The system SK CRIS, scientific publications and theses – mirror of Slovak science

Ján Turňa, Juraj Noge, Danica Zendulková
Slovak Centre of Scientific and Technical Information, Bratislava, Slovakia

Summary

The Slovak Current Research Information System (SK CRIS) as a part of the Central Information Portal for Research, Development and Innovation (CIP RDI) has the status of Information system of public administration, defined by the Act 275/2006 on information systems of public administration. The Slovak Centre of Scientific and Technical Information (SCSTI) is responsible for operating, maintaining and providing technical support to this system on behalf of the Ministry of Education, Science, Research and Sport of the Slovak Republic (ministry).

The new R&D information system, the SK CRIS, is being developed within the Activity No 4 of the NISPEZ[1] national project: *Enlargement of Central Information Portal for Research, Development and Innovation (CIP RDI) with new functionalities complying with EU standards.* The system will be focused on effective work with research projects including a connection with other, science related information systems. In the process of system building, the experience of SCSTI with the R&D information system building and operation is being utilised and generally accepted principles of best practice for current research information systems are followed.

1 SK CRIS content

The content structure of SK CRIS is based on current CIP RDI system. It means that all types of data collection will be maintained. But, the content of SK CRIS will be more complex and more specified (Figure No 1) than in the actual system. The details of core content entities of our forthcoming R&D system SK CRIS are presented below (Cvik, O. 2010).

1.1 Registry of Projects

The Research and Development Project is considered as main object (entity) of current research information system. It contains the most relevant information for users from different target groups. Other entities (mainly organisations, researchers, publications, but also events, facilities, equipment, services etc.) are entered into the system mainly via the relationship to any R&D project.

[1] National Information System Supporting Research and Development in the Slovak Republic – Access to Electronic Information Resources, http://nispez.cvtisr.sk/

The Registry of projects has been built up since 2000 as part of different versions of R&D information systems. Currently, it contains the data about approximately 7000 projects. There are still some R&D projects implemented by Slovak organisations that are not in the system. However, we believe that in the new system we will be able to acquire data about each project funded by public resources, mainly due to interconnection with other relevant science related systems, as mentioned hereafter.

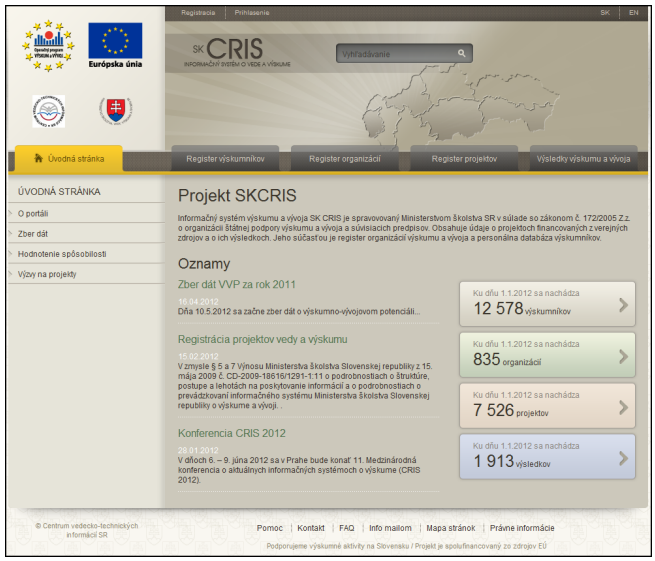

Figure 1 The home page of SK CRIS portal

At the moment, we are in the process of developing the integration interface interconnecting SK CRIS with other systems collecting the data about different categories of projects. This will allow an automated data import from external systems to SK CRIS. The aim is firstly to have a comprehensive data about research projects, and secondly there is an effort to avoid duplicity of data entries from different systems. The systems that will be connected to SK CRIS are mainly the following ones: system of the Slovak Research and Development Agency (SRDA); Scientific Grant Agency VEGA; and the system of Incentives for Research and Development. Nevertheless, the ambition is to integrate data from other relevant systems as well.

