

May 2015

R&D Evaluation Methodology and Funding Principles

Background report 9: The RD&I Information System as an information tool for evaluation



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

R&D Evaluation Methodology and Funding Principles

Background report 10: The RD&I Information System as an information tool for evaluation

May 2015

Jan Dvořák, Tomáš Chudlarský - InfoScience Praha

Gunnar Sivertsen - NIFU

Tomáš Vondrák - Technology Centre AS CR

Table of Contents

1. Introduction	1
2. Contents and Architecture of the IS VaVaI	2
2.1 National Research Information Systems in the international context	2
2.2 The IS VaVaI	3
2.2.1 Overview	3
2.2.2 Information contents of the IS VaVaI	5
2.2.3 Data collection into the IS VaVaI	14
2.2.4 Presentation of IS VaVaI data	16
2.2.5 SWOT analysis	17
3. The reliability and usability of the IS VaVaI	19
3.1 Introduction	19
3.2 Quality and reliability of the information	20
3.2.1 The interviewees' assessment	20
3.2.2 Detailed discussion of cases reported by the EvUs in the SPE	21
3.3 User-friendliness and efficiency of the data entry process	22
3.4 The extended use of the IS VaVaI	23
3.5 Conclusions	24
4. Options for the use of external information and analytical tools	26
4.1 The connectivity of the IS VaVaI	26
4.2 Availability and suitability of information resources	27
4.2.1 Sources for bibliographic data	27
4.2.2 Sources for data on IP outputs	35
4.2.3 Going beyond research outputs	36
4.2.4 Enhancing the connectivity of the information	37
4.2.5 Connectivity and connectability	39
4.2.6 Machine access to the IS VaVaI data	39
4.3 Conclusions	41
5. Proposal for the use of the IS VaVaI as an information tool for evaluation	43
5.1 The use of the IS VaVaI for the Evaluation Methodology implementation	43
5.1.1 Introduction	43
5.1.2 Providing data for the self-assessment reports	44
5.1.3 Handling the evaluation process	47

5.2 Options for the enhancement of the IS VaVaI	50
5.2.1 Subject classification	50
5.2.2 Improve RIV compatibility with established bibliographic conventions	53
5.2.3 Introduction of a RIV master list of publishing channels	54
5.2.4 Revision of the RIV output types classification	54
5.2.5 More comprehensive support for funders	55
5.2.6 Improved presentation of the IS VaVaI data and analytical outputs	58
5.2.7 Tracking research output lifecycles and impacts	59
5.2.8 Improve support for projects with postponed ex-post evaluation	59
5.2.9 Collect more detailed information about the cost structure of projects	60
5.3 Cost estimates	60
5.3.1 Direct support of the EM implementation: Handling the evaluation process	61
5.3.2 Enriched IS VaVaI: EM compatibility, new components, data quality enhancements	65
5.3.3 Improved accessibility of the IS VaVaI data	69
5.3.4 Other enhancements of the IS VaVaI	71
5.3.5 Licences for citation databases	75
5.3.6 Estimates of the total costs of the adaptations and enhancements of the IS VaVaI outlined in this report	75
6. Conclusions	77
Appendix A - Interview questions with answers	78
Appendix B - IS VaVaI fields mapped to the OECD FOS field categorisation	83

Table of Figures

Figure 1. Basic types of objects in the IS VaVal.....	5
Figure 2. State budget funding of RD&I projects over the past 10 years.....	7
Figure 3. The RD&I subsidy of an organisation.....	11
Figure 4. The flow of the RIV data collection process with approximate volumes	16
Figure 5. Proportion of Open Access documents 2012-2014 (InCites)	29
Figure 6. Coverage of Czech journal articles (RIV 2012) in international citation databases – WoS and SCOPUS	30
Figure 7. Coverage of Czech journal articles and proceeding papers (RIV 2012) in international citation databases – WoS and Scopus.....	31

List of Tables

Table 1 Research output types in the current RIV, with counts of outputs in the years 2008-2013.....	8
Table 2. Data items about persons collected in the IS VaVal.....	12
Table 3. Most frequent names of Czech researchers (creators of research outputs in the RIV, years 2001-2013, names of at least 20 distinct persons).....	13
Table 4. Name ambiguity of researchers in the years 2001-2013 (creators of research outputs in the RIV)	14
Table 5. Frequency of funders reporting outputs (research outputs of all types, years 2009-2013)	57

1. Introduction

This report constitutes a background report to the Final report 3 – *The Small Pilot Evaluation and the Use of the RD&I Information System for Evaluation*. It describes the present state of the Czech RD&I Information System (IS), looks into ways how it can be made more useful in general, and how it can best support the proposed Evaluation Methodology in particular.

The structure of this report reflects the objectives and covers the various relevant tasks of the study as described in the Terms of Reference and the proposal, i.e.:

- In Section 3 we look into the current state, reliability and usability of the R&D IS and the data it contains, as a premise for its potential use as information tool for evaluations. This includes investigation of the effectiveness and efficiency of the data entry processes and the current and envisaged quality management, as well as the structure of the R&D IS in the Czech Republic (RIV) itself
- The outcomes of the analysis related to the availability and suitability of external information sources and the potential connectivity of the R&D IS with these sources and related analytical tools is reported in Section 4
- A detailed proposal for the use of the RD&I IS as a tool of information support for i evaluation, including its necessary extensions and adaptations in order to fulfil the role of a primary data tool for evaluators, is provided in Section 5
- Overall conclusions are drawn in Section 5.3

We start this report with an overview of the RD&I IS features and characteristics in Section 2.

This report also reflects and references items in the IPN KA1 *Proposal for R&D&I IS Improvement* (IPn-KA1).

In the remaining of this report, we refer to the Czech RD&I Information System with its official abbreviation: **IS VaVal**.

2. Contents and Architecture of the IS VaVaI

In this chapter we describe the main features and characteristics of the Is VaVaI, setting it also in the context of international practice.

2.1 National Research Information Systems in the international context

Today, most research information systems are implemented at universities and other research performing organisations. The principal motivation for their introduction is the need for a consolidated base of information on the organisation's structure, staff, research activities, and research outputs to support reporting both inside and outside the organisation. For many organisations, this is the only way to make the administrative burden on researchers bearable.

Similarly, research funding organisations use research information systems to handle the processes of receiving, evaluating (ex-ante) and selecting the proposals to fund, the grant administration and the interim and final reporting and ex-post evaluation. Funders also need to demonstrate the outputs, outcomes and impacts their funding contributes to. A funder's research information system provides the consolidated base of research information about the activity of the funder, and about the outcomes thereof.

Also at the level of governments that have the responsibility over the strategic management of research and innovation¹, the need for a consolidated research information base is widely recognized. The motivations include the need for informed policy making and strategic management at the government level, advertising to attract hi-tech investment, and supporting knowledge transfer, among others. Increasingly often, research assessment and in particular the need for information informing performance-based research funding systems (PRFS) is among the drivers for government research information systems.

Throughout Europe and internationally, there are only a few national research information systems that contain a similar level of comprehensiveness and internal interconnection as the IS VaVaI. Among the most visible ones are CRISTIN (Norway), SK CRIS (Slovakia), FRIS (Belgium/Flanders) and ETIS (Estonia). Other systems in operation do not cover research projects as well as research outputs - e.g. SICRIS (Slovenia), SweCRIS (Sweden), BulCRIS (Bulgaria), Gateway to Research (the U.K.), POL-on and the Polish Scholarly Bibliography (Poland). In most other countries in Europe, the development of national research information system is currently planned or under way².

¹ In European federal states this responsibility seems to be more commonly located with the governments of the constituent parts (the federal states in Germany, the autonomous communities of Spain, the two regions of Belgium), whereas the national governments may but need not play a coordinating role.

² See also: Mahieu, B., Arnold, E., Kolarz, P. (2014) Measuring scientific performance for improved policy making - The feasibility and desirability of a European research information system. Study for the European Parliament - Science and Technology Options Assessment (ISBN 978-92-823-5531-2)

2.2 The IS VaVaI

2.2.1 Overview

The IS VaVaI is the national research information system of the Czech Republic. It is run by the Council for Research, Development and Innovation, the advisory body to the government of the Czech Republic for the domains of research, development and innovation. The institutional responsibility for the information system resides with the Office of the Government of the Czech Republic. The information system is anchored in the Act on public support of RD&I³.

From a national RD&I policy perspective, the information system has a double function: it acts as an operational management system

- For the public competitive funding in the Czech Republic, including information on input (funding, focus of the programmes, organisations/researchers involved etc) as well as the outputs (i.e. the related research results)
- For the public institutional funding of research organisations, including input and output data, related to the institutional research plans and for the calculation of the funding criteria defined in the PRFS (i.e. Metodika)

On the one hand, the system documents the public expenditures on RD&I in the Czech Republic⁴, the programmes and project calls, funding allocations, and the outputs from that funding and of Czech research organisations in general. It has a record that reaches twenty years back and thus comprises the whole period of competitive funding in the country.

On the other hand, it documents two systems of institutional funding: the institutional research plans (also known as *výzkumné záměry* = “research intentions”, 1999-2013), followed by the current performance-based research funding system that has been in use since 2008.

It is important to note that the IS VaVaI **does not stand in isolation**. It relies on research information feeds from funders and from research performing organisations. The bigger and the mid-tier players among them have implemented or are at various stages of implementing their own research information systems.⁵ The research information landscape in the Czech Republic is shaped by the Act on public support of RD&I. As for research information, it sets the following mandates:

1. Public RD&I funding providers must publish the information on the calls they open through the IS VaVaI (among other means).
2. Public RD&I funding providers must report into the IS VaVaI on the projects and other RD&I activities they fund.
3. The beneficiaries of the public RD&I support must report into the IS VaVaI on the outputs of the research they have been performing. This includes publications,

³ Act no. 130/2002 Coll., as amended

⁴ ~26 bn. CZK/year; between 0.9 and 1.0 bn. €/year depending on the exchange rate

⁵ While the default data entry tool of the IS VaVaI works well for smaller and many of the mid-tier players, it cannot possibly satisfy all the diverse requirements of research-intensive organisations or large research funders.

patents and a variety of other, non-scholarly research outputs; the research outputs must be linked to funding.

Also, the IS VaVaI integrates information from other information systems within the Czech Republic (e.g. the register of legal entities, data from the Industrial Property Office) and outside it (e.g. CORDIS, ISBN International).

Finally, information from the IS VaVaI is being used by research information systems: mostly within, but to a growing extent also outside the country.

The high importance of the system is also reflected in the official *Concept of the IS VaVaI for the years 2012-2015*⁶, Section 4:

Further direction and development of the IS VaVaI is dictated by the growing requirements on its maximum accuracy, dependability and reusability of the collected and published data. These requirements stem from the growing usage and importance of the IS VaVaI which is to an increasing extent used for audit at the side of the funders and of the beneficiaries, for various analytical studies, for disseminating information about research among fellow researchers, in the general public, and also for usage in other state administration bodies.

The fundamental departing point for the further development of the IS VaVaI is the fact that the IS VaVaI will affect the national research policy to a growing extent. The contents of the IS VaVaI - and its credibility - will impact the assessment of its influence on the whole RD&I domain and feedbacks in the evaluation of national research activities. The trustworthiness of RD&I evaluation systems and the quality of strategic decisions will depend on its reliability and data accuracy.

The information in the IS VaVaI underlays the distribution of the institutional support and simultaneously it is used for the audit activities. To fulfil this role the IS VaVaI and related applications must be adapted to guarantee a yet higher trustworthiness of the data which the funders and the beneficiaries of the public support insert into the IS VaVaI. The data can be taken as authentic and credible only if the funders and beneficiaries deliver the data strictly in accordance with the prescribed law i.e., valid and complete.

The IS VaVaI technologies and information background are also conveniently used in the related agendas of the Council for Research, Development and Innovation: the **state budget proposal preparation** and the collection and processing of data about contract research (the data is collected in the framework of the current PRFS). The elaborate system of validations of inputs and the rich information context enables both of these agendas to collect information that has a much higher level of reliability than any *ad hoc* data collection could achieve. The processing of the information in the same environment as the rest of the research information from the Czech Republic gives numerous advantages in the additional context and interconnectedness. The processed information is not displayed to the public, but is used internally at the Section for Science, Research and Innovation of the Office of the Government of the Czech Republic.

