Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 00 (2014) 000–000

Procedia Computer Science

www.elsevier.com/locate/procedia

CRIS 2014

CERIF – Is the standard helping to improve CRIS?

Carlos Sousa Pinto[a]*, Cláudia Simões[a], Luis Amaral[a]

[a]*Universidade do Minho, Campus de Azurém, 4800 Guimarães, Portugal*

Abstract

Governments and organizations are creating Current Research Information Systems (CRIS) to follow the growth of the amount of research data, providing tools to collect, preserve and disseminate that data. At the same time, we are facing the appearance of standards designed to regulate CRIS development. Common European Research Information Format (CERIF) is a standard for managing and exchanging research data. There are several types of CRIS – institutional, regional, national and international. In this work we have just considered the national and international types of CRIS worldwide. Only seven of them were CERIF-compliant. The aim of this study is to conclude if the use of CERIF is increasing the number of features in CRIS and how deep CERIF-compliant CRIS are adopting CERIF. Applying all the criteria considered in our methodology, only ten CRIS were analyzed, four of which are CERIF-compliant. CERIF tends to increase similarities between CRIS, in terms of its features and its data models. However, the need for customization of such systems leads to various implementations of the standard, creating an opposite effect of the one referred before. CRIS non CERIF-compliant have as central focus the researchers. The CERIF takes CRIS to focus also on projects and institutions of the research domain. With this exception, the CERIF doesn't show an increase of the number of features. We also consider the use of Dublin Core to increase interoperability between CRIS.

Peer-review under responsibility of euroCRIS.

Keywords: Common European Research Information Format; CERIF; Current Research Information Systems; CRIS; Dublin Core.

1. Introduction

In the last decade, the number of researchers increased progressively[1]. Large companies are investing large amounts of money in R&D. Annually, The Economics of Industrial Research & Innovation (IRI) gives the results of

* Carlos Sousa Pinto. Tel.: +351253510310; fax: +0-000-000-0000 .
E-mail address: csp@dsi.uminho.pt

Peer-review under responsibility of euroCRIS.

1500 companies and their investments in R&D. In 2012, these companies spent 510,7 billion euros in this investment, representing an increase of 4% over the previous year[2].

Governments have to intervene and support R&D, because of the direct effects it has on the progress of economy, technology, knowledge and society. These values require governments to plan strategies for this scenario to be sustainable. One approach adopted was the creation of information services as technological support for science, technology and innovation (STI). These systems are referred using different designations. Among others, we can find in the literature references to Current Research Information System (CRIS), scientific portals[3], research portals[4], research management systems, online information services for science and technology[5], research information systems, or scientific information systems. In this paper, all these systems are referred as CRIS.

According to Bittner and Müller, CRIS are "software tools used by the various interveners in the research process"[6]. euroCRIS's vision is that a CRIS should be understood as a tool that provides access to and disseminate research information[7]. Generally, a CRIS provides a context for research[8]. This means that these systems have information that supports STI, and sensitize society to R&D. This way, governments have an opportunity to justify their investment in STI[6]. Research results are made public, bringing society and STI closer.

Attempts to reduce this plurality of research information have emerged. One of the most significant has been the standard Common European Research Information Format (CERIF) which aims to standardize the management and the exchange of research data handled by CRIS.

Currently, it is not possible to identify in the literature any study comparing the various existing national or international CRIS, identifying their similarities and differences. This work aims to cover this gap and, at the same time, answer the following two questions: (1) Does the use of CERIF lead CRIS to implement more features? and (2) How deep CERIF-compliant CRIS are adopting CERIF?.

In section 2 we describe the methodology used in this research. Section 3 identifies the existing national and international CRIS. Section 4 discusses the CERIF standard as a solution for the heterogeneity of CRIS. Section 5 identifies CRIS compliant with this standard and compares these CRIS with other ones non CERIF compliant. Sections 6 and 7 correspond to the discussion of the results and conclusions, respectively.

