise data warehouse (EDW) build the main data storage units. Processing the data mainly involves transformation, standardization and matching steps and if necessary also cleaning steps. From the operational data store data are loaded into the EDW where fact data (numerical, continuously valued and additive measurements) are kept together with dimensional data (usually used to qualify, categorize, or summarize facts in queries, reports, or graphs) (Hoffer et al., 2011; Kimball and Ross, 2013). The end user presentation tools comprise the internal web application which steers all Extraction, Transform and Load (ETL) management and analysis processes and the external web application which allows access to the physical data mart (PDM) on the internet.

2.3. Data stores

2.3.1. Operational data store

The operational data store (cf. Fig. 1) enables the efficient storage of all NAFIDAS relevant raw data coming from internal and external data sources. It is built on a relational data model with high level of normalization which ensures referential integrity and atomicity. Furthermore, a sustainable logic with strict rules for input, storage and output of data was implemented. Categorical values are modelled in separate entities with referential constraints. This allows input validation as well as handing out presetting data for software applications. Furthermore, specific XML file exchange formats are defined in order to control the quality of the data flow including quality checks. Access to this database is limited to specially trained staff.

Tree data collected from terrestrial field plots build the vast amount of source data, followed by data from aerial photography interpretation, GIS data and records from forest service questionnaires (Brassel and Lischke, 2001). Spatial vector datasets are an additional type of source data persistently stored in the operational data store. They originate from earlier NFI cycles and also from different sources like the Swiss national government or the Swiss cantons. This vector data are used for the estimation of population

[1] Modelled, transformed.

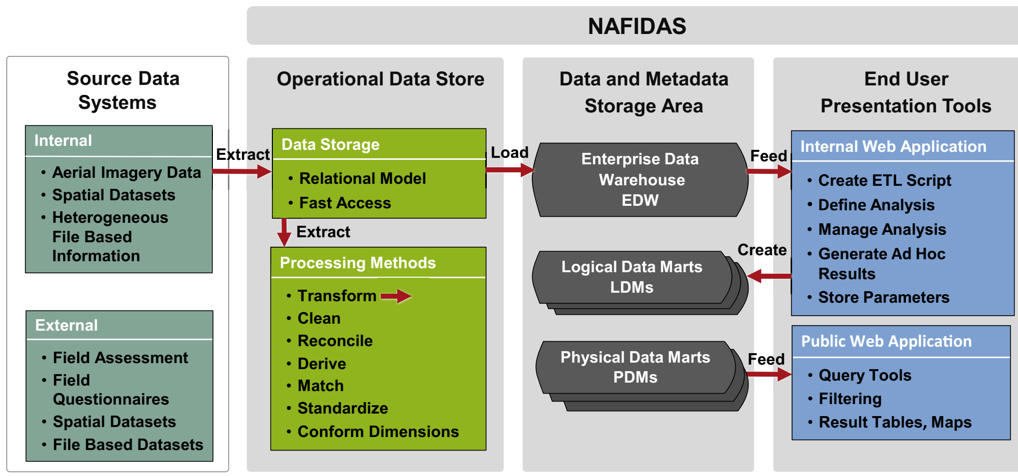

Fig. 1. Architecture of the NAFIDAS (EDW: Enterprise Data Warehouse, PDM: Physical Data Mart, LDM Logical Data Mart). The system includes the three main components: Operational Data Store, Data and Metadata Storage Area and End User Presentation Tools (adapted according to 'dependent data marts and operational data store architecture' as described in Hoffer et al., 2011).

totals, forest road survey analysis as well as for map production. This composition of several data types including GIS data is typical for multiresource inventories. Arvanitis et al., 2000 have described a system based on relational databases and a web browser environment. The architecture is quite similar to NAFIDAS and was used for the same reasons as: reducing data redundancy, raising performance efficiency and assuring data integrity.

2.3.2. Data and metadata storage area

The EDW is the control point and single version of the truth for data mart creation. It consists mainly of two type of tables: Analysis tables which are a combination of fact and dimensional tables (Kimball and Ross, 2013); and formalized metadata which can be roughly divided into data metadata, code tables and administrative metadata. The EDW is exclusively loaded by ETL scripts which read data from the operational data store and from the EDW itself. Direct access to the EDW and all data marts is restricted to internal users and controlled by user authentication. With focus on the data processing chain, the EDW is positioned between the operational data store and the internal web application. Data analysis scripts (cf. Section 2.4.2) in contrast to the ETL scripts retrieve data exclusively from the EDW, i.e. the operational data store will never be accessed in data analysis processing. Thus all data relevant for data analysis must be stored in the EDW, even in case that raw data could be directly used without transformation. This is a very important measure to optimize transparency and reproducibility of analysis results.

The logical data marts are derived from the EDW and are the basis for data analysis and especially for the calculation of forest population estimates and their sampling errors based on the statistical methods of the Swiss NFI (Kaufmann, 2001; Köhl, 2001). The physical data marts accessible via the public website contain a dedicated set of calculated estimates, dimensional variables and metadata. Unlike a typical data mart or OnLine Analytical Processing (OLAP) cube (Hoffer et al., 2011; Kimball and Ross, 2013), these tables contain estimates, which are derived from complex statistical analysis rather than just aggregated, i.e. summarized fact data.

All analysis tables of the EDW are strictly defined and reflect a semantical concept like 'recorded plots' or 'recorded trees'. They contain several variables, which may refer to fact data and to dimensional data. Each variable is uniquely defined by the kind of information to be stored and is assigned to exactly one column. It is indispensable that the definition of variables must be concise and applicable to all subjects (rows) of the analysis table e.g. to all trees ever stored. This implies that the variable definition and thus the technical column specification (such as type etc.) is inventory-independent and thus forms the basis of a generalized data model for forest inventory data sampled at successive occasions. The term 'inventory' specifies the sample plots and the set of fact and dimensional variables which are used for a certain type of data analysis. An inventory may define data from two inventory cycles (such as for change estimation) or may refer only to a subset of sample plots of one certain inventory cycle. The variable 'inventory ID' uniquely represents each inventory. The issues of developing a generic data model for longitudinal forest resource data was also addressed by Rasinmäki (2009) who developed an object oriented generic data repository.