⁶ *Koncepce Informačního systému výzkumu, experimentálního vývoje a inovací na období 2012 až 2015*, <http://vyzkum.cz/FrontClanek.aspx?idsekce=605357> (available in Czech only)

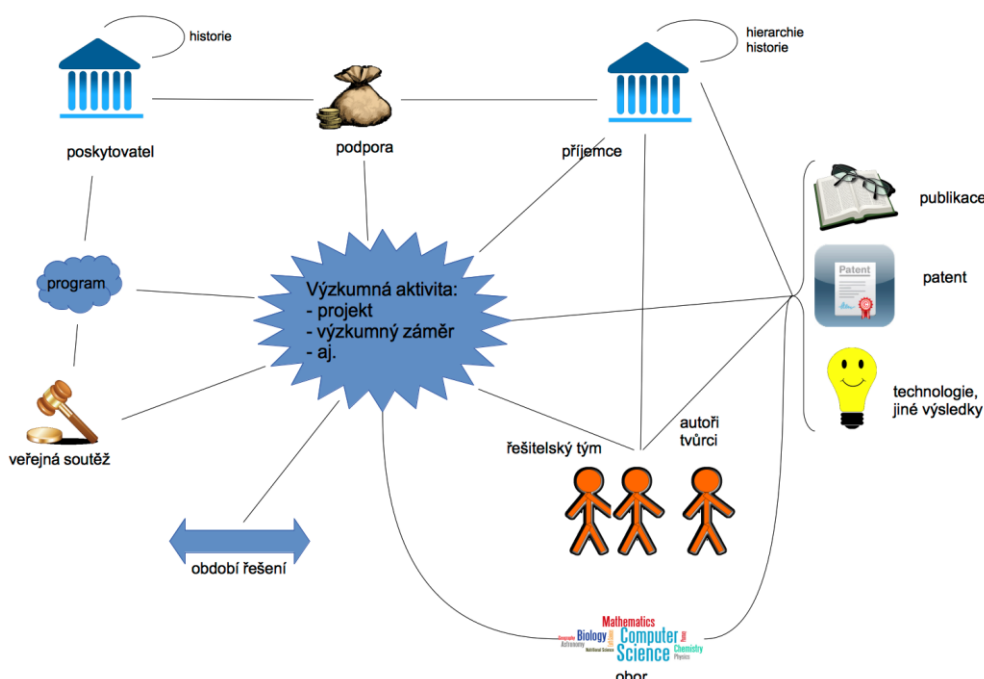
In the sections below we introduce the current (end of 2014) state of the information system. We do not go into the many details that accumulated throughout its long history: it's towards the future that this report is intended.

2.2.2 Information contents of the IS VaVaI

One of the major strengths of the IS VaVaI is in the interconnected and integrated structure of the information that it contains.

Figure 1 gives an overview of the basic types of objects that are present in the system.

Figure 1. Basic types of objects in the IS VaVaI



RD&I Activities

The *RD&I activity* is a general notion for a type of funding that is put into the RD&I system from public funds. It represents the following forms of funding:

1. Funding programmes
2. Grant projects (in groups, such as the Standard projects of the Czech Science Foundation)
3. Contracts in RD&I
4. Funding for university specific research
5. Funding of large infrastructure
6. Block funding of research organisations (the “long-term development of research organisations based on the evaluation of their results”)
7. Operational programmes in RD&I

8. Funding for international collaboration in RD&I: membership fees in international organisations (such as the European Space Agency) or programmes (such as the Fulbright Commission)
9. Awards for outstanding achievements in RD&I such as *Česká hlava* (“the Czech head”)⁷
10. Costs of organizing RD&I public tenders
11. Running cost of the agencies (Czech Science Foundation, Technology Agency of the Czech Republic), of the RD&I Council secretariat and of the Academy of Sciences of the Czech Republic

The IS VaVaI tracks the global funding amounts for the activities in its **CEA (Activities of RD&I)** CEA component. Cases 1-3 are further detailed in the CEP (Registry of RD&I Projects) component with the per-project per-beneficiary per-year granularity. Cases 4, 6, and 8 are detailed in the CEA itself with the per-beneficiary per-year granularity. The institutional research plans (a.k.a. research intentions, 1999-2013), which were tracked in the CEZ (Registry of institutional research plans) component, are also summarized in the CEA.

The information is displayed from two perspectives: that of the funder and that of the beneficiary (the funded organisation). For instance, the RD&I activities of the Ministry of Education, Youth and Sports are presented on <http://www.isvav.cz/funderDetail.do?rowId=MSM>. The various types of funding the Charles University in Prague gets are listed, together with the amounts, on <http://www.isvav.cz/organizationDetail.do?rowId=ico%3A00216208>.

RD&I Tenders (Calls)

The funding in funding programmes and groups of grant projects is awarded in public RD&I tenders, or calls for project proposals. The **VES (Public tenders in RD&I)** component keeps tracks of all public tenders in RD&I. This information contents includes two phases:

1. The call information (made visible on the IS VaVaI website on the day the call is published)
2. The call evaluation information (summary information about the call: how many project proposals were received, evaluated, and selected for funding, or possibly the reasons why the call was cancelled)

As projects in the CEP reference the call in which they were awarded funding, the list of funded projects in a call becomes visible after the projects enter the system.

Information about a tender is presented on the web portal too, e.g.

<http://www.isvav.cz/tenderDetail.do?rowId=SMSM2014LG4> represents the call for proposals in the LG programme that closed on September 6, 2013 at 23:59.

⁷ <http://vyzkum.cz/FrontClanek.aspx?idsekce=31238>

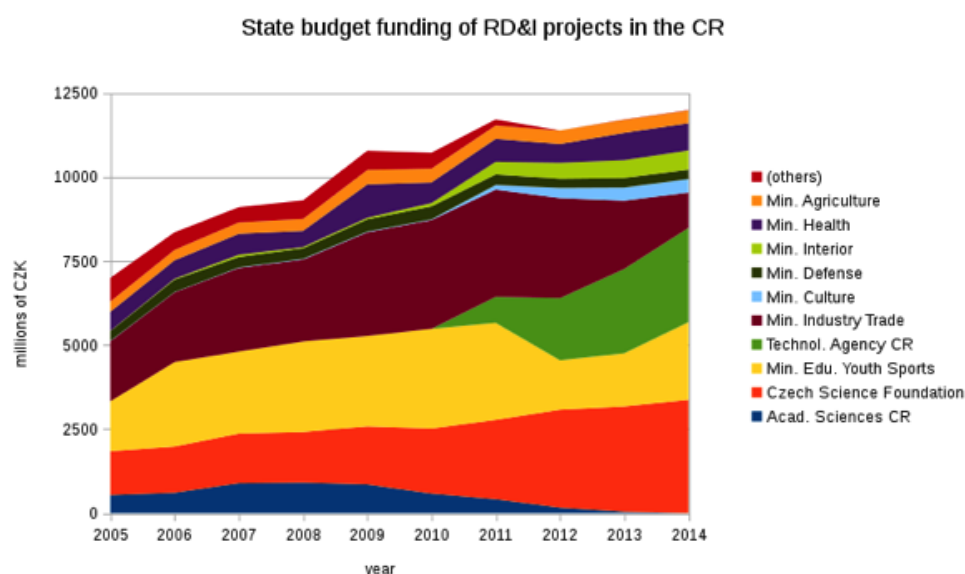
RD&I Projects

The tracking of RD&I project data goes to the year 1993, this was the core of the IS VaVaI functionality. Over the whole history over 41 thousand of project was collected in over 150 thousand of their yearly snapshots. The total funding to projects is shown in Figure 2, below.

All of the following categories of project-based funding are tracked in the CEP:

- National competitive funding
- National infrastructural funding (some, the rest is under the “other RD&I activities”)
- EC framework programmes co-funded by the Czech Republic
- Operational programmes (structural funds)
- International cooperation programmes
- Funding by the regional bodies in the CR

Figure 2. State budget funding of RD&I projects over the past 10 years



Very similar in the structure of the data, though different in their function in the RD&I system of the country, were the institutional research plans (a.k.a. the research intentions) that were collected in the **CEZ (Central registry of institutional research plans)**. The component was active in the years 1998-2014 (the institutional research plans ran 1999-2013). Technically, the CEP and the CEZ shared most of the data structures of the IS VaVaI database.

The detailed information about a project is accessible on the IS VaVaI web portal at URLs such as <http://www.isvav.cz/projectDetail.do?rowId=GAP101%2F12%2F1271>. Similarly, also information about an institutional research plan is available, e.g. <http://www.isvav.cz/researchPlanDetail.do?rowId=MSM6840770010>.

Research outputs tracked in IS VaVaI

Research outputs in the **RIV (Registry of RD&I Outputs)** are classified using a hierarchic classification of output types. The top level of the classification has three groups: publication results, results of applied research and other results. The definitions of the output types are given in a section of 9 terse pages in the RIV input data description document⁸. Table 1, below, lists the hierarchy and gives the counts of outputs.

Each research output type has a specific set of attributes that are relevant to the type. Each output also includes identification of all submitting organization alongside with the full list of their authors. The entry of other authors (who are not affiliated with any Czech organization) is optional. The total number of all authors is present in all cases. Also structured links to funding activities are present.

Table 1 Research output types in the current RIV, with counts of outputs in the years 2008-2013

Output type	2008	2009	2010	2011	2012	2013	Total
All results	54 067	53 169	56 451	60 605	63 569	57 879	345 740
Publication results	46 178	43 586	45 450	47 254	47 384	44 002	273 854
J - Article in a peer-reviewed journal	21 841	21 693	23 732	24 876	26 122	24 426	142 690
B - Monograph	1 655	1 489	1 621	1 889	1 813	1 700	10 167
C - Chapter in a book	4 072	4 259	4 770	5 331	5 381	4 261	28 074
D - Article in proceedings	18 610	16 145	15 327	15 158	14 068	13 615	92 923
Results of applied research	3 372	4 717	5 153	5 905	8 160	6 795	34 102
P - Patent	86	150	167	203	265	282	1 153
F - Utility model or industrial design	217	356	371	690	883	778	3 295
F/P - Industrial design	22	22	28	95	130	101	398
F/U - Utility model	195	334	343	595	753	677	2 897
Z - Pilot plant, verified technology, breed	478	579	470	467	498	414	2 906
Z/A - Pilot plant	69	58	81	79	76	83	446
Z/B - Verified technology	275	486	372	340	376	300	2 149
Z/C - Plant breed	39	38	17	48	47	31	220
Z/D - Animal breed	0	0	0	0	0	0	0
G - Prototype, Functional sample	1 113	1 477	1 731	1 874	2 070	1 560	9 825
G/A - Prototype	363	410	408	285	311	308	2 085
G/B - Functional sample	780	1 072	1 325	1 590	1 761	1 256	7 784
H - Result realized by the funding provider	53	79	67	133	102	92	526
H/A - Projected into legislation or norms	35	53	48	94	83	68	381
H/B - Projected into documents of non-legislative nature	16	26	18	22	4	18	104
H/C - Projected into strategic documents	2	0	1	17	15	6	41
N - Certified methodology, procedure, map	521	975	1 042	1 659	2 334	932	7 463
N/A - Realized certified methodology	343	337	336	309	376	363	2 064

⁸ Předávání údajů do Informačního systému výzkumu, experimentálního vývoje a inovací: RIV – Rejstřík informací o výsledcích 2015. Available (in Czech only) from <http://vyzkum.cz/FrontClanek.aspx?idsekce=986>

N/B - Medical treatment procedure	3	4	0	0	0	1	8
N/C - Historic heritage preservation procedure	0	0	0	0	0	0	0
N/D - Specialized map with expert contents	175	635	706	1 350	1 958	568	5 392
R - Software	727	1 100	1 287	850	1 147	952	6 063
V - Research report	8	7	18	30	864	1 785	2 712
V/S - Summary research report	0	1	15	19	842	1 735	2 612
V/U - Research report containing classified information	1	6	3	11	22	50	93
Other results	4 517	4 866	5 848	7 446	8 025	7 082	37 784
A - Audiovisual work or electronic document	843	604	553	750	826	374	3 950
Organizing an event	1 336	1 169	1 374	1 507	1 395	1 422	8 203
E - Organizing an exhibition	145	180	191	166	178	186	1 046
M - Organizing a conference	687	549	538	606	565	541	3 486
W - Organizing a workshop	504	440	645	735	652	695	3 671
O - Other result	2 338	3 093	3 921	5 189	5 804	5 286	25 631

Each research output is reported in the RIV by all Czech organisations that have their affiliated authors. Records that refer to the same output are identified in the central RIV database and interlinked in the web presentation – for instance the article

Abazov V. M. et al. (2013). *Evidence for s-channel single top quark production in pp collisions at $s = 1.96$ TeV*. Physics Letters. B 726(4-5) 656-664.
DOI <http://dx.doi.org/10.1016/j.physletb.2013.09.048>

was reported in the RIV by all three Czech institutions that had “their” authors:

- The Institute of Physics of the AS CR:
http://www.isvav.cz/resultDetail.do?rowId=RIV%2F68378271%3A_____%2F13%3A00424404!RIV14-AVo-68378271,
- The Faculty for Mathematics and Physics of the Charles University in Prague:
http://www.isvav.cz/resultDetail.do?rowId=RIV%2F00216208%3A11320%2F13%3A10193102!RIV14-MSM-11320____,
- The Faculty of Nuclear Sciences and Physical Engineering of the Czech Technical University:
http://www.isvav.cz/resultDetail.do?rowId=RIV%2F68407700%3A21340%2F13%3A00210632!RIV14-MSM-21340____.

