OPEN ACCESS TO RESEARCH DATA ON FOREST ECOSYSTEMS IN POLAND

DOROTA GRYGORUK

Department of Forest Ecology, Forest Research Institute
Braci Leśnej 3, Sękocin Stary, 05-090 Raszyn, Poland

(received: 21 August 2017; revised: 20 September 2017;
accepted: 27 September 2017; published online: 11 October 2017)

Abstract: Studies of forest ecosystems enable gathering important information on the natural environment the development of which is more and more disturbed by the global climate change. The current research on the ecosystem functioning provides data that may be of much value for future analysis and prognostic studies. Modern measurement techniques used in the forest research have a significant influence on the increase in the database resources, especially those concerning the spatial data. Big data requires the use of advanced analytical technologies, such as data warehouses, computer clusters or cloud computing. Consequently, cooperation of specialists from various scientific disciplines, including forestry, geography, climatology and computer science, has become increasingly necessary. The IT system of the Forest Research Institute (FRI) was modernized within the framework of the Operational Programme – Innovative Economy 2007–2013. Its functionality allows integrating, storing and analyzing ever more big databases from dispersed sources. The idea of open access to data is realized by the FRI mainly through publication of research results in domestic and foreign scientific journals, in specialized information services and on scientific portals. On the other hand, open access to raw data still raises a lot of concern and controversies in the scientific community, especially in the context of copyright infringement.

Keywords: forest ecosystem, open access, open science, raw data, research data
DOI: https://doi.org/10.17466/tq2017/21.4/w

1. Introduction

Many natural processes of plant and animal growth in forest ecosystems have been disturbed by the global climate changes [1]. In recent decades, a definite increase in the frequency and intensity of extreme weather events (hurricanes, floods, droughts) that cause enormous losses and destruction in the natural environment, all over the world including Poland, has been observed. Studies on natural environment changes provide much important data for future comparative analyzes and prognostic studies.

The Forest Research Institute conducts studies on forest ecosystems both during short-term research periods (from one to several years) within the implementation of the framework of specific project tasks and during long-term study cycles (about several decades) *e.g.* on research plots established by German researchers in forest stands in western and northern Poland in the mid-1800s to investigate the growth of native forest tree species (Schwappach's plots) or forest monitoring plots which have belonged to the European Forest Monitoring Network (ICP Forests) for almost 30 years [2]. In the last decade, the data resources of the Institute, especially the spatial data resources have increased, which is largely due to the use of modern measurement equipment and techniques, such as temperature and humidity sensors, infrared cameras, laser scanners and aerial photography of forest areas. Big databases require the use of advanced analytical technologies, including data warehouses, computer clusters or cloud computing. Modern research of forest ecosystems requires cooperation of specialists from different scientific disciplines, including forestry, geography, climatology and computer science.

The FRI's modernized IT system (within the framework of OPIE 2007–2013) enables integration and storage of big data from many dispersed sources. In addition, big data can be analyzed by the Business Intelligence technology (data warehouse) or the clustered technology. The FRI's research data is mainly shared in peer-reviewed scientific publications, according to the Open Access idea, whereas open access to raw data still raises a lot of concern and controversies in the Institute's scientific community.

The aim of this paper is to present examples of the use of modern measurement equipment and technology in forest studies in the context of a rapid increase in the data resources and problems with sharing research data and raw data in view of the legal regulations. It is also to present some results of the questionnaire survey on Open Science, which was conducted in the FRI in 2015.

2. Modern measurement techniques and database resources of the Forest Research Institute

Depending on the forest investigation methodology, data can be gathered during field measurements or under laboratory conditions. Until recently, traditional measurements of trees (*e.g.* DBH, height, crown size) taking into account the division into tree species and the forest stand structure were made by research teams composed of several people. The measurements taken by these teams were characterized by high work intensity, regardless of the specificity and diversity of the collected data, whereas the use of modern equipment can definitely reduce the time of work in the field, *e.g.* a terrestrial laser scanner (operated by 1 person) reduces 4 times the tree measuring time on a research plot of 500 m^2 (Figure 1).

Methods of remote sensing and geomatics are increasingly used in forest studies, and a consequence of introducing new research methods is a rapid increase in the amount of data, especially spatial data [3]. With the fall in the prices of

Figure 1. Terrestrial laser scanner parameterization
(project LIFE + ForBioSensing PL; photographer A. Modzelewska)

aerial and satellite photographs, their availability for forest studies has increased, *e.g.* for studies on the dynamics of forest tree growth. In 2015, the FRI bought about 2500 aerial high-resolution photographs for forest areas covering about 60 000 ha; totaling several hundred GB in size.