The specific level of granularity of the EDW analysis tables is defined by the primary key spanning those columns that identify the records stored (cf. Fig. 2, left part). The most important key is the inventory ID. Three main levels of granularity can be distinguished:

- Level 1: The granularity is determined by the geographical location of the sample plots. The inventory ID (InvID) is set to zero, thus these tables contain one row per PlotID which identifies the spatial position represented by x- and y-coordinates. The most important table (CLA) encloses several regional dimensions as well as geometrical sample grid information. This table contains predominantly dimensional data.
- Level 2: This level extends level 1 by a distinct inventory ID. Level 2 tables have one row per PlotID and per InvID. For the purpose of analyzing change estimates, special inventory IDs combining two inventory cycles are used. The two major analysis tables contain predominantly dimensional data collected from terrestrial sample plots (WAA) and from aerial photography interpretation (LBAUFNA). Some few fact data such as information on forest type are available as well.
- Level 3: The granularity is based on level 2 extended by information on multiple individuals per plot such as single trees, pieces of deadwood, assortments etc. The most important table (BAA) has one row per PlotID, per InvID and per individual subject ID. It contains the majority of factual tree data such as tree diameter and transformed data such as tree volume combined with a large set of dimensional data such as tree species group, etc. This approach supports the creation of logical data marts

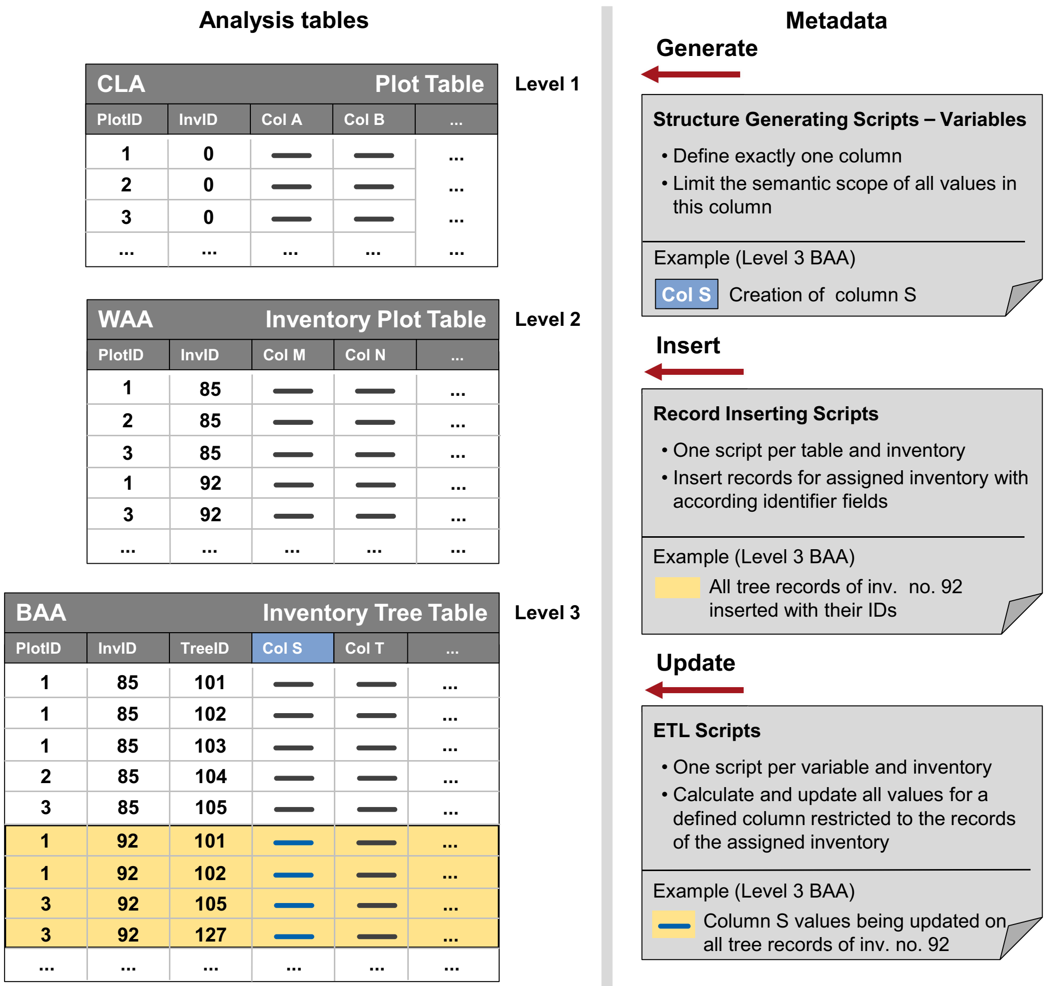

Fig. 2. On the left, the three most important EDW analysis tables with their different levels of granularity are illustrated. On the right, the essential types of EDW data definition management scripts are explained. Each type of script applies to all types and levels of analysis tables; the content of scripts is stored in metadata and will be executed as functions.

significantly since the analysis algorithms used for calculation of population estimates can directly access tree volume figures without applying complex tree volume calculation routines during run-time.

The EDW data definition management (Fig. 2, right part) involves three types of scripts:

- Structure-generating scripts which create new columns. Example: In a first step a variable is semantically specified as tree diameter of a tally tree in centimeters and column S is created by means of its structure-generating script on table BAA in order to contain integer values of this variable.
- Record-inserting scripts which fill the EDW tables with empty records defined for a certain inventory. Example (continued): In a following step all tree records assigned to inventory ID equal 92 are being inserted into table BAA by the record-inserting script and populated with the necessary key IDs.
- ETL scripts which are specific to variables and inventory IDs (one script per variable and per inventory ID). They load data into exactly one column of an analysis table and update only the rows of the according inventory ID. Data sources are restricted to the operational data store or analysis table of the EDW. Any transformation steps are applied to the extracted data on the fly before uploading. ETL scripts are primarily based on the Structured Query Language (SQL). Example (continued): Finally, column S is updated according to the specifications of its tree-diameter ETL script for all trees where inventory equals 92.

ETL scripts are stored entirely in the administrative metadata tables, every single EDW data management process starts by reading the script code followed by its execution.

It must be emphasized that ETL scripts will never change the structure of the EDW or EDW tables. This strict concept protects the database structure and guarantees data consistency while at the same time allowing for a fully automated population of the EDW. The transformation algorithms are developed by scientific staff that have special skills in the subject matter and are responsible for the subject-specific content. The technical implementation and computation is done by staff specialized in writing the ETL scripts. Logical cross checks are an important quality assurance measure which will be discussed before the transformed data will be productively generated.

The EDW must be updated regularly by means of automated ETL processing to reflect upload of new raw data and changes in the operational database. To sustain consistency, all ETL scripts will be handled in a certain sequence defined by their hierarchical dependency i.e. which ETL script requires input data from other ETL scripts. The dependency vectors are saved for each ETL script and since the algorithms and the calculated data are deterministic, a dependency tree can be built. This tree always represents a Directed Acyclic Graph (DAG) (Thulasiraman and Swamy, 1992, pp. 97–125) which allows for an automated update of the correspondent variables in the analysis tables.

Snapshots of the EDW are taken on a monthly basis. Continuous archiving is applied to ETL scripts and most relevant metadata, thus any changes are registered in the system. Regular information about this data can be queried for any point of time starting from 2012, when the feature was activated.