Each such report represents a particular view of the research output from the perspective of the reporting organisation: the specific set of affiliated authors and the specific set of research activities that funded the research output. By merging these views, the total Czech contribution to the research output is seen from the RIV.

Organisations

The IS VaVaI keeps structured information about the following organisations:

- Funding providers;
- Research-performing institutions, including those that enjoy the status of Research Organisation.

There are currently **11 funding providers** that award RD&I funding from the state budget of the Czech Republic and/or distribute EU funding. Some of the 14 higher-level territorial administrative units of the Czech Republic have their funding

schemes, too. In total, through the 20 years' history of the information system, there have been 31 funding providers.

Funding providers are identified by three-character codes. Example: the Grant Agency of the Czech Republic (GA ČR, also known under the name of Czech Science Foundation) is identified by the code "GAO" (G-A-zero). Details of the funding provider and its funding schemes, the calls, and the projects it funds can be retrieved from <http://www.isvav.cz/funderDetail.do?rowId=GAO>.

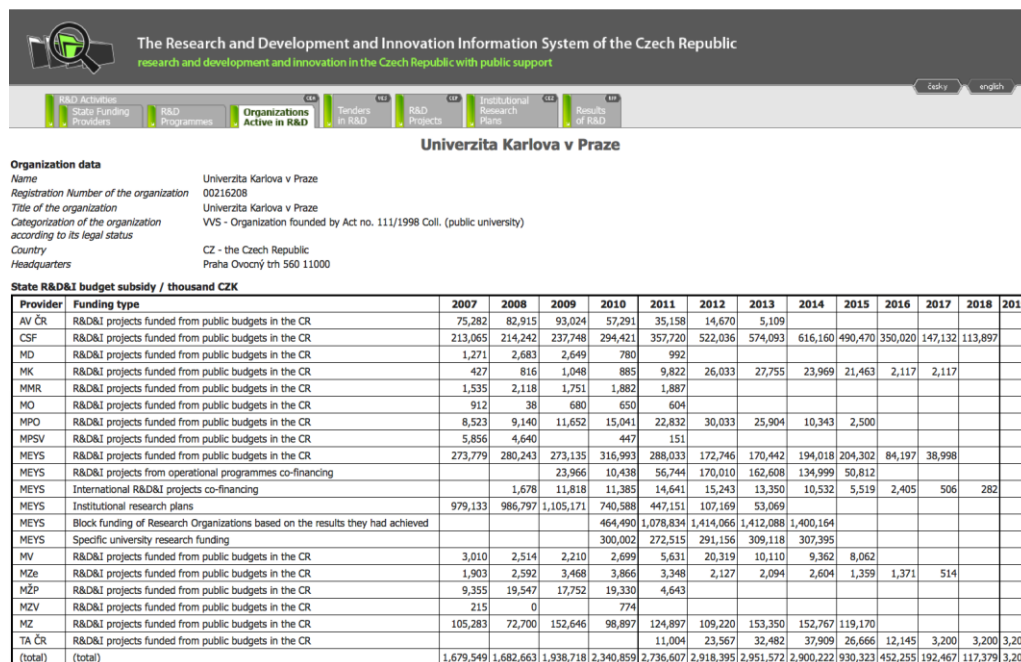
There are around **5,500 research-performing institutions**. These include the following organisations:

- The Academy of Sciences of the Czech Republic (53 institutes);
- 27 public Higher Education Institutions in the Czech Republic;
- Several tens of other HEIs;
- Sectorial research establishments;
- Museums, libraries, archives, ..., and other infrastructures;
- Companies;
- Some foreign organisations too (those that collaborated on funded projects).

This represents around 3,900 legal entities from CR. They are identified using the national legal entity identifier IČO.⁹ For example the *Institute of Mathematics of the AS CR, v.v.i.* is identified by the string "ico:67985840". Details of the organisation and the funding it received can be found at <http://www.isvav.cz/organizationDetail.do?rowId=ico%3A67985840>. See Figure 3 for an example of such display.

⁹ IČO ... *identifikační číslo osoby*, 8-digit numbers (one check digit), in conformance with the Act on the basic registers (no. 111/2009 Coll.), effective since June 1st, 2010, which extended an earlier usage of these numbers

Figure 3. The RD&I subsidy of an organisation



Foreign legal entities are identified using the country code of their origin and a portion of their name. Example: the *Royal Botanic Gardens* in the U.K. are identified by the string “ocs:GB:Royal Botanic Gardens” and the details of the organisation (as seen from the perspective of publicly funded RD&I in the Czech Republic) can be accessed at <http://www.isvav.cz/organizationDetail.do?rowId=ocs%3AGB%3ARoyal+Botanic+Gardens>.

Faculties of the public HEIs form a second tier in the organisational hierarchy - the subordinate organisational units. Currently there are 150 faculties and 260 other subordinate units of the public HEIs. They are identified using a “departmental identifier”¹⁰ assigned by the Ministry of Education, Youth and Sports (MEYS). Example: The *Faculty of Theatre* of the *Janáček Academy of Music and Performing Arts Brno* is identified by the string “orjk:54530”. The details of the faculty, the link to the whole university, as well as the faculty-specific part of the public funding can be accessed at <http://www.isvav.cz/organizationDetail.do?rowId=orjk%3A54530>.

The Ministry of Defence and the Ministry of Interior are the overarching legal entities for defence and security HEIs and research institutes which are represented as organisational units, too. These are identified using codes assigned internally in the IS VaVaI. Example: The Protection of Population Institute of the Fire and Rescue Service under the Ministry of Interior is identified by the string “orjk:K13”, its details can be accessed at <http://www.isvav.cz/organizationDetail.do?rowId=orjk%3AK13>.

¹⁰ RID ... *resortní identifikátor*, 5-character alphanumeric codes assigned by the MEYS.

Eventually, self-employed researchers are sometimes among the beneficiaries of small-scale projects (this is most common in the fields of arts, cultural heritage, labour policies, international relations and similar). For the purposes of the IS VaVaI they have a dual character an organisation (sometimes they are even assigned IČOs) and a natural person. 11

Persons

Information about natural persons is recorded in the IS VaVaI in the following contexts:

1. Beneficiary (including coordinator) of a project;
2. Principal investigator of a project;
3. Team member(s) of a project;
4. Creator(s) of a research output (including authors of publications);
5. Contact person for a call;
6. Technical contact for a data submission.

Table 2 summarizes the data items collected in the different contexts.

Table 2. Data items about persons collected in the IS VaVaI

Person role vs. data collected	Family name(s)	Given name(s)	Person identifier	Titles	E-mail, phone no.	Home address
Project beneficiary	✓	✓	✓	✓	✓	✓
Project Principal Investigator	✓	✓	✓	✓	✓	
Project team member	✓	✓	✓	✓		
Output creator affiliated with the submitting institution	✓	✓	✓			
Output creator external to the submitting institution	✓	✓ ¹²				
Call contact person (with a funder) ¹³	✓				✓	
Data submission technical contact ¹⁴	✓				✓	

¹¹ See <http://www.isvav.cz/organizationDetail.do?rowId=ico%3A72815035> and http://www.isvav.cz/organizationDetail.do?rowId=orko%3A51J8PoUQ_IZM3SE94_D9FTFIAJ_TISM11S8 for examples.

¹² Initials of the given and middle names suffice

¹³ Displayed on the IS VaVaI web for one year after the call closes, then withdrawn

¹⁴ Not displayed on the IS VaVaI web

The following person identifiers are tracked in the information system:

- The “birth number” (*rodné číslo*) that is assigned to the citizens of the Czech Republic and persons with long-term residence permits;
- A substitute “identification number” (*identifikační číslo*) for the cases where a person identifier is required, but the “birth number” is not available.

These identifiers serve the purpose of distinguishing among researchers of the same name. This is necessary in general, but also in the proposed EM where Research Units are formed across the organisational hierarchy, as groups of researchers. To illustrate the scale of the problem at the level of the country, Table 3 shows the most frequent names (the given name and the family name combined) of researchers in the Czech Republic. The need for reliable identification of researchers is clear from Table 4 that illustrates the trends for the names and the real researchers who use them in the country.

In 2014 this challenge was addressed by introducing a **new, countrywide identifier**: the researcher’s identifier (acronym: *vedidk*) of the form of 7-digit numbers. It is derived internally for a major subset of researchers represented in the IS VaVaI. The identifier is much more reliable than those that are based on either researcher self-registration or profile construction using machine-learning algorithms of identity recognition from bibliographic data.

There is a group of researchers to whom the identifiers have not been assigned yet: these are mostly researchers of foreign origin who became affiliated with research institutions in the Czech Republic. One of the findings of the Small Pilot Evaluation is that the identifiers need to be extended so that all occurrences of also these researchers are covered.

Table 3. Most frequent names of Czech researchers (creators of research outputs in the RIV, years 2001-2013, names of at least 20 distinct persons)

Given name	Family name	# distinct persons	# covered subject categories ¹⁵	# outputs	# affiliation org units
Petr	Novák	41	68	588	56
Petr	Dvořák	27	67	529	53
Jan	Novák	26	45	400	32
Pavel	Novák	26	48	1 010	34
Jiří	Svoboda	26	45	486	32
Jan	Dvořák	24	36	200	28
Petr	Svoboda	23	46	360	29
Jiří	Novák	22	32	646	22
Jan	Novotný	22	42	473	30
Pavel	Svoboda	22	33	464	24
Petr	Kučera	20	33	298	23
Jana	Nováková	20	28	420	21

¹⁵ Of the total 123 categories in the IS VaVaI subject classification

Given name	Family name	# distinct persons	# covered subject categories ¹⁵	# outputs	# affiliation org units
Tomáš	Novotný	20	21	171	24

Table 4. Name ambiguity of researchers in the years 2001-2013 (creators of research outputs in the RIV)

Year	# persons	# distinct names	% persons with an ambiguous name	# outputs	# affiliation org. units
2001	14 478	13 747	5.0%	31 436	558
2002	15 909	15 116	5.0%	35 229	576
2003	18 554	17 505	5.7%	40 255	627
2004	20 368	19 210	5.7%	44 004	647
2005	21 516	20 193	6.1%	42 170	642
2006	23 637	22 152	6.3%	47 383	697
2007	26 927	25 086	6.8%	55 638	720
2008	28 316	26 404	6.8%	53 742	728
2009	29 599	27 695	6.4%	52 764	715
2010	33 614	31 375	6.7%	55 915	752
2011	36 116	33 670	6.8%	59 931	729
2012	37 317	34 372	7.9%	62 629	717
2013	36 603	33 677	8.0%	56 796	710
Throughout	83 927	77 687	7.4%	635 056	2 010

2.2.3 Data collection into the IS VaVaI

The data collection process into any of the IS VaVaI components always involves the funders. The system is batch-oriented: every submission has a defined scope.¹⁶ A submission is expected to contain complete, accurate, and up-to-date data about objects within its scope.

The submissions are XML¹⁷ files with the specific ending “.vav”. The format of the files is defined by an XML Schema and is documented.¹⁸ In addition to the XML Schema, around 5,000 integrity constraints for the XML files are defined and enforced through the Checking web service¹⁹ that is mandatory for all submissions to pass. This open format and freely accessible integrity verification service have - since their

¹⁶ E.g. “all standard projects funded by the Czech Science Foundation in 2014”

¹⁷ The *eXtensible Markup Language* standardized by the WWW Consortium

¹⁸ See e.g. <http://www.isvav.cz/xml/2015/struktury-XML-2015-A.pdf> for the most current definition (available in Czech only)

¹⁹ An on-line service accessible at <https://www.isvav.cz/kontrola/> for any user

introduction in 2002 - given rise to a whole market segment of third-party tools and research information system offerings.