1.2 Registry of Organisations

Registry of organisations has been built from several information sources. The first one is annual additional statistical survey of R&D Potential of the Slovak Republic. The information about which organisations carry out research and development is also provided by the Slovak Statistical Office, which receives information about activities of organisations from statistical reporting.

The second information source on organisations comes from historical data in our system. Whereas the additional statistical survey of R&D Potential has been carried out since 2000, there

is a significant number of forms filled in the past. This data is stored in the system and it needs to be updated annually.

The part of entries into the Registry of organisations is being acquired also during the process of Certification to perform R&D, which is administrated within the information system. As certification is compulsory for each organisation applying for the support for research from public sources, the registry includes not only universities, Slovak Academy of Sciences and state research organisations, but also a significant number of research organisations from non-for-profit and private sector.

Currently, the Registry of organisations contains around 800 subjects while it is expected that by the end of 2012 it will contain more than1000 R&D organisations.

1.3 Registry of Researchers

Registry of researchers has been established on the basis of data from actual database of experts – evaluators of R&D projects funded by the ministry. The database currently contains more than 1100 personal records.

The overall goal is to have all scientists (about 23000 researchers) in the registry. However, as reaching this objective is a rather challenging task, in first phases of the new system operation the intention is to cover significant and representative part of scientific community - researchers from universities, Slovak Academy of Sciences (SAS) and from state operated research institutes. The rate of registry filling with relevant data will depend mainly on successful automated import of entries from other relevant information systems (registry of employees in higher education sector, registry of SAS employees etc).

The other data source on researchers is the registration of research projects. It is possible to extract relevant researchers' data (name, affiliation, role in the project) from successful grant applications administrated by different ministries and grant agencies.

Except for the data sources described above, the basic data (name, surname, organisation) about researchers from universities can be extracted also form the Central Registry of Publication Activity, which is operated by the SCSTI (more information below).

We assume that the base records about researchers created on the basis of publicly accessible sources or imported from other systems will be continually updated by respective researchers, or by repetitive imports of updated data. The result will be the bilingual up-to-date database of researchers containing professional profiles that will serve for searching experts according to scientific area, but also for presentation of scientific capacity in Slovakia and abroad.

1.4 Registry of R&D Results

The registration of R&D results as an independent entity is new in the Slovak research information system. Up to now, the results of projects have been registered, but they are not published on web. The similar situation applies for registration of researcher's publication activity. The publications inserted by researchers into the database are not accessible to public.

In SK CRIS system all results will be published. It will not be only bibliographic data about publications, but also data about intellectual property objects (referred to as patents, trademarks,

designs, utility models etc.) and data about other results of R&D activities (e.g. products and innovations).

For obtaining relevant data, the interconnection of SK CRIS with Central Registry of Publication Activity is being prepared. There is also an intention to use the data collected as R&D results (project results, researchers' publications and results) inserted by organisation in applications for Certification to perform R&D. The results will be continuously updated.

1.5 Other information

Other information means additional second level entities that substantially supplement the core and result entities mentioned above. These are the calls for projects, finances, laboratory infrastructure (facilities and equipment; offered services), but also events, awards and different full text attachments, i.e. CV of researchers.

2 SK CRIS data model

Implementation of the data format CERIF will simplify and clarify the data entry, but also their presentation as mutually related objects. CERIF (Common European Research Information Format) is XML data format to support the management of Research Information. It is recommended by the European Commission as standard for automatization in area of R&D information (CERIF 2012). The CERIF data format is based on data model, which allows metadata representation of research entities, their activities, interconnections and their output (results). The CERIF elements have defined the core structure, semantics and link entities and they are divided into hierarchical categories (Jörg, B. 2009).

Besides core entities (Organisation, Project, Person) and result entities, SK CRIS will contain statistical data (R&D potential and selected indicators about projects and laboratory infrastructure). Part of it will also be the administration of applications for Certification to perform R&D. The online application form contains several indicators outside the data format CERIF scope. Data model SK CRIS is schematically shown in Figure 2.