2.4. Web applications

The end user presentation tools consist of graphical user interfaces (GUI) which offer access to the NFI data by means of web applications hosted on two different websites: The web application on the internal website allows complete access to the EDW and to the full set of management and analysis tools. The web application on the public website offers querying of the physical data marts and provides a comprehensive set of Swiss NFI data (physical data marts) along with other general information such as publications and a glossary without any access limitations other than an internet connection and a current web browser. Public users neither need to register for an account nor install any further software on their computers. The user interface to query and display forest statistics follows the principles of an online shop. It uses filtering to limit the possible set of results, paging to display them and a temporary shopping cart to accumulate potentially interesting tables and maps. These conventions are generally known to the user from visiting websites such as Amazon or eBay.

2.4.1. Processes and data flows

Three main groups of NAFIDAS processes controlled by the internal web application may be distinguished as illustrated in Fig. 3:

(1) ETL-processing and also ETL script management which leads to a high degree of automation and transparency. (a) Extract data from raw database, (b) transform and load data to the analysis table(s) in the EDW and save according process metadata.
(2) Real-time analysis of NFI data (Data Analysis Application). Forest population estimates can be calculated ad hoc and are displayed in tables and maps in the web browser or could be saved to an XML file. (c) Define analysis and post parameters to the data analysis application, (d) import data to create a specific logical data mart, (e) send estimates as HTML table to the web application. Details are described in Section 2.4.2.
(3) Data publishing process. (f) Save the physical data mart including the defining parameter set, (g) visualize saved tables, (h) aggregate saved information for displaying on web pages.

Other tools of the internal web application provide a wide range of information about operational and EDW data and full access to all kinds of metadata including the documentation of the entire system.

2.4.2. Calculation of estimates

Statistically correct calculation of estimates requires the logical correctness of the parameter combination that defines the analysis. This is controlled by the EDW data model and metadata in such a way that each selected parameter defines a restricted set of remaining parameters to be offered.

The data analysis definition process is illustrated in Fig. 3c–e. It starts with the definition of the inventory by setting the inventory ID, which basically defines the logical data mart to be created during analysis computation. After that, the user is guided through further steps, where all relevant parameters such as state or change analysis, type of population estimates and other dimensional variables are determined. He can further decide whether the statistics are calculated as a total value (such as growing stock) or as ratio estimator (such as growing stock per ha of forest area) (Kaufmann, 2001). Moreover, various types of restrictions can be set to customize contents and layout of the resulting tables and maps.

The calculation of estimates starts in the background after all required parameters have been selected. First, all relevant analysis tables of the EDW are combined into a logical data mart based on the input parameters. This data mart contains all factual and dimensional data relevant for the analysis. Next, an aggregated multidimensional view, comparable to an OLAP cube, is created. This view contains the calculated forest population estimates and standard errors (Köhl, 2001) on all aggregation levels determined by the number of the dimensional attributes selected. The final tables are posted back as HTML within seconds.

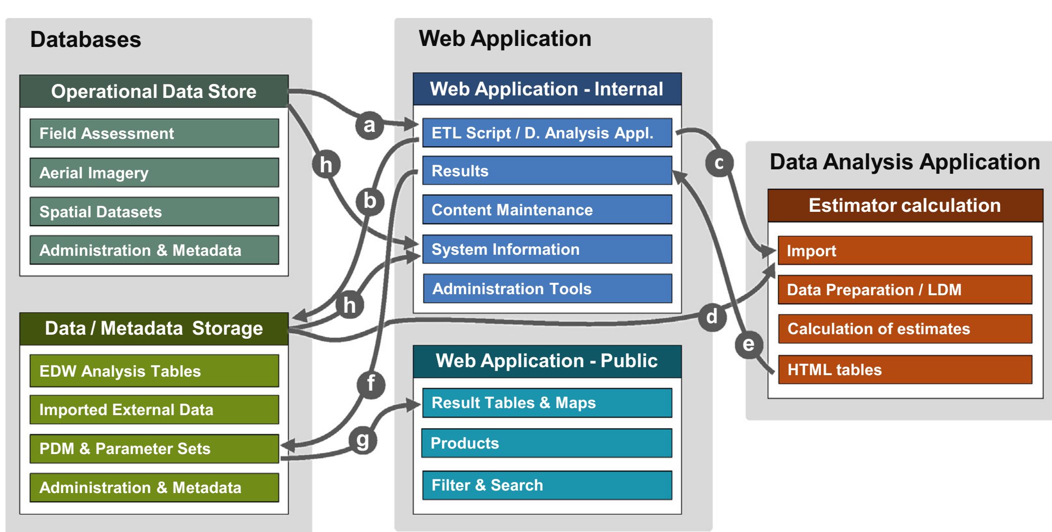

Fig. 3. NAFIDAS processes and data flows (EDW: Enterprise Data Warehouse, PDM: Physical data mart, LDM: logical data mart, ETL: Extract Transform and Load).

The data analysis scripts were designed as strictly generic and modular code generators which translate the source code into executable code based on the parameter set and other metadata from the EDW. This approach works without any user interaction and allows for high flexibility and complete control of the analysis by the parameter set and EDW metadata. For the sake of transparency and simplicity, exclusively SQL is used for all kinds of data management and calculations. The utilization of SQL aggregation functions is restricted to the sum function and mathematical operations are restricted to basic algebraic functions.

The user may save the final set of parameters permanently in metadata tables and rerun his custom retrievals at any time. All estimates are based on the most current EDW data. The user can also compile a number of parameter sets into packages defining a complete physical data mart. The physical data mart is materialized by execution of this package; according XML files are stored permanently along with a reference to the parameter set. The latter enables backtracking to the variables in the EDW at any time.

2.5. Technical solution

In technical terms, the NAFIDAS is a high-performance and highly available (24/7; 8/7 user support) framework. The system is based on two dedicated UNIX/LINUX OS servers: One runs the database; another Apache web server with PHP hosts the internal and external web application both for the productive and the development environment. A failover system is only available for the database due to complexity, expenditure and operating expense. The maintenance tasks are shared between central IT services of the institution (server infrastructure, OS, databases) and the research units (ETL, web and data analysis programming).

The web application is designed as a client-server architecture, executed on multiple tiers. The public web application follows a classic three-tier model, whereas the internal web application uses an additional tier which handles processing of selected analysis parameters to calculate forest population estimates (cf. Fig. 4 right). The design principle 'Separation of Concerns' (Laplante, 2007) is strictly followed.

The web application follows Representational State Transfer (REST) principles to a far extent, but not strictly. It makes use of cookies to identify users on the server across the entire site, which violates REST's original model of application state (Fielding, 2000). Client-side programming in the browser is done in JavaScript, HTML, SVG and CSS often using the dojo toolkit (http://dojotoolkit.org/) and the AMD format for modules (https://github.com/amdjs/amdjs-api/wiki/AMD). On the Apache web server PHP is used for programming as well as XML for parameter storage. Data between the two is generally passed as JavaScript Object Notation (JSON) (www.json.org), XML or directly as HTML. In order to save and to visualize the spatial context of several thousand tables spatial extension is used on the database side. The saved spatial datasets are then published as web services through a geometry server. The web based map application showing the geographical context of the forest statistics consumes the prepared services using REST interfaces.