The IS VaVaI offers a default, freely available set of tools - the “Vklap”²⁰ - that supports editing and managing information in the “.vav” files. The tools are co-located in a desktop Java application. The built-in editors perform the same integrity checks as the Checking web service: the users get immediate feedback about the data items they need to add or modify. The users are therefore not susceptible to the “surprise further down the road” effect other tools may produce. The Java application is distributed to users via the Java Web Start technology with digital signatures of the code and resources that guarantee that the users are using original code. The application is updated about 50 times per year to reflect changes in the underlying data for the built-in integrity checks.

The data into the **CEA (Activities of RD&I)** and **VES (Public tenders in RD&I)** components are provided by the funders themselves. These are low volume, low frequency data (less than a hundred records for the CEA per year, two records into the VES per call). The funders use the Vklap tools to produce and manage these submissions.

The collection of data into the **CEP (Registry of RD&I Projects)** works with mid-volume, mid-frequency data submissions. Depending on the existence and functionality of the funder’s CRIS, the submission is either produced automatically, produced once and maintained outside the CRIS afterwards, or produced using the Vklap default toolbox from the very beginning.

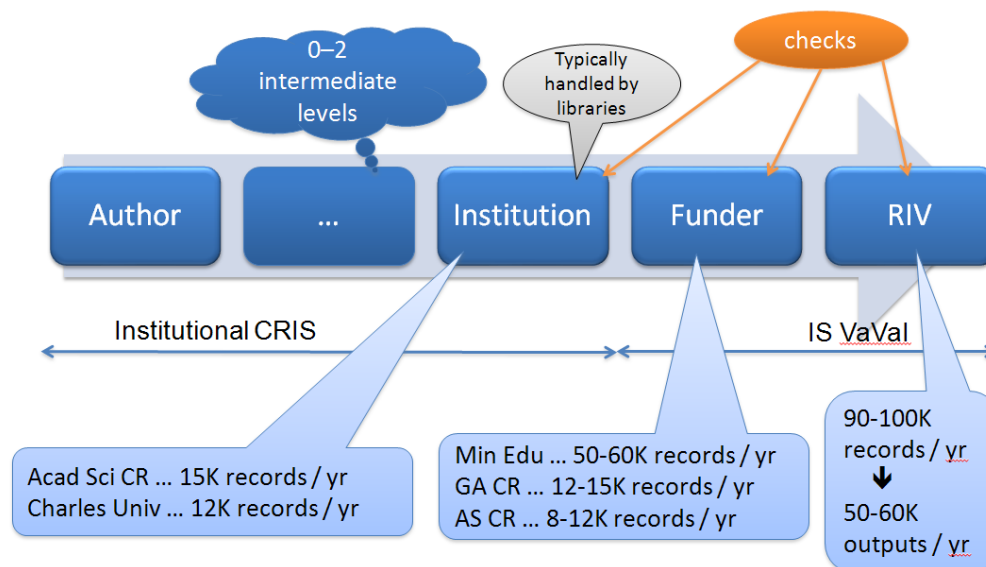
If the Vklap is to be used to prepare the CEP submission, the funder has two options: (a) copy and paste all the information by themselves (most of the information is found in the grant agreement and the project proposal); or (b) ask the principal investigators of the projects to supply the data about the projects and then combine those partial submissions. The Vklap can also support the funders in adjusting the parameters of the projects as they run, as well as in keeping structured information needed for the final evaluation and reporting about the project into the CEP.

The update regime is set forth in the Act no. 130/2002 Coll. as follows: The funder is to supply the first information about an RD&I project into the CEP within 50 days after the grant agreement comes in force. Regular updates are due until the 50th day²¹ of every year the project runs. The Funder shall not transfer the funds until their data submission into the CEP is acknowledged. Any modifications of a project that affect its representation in the CEP are to be reported within 30 days. When the project ends, a final statement is due by July 1st of the year after the project ended: the data includes the finalized financial statement and the result of the ex-post evaluation of the project.

²⁰ See <http://www.isvav.cz/vklap/>, since 2006

²¹ February 20th

Figure 4. The flow of the RIV data collection process with approximate volumes



The **RIV (Registry of RD&I Outputs)** data collection accounts for most of the data in the IS VaVal. The data in the RIV accumulate in about 1,500-2,000 low frequency submissions (typically reported once or twice a year) with sizes ranging from a few records to ones that exceed one thousand records.

The data collection starts with the research output creators (the authors in the case of publications) registering their outputs in an institutional CRIS (the Vklap as the default tool for small organisations). This aggregates to the level of the institution through the organisational hierarchy. Large universities generally follow the levels of department - faculty/institute - university. The Academy of Sciences has the parallel levels of department - AS CR institute - AS CR Library. Mid-size institutions use a single intermediate level hierarchy while the smallest organisations typically handle the data collection directly.

From the level of the institution the data submissions are transferred to the funders of the research activities that supported the creation of the research outputs. The funders do their verifications of the data and pass it to the RIV central database. Figure 1 shows the participants of the data collection process with approximate volumes of the data flow.

Reporting of the research outputs into the RIV is mandated by the Act no. 130/2002 Coll. Research organisations report their research outputs on a yearly basis (fundors shall submit the data until the 150th days of every year). This frequency ensures an adequate degree of timeliness of the research output data in the RIV.

2.2.4 Presentation of IS VaVal data

The public part of the IS VaVal data is presented using a WWW portal accessible at <http://www.isvav.cz>. The presentation allows six types of objects (programmes, calls, organisations, RD&I projects, institutional research plans, research outputs) to be searched by several tens of criteria, possibly in combination. Additionally, it gives a list of funders with detail pages for each funder.

The presentation is strictly bi-lingual (Czech and English). Most of the data is available in both languages, certain attributes (such as the research output title) are available in the original language (which can be any language).

The portal displays details of the individual objects, as well as their links (in both directions). For instance, the Czech Science Foundation call for standard projects in 2006 is presented on the page <http://www.isvav.cz/tenderDetail.do?rowId=SGAO2007GA-ST>.

That page contains links to the supported projects (such as <http://www.isvav.cz/projectDetail.do?rowId=GA101%2F07%2F0588>).

There is the same kind of bi-directional linking between projects and their research outputs, between organisations and projects they participate in, etc. This shows the micro-data. Overall, the portal offers about 2,4 million²² detail pages of IS VaVaI objects²³, with an increment rate of ~ 200,000 new pages per year. All of these pages are accessible to anyone on the web, including search engines.

Almost all funders rely on the CEP to disseminate the lists of their current and past projects.

2.2.5 SWOT analysis

Based on an earlier SWOT analysis in the official Concept of the IS VaVaI for the years 2012-2015²⁴, which was updated with the relevant outcomes of the interviews we conducted (see Section 3) and with information about the state-of-the-art in research information systems that we acquire through the international professional networks we are active in, the following are the strengths, the weaknesses, the opportunities of the IS VaVaI and the threats to it:

Strengths

- An integrated environment covering the whole research management cycle
- Continuously updated
- Reliable macro-data (based on micro-data)
- Centralized, authoritative, verified database of research outputs that comprises a broader range of research outputs than either national or commercial bibliographic resources
- An effective tool for budget planning and verification
- Basic data source for any evaluation of research
- Clear roles and responsibilities of actors, a reliable protocol for data submission
- Reliable IDs of researchers that can bridge to global initiatives
- Presentation is strictly bi-lingual (Czech and English)
- Cornerstone for transparency and trust in the Czech RD&I domain
- Generally accepted

²² As of end 2014

²³ ~ 1,140,000 output records, ~ 40,000 projects, a few thousands of other objects; each object has a Czech detail page and an English one

²⁴ See also [IPn-KA1.Annex]

Weaknesses

- Only coarse-grained information on foreign financial support
- The current structure of the research output record is incomplete as a bibliographic reference
- No cross-checking of submitted output category against other sources
- Somewhat difficult processing of the exported data for external users
- Absence of an interface to allow more complex queries
- Query form too complex for entry level users
- Absence of support for tracking the lifecycle of research outputs
- Currently insufficient support for tracking identifiers of research outputs and researchers in other data sources

(We address most of the weaknesses in the proposed extensions in Section 5.)

Opportunities

- Analytical outputs from the data
- One-stop shop for information on the Czech RD&I system
- Source of quantitative indicators for future evaluations
- More extensive validation of the data
- Improved interconnectivity with other existing systems
- Refine and extend the research field classification to improve support of the EM
- Support for tracking the lifecycle of research outputs
- Improved search facilities

Threats

- Underfinancing of further development of the system
- Insufficient human resources and expertise
- Inability of some funders to provide timely information on projects
- Loss of motivation for the research outputs entry in the system
- Research field classification manipulated to game an evaluation system
- Instability caused by unpremeditated modifications

3. The reliability and usability of the IS VaVaI

In this Chapter we report on our investigation in the context of this of the reliability and usability of the IS VaVaI. The study methods included interviewing the IS users and a testing of the data completeness and reliability during the Small Pilot Evaluation.

3.1 Introduction

An assessment of the reliability and of the usability of the IS VaVaI was carried out using two main approaches:

1. Interviews with selected IS VaVaI users
2. Checks of data quality as a part of the Small Pilot Evaluation (SPE)

The information in the sections below is predominantly based on the input from these interviews. In this chapter, reflections based on the use of the RD&I IS information for the SPE are made in particular related to the quality and reliability of the information in Section 3.2, below.

For the interviews we contacted 21 users of the IS VaVaI from the following three target groups:

1. Research organisation representatives (current or past deans, vice-deans), members of the RD&I Council and of its Research Evaluation Committee.
2. State officials at funding providers (agencies, ministries) or at the RD&I Council Secretariat.
3. Information workers responsible for the reporting of research outputs in the IS VaVaI at the levels of universities, university faculties, research institutes, enterprises - most frequently from the organisations' libraries.

Of these users, 17 responded and made themselves available for an interview.

In the interview we asked 76 detailed questions structured to 40 topics (see Appendix A). The options offered were on a four-grade scale (assuredly positive | rather positive | rather negative | entirely negative) or simplified yes/no. Additional comments were encouraged.

Most of the interviews were carried out within a month's range (September-October 2014). The interviewees were given an opportunity to see the questions in advance of the interview. Some of the interviewees, most frequently those from the first group above, brought in additional persons from their institutions which helped them answer the more technical questions. The interviews were carried out by telephone or in person, depending on the interviewees' preferences and time available. Almost all respondents actively use the IS VaVaI and consider themselves very experienced users of the system (the highest grade on 4-grade scale), three regarded themselves as experienced (the second highest grade). All respondents use IS VaVaI very often or often.

An overview of the interview questions and answers is provided in Appendix A to this report.

3.2 Quality and reliability of the information

3.2.1 The interviewees' assessment

The interview respondents regard the quality of IS VaVaI data as adequate. Verifications are performed by all participants in the data collection process.

Completeness. The interviews indicated that about a half of respondents know about research outputs that could not be reported in the RIV. These include the following types of outputs that the respondents consider missing: anthologies, specialist translations (in the fields of philosophy or oriental languages), expert reviews, prestigious presentation at a world congress or invited talks that do not involve a paper in the proceedings, awards.²⁵

Another reason for outputs of research not being reported are the intellectual property protection aspects. In rare cases of outputs of applied research, the outputs were not reported because (a) of contractual conditions of an accredited laboratory; (b) the idea was considered too precious and describing it for the RIV would give a hint to the competition.

The last reason for incompleteness is the occasional lack of discipline in reporting research outputs from the side of the authors. Sometimes researchers may feel unmotivated to report an output if they know in advance that it will not affect the organisation's performance score in the current PRFS.

Research outputs missing for any of these reasons were observed in very low numbers only.

Accuracy. Verifications are being performed by all interviewed organisations. The least frequently verified aspect is the subject classification of the research outputs (50 %) - this is obviously very difficult to check, especially at the nodes in the data collection process where the numbers of outputs are high. Apparently the trust is put in the authors to classify their works appropriately; however, they are not on their own here either: it's the expert committees of the RD&I Council in the past evaluations until 2012 and verification panels in the present EM 2013-2015 where the subject classifications are checked.

