Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 106 (2017) 87 – 94

Procedia Computer Science

13th International Conference on Current Research Information Systems, CRIS2016, 8-11 June 2016, Scotland, UK

R&D Statistics Information System: An Interoperability Tail between CERIF and SDMX

Dimitrios C. Karaiskos[a]*, Dimitrios Xinidis[a], Vasilis Bonis[a]

[a]*National Hellenic Research Foundation/National Documentation Center, 48 Vassileos Constantinou Ave., Athens, 11635, Greece*

Abstract

Research and Development statistics (R&D statistics) provide valuable information on the expenditure spent and personnel engaged in R&D activities in a country, knowledge that facilitates the understanding on how R&D output contributes to economic growth and societal wellbeing. This endeavor requires a sound evidence base which is succeeded through internationally comparable statistics and a common survey methodology and conduct per country as part of its national official statistical program. For this purpose the National Documentation Centre of Greece (NDC or EKT using the Greek abbreviation), the designated organization for the collection and compilation of the Greek R&D statistics, build the R&D Information System to automate and support this specific business activity. This paper aims to provide an overview of the implemented R&D Information System but especially to focus on the adoption of CERIF and SDMX standards and their integration. CERIF was selected as the systems' data model for its metadata representation capability and its high flexibility in forming semantic relationships while SDMX was adopted as the statistical data and metadata exchange standard. The integration of the two standards and their interoperability enables data and metadata quality maintenance, archiving and access while at the same time ensures valid and automated interchange of statistical information with national and international statistical offices.

Peer-review under responsibility of the Organizing Committee of CRIS2016

Keywords: CERIF; SDMX; R&D Statistics; Systems Interoperability

* Corresponding author. Tel.: +30 210 7273 945
E-mail address: karaiskos@ekt.gr

Peer-review under responsibility of the Organizing Committee of CRIS2016
doi:10.1016/j.procs.2017.03.039

1. Introduction

Research and Experimental Development statistics (R&D statistics) provide valuable information on the expenditure spent and personnel engaged in R&D activities in a country, information that is vital for understanding how R&D outputs contribute to economic growth and societal wellbeing. This endeavor requires a sound evidence base which is achieved through internationally comparable statistics and a common survey methodology and conduct per country as part of its national statistical program (based on Commission Regulations 995/2012). In this context, the collection and compilation of the Greek R&D statistics, has been assigned to the National Documentation Centre of Greece (www.ekt.gr) and is supported and carried out by a data management system, namely R&D Information System, implemented in-house specifically for this purpose. The R&D Information System was based on relevant international standards -such as CERIF and SDMX-, robust technologies and best practices. This paper aims to provide an overview of the developed R&D Information System but especially to focus on the adoption of CERIF and SDMX and the integration implemented between them. The next sections will attempt to shed light on the usage of both standards and will describe with detail the integration and interoperability between the two standards.

2. Background

2.1. R&D Statistics

The aim of the R&D (Research and Experimental Development) survey is to produce statistics about (intramural) R&D expenditure and R&D personnel covering R&D performing entities in the private and public sectors and for the country as a whole[1]. The R&D data collection is promoted by EC regulations 995/2012 (from reference year 2012 onwards) and it is carried out on a mandatory basis by each country, each year, according to EU regulations[2, 3, 4]. Subsequently, the country level R&D statistics are relayed to major statistical bodies, such as Eurostat and OECD, who aggregate and compose statistics reports regarding EU and/or the world in various levels of detail. For the R&D statistics aggregation to occur, internationally comparable statistics and a common methodology and language is required. This is provided by the Frascati Manual[1] which outlines all the basic statistical concepts and definitions, standard classifications and guidelines for the production of R&D statistics.

R&D statistics have gained momentum and have increased international interest due to the fact that the financial resources devoted to the implementation and support of R&D activities have been linked to the economic development of a country. A concrete example is the EU2020 headline target of investing 3% of the EU GDP on R&D which is monitored by the 'R&D intensity' indicator (R&D expenditure as a percentage of GDP)[5]. Furthermore, the decision to treat expenditure on R&D as a capital investment in the System of National Accounts (SNA)[6], i.e. accounted positively in national GDPs, has also attracted great attention to the flows of funds for R&D and its robust tracking and measurement. Thus, R&D statistics form the basis for the development and monitoring of policies at European and national levels.

2.2. SDMX Standard

The main purpose of SDMX (an ISO standard ISO 17369:2013 since 2013) is to tackle the problem of a common model for the representation of statistical data and metadata and to provide a set of guidelines and standards for exchanging statistics information between peers, e.g. international organizations and states. The guidelines of SDMX are concerned not only with data but also with metadata which are introduced as an integral part of the standard. Regarding information exchange, SDMX describes the structure of the content and the supported system architectures. It provides the flexibility for each organization that chooses to use it, to structure its data according to their specific informational needs and usage scenarios. The SDMX message formats have two basic expressions, SDMX-ML (using XML syntax) and SDMX-EDI (using EDIFACT syntax and based on the GESMES/TS statistical message). EKT uses the SDMX-ML format since it is the format that Eurostat's web services support and because of the extensive usage of XML in the systems that EKT develops and uses.