Changes to the source code of the internal and public website have been tracked with Apache Subversion (SVN) (https://subversion.apache.org/), since beginning of the fourth NFI cycle in 2009. The development server of the internal and public website is synchronized with the trunk branch; the productive (live) server is synchronized with the release branch of the SVN repository. Release versions are marked with tags to reference a certain point

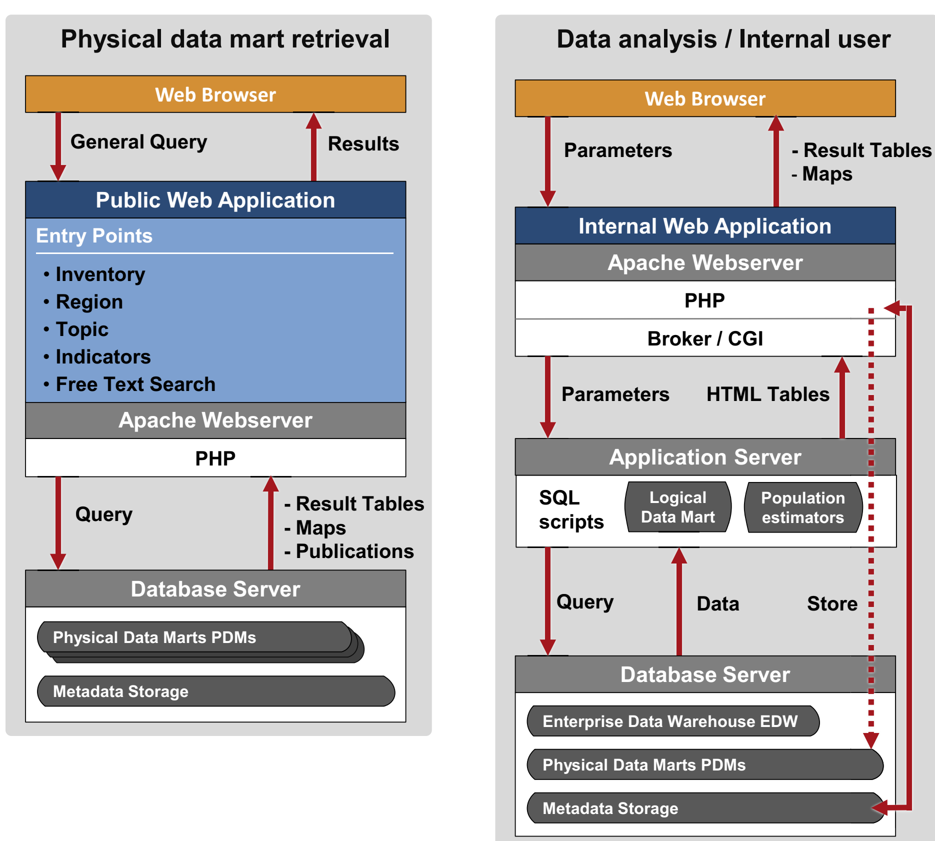

Fig. 4. Multi-tier client-server architecture of NAFIDAS. Classic three-tier architecture of the public web application consisting of presentation tier, logic tier and a data tier (left). Additional analysis tier allows estimator calculation based on data stored in the EDW (right).

in time. Changes are manually uploaded to the web servers by the programmer to allow for greater control and flexibility in publishing updates. A changelog is maintained by the developers to report on updates and new releases. It is the user's responsibility to get the most current information on changes; no proactive broadcasting process is established.

The internal web application is only accessible from the institute's intranet or over VPN. It is protected by a firewall and access is controlled by user accounts. To limit the amount of programming and testing of the internal web application, only the Firefox browser is supported. Users are generally assigned to the standard users group, who are responsible for data analysis, the management of analysis definitions and preparation of ETL scripts. Intermediate knowledge of the NAFIDAS framework, the Swiss NFI methods and SQL is a prerequisite. Four users with advanced knowledge of the NAFIDAS and SQL belong to the system administrators group, which has full access to the metadata and management of the ETL system. They are responsible for the classic database operations Create, Read, Update, and Delete (CRUD).

The public web application uses the same architecture and programming languages as the internal web application. However, it has to serve a wider range of users with different levels of knowledge. It runs in a less controlled environment where the application has to support different browsers of different generations and operating systems. Furthermore, security issues are of a much bigger concern, mainly compromising the database by SQL injection. The whole query system is implemented in such a way that search engine crawlers can simply follow links leading to complete indexing of all inventory tables. As a consequence search engines such as Google or any other can be used to query for a table or map instead of the website's own search function.

In contrast to the tables, maps are constructed on-the-fly from different REST endpoints by using spatial services whereas thematic data is served by parsing the HTML of the tables stored in the database and returning it as JSON.

2.6. Quality and risk management

High data quality requirements demand a system which performs its functions in a consistent and correct way over repeated use. To cope with this requirement, quality assurance (QA) measures with focus on the prevention of defects (IEEE, 1990) were applied as well as quality control (QC) measures with focus on defect detection (Tian, 2005) throughout all development steps of NAFIDAS.

Risk management (McDowall, 2005; Tian, 2005) has been applied following the steps

1. Risk assessment including risk analysis and evaluation
2. Risk treatment particularly consisting of risk mitigation
3. Risk acceptance and risk communication

Since the resources to be spent to avoid risks must be balanced with their risk priority, risk management in NAFIDAS means that processes and software modules with a high risk priority (McDowall, 2005) will be identified and prioritized at any time during development and when changes are made to the system. This translates into measures such as:

- Applying straightforward methods and programming languages and code style conventions for ETL scripts which allows efficient peer review and maintenance. Modularization and the application of the 'separation of concerns' principle on all levels of the design were further measures to minimize risks.
- Verification and plausibility checks on calculated estimates, reprogramming of complex queries and algorithms by independent programmer, code review (four eyes principle).
- Defining secure user interfaces and secure processes.
- Supporting users with comprehensive meta information and allowing the system to check user input to e.g. prevent any data analysis that could lead to statistically incorrect estimates.
- Ensuring that risks are detected should they occur.

Low impact risks will be treated by measures such as output verification of generated estimates which plays an important role when creating the physical data marts. Spending a lot of resources to detect these types of risks is, however, not justified. Residual risks, mainly connected with the misinterpretation of estimates, may remain. According documentation on statistical methods should keep this type of risk on a minimal level.

Quality assurance is mainly ensured by maintaining a high level of transparency of the data and process flow including clearly defined interfaces between the components. The process chain from generating single raw data entities to final forest statistics is supported by the highly automated and metadata controlled processes in combination with high-performance computer systems. The interactions by humans and thus the vulnerability to failures are reduced to a minimum. The technical specifications of the system are well documented; user requirements are under continuous review.

