

Preface

Towards standardisation, harmonisation and integration of data from heterogeneous sources for funding and evaluation purposes

Wolfgang Glänzel^{1,2} · Hans Willems³

Received: 8 December 2015 / Published online: 28 December 2015
© Akadémiai Kiadó, Budapest, Hungary 2015

Funding and performing organisations and other entities use data from heterogeneous sources for evaluating research performance or allocating funding. Data on research and technology output, scientific personnel and research projects that are collected and obtained using various approaches including centralised and decentralised, top-down and bottom-up, open and proprietary data. Furthermore money for financing research, innovation and technological development is spent at various levels, ranging from supranational organisations over governments down to the regional and local, intra-institutional level. Appropriate quality of these data, compatibility, interchangeability as well as their connectability with related data is increasingly a need and a necessary criterion. Proper definition, suitability for indicator building and reporting and the various application purposes are the determinants in the process of standardisation, harmonisation and integration. But this is also the reason why full compatibility cannot be imposed and complete concordance of structures cannot be granted.

At the 20th International Conference on Science and Technology Indicators held in Lugano (Switzerland) on 4–9 September 2015, we had the opportunity to organise a panel session with special focus on the harmonisation and integration of data relevant for the evaluation of research and current information systems. The idea goes back to an initiative of the work group *Research Policy and Programme Evaluation* launched by Science Europe. Unlike the special session on “Grand challenges in data integration for research and innovation policy: handling big data, coping with quality issues and anticipating new policy needs. State of the art and future perspectives” organised during the ISSI 2015

✉ Wolfgang Glänzel
wolfgang.glanzel@kuleuven.be

¹ ECOOM and Department of MSI, KU Leuven, Leuven, Belgium

² Department of Science Policy and Scientometrics, Library of the Hungarian Academy of Sciences, Budapest, Hungary

³ The Research Foundation – Flanders (FWO), Brussels, Belgium

conference in Istanbul (see <http://www.issi2015.org/en/Workshops.html>), where data integration was discussed in a broader scope, for the STI/ENID panel session we have collected four contributions reporting on specific issues and efforts recently made in Europe in this context. Different approaches have been outlined that showed perspectives and plotted roadmaps in data standardisation, harmonisation and integration, but also pointed to challenges, caveats and limitations of these processes. The presentations were followed by a panel discussion with the participation of the speakers and the attendees of the session.

In the following we briefly summarise the topics of the four contributions presented in Lugano. The four papers are part of a special section of this journal issue dedicated to the STI 2015 panel session on data standardisation, harmonisation and integration.

The first contribution presented by Glänzel et al. was prepared by a task group launched by Science Europe within the work group Research Policy and Programme Evaluation. The goal is the identification of issues experienced by the Member Organisations (MO) regarding collection, standardisation and treatment of data related to the analysis and ex-post evaluation of activities funded or performed by MOs and to propose solutions for these issues. This objective is directly related to their management needs, for example, to difficulties with the implementation of unique identifiers, output classification in interdisciplinary research areas or in the social sciences, arts and humanities. In order to achieve the objective, the project aims first at mapping the state of affairs in data collection and their use at European funding and research organisations. This is implemented through a survey sent to the MOs on “Data collection and use in research funding and performing organisations”. Questions are organised in nine thematic blocks. The survey is analysed with special attention to the particular needs of funding and performing organisations. Based on this survey, the task group will work further on methodological advice and standardisation of data for evaluation purposes.

This talk was followed by a report (Biesenbender & Hornbostel, IFQ, Berlin) on the standardisation of research information within the framework of German federal standardisation project for research information. The decentralised research system in Germany has actually led to the emergence of different information systems. The “Research Core Dataset” (RCD) based on recommendations adopted by the German Council of Science and Humanities therefore focusses on providing standardised rules and approaches for the operationalisation and measurement of research information to support universities and research institutions in the design and organisation of their individual research information systems. For the implementation of the project four working groups were established, particularly, *Definitions and data formats*, *Technics and interfaces*, *Bibliometrics* and *Subject classification*.

The third contribution presented by Sivertsen (NIFU, Oslo, Norway) used the example of Scandinavia to focus on the aspect of data integration in the context of national current research information systems. During the last decade Scandinavian countries have made an effort to integrate bibliographic data sources with their national research information systems. The “Current Research Information System In Norway” (CRISin) was one of the pioneering systems and aims at recording and promoting publication data, projects, units and competency profiles. The system or parts of the system have served as a model for national research information systems in Scandinavia but also in other countries and regions of Europe. The presentation reports on challenges, among others, on “the integration of institutional current information systems on the national level” and “the integration of current research information systems with systems for project steering in funding organizations”.

The fourth presentation by Daraio et al. (Sapienza University of Rome, Italy) introduced the Italian ontology project that aims at providing the groundwork, platform and tool for the efficient integration of heterogeneous data for the purpose of research assessment. In particular, the *Ontology of the Multi-Dimensional Research Assessment* with its underlying Ontology Based Data Management (OBDM) approach is described as a powerful tool for the coordination, integration and maintenance of various data needed in the framework of Science, Technology and Innovation policy. The OBDM approach, which is implemented in the Sapienza ontology, provides a transparent platform for the evaluation process including the unambiguous definition and specification of indicators for evaluative purposes and to track their evolution over time. It allows to measure and analyse repercussions on scientists' behaviour and to monitor the changes in the established evaluation criteria and their consequences on the research system as well. Finally, the approach might also be able to foster the scientists' involvement in the evaluation process.