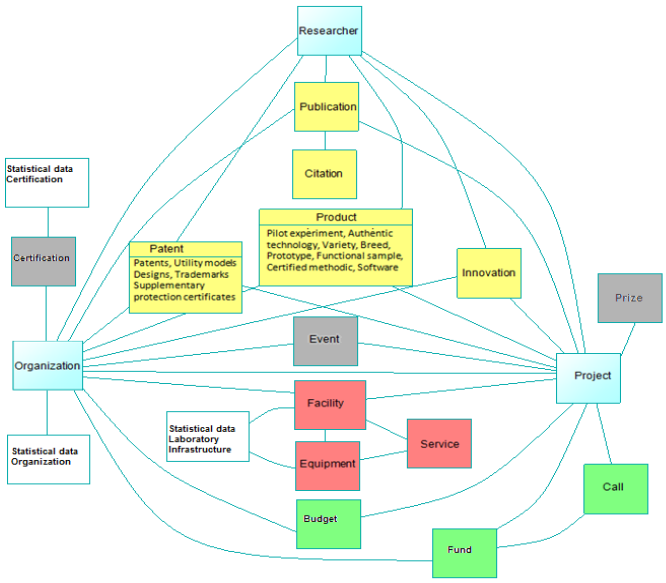

Figure 2 Data model SK CRIS (Valkovič, L. 2011)

The part of SK CRIS data model is also CERIF semantics, including different code lists and classification schemes. Their use results from Slovak legislation. The example of semantics and link entities between core and result entities is shown in Figure 3.

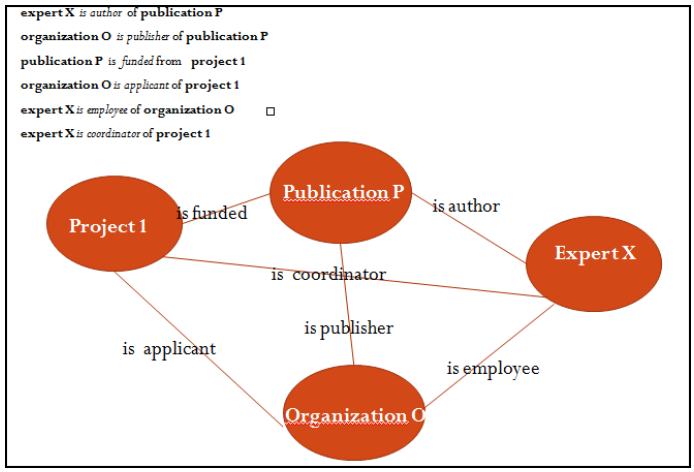

Figure 3 Linkage between entities and semantics use schema

3 Technical specifications of SK CRIS

The portal part of the software solution of SK CRIS system will be based on Liferay portal. This technology is suitable for portal development and for implementation of portals including publishing systems. However, it should also serve as development platform, as integration tool for different applications and as presentation layer.

The presentation layer functionality of SK CRIS system including CMS system will be ensured by Liferay portal. Portlets implemented through portal technologies will ensure the functionality and configuration for electronic online forms for data entry. Workflow solution allows configurable process implementation associated with electronic form processing. The support for document store and process is also integrated in Liferay portal (Jackrabbit). Data persistency will be implemented by using Hibernate libraries. The structure of solution is shown in Figure 4.

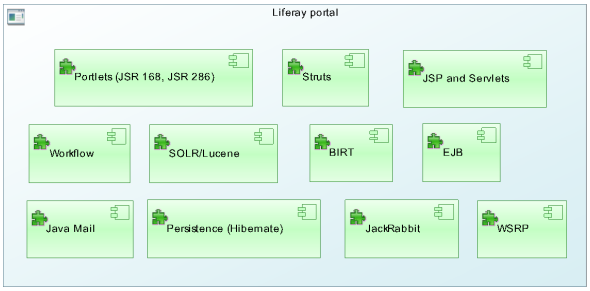

Figure 4 Liferay portal solution (Valkovič, L. 2011)