3. Results

In Switzerland, the group 'Scientific Services NFI' of the Swiss Federal Research Institute (WSL) is responsible for planning, data collection and analysis as well as for the scientific interpretation of the Swiss NFI. National and regional reports on forest statistics are based on data analyses carried out with NAFIDAS by NFI specialists which are part of the standard users group. Most recent examples are the publications on the internet (Brändli et al., 2015), the contributions to the Swiss Forest Report of 2015 (Rigling and Schaffer, 2015), the report to the Swiss greenhouse gas inventory (FOEN, 2016) as well as national reports (Brändli, 2010). Furthermore, data were provided to the Forest Resources Assessment (FRA) (FAO, 2015) and the State of Europeans Forests (SoEF) report (FOREST EUROPE, 2015).

In addition to and based on these reports a big physical data mart consisting of all possible and relevant combinations of forest population estimates was created with NAFIDAS. The resulting tables and maps are made available in four languages on the public website (http://www.lfi.ch/resultate/anleitung-en.php) (Fig. 5). This huge set of statistics on Swiss forests (for example forest area, growing stock, yield, increment, mortality and biodiversity indicators) is updated every 4–5 years and serves as an environmental planning and information system on national and regional (cantonal) level.

To make finding the right tables and maps as easy as possible, the user is given several entry points from where he can query the data mart as showed in Figs. 5 and 6. The choices are: start by (1) full text search with real-time suggestions, (2) theme such as volume, (3) regions such as Swiss cantons, (4) inventory such as the period 2009–2013, or (5) criteria and indicators (Abegg et al., 2014).

These entry points act as a filter to reduce the amount of possible search results. Each individual filter setting comes with its own description to facilitate choosing the right one. The filter and the accompanying description are generated directly from metadata in the database. At the end of the process a list of tables and maps is presented. Each item in the list leads either to a table with additional meta information, an interactive map or can be added to the shopping cart. The tables and maps can also be exported to various formats as well as printed.

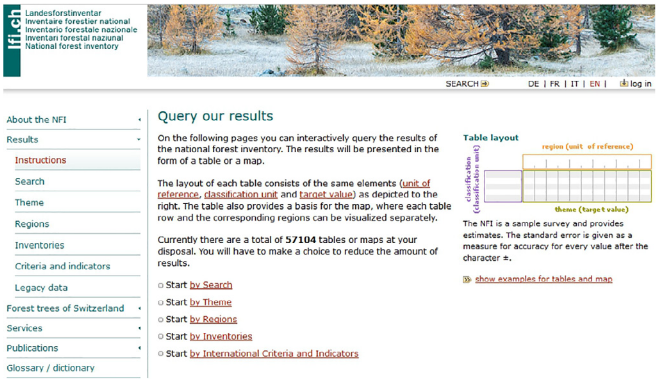

Fig. 5. NAFIDAS entry points to statistics on Swiss forests. The system offers entry points by (1) full text search, (2) theme, (3) regions, (4) inventories or (5) international criteria and indicators.

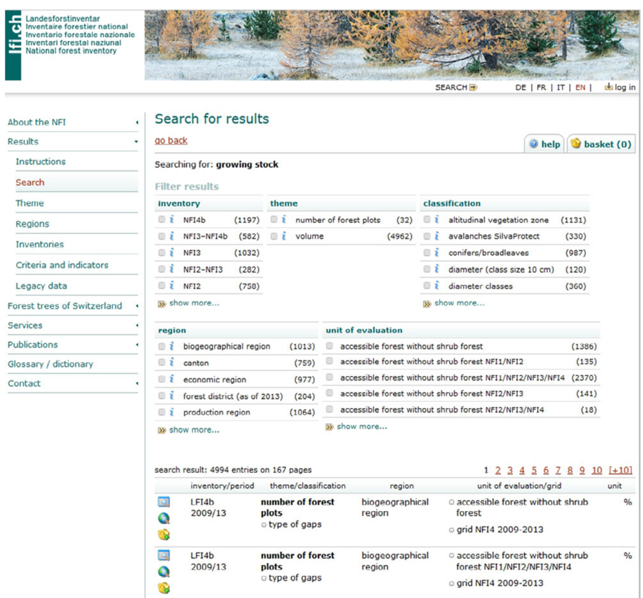

Fig. 6. Full text search and filtering of forest population estimates derived from the Swiss NFI. Example: Query results offered by submitting the search term 'growing stock'. Filtering topics: Inventory (inventory cycle to be selected), theme (forest population parameter), classification (topical dimensions), region (regional dimension), unit of evaluation (type of forest such as accessible forest without shrub forest).

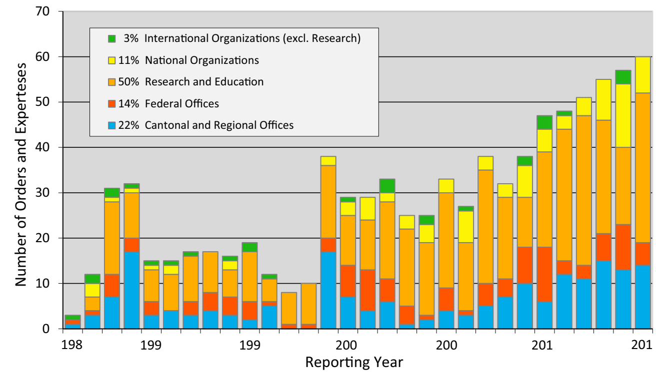

Fig. 7. Number of orders processed with NAFIDAS by reporting year and separated into five groups of stakeholders: International organizations excl. research (e.g. data delivery to European National Forest Inventory Network ENFIN and Forest Resources Assessment FRA), National organizations (e.g. Engineering offices, Biodiversity Monitoring Switzerland (BDM)), Research and education (e.g.: Swiss Federal Institute of Technology Zurich (ETHZ), University of Yale), Federal offices (e.g. The Federal Office for the Environment FOEN, Federal Office of Topography swisstopo, Federal Office of Meteorology and Climatology MeteoSwiss), Cantonal and regional offices (e.g. forest service of the canton Grisons).

The Swiss NFI also offers quick service of standard products and specific requests not covered by the public website. These deliveries of tailor-made NAFIDAS analyses and data base extractions support decision making, planning and research of various stakeholders such as the scientific community, administration bodies, universities, wood and energy industry or non-governmental organizations. The development of requests for specific analyses and data supply by NFI stakeholders is illustrated in Fig. 7. Since the beginning of the year 2000 an increase in orders from about 30–60 orders per year can be observed. The number of orders from cantonal offices and from research organizations increased substantially, whereas the number of orders from federal offices and international organizations remained stable.

In addition to data delivery services, the NFI staff offers methodological and organizational support services for regional inventories that are based on the Swiss NFI method carried out by local authorities (Gordon et al., 2000). This includes particularly the hosting and analysis of the collected field data. Requests that require development of additional software will only be accepted in rare cases or upon financing the necessary resources.

