

CRIS 2014

The implementation of CERIF based data model for statistical survey of Research and Development Potential within the SK CRIS

Danica Zendulkova*

Slovak Centre of Scientific and Technical Information, Lamačská 8/A, Bratislava, 84103, Slovakia

Abstract

National information system about research, development and innovation SK CRIS is complex research information system which has been based on the CERIF data model. It collects selected information on organisations, projects, researchers, R&D results and research infrastructure acquired with the support of public funds. Its realization has been assured by the national project NISPEZ funded from EU structural funds. The SK CRIS was put into operation in January 2013. The SKCRIS system provides also some modules implemented outside the actual version of CERIF data format apart from the mentioned entities and their attributes. Additional Statistical Survey of Research and Development Potential is one of them.

© 2014 The Authors. Published by Elsevier B.V.

Peer-review under responsibility of euroCRIS.

Keywords: Slovak Centre of Scientific and Technical Information; Slovak Current Research Information System (SK CRIS); CERIF; statistical indicators; statistical survey; research and development potential;

1. Summary

A Data model of the SK CRIS consists of two parts: database CERIF is compatible with CERIF 1.3 data format[6] and CERIF inspired database STATISTICS contains indicators out of actual CERIF frame. The data

* Tel.: +421-918-976-304;

E-mail address: danica.zendulkova@cvtisr.sk

about Research and Development Potential on national level, yearly collected to this database are used by decision makers for set the parameters of state science policy and aggregated reports are accessed to public.

2. Statistical survey in nutshell

Statistical Survey of Research and Development Potential follows and complements statistical surveys of Slovak statistical Office in area of science and it is announced by the guarantee of the SK CRIS – the Science and Technology Division of Ministry of Education, Science, Research and Sport of the Slovak Republic. The survey is administrated by the system SK CRIS provider - Slovak Centre of Scientific and Technical Information.

All organizations engaged in scientific research activities funded from state resources are required to complete and submit the form "Annual Report on research and development potential". The data collection runs in electronic form on the SK CRIS website [1].

3. Indicators

The survey focuses on several groups of indicators about R&D organizations [4] in the reference year. They concerns mainly of innovations as results of research and development process, its objectives, selected economic indicators, structure of R&D activity (in FTE and natural persons), substantive focus of R&D activity, laboratory equipment, structure of outcomes and international co-operation: The form consists of the following parts:

- Innovations
- Selected economic indicators of legal / natural person research and development (R & D) in the year
- The structure of R & D activities
- The substantive focus of research and development activities
- The level of laboratory equipment
- Structure of research and development outputs
- Publications, citations and patents
- Participation in the international governmental or non-governmental organization in the field of science and technology
- International scientific and technical contacts and cooperation: international R & D projects
- Funding for international R & D projects
- Professional stays abroad - length of stay and fields of science and technology (S & T)
- Professional stays abroad - the country of residence
- Professional stays from abroad - stay and sort by field of S & T
- Professional stays from abroad by country of residence

4. Data model

For statistical survey the separate data model inspired by CERIF attributes was defined. The goal is to collect to SK CRIS all required information, assure the maximal possible quality and completeness and to avoid of duplicate insert of data already put to the system at the same time. Consequently, the communication between statistical module and core entities, and especially registry of organization is very close (Figure No 1). Applies that CERIF data entities are independent of other data structures. This means that CERIF data model without additional data structures may work, but the additional modules do not work without the connection on CERIF entities. Some data from the R&D Potential data collection, however, are put to the database as CERIF data, because fully meet its structure. For example, the indicator "membership of R&D subject in international organizations" is interpreted as link entity *OrgUnit_OrgUnit*.

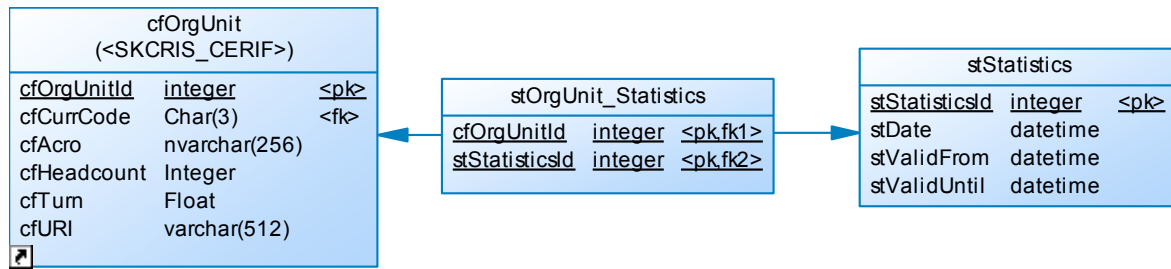


Figure 1 Link of Form R&D Potential to Organisation as CERIF entity [3]

Data model for statistical survey contains the definition of category groups (*stCatGroup*, *StCatGroupDesc*), which correspond to the range of data, i.e. to tables of form. The value of item in classification schemes is defined as category (*stCategory*, *stCategoryDesc*), see Figure No 2.