In recent times, the ROXAS 3.0 application has been used in the dendrochronology studies of forest trees. ROXAS is a specialized image analysis tool for quantifying the xylem anatomy in circular and linear samples of angiosperms and conifers [4]. A key feature of ROXAS analysis is to process images of entire samples in one step, including up to 1 000 000 cells and over 100 annual rings of the measured tree (Figure 2). The results of analyses obtained are saved in MS Excel or text files. Dendrochronology analyses enable the study of forest history and dynamics,

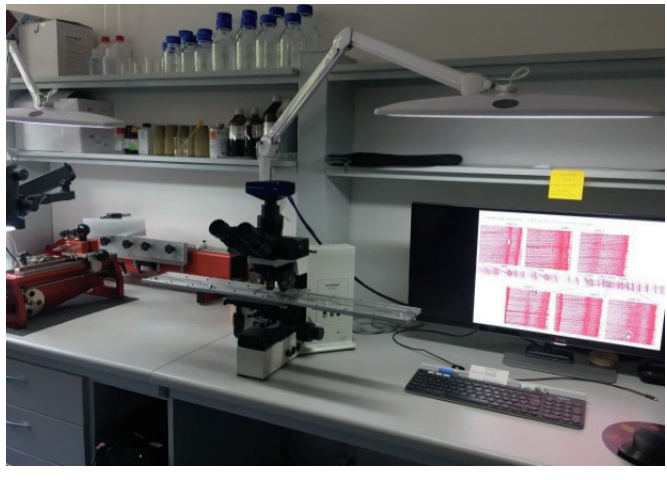

Figure 2. Measurement of xylem anatomy of forest trees – example of using Roxas 3.00 application; photographer D. Grygoruk

including precise determination of the tree age and disturbance history, *e.g.* fire history.

The modern equipment of the Laboratory of Natural Environment Chemistry of the FRI enables performance of a wide range of chemical analyses of plant, water and soil samples. The analysis results have been providing more and more data on forest functioning. The laboratory is certified by PCA AB 740 [5]. Using the atomic ICP-OES THERMO SCIENTIFIC iCAP 6000 emission spectrometer only, it is possible to obtain nearly 5000 analytical items of data during the 3-minute-long measurement of samples.

3. Open Access to data and legal regulations

According to the European Commission, open access to the research results are actions aimed at improving the quality of data, accelerating the scientific progress, and supporting the fight against abuses in science. Data sets should be labeled in such a way that they can be easily identified and integrated with other databases [6, 7]. The Horizon 2020 program framework includes the following requirements: open access to scientific publications and promotion of open access to scientific data generated owing to the publicly funded research under the program [8]. The guidelines of the Polish Ministry of Science and Higher Education call attention to sharing research data in open access, in compliance with the copyright and taking into consideration information management principles [9]. At the same time, according to the code of conduct on research and research funding issued by the National Science Center, it is crucial to "preserve the starting material and primary data and store it in a proper way, in a form allowing verification of the results, the same applies to reports and procedures so as to enable repetition of research" [10]. In the legal regulations, there is no information about the standards for sharing raw data, even though in light of extremely fast development of the computing technology, raw data is more than ever valuable for future analyses, models and forecasts.

Since the beginning of its activity (1933), the Forestry Research Institute has been conducting publishing activities. Scientific papers on forestry and related sciences of Polish and foreign authors are published in two periodicals of the Institute: *Folia Forestalia Polonica*, series A – *Forestry* (co-published with the Polish Academy of Sciences) and *Forest Research Papers* [2]. The research results conducted by the Institute have been also available through professional information services according to national and international agreements, as well as they have been presented on scientific portals [11, 12]. Furthermore, the Institute shares the research data under a license, however, mainly for the scientific partners within the framework of cooperation.

The questionnaire survey on methods of collecting, processing and sharing of data was carried out among the Institute's scientific staff in 2015. The first part of the questionnaire concerned the professional experience of the respondents. The second part of the survey included detailed questions related to the research

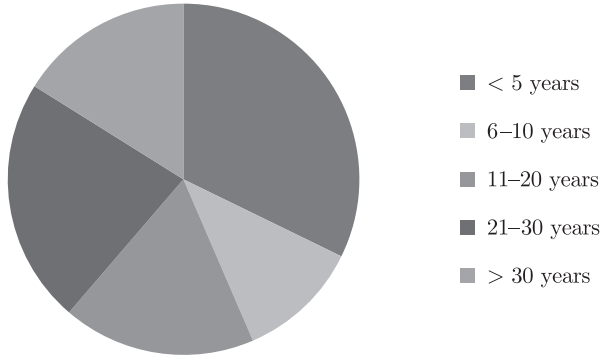

Figure 3. Professional experience (years of work) of survey participants

methodology and research data sharing after closing a research project. The survey participants were characterized by diverse professional experience (Figure 3).

The most numerous group (32%) was the study staff with a minimum of 5 years of employment. Persons with the period of employment within the range from 21 to 30 years accounted for 23% of the respondents. The remaining respondents had 6–10, 11–20 and over 30 years of professional experience (11%, 18%, 16%, respectively). Furthermore, the survey participants represented all academic degrees (Figure 4).