2. Methodology of the Study

We established five steps to answer the original research questions. The first step concerns the search of existing CRIS. This search was done using Google, the euroCRIS's website and by consulting several scientific works. Commercial and institutional CRIS were rejected, and national and international CRIS were considered. Commercial and institutional CRIS were rejected because the access to these systems was restricted to enrolled members. As a result of that step we obtained a list with 43 CRIS. In the second step we just classified the CRIS belonging to the initial list as CERIF-compliant or not CERIF-compliant. In the third step the 43 identified CRIS were required to support the following languages: Portuguese, Spanish or English. The CRIS with largest number of registered researchers were selected, and in case of a tie, the one with the higher number of institutions involved was considered. Using these rules, a list of ten systems was obtained. In the fourth step, CRIS previously selected were compared. This comparison was based on: (1) types of actors in the process, (2) researcher's personal information, (3) researcher's curricular information, (4) levels of interoperability, (5) availability of indicators, (6) information search facilities, (7) availability of institutional information, and (8) information about research projects. The last step discusses the similarities and differences between the analyzed CRIS, based on the results obtained in the fourth step, and allowed to answer the research questions identified earlier in the beginning of this article.

3. National and international CRIS

Regarding its scope, there are four types of CRIS: institutional, regional, national and international. Institutional CRIS includes information of just one institution. National CRIS manipulate STI information from many (or all) institutions belonging to a country. Regional and international CRIS involve more than one country. There are also CRIS that include STI information by area/subject (agriculture, health).

In the first step of this research, we identified 43 national and international CRIS (see Tab. 1) all over the world.

Table 1. National and international CRIS

Country	System	Acronym
Belgium	Flanders Research Information Space Research Portal	FRIS
Bulgaria	The Bulgarian Current Research Information System	BULCRIS
Czech Republic	The Research and Development and Innovation Information System of the Czech Republic	IS R&D&I
German	German Project Information System	GEPRIS
German	Research explorer	(ReX)
Estonian	Estonian Research Portal	ETIS
Finland	Finnish science and technology information service	Research.fi
France	CV Science	CV Science
Slovenia	Slovenian Current Research Information System	SICRIS
Slovak	Slovak Current Research Information System	SK CRIS
Uruguay	CVuy System	CVuy
Colombia	COLCIENCIAS	COLCIENCIAS
Mexico	Integrated Information System of Scientific and Technological Research	SIICYT ou CvU
Argentine	Information System of Science and Technology at Argentine	SiCyTAR
Spain	*Sistema de Informação Científica de Andaluzia*	SICA2
Italy	DAVINCI Database	DAVINCI
Brazil	Plataforma Lattes	Lattes
Canada	The Canadian Common CV for Researchers	CCV
Portugal	*Plataforma DeGóis*	DeGóis
Venezuela	*Registro Nacional de Innovación e Investigación*	RNII
Japan	Directory Database of Research and Development Activities	ReaD
Paraguay	*Sistema CV Paraguay*	CVpy
El Salvador	*El Registro de Investigadores Científicos Nacionales*	Redisal
Netherlands	National Academic Research and Collaborations Information System	NARCIS
Turkey	Researcher Information System	ARBiS
Singapore	Singapore researchers database	
Norway	Current Research Information System in Norway	Cristin
Denmark	Danish National Research Database	
Chile	*Sistema Información Ciencia Tecnología e Innovación*	SICTI
Ecuador	*Directorio de Currículum Vitae en Ciencia y Tecnología*	CVLAC
Panamá	*La Secretaría Nacional de Ciencia, Tecnología e Innovación*	SENACYT
Peru	*Red del Sistema Nacional de Ciencia, Tecnología e Innovación*	Red SINACYT
Bolivia	*Sistema Boliviano de Información Científica y Tecnológica*	SIBICYT
Costa Rica	*Consorcio Registro Científico Tecnológico*	RCT
Switzerland	ARAMIS Information System for Research and Development Projects in Switzerland	ARAMIS
Austria	Austrian Research Information System: Multimedia Extended	AURIS-MM
Russian	CRIS of Russian Academy of Sciences	RAS CRIS
Sweden	Sweden ScienceNet	
Hungary	HunCRIS	
Poland	Nauka Polska	
International	IST World	istworld
	EuroRIs-Net+ Research Infrastructures Observatory	EuroRIsNet+ Observatory
	Socionet - Russian Research Community CRIS	Socionet

Europe, Central and South America stand out as regions with more national CRIS. About 53% of national CRIS are European. Considering the Asian continent, only in Russia, Japan and Singapore were found national CRIS, while in North America, national CRIS was found only in Canada.