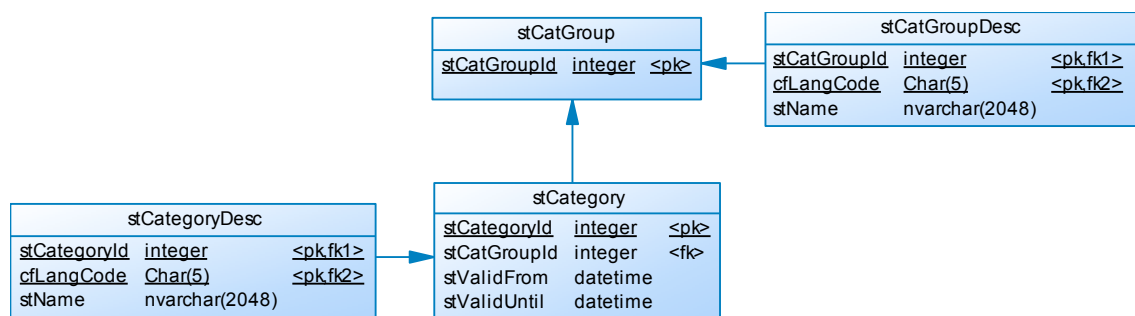


Figure 2 Category groups and Categories [3]

The data model can be illustrated on a form table: *Implemented innovation*. The data completed to form and same data from SQL table *stInnovFunding* are shown in Figure No 3.

	stStatisticsId	stCategoryId	stAmountPublic	cfCurrCodePublic	stAmountOwn	cfCurrCodePublicOwn	stAmountOther	cfCurrCodePublicOther
1	2	1101001000001	2553.79	EUR	0	EUR	0	EUR
2	2	1101001100001	1297.92	EUR	0	EUR	0	EUR
3	2	1101001200003	10	EUR	0	EUR	0	EUR
4	2	1101001300001	10	EUR	0	EUR	0	EUR

Realizované inovácie			
Typ inovácie	Typ finančného zdroja		
	Financovanie z verejných zdrojov (v tisíckach EUR)	Financovanie z vlastných zdrojov (v tisíckach EUR)	Financovanie z iných zdrojov (v tisíckach EUR)
Produktu	2553.79 EUR	0.00 EUR	0.00 EUR
Procesu (technologická)	1297.92 EUR	0.00 EUR	0.00 EUR
Organizačná (nové metódy organizovania vonkajších vzťahov s ostatnými organizáciami)	10.00 EUR	0.00 EUR	0.00 EUR
Marketingová (významné zmeny dizajnu alebo obalu výrobkov a služieb)	10.00 EUR	0.00 EUR	0.00 EUR

Figure 3 Realized Innovation: Form and SQL data

On the Figure No 6 we can see that the organization has filled four types of innovation, funded entirely from public sources. Field *stStatisticsId* expresses assignment of these innovations (categories) to one form. The *StCategoryId* field contains identifiers for following type of innovations: [2]

- Product
- Process(technological)
- Organization (new methods of organizing external relations with other organizations)
- Marketing (significant design changes or pack of goods and services changes)

5. Process of data collection

At the beginning of data entry the user must create an account in the system SK CRIS to give a record of the researcher containing his personal profile. Subsequently, a record of the organization that registered person represents and for which it is responsible is assigned to the researcher record. The system is able create the link between the organization and the researchers automatically in case that this information can be found in the existing database. Otherwise, user has to register his organization itself through the application form dedicated for registration of the organization. Filled form of Research and Development Potential is put to personal space of a registered organization.

The screenshot displays the 'Osobný priestor organizácie' (Organization's personal space) interface. On the left is a sidebar menu with categories like 'Kmeňové dáta', 'Publikácie', 'Citácie', 'Inovácie', 'Patenty', 'Produkty', 'Podujatia', 'Laboratória', 'Služby', 'Zariadenia', 'Certifikáty', 'Dáta VVP', 'Hodnotenie spôsobilosti', 'Ocenenia', and 'Dokumenty organizácie'. The main area shows the 'VVP' (Research and Development Potential) form. At the top right of the main area is a button labeled 'Nový formulár'. Below it is a table with columns 'Dátum podania', 'Rok', and 'Stav'. The table contains one row with the values '27.09.2013', '2012', and 'odoslaný'. To the right of the table are buttons for 'Formulár' and 'Detail'. Below the table are navigation controls including arrows and a yellow button with the number '1'.

Figure 4 Personal space of organization – Filled form saved on Portal SK CRIS

From this part of the user can access historical forms filled in the past by pressing the "form" approach (Figure No 4), open saved (unfinished) form, and also could create a new form. It is necessary to create a new form every year.