Of the verification mechanisms the automated ones were considered complete; the verifications performed by the expert committees had the least favourable reception. Most data collection participants have experienced cases where they feel a committee's decision was not correct, but those are singular cases overall.

Data quality probes alongside the SPE. We carried out checks of data completeness of the IS VaVaI alongside the Small Pilot Evaluation. These probes asked whether the set of research projects on the one hand and of research outputs on the other hand for Research Units (RUs) were complete.

For the 17 RUs in the SPE, 614 projects and 5,894 research outputs were retrieved from IS VaVaI. The RUs went through their lists and identified potential missing items.

The main finding is that just a few cases of information missing from the IS VaVaI were encountered: one missing and one inaccurate project records (3.3 ‰) and one missing research output report by an EvU (0.2 ‰).

²⁵ It is questionable whether an award qualifies as an output of research. In any case, there certainly is a strong interest to report them.

It seems that if attention of the EvUs is drawn to the importance of all aspects of their image in the IS VaVaI,²⁶ they will check the information and have it corrected.

3.2.2 Detailed discussion of cases reported by the EvUs in the SPE

(#1) 1 project was present in the IS VaVaI (the CEP component) and was not listed to the RU. The reason was that the principal investigator of the project from the EvU was not on the RU's researcher list. The EvU asked additionally for the researcher to be included in the RU. No deficiency about the CEP data was reported.

(#2) 1 project was present in the CEP and was not listed to the RU. It was found that the data in the CEP failed to reflect the researcher's move from another institution to the EvU: both institutions were participating in the project (which ended in 2011). The data in the CEP was obsolete for several years, probably as a result of the change not having been reported to the funder. When the EvU asked the funder to correct the data in the IS VaVaI, this was done and a correction data batch entered the information system. Overall, the corrective submission propagated to the public IS VaVaI website within a week's time.

(#3) 1 project was missing from the CEP: the funder had not reported it. At the date of the Small Pilot Evaluation underlying data freeze²⁷ the funder was already 58 days in delay with delivering the data.²⁸ The funder eventually delivered the data about the project with a total delay of 152 days.

(#4) 1 research output was present in the RIV, but it had been reported by only some organizations. It turned out it wasn't submitted by the EvU whose RU was claiming it. This was an omission on the side of the EvU: it is every organisation's responsibility to report their research outputs.

(#5) An undetermined number of research outputs were not listed to the RUs because of the matching between the identities of the researchers. The search for the projects and research outputs was done via the IS VaVaI researcher identifier ("vedidk") with a fall-back to person names matching where the identifier was not available. Where the research output creators were not citizens of the CR, the method didn't find all the projects/research outputs. This could be overcome by implementing an internationally recognized researcher identification scheme, and collecting the identifiers in the IS VaVaI. Then one can perform a consolidation to disambiguate researcher names and identities in the IS VaVaI. This is one of the extensions we are proposing in Section 5. None of the outputs discussed here were missing from the RIV.

(#6) An undetermined number of research outputs were not listed to the RUs because of a restriction on the field of the RU that was present in the beginning of the SPE exercise. When the RU definition was modified and the said constraint lifted, the three EvUs that requested this were satisfied. This comes at the cost that the panel for a RU will have to assess research outputs from other fields than the field of its focus. This did not challenge the completeness of the IS VaVaI data.

(#7) One RU indicated they were missing 16 research outputs from the RIV. Out of this number, 4 were found in the current RIV, so these were in fact not missing. 3 outputs were found in past states of the RIV, but were excluded for insufficient data

²⁶ In the current PRFS it was mostly the RIV component that was receiving attention.

²⁷ October 17, 2014

²⁸ The deadline is set by the Act no. 130/2002 Coll., as amended, to 50 days after the grant agreement takes effect.

(neither ISBN nor ISSN was present in the conference proceedings article records). The organisation could have re-entered them in the subsequent years, but hasn't done so. We do not consider these publications as missing.

Then, 7 outputs were found on the organisation's own web publication list in categories of "popularisation publications", "non-reviewed presentations", or "presentation without a publication". These were most likely never reported in the RIV. The 2 remaining outputs were not found in either the RIV or on the organisation's list of publications. None of the outputs mentioned in this paragraph was a full scientific publication, so the likely explanation is that the EvU chose not to report the outputs in the RIV.²⁹

3.3 User-friendliness and efficiency of the data entry process

The data collection process has a clear division of roles and responsibilities. Every submission into the IS VaVaI has a defined scope (e.g. all standard projects funded by the Czech Science Foundation in 2014). A submission is required to contain complete, accurate, and up-to-date data about objects within its scope. There is unambiguous responsibility for each submission.

This is an essential condition for the collected data to be authoritative and reliable to support research evaluation. The public accessibility of most of the data is a very strong driver for data quality. This creates an environment of transparency and contributes to the trust among the stakeholders.

The CEA & the VES. Data into the CEA and VES components are provided by the funders themselves. The funders use the Vklap tools to produce these submissions. They find the tools adequate for the job (experience from user support, confirmed by interviews).

The CEP. The data about RD&I projects is provided by the funders, too. The funders can solicit co-operation from the beneficiaries in keeping the information up-to-date (usually as a part of the yearly reporting in the projects). Where the funder has an information system, it can export the CEP submission. Where the funder administers a low number of projects, they use the Vklap to produce the submission.³⁰ CRIS-less funders with a larger number of projects delegate the updates to the beneficiaries who can use the Vklap.

RIV. The data into the RIV component is collected through a process that involves the creators (authors in the case of publication outputs), the institutions, and the funders of research activities that contributed to the creation of the research outputs.

The Vklap toolset supports all the steps in the data collection process. It is not an on-line application, and it doesn't use any central storage. It processes files stored locally; users use their standard means of electronic communication to exchange the files. The fact that the application runs on the users' devices eliminates performance bottlenecks that are often experienced with on-line application.

However, the Vklap is used mainly by smaller organisations that do not have many research outputs to report. Larger, research-intensive organisations build or acquire their own institutional CRISs to support them in collecting and managing their

²⁹ The EvU might have also chosen to report those publications as type O - Other.

³⁰ The Vklap can take a previous year's submission and convert that into one for the current year. The user only needs to update the financial amounts and record project modifications (if there were any).

research output metadata, including reporting them into the RIV. Examples of these systems include the following (in no particular order):

- ASEP at the Academy of Sciences of the Czech Republic (a bespoke system)
- OBD at several universities and other institutions, most notably the Charles University in Prague (a commercial offering)
- VVVS at the Czech Technical University (a home-built product)
- IS MUNI at the Masaryk University (a home-built product)

As a result, most research output authors do not enter their records directly into the RIV. The collection rather happens in the institutional CRIS. As a rule, these institutional CRISs contain a broader range of research outputs than the RIV and use a finer-grained output typology, thus catering for their specificities. Once the data is collected in the institutional CRIS, it is used many times to satisfy various reporting needs locally.

The RIV provides a common platform for the integration and presentation of research output from all of the Czech research performing organisations. The institutional CRISs are adapted for and integrated with the RIV.

Most of the respondents in the interview reported that they import data from external sources. Most frequently they referred to the Web of Science followed by Scopus, sometimes also accompanied by other sources: the National Library Catalogue, the database of the Industrial Property Office, PubMed, INIS, and DOI.

We can conclude that at most institutions, research output metadata have to be filled in at most once. In many cases the records can be imported.

Almost all respondents in the interviews reported that they have sufficient staff for managing the data collection, which is most often the responsibility of the libraries and appointed staff members at departments. Most data administrators have several years of experience.

When asked about problems they meet when realizing the data collection, the respondents mentioned the following ones (ordered from most frequent to the most rare):

1. Timing. Several of the past evaluations required additional input and reactions from the institutions during the summer season. It was problematic to get in touch with the relevant researchers in the (many) cases their co-operation was needed.
2. Methodological ambiguities. In some border cases it was not entirely clear how to describe the research outputs at hand using the RIV data fields.
3. Technical difficulties mostly connected to the institutional CRISs.
4. Low reporting discipline of some researchers.

All respondents expressed satisfaction with the scope and quality of the IS VaVaI user support provided by the help desk.

3.4 The extended use of the IS VaVaI

While the RIV is the most frequently accessed component, the other parts of IS VaVaI are also actively used: most notably CEP (RD&I projects), followed by VES (calls) and CEA (other RD&I activities).

Research organisations use the information for their day-to-day operations:

- Search of funding opportunities;
- Grant application support;

- Search for potential partners;
- Cross-checking their list of active projects and their obligations therefrom, as well as for their strategic management;
- Getting an overview of current RD&I in the fields that are relevant to the organisation;
- Benchmarking with other organisations;
- Analytical studies.

For the funders and the national strategic management bodies these parts of IS VaVaI represent a fundamental and integrated information source that reflects the funding flows in the RD&I domain and give the details on targeted funding in the Czech Republic.

Data on the research outputs produced in an organisation is collected not only in order to be reported in the RIV, there are many alternative uses of the information: for strategic management, for the evaluation of constituent organisational units, to support funding applications, and for dissemination in the expert communities and for the general public.

The quality of IS VaVaI data is regarded as adequate. Total error rates are estimated in the order of a few (at most tens of) cases per year. These are not seen as a major problem. While automated checks of the metadata are considered sufficient, some respondents would support more thorough human verifications.

Applications of an organisation's research outputs are tracked mostly with applied research outputs where they can generate income. Societal impacts are tracked by only a minority of organisations.

The respondents made the following recommendations:

- Add more output types (but they recognize the difficulty of reaching consensus on which ones);
- Track infrastructures;
- Track the usage of research outputs;
- Track awards and other recognitions;
- Improve the search interface of the IS VaVaI web presentation;
- Provide more support to funders in the RIV data collection process;
- Make it mandatory to enter all authors (except for cases with extremely long author lists); make the first author mandatory in any case;
- Track the countries of the authors' affiliate institutions;
- Provide analytical tools for the IS VaVaI data.

3.5 Conclusions

The main conclusion from the interviews is that all respondents recognize the IS VaVaI as an important and valuable information service.

Most large research performing organisations would be collecting the information on the outputs of the research they perform in any case, whether it was required for the RIV or not. All of the responding institutions have the collection processes in place, backed by the necessary infrastructure and qualified staff.

They supply information from their institutions, and they use the aggregated information from all other institutions for their analyses, too. In this sense the

IS VaVaI is providing a service to them. Imprecisions and errors in the RIV are regarded as a relatively minor problem that could be remedied by a more thorough validation.

It is an information system in operation, all stakeholders have adapted to it. It contains authoritative, continuously updated data about Czech RD&I. The *collect once, use many times* principle of good data management is realized. The overall benefit of the IS VaVaI is mostly seen in having a publicly available one-stop shop with research information from the whole country; the key role of the information system for transparency is widely recognized.

The IS VaVaI information was found sufficiently reliable. This is the outcome of the interviews as well as of the checks that were performed alongside the Small Pilot Evaluation in line with the proposed Evaluation Methodology.

4. Options for the use of external information and analytical tools

In this Chapter we cover sources of information that are external to the IS VaVaI and the options for linking them. This includes citation databases, Open Access repositories, article metadata stores, as well as journal databases. We tackle the issue of insufficient coverage of the Social Sciences and the Humanities in traditional citation databases and suggest a possible way of countering that.

The discussion is not limited to research outputs only – person and organisation identifiers are discussed as well.

4.1 The connectivity of the IS VaVaI

Current Research Information Systems (CRIS) such as the IS VaVaI generally tend to be the **central information systems** in research-intensive environments. While they contain the basic information about the important objects in the information domain, they typically tend not to record every detail: instead they refer the user for detailed information to a supporting information system such as an Open Access repository, a bibliographic database, a project management system, a financial system, a human resources system, etc.

CRISs generally use the following two mechanisms to realize these links:

1. Pointers to the external systems' interface (e.g. a URL of the publication metadata page in an OA repository, a URL of a Framework Programme call detail page) in case it is directly addressable
2. External identifiers (e.g. the CORDIS call identifiers, project numbers, the ISBN, ISSN, DOI³¹, etc of a publication, or the patent number, or a researcher identifier - an ORCID³², an agency-assigned number³³, a staff number within an institution)

The former approach provides for navigation to a particular information service, while the latter approach supports a multitude of relevant information services and leaves the choice on the user. Both approaches have their merits.