4. Discussion

4.1. Architecture and technical solution

The NAFIDAS provides a scalable, sound and stable backbone for the storage and analysis of Swiss NFI data. The interaction between the NAFIDAS components via the Apache web server and databases has proven to be very efficient and stable. The program logic is concentrated on one physical server and limited to two programming languages, which leads to improved security and less complexity. Due to its modular and scalable design, the maintenance and future development of the system is very efficient even for the ever increasing amount of data and functionality. The system may be enhanced in any direction within the framework, thus, the concept allows for fast programming and development of changes or extensions to the system. Even moderate resource investments will yield large effects without compromising the underlying paradigms.

The complexity of the system however demands the application of state of the art technology and highly skilled staff. Currently, the NAFIDAS team of administrators consists of four team members, which is the minimum in regards to workload deputizing and maintaining a system in a 24/7;8/7 user support framework and in the context of QA and risk management. Setting up of a new inventory in the EDW generally demands some effort to clone and adapt the ETL scripts, since they are strictly related to an inventory ID (cf. Section 2.3.2). Although this circumstance inhibits the seamless access to data spanning a range of inventory cycles, the inventory ID is a fundamental element of the structure of the analysis tables in the EDW and thus will be kept also in the future.

The performance and availability of NAFIDAS has proven to be highly satisfying. So far, the system has not shown any extraordinary downtime or data loss, nor were wrong data published. Several data warehouse features (Inmon, 2005; Kimball and Ross, 2013) form a transparent and reliable system that allows for efficient hosting as well as for very flexible dissemination of tables and maps. Special features such as the update of the EDW controlled by dependency trees (cf. Section 2.3.2) are invaluable for the consistency of the database. No less important is the metadata-controlled definition of the data analysis which ensures statistically correct calculation of estimates. Also the capability to produce and publish several thousand inventory tables (physical data mart) within a few days is highly satisfying to our stakeholders.

The most recent improvement on the public website was the extended filtering function by entry points such as inventory cycles or indicators which allowed for a more intuitive access to forest statistics. The inclusion of a semantic search function is currently under development and will add further value to the querying system.

General principles of risk management and computer system validation (McDowall, 2005; Tian, 2005) have been applied during the development of NAFIDAS. The design of the applications and the source code includes various risk mitigation measures. The ETL and data definition scripts and the data analysis modules are SQL based, transparent and well understandable due the application of the 'separation by concerns' principle (Laplante, 2007) wherever possible. Implementing a professional test suite for integral unit testing is an example of an inevitable QA requirement to be met in the near future. However, most important from our perspective is the establishment of a more formal process for change requests to avoid misunderstandings and inefficient cyclic discussions between programmers and users. This holds particularly true for changes to or cloning of ETL scripts. In this context it is crucial that users have a clear understanding of the resources needed to implement changes or new features. This information must be provided by the system architects and the programmers which requires a good understanding of the user needs supported by concise and well written requirement specification documents. That the users and system maintainers (programmers, database managers and statisticians) speak a common language is key for appropriate decision making on both sides. Obtaining an adequate mutual understanding of NAFIDAS related terms is time consuming and semantics have to be discussed repeatedly. And last but not least, a formal approval of user requirements by the customer is essential with regard to responsible handling of programmers' resources and for the sake to discern between wants or needs and requirements.

Several examples of wrong tracks or shortcomings can be found in the first solutions of Swiss NFI databases and analysis software developed in the early eighties. From a today's view, these stand-alone predecessors of NAFIDAS tended to be error-prone particularly when utilized by users with no special training and experience in data retrieval and statistical programming. One reason was, that these systems required extensive parameterization or even additional coding such as SQL statements to define constraints in SQL WHERE-clauses. The effort and resources needed for user support and training on these systems were repeatedly underestimated.

The current NAFIDAS aimed at a complete integration of subsystems and had the objective to also open up the system for non-expert internal and public users. Especially the latter was only possible with the allocation of the necessary resources. Maintenance and further development of NAFIDAS is an ongoing task and implementation under time constraints can never completely be avoided. Users in particular but also programmers sometimes tend to underestimate the complexity of underlying processes. Substantial refactorings resulting from insufficient specifications of interfaces are an inevitable consequence of fast-track solutions which is one of the most severe issues we experienced. Other examples of major shortcomings will be described in brief. (1) An XML schema definition has been introduced only recently. This should have been done right from the beginning to allow for proper validation of parameter sets stored as XML. (2) Any hard coded metadata in the source code turned out to be of high risk in terms of tracking changes. Although this conflicts with good programming practice, it does happen from time to time, typically under time pressure. (3) Multiple types of meta information were stored in non-normalized metadata tables. These were hard to maintain and culminated in tough and long lasting cleaning tasks. (4) Internationalization for the localization of the output in German, French, Italian and English was not taken into account right from the beginning. (5) Finally we strongly recommend establishing a versioning and archiving system for the source code even before writing the first line of code.

4.2. Business processes and stakeholders

NAFIDAS seems to be very suitable for both the users of the public website and the internal users working at WSL; novice users with solid knowledge of the Swiss NFI will acquire the skills required to successfully operate the web based system within days.

The reporting duties to national stakeholders (FOEN, 2016) and international protocols (FAO, 2015; FOREST EUROPE, 2015) can be accomplished quickly, either based on the execution of standard analysis definitions or by querying the EDW directly. It should however be emphasized, that the utilization of the system needs in-depth knowledge of both, the Swiss NFI methodology (Brassel and Lischke, 2001; Lanz et al., 2010) and the NAFIDAS functionality itself.

The set-up of regional inventories can be realized in NAFIDAS by just cloning and adapting a set of already existing ETL scripts. Provided that the regional inventory is conceptually based on the Swiss NFI methods, there is no need to change any structure of the operational database, the EDW or the tool for data analysis. This is very important in regards to resources and to risk management. The final set of forest population estimates of regional inventories could be generated within a few weeks after the field data collection is completed. Thus, regional inventories such as the inventory of the canton Grisons (Gordon et al., 2000) can profit substantially from well-established and verified NFI methods. The cloning itself, however, demands some efforts as mentioned in Section 4.1.

The public website has gradually developed to a comprehensive information system which provides general and specific information about the Swiss NFI on a national and regional level. It serves a broad range of national and international stakeholders; users generally express their satisfaction with the system since they benefit from access to currently more than 57,000 tables and maps in four languages. The guided querying and filtering system facilitates finding the right tables and maps quickly. The customer is only left with the correct interpretation of the inventory tables. Proactively seeking user feedback over different communication channels allowed development very close to user needs and requirements. User satisfaction is based on the high level of data quality, which is even more important than the sheer amount of information. Nilakanta et al., 2008 have referred to this fact as well and emphasize that the users' confidence is directly depending on the information quality provided by the INARIS data warehouse.