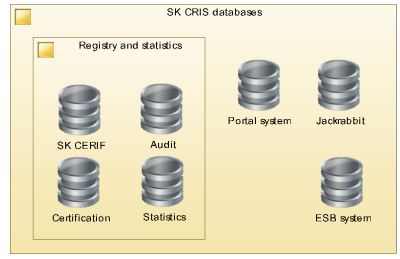

Figure 5 Databases SK CRIS (Valkovič, L. 2011)

Databases are divided into two groups (Figure 5):

- Registry databases (SK CERIF, Certification, Statistics, Audit)
- System databases (system database of Liferay portal, database for document storage system, ESB database)

SK CRIS database is based on the standard CERIF 2008 v1.2. Databases Certification (applications for Certification to perform R&D) and Statistics (additional statistical survey of R&D Potential) contain data collected about each entity outside the scope of CERIF data model SK CRIS.

The functionality of prepared system is user oriented. The search interface contains the possibility of use of different search techniques: simple, advanced, but also full text searching and faceted navigation. Bilingual Slovak and English user interface with selected data by CERIF requirements will access the information about Slovak science to users from different countries worldwide. The specification of user interface functionality was inspired by the results of European CRIS systems desk research conducted as part of the SK CRIS analysis in 2009-2010 (Zendulková, D. 2011b).

The data already present in the system will not be required to be entered again. When online form opens, the relevant archived data will be loaded automatically. The system allows to track the history of data that can be changed (surnames, organisation names, etc.)

The search results will present mutually related information. Figure 6 contains following example: We started searching in project registry. The detail of record about selected project contains also the information about responsible organisation and project team. The personal and institutional names are clickable; it means that they are linked with detail also visible in Figure 6. The mutually related objects are linked also in scope of organisation detail (link to solved projects and researchers) and in scope of researcher detail (link to solved projects and employer's organisation).

4 Link with Registry of publications and Registry of theses

There are also other systems being used beyond the SK CRIS, which contain a lot of important information related to Slovak science performance. It is mainly *The Central Registry of Publication Activity* and *the Central Registry of Theses and Dissertations*. The SK CRIS aims to include and present this information in relevant context as well. Due to this fact, the system will provide a comprehensive overview of the Slovak science.

4.1 CRPA

The Central Registry of Publication Activity (CRPA) accounts for a unique source of information on publications of Slovak universities, which is available to both professional and non-professional public. The SCSTI has been responsible for processing and verification of the data, formal control of bibliographical records of publication files and correctness of categorization of publications. The data from registers serves the ministry in the process of calculating and allocating public funds for the state universities, as well as for mapping results achieved in scientific research carried out at Slovak universities.

Currently, 28 Slovak universities contribute to the CRPA database (20 of them are public, 2 are state operated and 6 are private). The database has been built since 2007. It is updated at the end of every year and it contains the data for the period 2007-2011, together more than 212000 publications (Hrčková, Ľ. 2011).

4.2 CRTD

The Central Registry of Theses and Dissertations (CRTD) has also been created following the initiative of ministry. The primary goal of the registry is to manage collection and archiving of final examination works and second doctorate works (theses, dissertation and qualification works) from universities and colleges in Slovakia. The CRTD, together with selected Internet resources serve as a comparative corpus for the System for plagiarism detection, called also the Anti-plagiarism system (APS). Once the work is checked for plagiarism, an output protocol is produced for each work, which serves the exam commission as a tool for deciding about the work originality (Noge, J. 2011). The system has been running since 2009 and the number of registered theses currently exceeds 160000 works.

The interconnection of SK CRIS with CRPA (its enlargement by publication activity of non-university R&D organisations is expected) will allow to get data about the significant volume of R&D results and also the core data about researchers – publications'authors.