4. CERIF as a Solution for the Heterogeneity of CRIS

Given the inevitable heterogeneity of CRIS, there are attempts to standardize these systems. Standardization is necessary not only to regulate the development of CRIS, but also to enable higher levels of interoperability between them. National standards (that is, developed by entities of a particular country) do not cover all needs, because they have limited scope. The international and regional (including several countries) initiatives are more complex and its adoption is more difficult because in that case, standards are transversal to governments, policies and countries.

The most widely referenced standard in the field of CRIS is CERIF. This standard is maintained by euroCRIS since 2002[9]. CERIF is an attempt to standardize the data manipulated and traded in these systems, partially by using

XML to provide a common format. This standard proposes a formal data model, including entities, attributes and relationships between entities. In its latest version, CERIF 1.6, the standard also implements a semantic layer that adds controlled vocabularies to the standard[10]. The European Union (EU) aims to make the research information homogeneous, by placing CERIF as a recommendation to member states[8, 9].

The detail and high scope of the standard make CERIF's understanding and use an arduous task[4]. The existence of 293 entities, 1814 attributes and 665 relationships in the version 1.6[11] of its data model don't help its usage.

5. Comparing CRIS

According to the latest data provided by euroCRIS[12] and other authors[13], and taking into account the previously identified CRIS (see Tab.1), we have just identified the following seven national and international CERIF-compliant CRIS: RAS CRIS, SK CRIS, Socionet, EuroRIsNet+Observatory, FRIDA (actual FRIS), HunCRIS (not accessible), and SICRIS. RAS CRIS and Socionet were not considered because they are not available in English, Portuguese or Spanish. We compared 10 CRIS that verified the original constraints, 4 of which are CERIF-compliant (see Tab. 2).

Table 2. Main indicators of analyzed CRIS

INDICATORS* / **CRIS**	Researchers	Institutions	Research groups	Projects	Research Programs	Scientific Activities	CERIF-compliant	Integrating CERIF in the future
SICRIS	14 438	978	1 528	5 854	451	NA	Yes	
SK CRIS	18 156	1 257	NA	9 998	NA	NA	Yes	
EuroRIsNet+ Observatory	718	1 909	NA	330	NA	NA	Yes	
FRIS	27 350	2 273	NA	26 987	NA	NA	Yes	
GEPRIS	55 402	23 763	NA	90 638	NA	NA	No	No
NARCIS	50 840	2 901	NA	NA	59 550	NA	No	Yes
Redisal	624	42	NA	1 340	NA	NA	No	No
DeGóis	19 113	70	NA	5 741	NA	NA	No	No
Lattes	2 601 696	NA	NA	NA	NA	NA	No	No
SICA2	51 994	15 458	NA	NA	NA	644 978	No	Yes

Legend: NA - Not Applicable | *Values obtained on the website of the CRIS in 16-12-2013

CRIS collect personal information from researchers, but it is in the curricular information that these systems are more specialized. The curricular information is captured with high granularity, especially in non CERIF-compliant systems. No CRIS collects data on the personal preferences of the researcher, the so called soft skills. This fact can be justified by the high formalism associated with these systems. Soft skills have been increasingly recognized as important[14, 15] and as a main component of the personal curriculum in what concerns employability[16]. In some contexts of employability, the soft skills can even override the technical skills[17].

Almost all the analyzed CRIS do not collect evidence about the veracity of the curricular information, except in the case of SICA2 which included a feature to collect certificates. This feature increases trust in CRIS.

We also concluded that the analyzed CRIS don't allow the customization of the personal or curricular information made public. What is public or private is determined by the systems, equally for all the enrolled researchers. One dimension also analyzed was the verification of the existence of a standard for knowledge areas. We can refer as two examples, the CERIF Schema and the FOS 2002 from the OECD (The Organization for Economic Co-operation and Development). The knowledge areas are very important, because they allow knowing what area in which researchers, projects, groups and programs are specializing in. All analyzed CRIS allow the association of a knowledge area to a scientific or technologic production, but they are not using a unique system of classification to do that. This may constitute a problem in what concerns interoperability.