Although there is already a substantial number of users who know and regularly use the system, one important issue is how to sustain or even increase the stakeholder's interest in the system. Future development of NAFIDAS will concentrate on features that facilitate the individual selection or definition of regional analyzes and stepwise development of an interactive OLAP tool. Certain features of the EVALIDator tool (http://apps.fs.fed.us/Evalidator/evalidator.jsp) developed by the U.S. Forest Service may serve as examples for these type of NAFIDAS extension. Two main issues in the context of this extensions, however, must be disentangled in advance: (1) What is the proper level of interactivity for certain user groups; and (2) what is the optimal design of the user interface aiming at simplifying matters as far as possible. And although bearing valuable information on regional (cantonal) level, it is important to emphasize that NAFIDAS and its public web site was never intended and also in future will not develop to a decision support system on forest management level such as the systems described by Zambelli et al. (2012) or by Kirilenko et al. (2007). Whereas these systems provide direct support for planning and management of forest stands, regional inventory data provided by NAFIDAS will not exceeding to flank such decisions. The challenges in terms of providing highly efficient interactive user interfaces and search tools however are quite comparable between NAFIDAS and the cited decision support systems.

Many attempts in this regard often reach their limits due to the system-inherent or factual complexity which cannot be eliminated. Furthermore, the majority of NFI stakeholders are absolutely satisfied with the current service of data supply. The above mentioned extensions could probably be too complex since they require too much statistical knowledge to be useful for many stakeholders.

5. Conclusions

The 2016 version of NAFIDAS goes a long way to cope with requirements on various levels. After more than 10 years of application development, continuous improvement and extension, the NAFIDAS has now reached a stable and mature state. Especially the multilingual query and filtering system of the NFI website that provides access to a huge amount of tables and maps has been continuously improved and extended. Due to its reliability and its sustainable support of high quality forest population estimates, the system qualifies as an information and decision support system with high potential on regional and national level.

The resources to be invested for any future changes to NAFIDAS should outweighed by their benefits for users. Whereas estimating the costs of maintaining the system is usually rather challenging, the costs of extending and improving the system can normally be estimated quite accurately, especially due to its modular architecture.

We see a great potential of NAFIDAS to succeed as a blueprint for any other systems to be developed with the objective to manage, analyze and disseminate data from sample based forest monitoring and inventory programs. Since the concept covers the complete data processing chain from field data collection to final tables and maps on forest statistics it has a strong power in efficient processing of forest inventory data. A further strength is the clear compliance to the basic requirements of these types of systems: to efficiently organize long-term and large-scale inventory data, to comply with high quality requirements and to transparently offer data to a broad user community in a timely manner.

Acknowledgements

The Swiss Federal Office of Environment FOEN and the Federal Institute for Forest, Snow and Landscape Research WSL fund the Swiss NFI and as part of it the development of NAFIDAS. We thank Prof. Marc Hanewinkel, who encouraged us to publish this article and who provided valuable suggestions to the structure of the manuscript. We also thank our reviewers at WSL and all external reviewers for their valuable input to the manuscript.

References

Abegg, M., Brändli, U.-B., Cioldi, F., Fischer, C., Herold-Bonardi, A., Huber, M., Keller, M., Meile, R., Rösler, E., Speich, S., Traub, B., Vidondo, B., 2014. Viertes Schweizerisches Landesforstinventar – Ergebnistabellen und Karten im Internet zum LFI 2009-2013 (LFI4b). Eidg. Forschungsanstalt WSL, Birmensdorf.

Arvanitis, L.G., Ramachandran, B., Brackett, D.P., Abd-El Rasol, H., Du, X.S., 2000. Multiresource inventories incorporating GIS, GPS and database management systems: a conceptual model. Comput. Electron. Agr. 28, 89–100.

Bachofen, H., Brändli, U.-B., Brassel, P., Kasper, H., Lüscher, P., Mahrer, F., Riegger, W., Stierlin, H., Strobel, T., Sutter, R., Wenger, C., Winzeler, C., Zingg, A., 1988. Schweizerisches Landesforstinventar. Ergebnisse der Erstaufnahme 1982–1986, Ber. Eidgenöss. Forsch.anst. Wald Schnee Landsch. EAFV (Eidg. Anstalt für das forstliche Versuchswesen), Bundesamt für Forstwesen und Landschaftsschutz, Birmensdorf.

Böhl, J., Bierer, D., 2004. Konzept und Prototyp der Auswertungssoftware für das dritte schweizerische Landesforstinventar. In: Wunn, U., Quednau, H.D. (Eds.), Die grüne Reihe. Deutscher Verband Forstlicher Forschungsanstalten, Sektion Forstliche Biometrie und Informatik, 15. Tagung, Freiburg, pp. 250–257.

Brändli, U.-B., Cioldi, F., Fischer, C., Huber, M., 2015. Swiss National Forest Inventory – Special assessments for the Forest Report 2015 in the internet. Eidg. Forschungsanstalt WSL CH-8903, Birmensdorf.

Brändli, U.B. (Ed.), 2010. Schweizerisches Landesforstinventar. Ergebnisse der dritten Erhebung 2004–2006. Birmensdorf, Eidgenössische Forschungsanstalt für Wald, Schnee und Landschaft WSL. Bern, Bundesamt für Umwelt, BAFU.

Brassel, P., Brändli, U.-B., 1999. Schweizerisches Landesforstinventar. Ergebnisse der Zweitaufnahme 1993–1995. Eidg. Forschungsanstalt WSL, Bundesamt für Umwelt BAFU, Birmensdorf, Bern, Wien, p. 442.

Brassel, P., Lischke, H., 2001. Swiss National Forest Inventory: Methods and Models of the Second Assessment. Eidg. Forschungsanstalt WSL, Birmensdorf.

FAO, 2015. Global Forest Resources Assessment 2015 Desk Reference. Food and Agriculture Organization of the United Nations, Rome.

Fielding, R.T., 2000. Architectural Styles and the Design of Network-based Software Architectures. University of California, Irvine, p. 162.

FOEN, 2016. Switzerland's Greenhouse Gas Inventory 1990–2014: National Inventory Report Including reporting elements under the Kyoto Protocol. Federal Office for the Environment FOEN, Climate Division, Bern, p. 629.

FOREST EUROPE, 2015. State of Europe's Forests 2015 Ministerial Conference on the Protection of Forests in Europe. FOREST EUROPE Liaison Unit Madrid, Madrid, p. 314.

Fridman, J., Holm, S., Nilsson, M., Nilsson, P., Ringvall, A.H., Stahl, G., 2014. Adapting National Forest Inventories to changing requirements – the case of the Swedish National Forest Inventory at the turn of the 20th century. Silva Fenn. 48, 1–29.

Gabler, K., Schadauer, K., Tomppo, E., Vidal, C., Bonhomme, C., McRoberts, R.E., Gschwantner, T., 2012. An enquiry on forest areas reported to the global forest resources assessment – is harmonization needed? Forest Sci. 58, 201–213.