Figure 4. Study staff (%) by academic degree

In the opinion of the majority of the respondents (82%) the research data after project closing can provide valuable scientific material, in both the planning stage of new study projects and analytical works (Figure 5).

Approximately 75% of the respondents accept the concept of open access, primarily with respect to the publication of research results (Figure 6).

The study staff are eager to use database resources of other scientific centers, but open access to one's own research data and databases raises a lot of controversy and concern, especially in the context of copyright infringement.

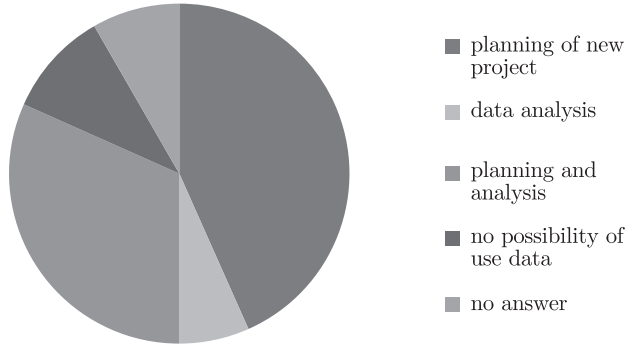

Figure 5. Possibilities of using archival data by survey participants

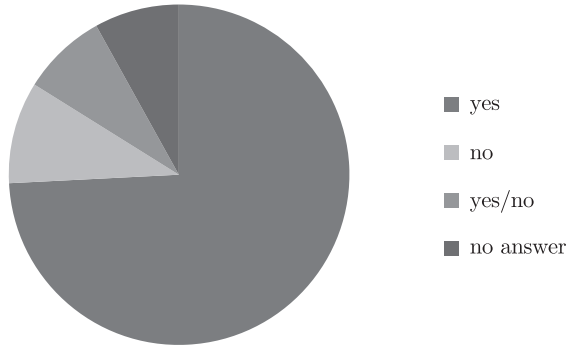

Figure 6. Do you support the idea of Open Access?

4. Conclusions

The modern measurement and analytical technology in the research of forest ecosystems has influence on the increase in the database resources. Open Access to scientific data and research results has long been accepted by the scientific community because the number of scientific publications and the sharing of the research results are related to the principles of the parametric evaluation of science institutes and the assessment of science achievements of employees.

On the other hand, sharing raw data is a great problem, for the most part due to the lack of clearly defined rules and standards [13]. A great part of valuable data has been stored only in personal computers that cannot guarantee secure storage and access [14, 1, 15]. The experience of foreign scientific centers with widely recognized achievements (*e.g.* University of Oxford, University of Cambridge) shows that „the first step" towards raw data sharing is to define the data management principles [16]. Raw data is "specific intellectual property". At the present time, analytical applications define the scientific value of raw data. In the future, the same data can play a new important role in analyses carried out using next-generation technologies, not yet fully defined. For that reason, the protection of raw data against loss and destruction becomes a challenge for the study staff.

References

[1] Malmsheimer R W *et al.* 2008 *Journal of Forestry* **10** 115

[2] [Online] available at: http://www.ibles.pl/

[3] [Online] available at: http://www.ibles.pl/web/guest/-/life-forbiosensing-pl-kompleksowy-monitoring-dynamiki-drzewostanow-puszczy-bialowieskiej-z-wykorzystaniem-danych-teledetekcyjnych

[4] [Online] available at: http://www.wsl.ch/dienstleistungen/produkte/software/roxas/index_EN

[5] [Online] available at: http://www.ibles.pl/web/cs/start

[6] [Online] available at: https://repin.pjwstk.edu.pl/files/Przewodnik_po_otwartej_nauce.pdf

[7] Commission Recommendation of 17 July 2012 on access to and preservation of scientific information (2012/417/UE)

[8] Regulation (Eu) No 1291/2013 of The European Parliament and of The Council of 11 December 2013 establishing Horizon 2020 – the Framework Programme for Research and Innovation (2014–2020)

[9] Directions of development of open access to publications and research results in Poland (MNISW 2015)

[10] National Science Center Code on the reliability of research and funding for research (NCN 2016)

[11] [Online] available at: http://icp-forests.net, http://www.gios.gov.pl, http://inspire.gios.gov.pl, https://www.bdl.lasy.gov.pl

[12] [Online] available at: http://www.dendrogen.pl

[13] Bednarek-Michalska B 2012 *Bulletin EBIB* **8** (135) ???

[14] Kędzierska E, Kavalchuk N and Stepniak J 2014 *Conference on Grey Literature and Repositories: proceedings 2014: the Value of Grey Literature in Repositories* [Online], National Library of Technology

[15] Stępniak J 2014 *Seminar Open science and open communication model*, Warsaw University of Technology

[16] Wilson J A J 2015 *Conference Open Access to Research Data as a Driver for Open Science*, The University of Oxford