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.