Most CRIS have a list of existing projects, and relevant information about them. In the case of CERIF-compliant CRIS, the entity project is deeply implemented, in contrast to CRIS as DeGóis, Lattes and SICA2.

In particular, the CERIF-compliant set of CRIS does not give much importance to curricular information such as events, evaluation panels and awards. This type of information is only captured by some non CERIF compliant CRIS. CERIF has been analyzed in order to identify if its data model provided elements related to the curricular

information. We concluded that there is a set of elements whose use is mandatory and would allow capture that information. We didn't find CERIF entities related to the researcher experience.

This set of CRIS also has a small number of individual or global indicators, which are called STI indicators. These indicators are presented as total values and we can't identify indicators that combine multiple perspectives. Also, in that set, imports and exports of data were not identified. This is somewhat strange because, as seen before, one of the main goals of the CERIF standard is to promote data exchange.

Most of these systems provide global STI indicators. This confirms the previous finding that these systems are privileged instruments to generate such indicators. However, most of the CRIS are not very ambitious about this functionality. Bibliometric and non bibliometric indicators are still not covered.

The system entities must be identified equally by the whole community involved in the STI production. To identify productions, it is used the DOI or the ISBN in the case of books, or the ISSN in the case of papers or journals. It is also possible to use the Accession Number to identify productions stored in the ISI Web of Science. In the case of researchers, the use of a unique identifier, like ResearcherID or ORCID iD, is seen as a solution to the problem of ambiguity in what concerns the authorship of a production[18]. For example, NARCIS uses an author digital identifier to identify uniquely a researcher[3], and that identifier is also used by the Dutch universities in a wide range of situations, including scenarios not related to scientific and technologic production.

CRIS like DeGóis, Lattes or SICA2, non CERIF-compliant, have little information about institutions and research projects, because they were designed considering the researcher as a central element of the information system. The information from other STI entities is partially obtained through the curriculum information of the researchers.

Considering the CERIF-compliant CRIS, SICRIS uses the concept of program or funding program which is not used by the other CRIS belonging to this set. CERIF, at this level, should play a role of normalization and clarification of the concepts of program and project. This customization can generate potential problems in the integration/interaction process, even among systems CERIF-compliant. In what concerns the areas of knowledge, it should be noted that only the SICRIS follows the classification scheme proposed in the CERIF. The remaining CERIF-compliant CRIS, normalize these areas, using other classifications. We concluded that CERIF compliant CRIS are adopting the CERIF data model according to their specific needs, and are not implementing all the elements of the standard.

It was possible to conclude that there are several national and international CRIS with no integration with other type of systems, in particular scientific databases like ISI, Scopus, Google Schoolar or SciELO. The integration of information about scientific and technologic productions available in those systems, allows to avoid the duplication of work. We concluded that DeGóis, Lattes, SICRIS and NARCIS can do that type of integration.

6. Discussion of Results

The analyzed CRIS show different stages of maturity (different number of functionalities, different levels of interoperability, etc). There are some CRIS in a pilot stage, such as SK CRIS, and CRIS with a high number of registered researchers, like Lattes (about 3.000.000). Strangely, countries like the United States, the United Kingdom or China, that have a strong investment in R&D[19], don't integrate the benchmarking of national CRIS.

Globally, it is costly to adopt standards. The SK CRIS reflects this reality. In this case, the implementation of CERIF was planned for six years (2008-2014).

If a new CRIS arises, it should adopt the standard, but if we consider existing systems, it will be very difficult to fully include the components of the standard. In these situations, it can be considered the use of Dublin Core, as an intermediate format between CERIF-compliant CRIS and non CERIF compliant CRIS (see Fig. 1). However, in these cases, some information would be lost because it is not possible to map all terms of CRIS using Dublin Core. The purpose of CERIF is highly desirable but there are few cases of national and international CRIS adopting the standard. In practice, interoperability between systems remains a challenge, even among CERIF-compliant systems. An extended or partial implementation of the standard can cause interoperability problems.

