Chapter 27
Conclusions and Outlook—Summary of Big Data in Forestry

Jukka Miettinen and Renne Tergujeff

Abstract In this chapter, we summarize the findings from the forestry pilots conducted during the DataBio project. Although the pilots demonstrated the functionality of big data in forestry through several practical applications and services, they also highlighted areas where further development is needed. More effort is needed particularly in ensuring smooth connections between the technical components of the processing pipelines, as well as designing the best business solutions within the big data service chain and between the service providers and users. Overall, the challenge for the coming years is to establish operational big data processing pipelines that meet the requirements and expectations of forestry stakeholders.

27.1 Introduction

As discussed in the forestry introduction chapter (Chap. 22), new technologies that have emerged over the past decade enable utilization of novel big data approaches in forest monitoring. At the same time, the requirements for forest monitoring information have widened. Indicators of carbon balance, biodiversity, and forest health, to name just a few, have an increasingly important role in forest management, alongside the traditional forest characteristics (e.g., height, volume, species distribution). The forestry pilots of the DataBio project set out to investigate and demonstrate ways to maximize the benefits of big data in forestry, providing users with timely datasets and analysis results that would meet their specific information requirements.

The preceding chapters (Chaps. 23–26) have presented four DataBio pilots utilizing big data for forest monitoring and management. The use of datasets varied from crowdsourced field data to satellite observations. The selected use cases included forest structural variable estimation (e.g., tree height and basal area), health monitoring as well as bark beetle and storm damage mapping. Geographic coverage of the use cases varied from local forest estate level to national level. Stakeholders involved in the pilots ranged from Earth observation (EO) service companies

J. Miettinen (✉) · R. Tergujeff
VTT Technical Research Centre of Finland, 02044 VTT, Espoo, Finland
e-mail: jukka.miettinen@vtt.fi

C. Södergård et al. (eds.), *Big Data in Bioeconomy*,
https://doi.org/10.1007/978-3-030-71069-9_27

and private forestry businesses to government organizations and academic institutions. Technical solutions included local processing, interlinked cloud storage and processing platforms, and online user interfaces.

As much as the DataBio project pilots confirmed the usability of big data in forestry and the functionality of already existing technical solutions, they also revealed some weak points in the value adding chain where more effort is needed to fully utilize the potential of big data in the forestry sector. In the following, the lessons learnt from the DataBio forestry pilots have been synthesized from the perspective of (1) technical solutions and (2) business solutions. In both of these areas, functioning solutions for individual components within big data value adding chain already exist, but more effort is needed in smooth connections between the components, as well as the interface between the big data service providers and users.

27.2 Lessons Learned from DataBio: Technical Solutions

Due to the high data volumes and processing requirements in big data analysis, traditional data processing and analysis approaches (i.e., image-by-image analysis on personal computers) are not sufficient to fully exploit the benefits of the data. Redesign of the processing and delivery pipelines was needed to match with today's data volumes and modern processing infrastructures. In the DataBio project, big data processing chains were divided into four main steps:

1. Acquisition and storage
2. Preparation
3. Analytics
4. Visualization and user interaction.

As presented in Chaps. 23–26, functioning solutions for all of these individual steps in the forestry sector were identified, developed further and piloted. The Finnish Forest Centre demonstrated the functionality of their crowdsourcing field observation application, allowing innovative data acquisition for big data applications concentrating on forest damage monitoring. VTT and Spacebel demonstrated the usability of their online platforms Forestry TEP (https://f-tep.com/) and EO Regions! (https://www.eoregions.com/) for data preparation and analytics. TRAGSA, Technical University of Denmark and the Forest Management Institute (Czech Republic) exhibited the usability of their data analytics algorithms with EO big data. Many of these activities were linked with Wuudis (https://www.wuudis.com/), a commercial service for forest owners, timber buyers and forestry service companies, providing, e.g., visualization tools and supporting the linkage between the users and big data service providers. Likewise, the Forestry TEP and EO Regions! have their own user interfaces, and the Forest Management Institute (Czech Republic) provided their maps through the online service "Kůrovcová mapa" (www.kurovcovamapa.cz).

From a technical perspective, the weakest link in big data utilization in the forestry pilots was considered to be the connection between different datasets, platforms, and

applications. As the legacy of traditional, localized processing approaches, many processing and analysis applications are optimized to work with locally stored datasets. Although in some cases, typically in large institutions with sufficient storage and processing power, big data-based operational systems may be set up locally, the only way to fully and effectively unleash the benefits of big data for the wider forestry stakeholder community is through interplatform connections. The aspects of interplatform operability will need to be developed further in the future to ensure that technical difficulties do not start to hinder further uptake of big data solutions by forestry stakeholders.