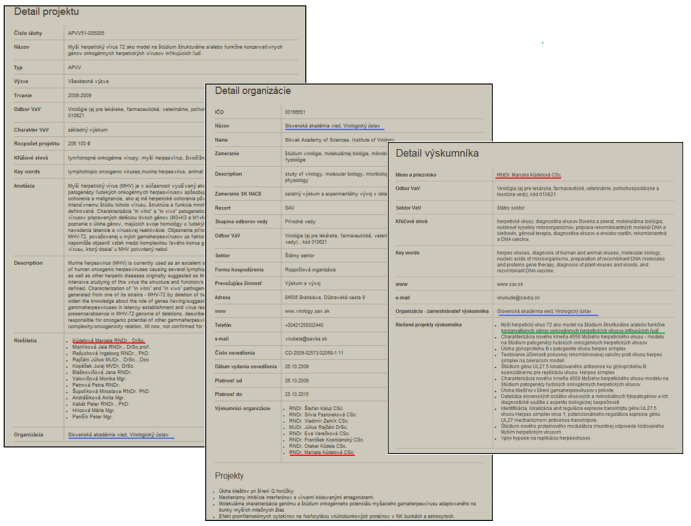

Figure 6 SK CRIS: Details of project, organisation and researcher. The related (underlined) objects are mutually hypertextually linked and it is possible to open them by click

The simplifying functionality for research results entry will be a part of SK CRIS. System user receives during the entry of research results the verified list of publications from CRPA. He/she identifies relevant publications, which eliminates duplicity of the same data entries to both systems.

For SK CRIS users, the data integrated from CRPA will complete the field of research results that have been continuously inserted to the SK CRIS. In Figure 7 we can see the publication registered to CRPA in UNIMARC data format (on the left). After that, as this publication was included into SK CRIS, it becomes the result of registered project, obtains the relationship to the project, responsible organisation and to authors – researchers. The figure also shows that four of six authors are employees of Comenius University - Jessenius Faculty of Medicine and two are not (by SK CRIS data) employees of university. The results will be shown in each project and publications are displayed by researcher. In the same way, it will be possible to know relevant information from the CRTD.

There are several aims of putting SK CRIS into operation. These could be divided into four groups (Zendulková, D. 2011a):

1. The presentation of information and new knowledge and its interchange between scientists and researchers in electronic environment.

2. Accelerate the application of knowledge into the practice (technology transfer), information about research results for enterprises and business and their involvement in active utilisation of this knowledge in practice.

3. To obtain general support for science and research. To provide information for decision makers and for state administration about scientific and research activities and results.

4. Popularisation of science and technology, including RTD results in a comprehensible way attractive for public. These activities should positively influence the perception of science by society.

The SK CRIS system supplied by the Slovak corporation InterWay s.r.o. will be put into operation in the second half of 2012. We believe the system will be welcomed by scientific community, decision makers, R&D management bodies, entrepreneurs, technology transfer stakeholders, but also by media and wider public. The new system has a challenging ambition to become the mirror of Slovak science.

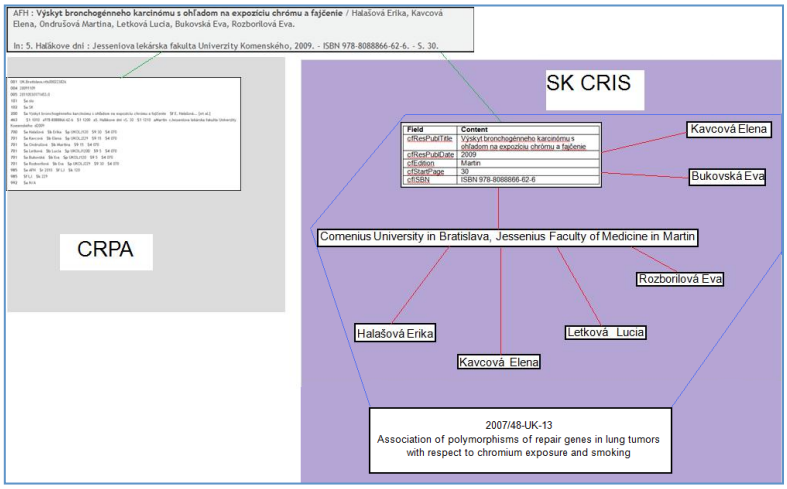

Figure7 Publication CRPA in context of the SK CRIS

Acknowledgements

Európska únia
Európsky fond regionálneho rozvoja
Operačný program VÝSKUM a VÝVOJ
Supporting research in Slovakia / Project co-financed from the EU funds