In this interconnecting function, the role of a CRIS in an organisation is largely that of an **information integration platform**. This is also true of aggregating CRIS such as the IS VaVaI. Connectivity of the IS VaVaI information therefore means recording a wide range of external pointers or identifiers. Connected information allows users to navigate to resources that give more details about a particular object (e.g. from the generic CEP project record to the project's own website - the URL of the website is the pointer). This way the information from the CRIS can reference its context. This is entirely in line with the proposal [IPn-KA1.II.3].

An important particular use-case here is recording the provenance of information. For instance, most research output records in the RIV originate from institutional CRISs. Where this is the case and where the CRIS has a separate internal identifier or a publicly accessible web page of the research output, it can (and should) be recorded.

³¹ Digital Object Identifier, http://en.wikipedia.org/wiki/Digital_object_identifier. The DOI is persistent in the sense that it remains fixed over the lifetime of the document, whereas its location and other metadata may change.

³² Open Researcher and Contributor ID, <http://orcid.org/>

³³ Such as the HESA (Higher Education Statistics Agency) Number in the UK

Apart from linking, this also allows for the information in the IS VaVaI to be enriched with data from external sources. The most prominent example involves a citation database: the RIV could contain the current number of citations of each article or proceedings paper that is matched in the Web of Science or Scopus database. Since the citation count is dynamic, it has to be updated regularly. The key components to implement this enrichment are the following:

- The appropriate identifier,³⁴
- Access to the relevant information service,³⁵ and
- The contractual arrangement that gives the permission to retrieve, store, process and disseminate the particular piece of information.

4.2 Availability and suitability of information resources

4.2.1 Sources for bibliographic data

Internationally, research assessments traditionally turn to **citation databases** to inform on scholarly performance and excellence. Citation databases are the most detailed databases about publications, as they include also the links in the citation network.

Across disciplines, albeit with largely varying levels of coverage, the following two competing information sources are available:

- The Web of Science database by Thomson Reuters – originally a product of the Institute for Scientific Information founded in the 1960s by Eugene Garfield, one of the founders of scientometrics and bibliometrics. ISI pioneered the approach of citation analysis.
- The Scopus database by Elsevier with citations available for works published from 1996 onwards.

Both of these databases are commercial products. For analytical purposes, both vendors offer “research intelligence”³⁶ products:

- The InCites analytical tool by Thomson Reuters. It gives analytical access to the Web of Science database.
- The SciVal analytical tool by Elsevier. It gives analytical access to the Scopus database.

With on-line scholarly publishing gaining momentum and the Digital Object Identifiers (DOIs) now being employed more and more commonly, the **CrossRef**³⁷ **infrastructure** is becoming an alternative source of citation information. While it cannot claim the same level of consistent coverage as the traditional citation databases, it certainly is a service on the rise.

³⁴ The WoS Accession Number (also known as the UT - unique tag) or the EID for Scopus

³⁵ The respective citation database web service interface for machine access

³⁶ Paraphrased from Business Intelligence, the transformation of raw data into meaningful and useful information about a business, or – more generally – about any activity.

³⁷ CrossRef (<http://www.crossref.org/>) is the leading DOI registration agency that specializes in scholarly publication

Another service that can help in searching for individual scholarly outputs is **Google Scholar**.³⁸ However, its records are not bookmark-able (linkable) and it does not offer an Application Programming Interface (API), so it can be used as an auxiliary data source in manual verifications, but not in a more systematic way. The citation counts displayed by Google Scholar are usually the highest, but the service has known problems with distinguishing different versions of what eventually ends up as one paper. That is probably a feature that is determined by the web crawling approach to data acquisition.

Citation databases are the most detailed databases about publications, as they include also the links in the citation network. But even without the citations, **article databases** are very useful. There are many discipline- or field-specific services. These include the PubMed³⁹, DBLP⁴⁰, SSRN⁴¹, or arXiv⁴² services. The former two contain publication metadata only and have very systematic coverage of their respective domains.

The SSRN and arXiv are examples of field-specific Open Access (OA) repositories. Other OA repository types are: institutional repositories (that include OA publications from a specific institution) and funder repositories – the most notable example here being the OpenAIRE⁴³ infrastructure to care for OA publications from Horizon 2020 and former Framework Programmes by the European Commission.

Figure 5, below, documents that the Czech Republic is among the leading countries in Europe in adopting Open Access – at least as regards the subset of scholarly communication that is tracked in the Web of Science database. This may look surprising. Funders in the CR make general declarations, but do not actually mandate or systematically monitor the OA status of the publications that are produced with their funding. OA is being popularized by national library infrastructures and by academic libraries.

Where an article database is not available, at least a database of sources (journals) can help in getting the basic information for an evaluation of institutions. The most prominent example here is **ERIH Plus**⁴⁴ database that covers Social Sciences and the Humanities – more details below.

Also the **List of peer-reviewed non-impacted journals published in the Czech Republic**⁴⁵ that is compiled and regularly updated by the Council for RD&I is a journal database for the purposes of distinguishing journals of national importance. It is used in the current system of evaluation.

³⁸ <http://scholar.google.com/>

³⁹ PubMed Central, <http://www.ncbi.nlm.nih.gov/pmc/>, a service of the National Library of Medicine in the National Institutes of Health, US

⁴⁰ DBLP computer science bibliography, <http://dblp.uni-trier.de/>

⁴¹ Social Science Research Network, <http://www.ssrn.com/en/>

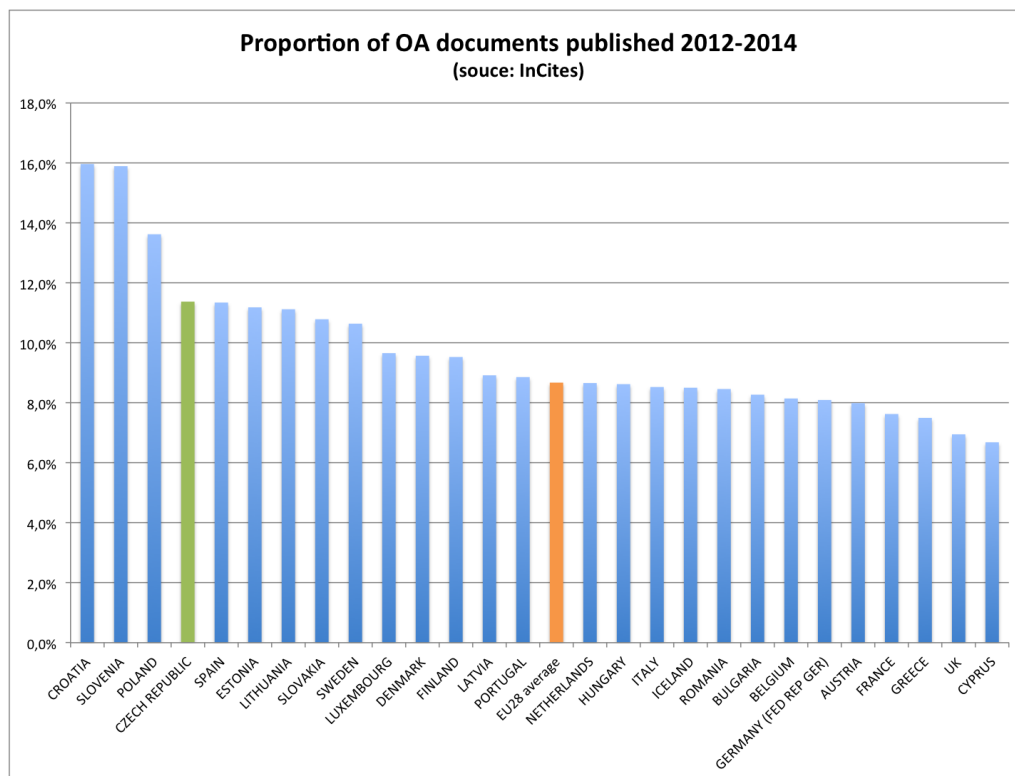
⁴² <http://arxiv.org/help/general>, operated by Cornell University

⁴³ <http://openaire.eu/>

⁴⁴ <http://erihplus.nsd.no/>

⁴⁵ <http://vyzkum.cz/FrontClanek.aspx?idsekce=495942>

Figure 5. Proportion of Open Access documents 2012-2014 (InCites)



An enriched or interlinked RIV

A first approach will result in a **consistent interlinking between the IS VaVaI and a bibliographic database**. We refer to the result of this interlinking as an enriched RIV or interlinked RIV. We believe this approach handles all the current issues.⁴⁶ In a basic form this approach has already been present in the current Metodika 2013-2015.

This approach makes all the attributes from the bibliographic database available to the IS VaVaI as well - the primary example here is the number of citations and the document type from a citation database.⁴⁷ Other examples include getting the form/genre classification from the National Library Catalogue.⁴⁸

Specifically, we recommend that support for the Scopus EID identifier be added so as to cover both leading citation database providers.⁴⁹ This will increase the

⁴⁶ In line with [IPn-KA1.II.3].

⁴⁷ In line with [IPn-KA1.II.6]; the link to the citation database is possible; any dissemination of the citation counts is subject to the license terms negotiated with the citation database provider.

⁴⁸ Alternatively, if such additional information was not deemed reliable enough, a finer distinction could be collected in the submitted metadata.

⁴⁹ In line with [IPn-KA1.II.5].

interoperability of the RIV and foster the ability to use citation data by improving the negotiating position of the IS VaVaI management.

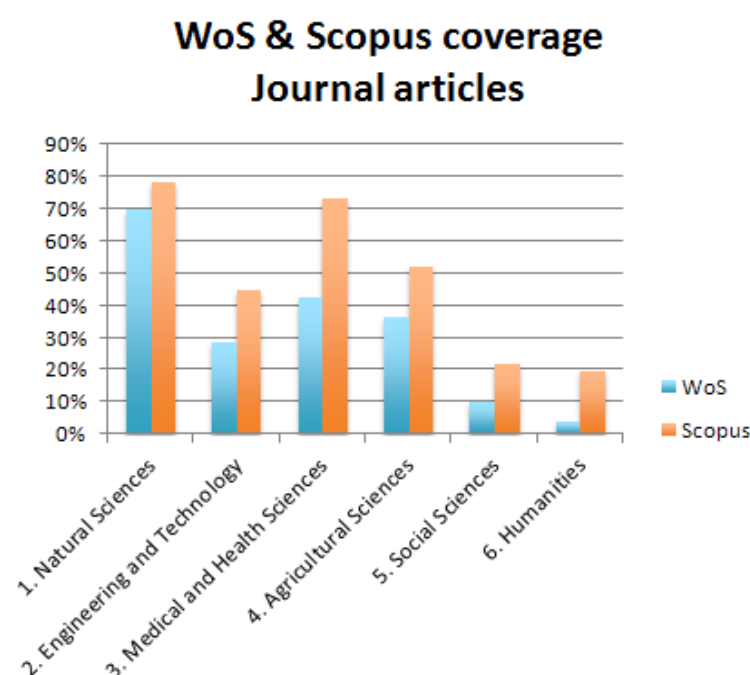
We also recommend that other, discipline-specific identifiers be supported, such as PubMed⁵⁰ ID, SSRN⁵¹ Article Identifier, URL of the article page at the ACM⁵² Digital Library or in the DBLP⁵³ service, etc. Some of these information services, as well as CrossRef, can provide alternative citation information, too. Of course, the citations these services can account for are limited by their scopes.

Creation of a master journal register categorising publication channels

We compared the information about publications registered in the IS VaVaI with the information available in the two international citation databases WoS and Scopus. The percentage of publications from 2012 as reported in the Czech R&D&I IS and matched in the 2 bibliographic databases are shown in Figure 6 and Figure 7, below.

The Scopus database tracks approximately twice as many sources as the Web of Science. However, Web of Science database is generally believed in the bibliometric community to contain more precise information than Scopus.