In the DataBio forestry pilots, the interplatform connections based on established infrastructural configurations, like the connection between Forestry TEP processing and analysis platform and the CREODIAS (https://creodias.eu/) data and storage platform, worked well. However, various technical problems were encountered in connections between independently operated processing, application or visualization systems. Further development of smooth interoperability of different platforms should be a key goal for technical development in the near future. The future big data solutions, covering storage, processing, analysis and visualization capabilities, would optimally lean on interconnected online platforms. Large storage and processing facilities on the cloud, like the Copernicus Data and Information Access Services (DIAS), will provide the core EO and supporting datasets. Other national or international databases may store, e.g., field data or other auxiliary datasets usable specifically in the forestry sector. Forestry application platforms, like the Forestry TEP and EO Regions!, will provide processing tools, algorithm development interfaces and ready-made products with user interfaces optimized for forestry stakeholders. Further still, these platforms can provide analysis results (e.g., on structural forest characteristics, damages or forest health) to various forestry services (like Metsään.fi or Wuudis), which utilize up-to-date data for forest management and operations planning and user interaction. This entire service chain would benefit from smooth interplatform operability.

27.3 Lessons Learned from DataBio: Business Solutions

In the DataBio forestry pilots, big data solutions were piloted with the users of an online forest management support platform Wuudis, forestry sector partners, associated partners and other stakeholders, to evaluate the business potential and end user interest in the products. The pilots demonstrated a high demand for frequently updated forestry information on forest structural characteristics (e.g., tree height, basal area), forest health, storm damages, and other. However, it also became evident that significant progress is needed in the business practices and market development. Need for improvement was identified in two major areas, before the full potential of big data in forestry can be efficiently unleashed:

1. Business practices within the big data value chains.

2. Operating practices in the forestry sector.

The big data economy with multiple operators working together in a single value chain is something new to most forestry sector stakeholders. Whether it be a national database that has opened up to the public, a commercial EO satellite operator, a processing and application platform operator or a private forestry company (to name only a few stakeholders), they all need to define the value of their work and information in a new way. This process takes time. A typical example of the new complications is the interplatform operations. They do not create only technical challenges, but they also require new types of business arrangements.

In the DataBio forestry pilots, progress was made in creating business connections between the platforms involved. A business agreement was set up between Wuudis Solutions and Spacebel, regarding the distribution of the Wuudis services to the forest users of the EO Regions! platform and the commercialization of Spacebel's earth observation forest products in the Wuudis platform. This type of business arrangements is needed between collaborating services or other big data providers who operate on a commercial basis, before the technical benefits discussed in the previous section will materialize. Delays in setting up business agreements will slow down the uptake of big data in the forestry sector as surely as technical problems.

The other major area of challenge in the field of business is the slow development of management practices in the forestry sector in many countries. Traditional management practices are largely based on manual field work and static management plans. It may take some time to convince forestry stakeholders of the benefits of big data for their operations. This is best achieved by providing high quality services and products that meet the requirements of the stakeholders. For this, it is essential to (1) know the requirements of the stakeholders to be able to provide the right kinds of products, (2) create smooth user experience when accessing and using the information/products, and (3) actively promote the possibilities of big data in the user community. Local promotional activities and locally designed services are in an essential role in marketing, due to the varying forestry practices in different countries. Over time, forestry stakeholders will realize the benefits of online service provision of frequently updating information based on big data.

27.4 Future Outlook

Overall, the DataBio forestry pilots (1) demonstrated the usability of big data in forestry through several practical applications and services and (2) highlighted areas where further development is needed to increase the benefits of data-driven solutions for forestry stakeholders. Although big data solutions in forestry are far from being fully developed, it is clear that big data is here to stay. The technological development that has already enabled the collection of massive data volumes from both remote sensing and field measurements, and their processing on online platforms, will only accelerate in coming years. The information available to be extracted from

the massive volumes of data is too valuable to be ignored by the forestry sector. The challenge for the coming years is to establish operational big data processing pipelines that meet the requirements and expectations of forestry stakeholders.

The future will tell how fast big data solutions will replace traditional practices in forestry. In some countries, this may require even legislative changes, to allow utilization of remote sensing based solutions in official reporting. In any case, the great benefits of big data to the forestry sector are clear. At the same time, the reporting and monitoring requirements are constantly increasing with growing demands, e.g., on forest carbon flux and forest management sustainability monitoring. Big data approaches through online platforms provide the means to answer these demands. Big data also provides possibilities for entirely new and exciting types of forest monitoring approaches based on artificial intelligence, which were not yet within the scope of the DataBio project. It is up to all of us forestry stakeholders to find the best solutions to make big data benefit the entire forestry sector, our common environment and the whole world.