Gillis, M.D., Omule, A.Y., Brierley, T., 2005. Monitoring Canada's forests: the national forest inventory. Forest. Chron. 81, 214–221.

Gordon, R., Bühler, U., Zinggeler, J., 2000. Änderung der Waldinventurmethode im Kanton Graubünden. Schweiz. Z. Forstwes. 5, 165–173.

Gschwantner, T., Schadauer, K., Vidal, C., Lanz, A., Tomppo, E., di Cosmo, L., Robert, N., Duursma, D.E., Lawrence, M., 2009. Common tree definitions for national forest inventories in Europe. Silva Fenn. 43, 303–321.

Hoffer, J.A., Ramesh, V., Topi, H., 2011. Modern Database Management. Pearson, Boston.

IEEE, 1990. IEEE Standard Glossary of Software Engineering Terminology. The Institute of Electrical and Electronics Engineers Inc., New York.

Inmon, W.H., 2005. Building the Data Warehouse. Wiley, Indianapolis.

Kangas, A., Maltamo, M., 2006. Forest Inventory: Methodology and Applications, Managing Forest Ecosystems. Springer, Dordrecht, p. 362.

Kaufmann, E., 2001. Estimation of Standing Timber, Growth and Cut. In: Brassel, P., Lischke, H. (Eds.), Swiss National Forest Inventory: Methods and Models of the Second Assessment. Swiss Federal Research Institute WSL, Birmensdorf, pp. 162–196.

Keenan, R.J., Reams, G.A., Achard, F., de Freitas, J.V., Grainger, A., Lindquist, E., 2015. Dynamics of global forest area: Results from the FAO Global Forest Resources Assessment 2015. Forest Ecol. Manag. 352, 12.

Kimball, R., Ross, M., 2013. The Data wareHouse Toolkit the Definitive Guide to Dimensional Modeling. Wiley, Indianapolis.

Kirilenko, A., Chivoiu, B., Crick, J., Ross-Davis, A., Schaaf, K., Shao, G., Singhania, V., Swihart, R., 2007. An Internet-based decision support tool for non-industrial private forest landowners. Environ. Modell. Softw. 22, 1498–1508.

Köhl, M., 2001. Inventory concept NFI2. In: Brassel, P., Lischke, H. (Eds.), Swiss National Forest Inventory: Methods and Models of the Second Assessment. Eidg. Forschungsanstalt WSL, Birmensdorf, pp. 19–44.

Köhl, M., Magnussen, S., Marchetti, M., 2006. Sampling Methods, Remote Sensing and GIS Multiresource Forest Inventory. Springer, Berlin, London.

Lanz, A., Brändli, U.-B., Brassel, P., Ginzler, C., Kaufmann, E., Thürig, E., 2010. Switzerland. In: Tomppo, E., Gschwantner, T., Lawrence, M., McRoberts, R.E. (Eds.), National Forest Inventories – Pathways for Common Reporting. Springer, Heidelberg, Dordrecht, London, New York, pp. 555–565.

Laplante, P.A., 2007. What Every Engineer Should Know about Software Engineering. Taylor & Francis, Boca Raton.

MacDicken, K.G., 2015. Global Forest Resources Assessment 2015: What, Why and How? Forest Ecol. Manag. 352, 3–8.

McDowall, R.D., 2005. Effective and practical risk management options for computerised system validation. Qual. Assur. J. 9, 196–227.

Nilakanta, S., Scheibe, K., Rai, A., 2008. Dimensional issues in agricultural data warehouse designs. Comput. Electron. Agr. 60, 263–278.

O'Connell, B.M., Conkling, B.L., Wilson, A.M., Burill, E.A., Turner, J.A., Pugh, S.A., Christiansen, G., Ridley, T., Menlove, J., 2016. Database Description and User Guide for Phase 2 (version 6.1). U.S. Department of Agriculture, Forest Service, p. 892.

Rasinmäki, J., 2009. XQuery as a retrieval mechanism for longitudinal multiscale forest resource data. Environ. Modell. Softw. 24, 1153–1162.

Rigling, A., Schaffer, H.P., 2015. Forest Report 2015 Condition and Use of Swiss Forests. Bundesamt für Umwelt, Eidg. Forschungsanstalt WSL, Bern, Birmensdorf, p. 144.

San-Miguel-Ayanz, J., Schulte, E., Schmuck, G., Camia, A., 2013. The European Forest Fire Information System in the context of environmental policies of the European Union. Forest Policy Econ. 29, 19–25.

Schuck, A., Andrienko, G., Andrienko, N., Folving, S., Kohl, M., Miina, S., Paivinen, R., Richards, T., Voss, H., 2005. The European Forest Information System – an Internet based interface between information providers and the user community. Comput. Electron. Agr. 47, 185–206.

Speich, S., Meile, R., 2013. Interactive Maps of Results of the National Forest Inventory NFI. Swiss Federal Research Institute WSL, Birmensdorf.

Thulasiraman, K., Swamy, M.N.S., 1992. Directed Graphs, Graphs: Theory and Algorithms. John Wiley & Sons Inc., pp. 97–125.

Tian, J., 2005. Software Quality Engineering: Testing, Quality Assurance, and Quantifiable Improvement. Wiley, Hoboken, N.J..

Tomppo, E., 2006. The Finnish national forest inventory. In: Kangas, A., Maltamo, M. (Eds.), Forest Inventory—Methodology and Applications. Springer, Dodrecht, The Netherlands, pp. 179–194.

Tomppo, E., Gschwantner, T., Lawrence, M., McRoberts, R.E., 2010. National Forest Inventories. Springer, Netherlands, Dordrecht.

Traub, B., Schnellbächer, H.J., 2001. Analysis software. In: Brassel, P., Lischke, H. (Eds.), Swiss National Forest Inventory: Methods and Models of the Second Assessment. Birmensdorf, Swiss Federal Research Institute WSL. Eidg. Forschungsanstalt WSL, Birmensdorf, pp. 288–296.

Vidal, C., Lanz, A., Tomppo, E., Schadauer, K., Gschwantner, T., di Osmo, L., Robert, N., 2008. Establishing forest inventory reference definitions for forest and growing stock: a study towards common reporting. Silva Fenn. 42, 247–266.

Wulder, M.A., White, J.C., Cranny, M., Hall, R.J., Luther, J.E., Beaudoin, A., Goodenough, D.G., Dechka, J.A., 2008. Monitoring Canada's forests. Part 1: Completion of the EOSD land cover project. Can. J. Remote Sens. 34, 549–562.

Zambelli, P., Lora, C., Spinelli, R., Tattoni, C., Vitti, A., Zatelli, P., Ciolli, M., 2012. A GIS decision support system for regional forest management to assess biomass availability for renewable energy production. Environ. Modell. Softw. 38, 203–213.