Figure 6. Coverage of Czech journal articles (RIV 2012) in international citation databases – WoS and SCOPUS



⁵⁰ PubMed Central, <http://www.ncbi.nlm.nih.gov/pmc/>, a service of the National Library of Medicine in the National Institutes of Health, US

⁵¹ Social Science Research Network, <http://www.ssrn.com/en/>

⁵² Association for Computing Machinery, <http://www.acm.org/>

⁵³ DBLP computer science bibliography, <http://dblp.uni-trier.de/>

Figure 7. Coverage of Czech journal articles and proceeding papers (RIV 2012) in international citation databases – WoS and Scopus

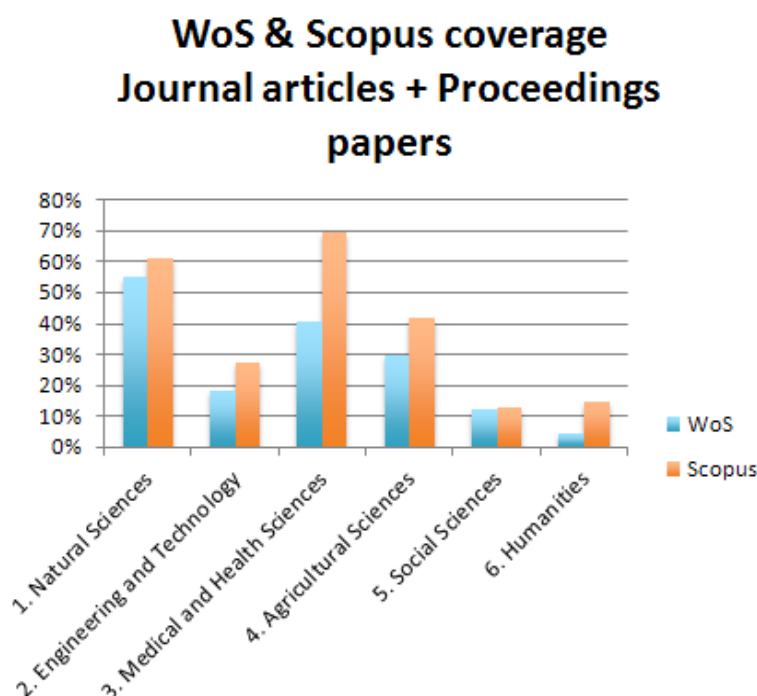


Figure 6 and Figure 7, below, confirm that the coverage of the **Social Sciences and Humanities** (SSH) in Scopus and Web of Science is limited also in the case of the Czech Republic. The general picture, which is similar when studying other CRIS-systems such as the Belgian and the Norwegian systems, is that although Scopus has a broader coverage, the two data sources follow the same pattern in the representation of major scientific areas. The literature shows that this limited coverage of articles and proceeding papers in SSH is mainly due to an incomplete coverage of the international journals in SSH and especially, a limited or no coverage of national disciplinary journals. In addition, but not represented in figures 6 and 7, there is a very limited coverage of scholarly books, which are just as important in the SSH.

These limitations are well known in the international environment and led to the development of the European Reference Index for the Humanities (ERIH), which was created by the Standing Committee for the Humanities of the European Science Foundation. The concept was to add value to output from the SSH domains through a grouping of journals into categories or quality levels, based on peer-review.

As of 2014, ERIH has evolved into **ERIH PLUS**, a dynamic register of approved peer-reviewed journals also covering the social sciences. To be included in ERIH PLUS, journals must meet the following minimum requirements:

- Established procedures for external peer review. The concept of external peer review refers to various forms of editorial procedures that differ between academic fields and scholarly journals, and which indicate that the manuscript has been evaluated by one or more independent experts on the subject matter.
- Academic editorial board (or an equivalent), primarily consisting of scholars from universities, research institutes etc.

- Valid ISSN code, confirmed by the [international ISSN register](#)
- Publication of all original articles with abstracts, as well as author and address information, in English (or in another language relevant for the field)
- International or national authorship, i.e. “less than two thirds of the authors published in the journal are from the same country”, or “more than two thirds of the authors published in the journal are from the same country”

Scientific journals with local authorship, i.e. “more than two thirds of the authors published in the journal are from the same institution”, will not be included in the ERIH PLUS list because independent peer-review cannot be expected to function satisfactorily for a journal that primarily represents an institution’s own researchers.

There also is no longer a categorization of the journals. This is a procedure that each country will be expected to perform by itself, relying on the register as an updated representation of well-performing scholarly journals in the SSH.

There are various examples of countries that have established such a categorization of the publication channels in their countries, at times with a specific focus on publications in the field of SSH. The best-known ones are the VABB-SHW developed in the Flanders (Belgium) and the system adopted in Norway, known as the ‘Norwegian model’.

We briefly present these further below before we use them as a background for presenting a recommendation concerning the creation of a **master journal register for the RD&I Information system** with two major purposes:

- In general, such a register is needed to enhance the analytical potential of the system for the evaluation methodology in all fields of research
- In particular, this register is needed to compensate the deficiencies in coverage in Scopus and WoS in more systematic manner than the present categorizations in the RD&I system allow for

In 2008, the Flemish Academic Bibliographic Database for the Social Sciences and the Humanities (Vlaams Academisch Bibliografisch Bestand voor de Sociale en Humane Wetenschappen, **VABB-SHW**) was created in order to capture non-WoS peer reviewed SSH output and thus to take into account the specific characteristics of the SSH when allocating research funds among the universities. It also explicitly tackles the language issue, i.e. the fact that publications by SSH researchers working in the Flanders are often published in Dutch. It contains the entire peer reviewed publication output in social sciences and humanities (SSH) in the Flanders, making the distinction between social sciences, arts and humanities.

An Authoritative Panel (AP) is in charge of the scientific management of the database, composed of 12 to 18 professors affiliated with Flemish universities and university colleges. The role of the AP is scientifically to select the publications that are allowed for inclusion in the VABB-SHW database (which does not imply an assessment of the intrinsic scientific quality of the publication) and to assign quality labels to the used publication channels, i.e. the journals and editors.

The so-called “**Norwegian model**” for ranking journals and book publishers in two levels is now used in Denmark, Finland, Norway, Portugal, and at individual universities in Sweden. In this model, publication channels are divided into two levels. The highest level is named “Level 2”. It includes only the leading and most selective international journals, series and book publishers.

Publication channels (journal, series, book publisher) should represent authors from several institutions and organise independent peer review of manuscripts before publication. The Norwegian system takes into account only original research publications, i.e. the focus is on research, and not on the full range of the researchers' publishing activities.

The list of journals, series and book publishers on "Level 2" is revised annually in collaboration with national councils in each discipline or field of research. These Norwegian councils propose changes to an interdisciplinary National Publishing Board, which governs the process on behalf of all institutions and has the final decision.

To sum up and to place this information in relation to the need for further development of the IS VaVaI for evaluation purposes, these are the main features of the above mentioned initiatives:

- ERIH PLUS is a *standardized register* of journals and series in the SSH only. It has three strengths: Clear definitions and quality control so that only proper scholarly journals are admitted; it is dynamic and open to suggestions for new journals; it is European-wide and free to make active use of by the Czech Republic. However, the journals and series are not ranked in quality tiers any more.
- VABB-SHW in Belgium is a combination of a *database* and a *standardized register* of scholarly journals, series and scholarly book publishers. The database is a *supplement* to the Web of Science in that it contains all scholarly publications in the SSH from Flemish universities that are not covered in WoS. The register only includes publication channels where Flemish scholars have published. There is no ranking of quality in the register, but it follows the same definitions and principles for quality control as the ERIH PLUS.
- The Norwegian model is also a combination of a *database* and a *standardized register* of scholarly journals, series and scholarly book publishers. Unlike in Belgium, the database is part of a national CRIS-system covering all scientific and scholarly publications in all fields. Here, the WoS-publications, which are imported, provide the supplement. The register of publication channels follow the same definitions and has the same quality control as in ERIH PLUS and VABB-SHW, but unlike these two, the Norwegian model has a division of the register in two quality levels. The highest level is so far defined on a national level in each of the countries in Denmark, Finland, Norway, and Portugal, while the Swedish universities are so far relying on the quality levels that are created in Norway. Except for the quality levels, the Scandinavian countries are collaborating in maintaining the registers. In addition, the ERIH PLUS register is now maintained by the same organization in Bergen, Norway, which maintains the parallel Norwegian register.

In the Czech Republic the **List of peer-reviewed non-impacted journals published in the Czech Republic** (the List) is compiled and regularly updated by the Council for RD&I. It is used for the purposes of distinguishing journals of national importance in the current system of evaluation. Articles in journals from the list only receive points under the current system if they are classified into the fields of the Humanities, Arts, and some Social Sciences.⁵⁴

Journals are eligible for inclusion in the list if they meet the following criteria:

⁵⁴ The "SHVa" and "SHVb" field groups, but not "SHVc".

1. The journal website displays the information on:
 - a) The mission of the journal,
 - b) The date of establishment of the journal,
 - c) The annual periodicity,
 - d) A statement that it is a peer-reviewed scholarly journal,
 - e) Identity of the publisher,
 - f) Composition of the editorial board: the editorial board shall have more than half of external members (not employees of the publishing institution),
 - g) Instructions for Authors (types of manuscripts, method of delivery, etc.),
 - h) The organisation of the review process.
2. Each article is unambiguously classified as either a scientific article or a review article, or some other kind of article.
3. Original scientific articles or reviews, if published in Czech language, have abstracts in English, French, German, Russian, or Spanish.
4. Manuscripts are published only after they successfully complete an independent review process with at least two reviews. Reviewers shall not be employed by an employing institution of any of the authors, or have any other conflict of interest regarding the manuscript they review. The publisher shall archive the reviews for at least five years.
5. Original scientific articles and reviews where at least one author is affiliated to the publishing institution shall not make more than one third of all original scientific articles and reviews published in the journal in a calendar year.

The eligibility of journals is checked by the Council's Committee for Social Sciences and the Humanities, based on a questionnaire submitted by the editorial board of the journal.

An inspection reveals that the ERIH PLUS criteria basically make a subset of the criteria for inclusion of a journal on the List. As a consequence, virtually all journals from the List satisfy the ERIH PLUS criteria.

In our view, the Czech Republic has one of the requirements of the Norwegian model – the national CRIS-system – but not the standardized register which creates the potential of performing bibliometric analysis in the data and establishing quality levels.

A starting point in Belgium, Denmark, Finland, and Portugal, was to acquire the Norwegian list (which has the ambition to be global) for free and then to expand it for the national needs. The ERIH PLUS list can now also be acquired for free from the same organization, the Norwegian Social Science Data Services, in Bergen.

Our proposal is therefore to **combine the List** of peer-reviewed non-impacted journals published in the Czech Republic **with ERIH PLUS**:

- The next update of the List will only admit journals that are listed in ERIH PLUS.
- Journals from the current List will be given a transition period of two years for inclusion into ERIH PLUS.
- The committee will only check the criteria that are specific for the List.
- Journals from the List shall be treated as meeting more stringent criteria than the ERIH PLUS ones.

The need for a distinction of different levels of SSH publishing channels can, at least as regards journals, be catered for in a very organic way: Articles in journals from the List

receive more attention and/or higher weights than those from journals that are in ERIH PLUS, but not on the List.

Important features in the ERIH PLUS, VABB-SHW and the Norwegian model are

- A standardized register opens up for a bibliometric analysis of the publication output which has so far only been possible in Scopus, WoS and other databases that are running on top of a standardized register.
- The standardization makes it possible to identify all publications in one and the same publication channel, thereby making it possible to distinguish between peer-reviewed original research publications and other material (e.g. editorials, textbooks) in the same publication channel.

Only with this option, it is possible to create CRIS data that allow for the identification of scholarly publications versus publications aimed at a broader non-scholarly audience, i.e. the distinction between **communication for research** and **dissemination**.

Scholarly publications are required to:

1. Present new insight
2. In a scholarly format that allows the research findings to be verified and/or used in new research activity
3. In a language and with a distribution that makes the publication accessible for a relevant audience of researchers
4. In a publication channel (journal, series, book publisher) which represents authors from several institutions and organizes independent peer review of manuscripts before publication.

Opening up the possibility of distinguishing between **quality levels** in the journals, series, and the (book) publishers – similar to the ERIH levels of authorship (see above) - provides also a basis for a more self-determined evaluation of the SSH in the Czech Republic than the present categories provide.

4.2.2 Sources for data on IP outputs

Patents, utility models, industrial designs, plant varieties, animal breeds – all of these outputs of applied research have the property of being registered by a specialized office or agency. Typically these organisations keep public databases of the IP they give protection to.

For instance, the Industrial Property Office (IPO) of the Czech Republic keeps a database of patents and utility models and a separate database of industrial designs. Currently these databases are integrated in the RIV to support the processes of data entry and of research output verification. Czech patents account for 80 % of all patents in the IS VaVaI.

A useful extension would be the inclusion of the following types of additional information from the IPO:

- Data on when the exclusivity in reality ended i.e., when maintenance fees stopped being paid.
- Data on transfer of the rights from the patent.