References

Centrálny informačný portál pre výskum, vývoj a inovácie. Informačný systém. (2012) [Online]. [cit. 05.03.2012] Retrieved from: <https://www.vedatechnika.sk/SK/ISVVP/Stranky/default.aspx>

CERIF: the Common European Research Information Format. (2012) [Online]. [cit. 05.03.2012] Retrieved from: <http://cordis.europa.eu/cerif/>

Cvik, O. (ed.) (2010): Národný projekt NISPEZ, aktivita 4.1 Analytická štúdia – Súhrnná správa. Bratislava, CVTI SR, 2010 . [Online]. [cit. 05.03.2012] Retrieved from: <http://nispez.cvtisr.sk/index.php?&menu=107>

EuroCRIS. (2012) [Online]. [cit. 05.03.2012] Retrieved from: <http://www.eurocris.org>

Hrčková, Ľ. (2011): CREPČ a CREUČ ako unikátny zdroj informácií o publikačnej a umeleckej činnosti vysokých škôl v SR. Inforum 2011. Konference o profesionálních informačních zdrojích. Praha, 24.-26.5.2011 [Online]. [cit. 05.03.2012] Retrieved from: <http://www.inforum.cz/pdf/2011/hrckova-ludmila.pdf>

Jörg, B. (2009) : CERIF: Spoločný formát informácií o výskume. Formálne kontextuálne vzťahy na prechod cez labyrint vedeckých poznatkov. Informačné systémy o vede v EÚ. Štandardizácia a kompatibilita. Zborník z medzinárodnej konferencie. CVTI SR, Bratislava 23. 4. 2009. [Online]. [cit. 05.03.2012] Retrieved from: < http://nispez.cvtisr.sk/userfiles/file/Aktivita%204.1/Konf2009%20prezentacie/Zbornik_konferencia_SK.pdf>

Noge, J. (2011) : Central register of theses and dissertations in Slovakia and document originality verification as a centraly provided service. ProInflow : časopis pro informační vědy [Online]. [cit. 05.03.2012]. - 2011, roč. 3, č. 2 , s. 110-120. - ISSN 1804-2406. Retrieved from:< http://pro.inflow.cz/central-register-theses-and-disserations-slovakia-and-document-originality-verification-centrally-pr>

Valkovič, L. (2011): Detailná funkčná špecifikácia systému SK CRIS. Bratislava, 2011. [cit. 05.03.2012]

Zendulková, D. (2011a): Centrálny informačný portál pre výskum, vývoj a inovácie a jeho úloha v oblasti sprístupňovania vedeckých informácií. Zborník príspevkov z XI. ročníka odborného seminára pre pracovníkov knižníc Virtuálna knižnica – knižnica tretieho tisícročia, ŠVKBB, 2011. - 18 s. [Online]. [cit. 05.03.2012] Retrieved from: <http://www.svkbb.eu/zbornik/zbornik2011/data/CIPVVI_BB2011.pdf >

Zendulková, D. (2011b): Štátne informačné systémy o vede v EÚ. CVTI SR: Bratislava, 2011.- 66 s. ISBN:978-80-89354-04-7 [Online]. [cit. 05.03.2012] Retrieved from: <http://nispez.cvtisr.sk/index.php?&menu=158>

Contact Information

prof. Ján Turňa, PhD.
jan.turna@cvtisr.sk

Juraj Noge
juraj.noge@cvtisr.sk

Danica Zendulková
danica.zendulkova@cvtisr.sk

Slovak Center of Scientific and Technical Information
Lamačská 8/a
81104 Bratislava
Slovak Republic

http://www.cvtisr.sk