The system can automatically retrieve data from existing records of organizations, thus eliminating having to re-entry data such as Organization name, address or Organization ID. It also eliminates the obligation to fill the entire table: the user chooses always the relevant items of the available options (Figure No 5). The SK CRIS system is also able to identify and locate missing or erroneous data on related parties of form by using validation tools.

Figure 5 R&D Potential Form, data entry, part Realized Innovation

The SK CRIS includes a tool that allows by simple database script to find out who has difficulties in completing the online form. SK CRIS user generally starts to fill the form but do not send it to the database because of errors. In case that user is unable to deal with the problem by studying the available documentation, he should contact the operator. Approximately 20 percent of those who was finally successful in form filling, contacted the operator only on e-mail reminder. In Figure 6 is a command in SQL environment to create a list of organizations that have successfully sent online form.

```
select s.stStatisticsId, stDate, o.cfOrgUnitId, p.cfPaddrId, skICO, cfName,
a.cfAddrline3, a.cfAddrline2, a.cfAddrline1, cfPostCode, cfCityTown
from stat.stOrgUnit_Statistics o, stat.stStatisticsState s, certif.cfOrgUnitName n,
cerif.cfOrgUnit_Paddr p, certif.cfPaddr a
where stcategoryID='1101019000002' and (o.stStatisticsId=s.stStatisticsId)
and (p.cfOrgUnitId=o.cfOrgUnitId) and (p.cfPaddrId=a.cfPaddrId)
and (o.cfOrgUnitId=n.cfOrgUnitId) and cfLangCode like '%SK%'
and skEndDate='9999-12-31 00:00:00.000' and p.cfenddate='9999-12-31 00:00:00.000'
order by cfName
```

Figure 6 Select of organisations just sending filled form to database

6. Aggregated report

As it is common in statistical surveys, also the result of mentioned survey is aggregated report [5] containing several tables of indicators mentioned in part 2 of this paper. Report generation means process of prepared SQL scripts and procedures running with year as parameter.

Data collection of R&D Potential for year 2012 was the first realized larger data collection in new SK CRIS. Within the process of collection 339 organisations inserted theirs statistical data. It means that 339 forms had been collected from which aggregated report was generated. One table, part of aggregated report is on Figure No 7.

7. Conclusion

Statistical Survey of Research and Development Potential is substantial tool to obtain information about Slovak R&D. Current use of collected data is following: Results of the survey have been presented on Central Information Portal for Research, Development and Innovation [5] in the form of aggregated tables. The reports are used by decision makers for set the parameters of state science policy. Selected data from this survey are also used to assess the potential of applicants for support in technology transfer [7]

2. Všetky právnické osoby VV a fyzické osoby VV podľa formy hospodárenia

Forma hospodárenia	Počet fyzických a právnických osôb
Neznáma	4
Rozpočtová organizácia	44
Príspevková organizácia	39
Nezisková organizácia	11
Hospodárska organizácia	112
Vysoká škola	19
Verejnoprávna inštitúcia	3
Fyzická osoba	7
Spolu	239

Figure 7 R&D Organisations by economy type

Implementation of data CERIF format to module of statistical indicators and its integration into the CRIS system would allow comparability of collected data, would increase their quality and contextual linkages that would open the next possibility of use of these data, especially in the field of measurement and evaluation of research.



References

1. The SK CRIS. Information system on Science and Research. [Online] [cit. 25.03.2014] Retrieved from: <<https://www.skcris.sk>>
2. Oslo Manual: Guidelines for Collecting and Interpreting Innovation Data . [Online] [cit. 25.03.2014] Retrieved from: <<http://dx.doi.org/10.1787/9789264013100-en>>
3. Valkovič, L. (2011): Detailná funkčná špecifikácia systému SK CRIS. Bratislava, 2011. [cit. 25.03.2014]
4. Valkovič, L.: Štúdiá o optimalizácii ukazovateľov o výskume a vývoji. Informačný systém o výskume a vývoji rešpektujúci odporúčaný dátový model CERIF . Bratislava, 2011. [cit. 25.03.2014]
5. Centrálny informačný portál pre výskum, vývoj a inovácie. Štatistika. [Online] [cit. 25.03.2014] Retrieved from: <https://www.vedatechnika.sk/SK/VedaATechnikaVSR/Stranky/StatistickeUkazovatele.aspx>
6. CERIF 1.3. [Online]. [cit. 25.03.2014] Retrieved from: <<http://www.eurocris.org/Index.php?page=CERIF-1.3&t=1>>
7. Shearman, A. a kol.: Metodika hodnotenia kvality vedeckých pracovísk v oblasti výskumu, vývoja, inovácií a transferu technológií. /Zost. A. Shearman. Bratislava, CVTISR, 2012. [Online]. [cit. 25.03.2014] Retrieved from: http://nitt.cvtisr.sk/buxus/docs/METODIKA_-Hodnotenie_kvality_VV_pracovisk_1.pdf