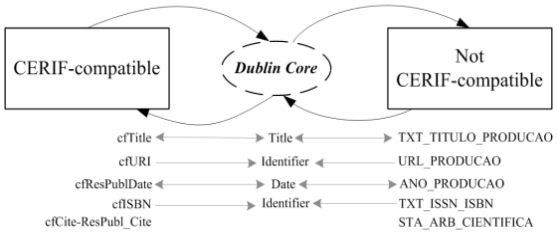

Fig. 1. Using Dublin Core to increase interoperability between CRIS

7. Conclusions

CRIS are part of national strategies to promote STI. Currently, more than 40 national CRIS can be found. These systems are the preferred instrument to contextualize STI in a country or region. The data model of a CRIS is closely related to its context. The CERIF standard tries to unify these data models to ensure interoperability. CERIF allows CRIS to have close functionalities and data models. However, the customization of CERIF - by extension or partial implementation – tends to deviate the systems from that goal. Therefore, the full adoption of CERIF would in fact lead to the increase of compatibility between CRIS, but this scenario is still far from being achieved. The complementary approach of using Dublin Core to increase interoperability between CRIS is a possible strategy to use in the short term, but in this case, losing some research information is inevitable, because it is not possible to map all terms of a CRIS using Dublin Core Metadata Element Set.

References

1. Eurostat. (2013a). Total researchers (FTE) by sectors of performance. Retrieved May 7, 2013, from http://epp.eurostat.ec.europa.eu.
2. Joint Research Centre, J. (2012). EU R&D Scoreboard: The 2012 EU Industrial R&D Scoreboard (pp. 1–124). Spain.
3. Dijk, E., Hogenaar, A., & Meel, M. (2010). Users in the spotlight: study on the use of the Dutch scientific portal NARCIS, 2009. Netherlands.
4. Spyns, P., Grootel, G. Van, Jörg, B., & Christiaens, S. (2010). Realising a Flemish government innovation information portal with Business Semantics Management. Proceedings of CRIS, 45–54.
5. Santos, L. (2004). Factores determinantes do sucesso de serviços de informação online em sistemas de gestão de ciência e tecnologia. Universidade do Minho.
6. Bittner, S., & Müller, A. (2011). Social networking tools and research information systems: Do they compete? In Koblenz (Ed.), In: Proceedings of the ACM WebSci'11 (pp. 1–4). Germany: ACM.
7. Leenheer, P. De. (2007). Artifact-centric Service Interoperability in the Flanders Research Information Portal. acsi-project.eu, 5–6.
8. Jeffery, K., & Asserson, A. (2009). Institutional repositories and current research information systems. New Review of Information Networking, 14(January), 71–83.
9. Ivanović, D., Surla, D., & Racković, M. (2010). A CERIF data model extension for evaluation and quantitative expression of scientific research results. Scientometrics, 86(1), 155–172.
10. EuroCris. (2010a). CERIF Introduction. Retrieved March 20, 2013, from http://www.eurocris.org.
11. EuroCris. (2013). Statistic Information. CERIF 1.6. Retrieved March 20, 2013, from http://www.eurocris.org.
12. EuroCris. (2010b). Directory of Current Research Information Systems (DRIS). Retrieved March 20, 2013, from http://www.eurocris.org.
13. Ivanović, L., Ivanović, D., & Surla, D. (2012). A data model of theses and dissertations compatible with CERIF, Dublin Core and EDT-MS. Online Information Review, 36(4), 548–567.
14. Bailly, F. (2013). The personification of the service labour process and the rise of soft skills: a French case study. Employee Relations, 35(1), 79–97.
15. Joseph, D., Ang, S., Chang, R. H., & Slaughter, S. A. (2010). Practical intelligence in IT: assessing soft skills of IT professionals. Communications of the ACM, 53(2), 149–154.
16. Andrews, J., & Higson, H. (2008). Graduate Employability, "Soft Skills" Versus "Hard" Business Knowledge: A European Study. Higher Education in Europe, 33(4), 411–422.
17. Boreham, N. C., & Lammont, N. (2000). The need for competences due to the increasing use of information and communication technologies (p. 61). Thessaloniki.
18. ResearcherID. (2013). What is ResearcherID? Thomson Reuters. Retrieved March 15, 2013, from http://www.researcherid.com.
19. Battelle. (2011). 2012 Global: R&D FunDing FoRecast (pp. 1–35).