These two types of additional information would make it possible to independently track the lifecycle of a patent after it was reported in the RIV.

Another extension would be to link the RIV page of a patent research output and the document at the IPO web portal. This extension requires co-operation from the IPO web maintainers. In principle also other patent offices could be interlinked with the

RIV in this way, but there are costs of constructing and maintaining the linking bridge infrastructure at the IS VaVaI side.

While patents are the most visible here, similar is the situation in the other types of IP research outputs.

4.2.3 Going beyond research outputs

The extended identifiers are **not limited to research outputs only**. There is a host of other identifiers to use. Funders of research are assigned FundRef⁵⁵ identifiers, there is the ISNI⁵⁶ standard to identify organisations that are involved in creative activities; projects receive numerous identifiers at different stages of their lifecycle.

Researcher identifiers provide persistent identities for humans. These relate to the efforts also to address the problem that a particular author's contributions to research outputs can be hard to recognize. There are the following reasons for that:

- Personal names are not unique,
- Personal names can change (such as with marriage),
- Personal names have cultural differences in name order,
- Personal names contain inconsistent use of first-name abbreviations and
- Personal names employ different writing systems.

There are many standards for the identification of researchers: ORCID⁵⁷, ResearcherID⁵⁸, Scopus Author ID⁵⁹, Digital Author Identifiers⁶⁰. On top of that, basically every digital library or open access repository has service-specific identifiers, e.g. the arXiv⁶¹ author identifiers, SSRN⁶² author identifiers, among many others. Most of these systems are based on research self-registration, sometimes complemented by bulk registration by employers (which can, however, lead to multiple identifiers for one person). ORCID is aspiring to play the role of a bridge identifier.

⁵⁵ See <http://www.crossref.org/fundref/>. E.g. the FundRef ID of the Czech Science Foundation (GA ČR) is <http://dx.doi.org/10.13039/501100001824>. FundRef IDs are similar in form and functionality to DOIs, but the responsibilities in their minting and updating are different.

⁵⁶ International Standard Name Identifier, <http://www.isni.org/how-isni-works>. E.g. the ISNI of the Czech Technical University is <http://isni.org/isni/0000000121738213>. The service is run by OCLC.

⁵⁷ Open Researcher and Contributor ID, <http://orcid.org/>. E.g. the ORCID of one of the authors of this report is <http://orcid.org/0000-0001-8985-152X>. The service is run by the not-for-profit ORCID Inc. organisation. Technically the service uses a reserved subspace (20-million entries) of ISNIs. One of the authors of this report, Jan Dvořák, is a member of the ORCID Technical Steering Group.

⁵⁸ By Thomson Reuters

⁵⁹ By Elsevier

⁶⁰ An initiative in the Netherlands that pioneered author identification, <https://www.surf.nl/en/themes/research/research-information/digital-author-identifier-dai/digital-author-identifier-dai.html>

⁶¹ arXiv, <http://arxiv.org/help/general>, operated by Cornell University

⁶² Social Science Research Network, <http://www.ssrn.com/en/>

In line with the general recommendations for maximum connectivity, the IS VaVaI shall of course allow for these identifiers to be recorded. However, none of them is ready to replace the current use of the national person identifiers (the “birth number” – see Section 2.2.2) and its anonymized form (the “vedik”). The reasons for this are:

- Incomplete and field-specific coverage;
- Lower reliability.

In relation to [IPn-KA1.II.1] we suggest that the “birth numbers” in the IS VaVaI are complemented with the above mentioned researcher identifiers, but not replaced.

4.2.4 Enhancing the connectivity of the information

In order to enhance the connectivity of the information, we propose to extend the IS VaVaI data structure with a generic mechanism to record **any number of pointers and identifiers**.

Along with each pointer/identifier the following information shall be recorded:

- The type of the identifier / the type of the resource at the pointer’s end
- The semantics of the link between the CRIS object and the pointer/identifier

Whereas a pointer references a particular information service, information about the object that is referenced by the identifier can be found in several systems. If the relevant systems are made known to the IS VaVaI maintainers, the appropriate linking mechanism can be constructed. For scholarly publications, systems of de-coupled linking services, such as OpenURL,⁶³ can provide a wider and more dynamic range of links to relevant resources.

Currently, one URL plus a few most relevant identifiers per object are collected in the IS VaVaI. The WoS Accession Number for output types **J** (article in a peer-reviewed journal) and **D** (conference paper) is collected since 2009. The optional DOI can be reported since 2012 for any output type.

Ensuring consistency in the use of pointers/identifiers

Allowing pointers/identifiers of many kinds to be reported is just one side of the extension. One could easily end up with a true hotchpotch of data. The pointers/identifiers are most useful if they are present **consistently** across a whole range of similar objects. For instance, books and book chapters can be reported in the RIV only if they have a valid ISBN.⁶⁴ Similarly, in the roles of project beneficiary, project principal investigator or team member, or research outputs creators affiliated with the submitting institution, researchers who are citizens of the Czech Republic can only be entered if their “birth number” is reported⁶⁵ (see Section 2.2.2 for more details).

The entry of a particular identifier can actually be mandated by the submission protocol, and enforced by the integrity rules implemented by the Checking web service. However, such a mandate should be introduced carefully, so that the principle

⁶³ See <http://en.wikipedia.org/wiki/OpenURL>

⁶⁴ There is a process to handle ISBNs that are technically invalid: upon submitting digital evidence of the publisher actually assigning the number to the reported document, such an ISBN is allowed in the system.

⁶⁵ Again, there is a mechanism in place to handle technically invalid birth numbers.

of fair and equal treatment is observed. For instance, there is a debate whether the submission of an article in an impacted journal⁶⁶ should only be allowed if a valid WoS Accession Number was reported with it. While we agree that having a consistent interlinking between impacted journal article metadata in the RIV and the WoS is very important and desirable, there are the following pitfalls:

- Access to the WoS database is provided on a commercial basis. Not all submitting organisations have access to the database (public HEIs and the Academy of Sciences institutes do). Mandating that could be regarded as incompatible with state aid rules.
- It is difficult to determine, at the time the data collection is actually done (the first quarter of the year following the publication of the article), which journals are going to be listed on the Journal Citation Report for the year of publication (that JCR being published only in the middle of the year following the publication of the article). It is therefore difficult to determine to which articles such an additional business rule should apply.
- There is a natural delay between the publication of the article and its indexing in the WoS database – the delay can be long especially for conference papers. Neither the journal publisher, nor the authors of the article, nor the authors' institutions can do anything about that delay. The delay could, however, undermine the institution's ability to report the article metadata. Effectively, the citation database provider would become part of the research outputs reporting chain, without actually being bound by any obligations.

Whereas the Web of Science database is the prominent example here, the same arguments apply to Scopus or any other bibliographic database. While the Web of Science Accession Number has been collected as an optional field since 2009 (i.e., with research outputs published since 2008), the corresponding Scopus EID identifier has not been introduced yet. Handling of possible additional information from the National Library Catalogue is also analogous.

This leads us to recommending a **hybrid approach**:

1. Not to mandate the identifier to be entered by the submitting institution, but to provide incentives for this to happen. There can be negative incentives, such as displaying a warning by the Checking web service when a submission contains an article without the identifier in a source that is likely to be listed in the database. If a certain proportion of funding was driven by a performance agreement, a possible positive incentive could be a small boost in the funding for each correctly entered identifier.
2. Establish the interlinking in a separate verification process in the RIV, after the publication metadata is submitted. This involves an independent verification of the identifiers that have been reported, as well as a search for the same identifiers on publications that should have the identifier, but none was reported. Naturally, there has to be sufficient time for the verifications to happen.

We believe this approach satisfies the needs in [IPn-KA1.II.11] without the pitfalls that are mentioned above; such verification is entirely in line with the proposal [IPn-KA1.II.4].

⁶⁶ A journal listed in the Journal Citation Report by Thomson Reuters

4.2.5 Connectivity and connectability

A reverse of connectivity is connectability, i.e. the feature that allows for supporting external references to information in an information system.

The IS VaVaI has traditionally held high standards in this aspect: the details of the main objects are exposed on web pages with URLs following a common, easy-to-understand pattern where the variable part involves the identifier of the respective object in the IS VaVaI.⁶⁷

This currently supports human consumers of the information. With machine information processing agents taking off, and in line with the general trend towards Government Open Data, we recommend to extend the portal by exposing the same information in at least some of the **Linked Open Data (LOD) formats**,⁶⁸ while maintaining the same URL structure. This way, the information is made accessible to LOD agents. (see Section 4.2.6 below.)

We shall also support the possibility for **interconnections between research outputs** to better track the lifecycles and usage of outputs of applied research and research datasets, wherever relevant and adequate.

4.2.6 Machine access to the IS VaVaI data

While humans remain the prime users of information in today's IT, they rely to an increasing extent on automated agents that take up support tasks such as searching, aggregating, restructuring and combining information from different sources. Machines as users of information from the IS VaVaI cannot and should not be ignored. Here we list several options for **machine-friendly interfaces** to access the data.

We do **not**, however, recommend an automated machine-to-machine communication to update the IS VaVaI data by the funders or the research organisations. The administrative dimension of the IS VaVaI update processes is not ready for such automation. The submission hand-over protocol is regulated by the Government Decree on the IS VaVaI. While there may be room for making the hand-over and the subsequent handling at the side of the Office of the Government more efficient, explicit actions by human actors in specific capacities cannot (and should not) be eliminated.

⁶⁷ For instance, the URL of the detail page of a research project with identifier X is <http://www.isvav.cz/projectDetail.do?rowId=X> (where the variable part has to be properly escaped so that the URL syntax rules are followed)

⁶⁸ The RDF/XML is a mandatory format to support, RDFa makes it possible to embed LOD markup into standard HTML pages, while formats such as Turtle or N-Triples are conventionally used to support large-scale LOD processing tasks.

CERIF for international accessibility⁶⁹

The IS VaVaI internal data structures share many aspects with the Common European Research Information Format (CERIF). CERIF is the standard data model and interchange format for research information, which includes information about research projects, research outputs, research funding, research infrastructures, organisations and researchers. CERIF was developed in the 1980s and 1990s; in 2002 the European Commission entrusted the maintenance and further development to euroCRIS⁷⁰, a not-for-profit association of research information management professionals.⁷¹ The EC recommended CERIF as the model for building research information systems.

The IS VaVaI development has always kept in mind that the need for CERIF-compliant exports will come. This will likely occur as a result of a future Europe-wide research information standardisation. That would be the ultimate impulse for implementing **CERIF on the output side** of the IS VaVaI, but requirements in this direction will most probably come earlier than that. With respect to CERIF, the IS VaVaI can play the role of a national hub that handles the translation. The CERIF API (now in development by euroCRIS) will be a standardized way of making the IS VaVaI information base accessible to the international audience.

One can contemplate whether CERIF should also be introduced into the data collection format. We cannot recommend that. There are little demonstrable benefits (and certainly no savings) that this transition could bring at the present state of the research information landscape in the country. The IS VaVaI XML format⁷² works for all players; it has achieved the goal of national integration. The funders and research organisations have invested considerable effort as well as financial amounts into the research information systems they are using now. A radical change of the exchange format would be costly and disrupt a functioning ecosystem. Moreover, certain aspects that are at the core of the IS VaVaI data structure⁷³ are only now being introduced into CERIF. We therefore recommend a path of gradual convergence of IS VaVaI input format with the CERIF syntax.

Simple presentation API

Through the user support desk as well as in the interviews, IS VaVaI users have requested an API (Application Program Interface) to offer the basic information about IS VaVaI objects for a quick display on their web pages. For instance, a faculty wants to display a list of research projects they are working on, or a researcher wants to have a dynamic list of publications (independent of their affiliation in the CR). In both cases

⁶⁹ Related to [IPn-KA1.IV]

⁷⁰ <http://www.eurocris.org/>

⁷¹ One of the authors of this report, Jan Dvořák, is a member of the euroCRIS Board and leader of the CERIF Task Group

⁷² The IS VaVaI XML data collection format was designed 5 years before CERIF got its CERIF-XML exchange format

⁷³ For instance, the complex relationship between a publication, its authors and the institutions they are affiliated with. This is a ternary relationship that can be projected in three the binary links CERIF presently offers, but in the general case it's impossible to reconstruct the ternary links back from the binary ones.

the users are interested in the underlying data while they would handle the presentation by themselves i.e., they want to use the IS VaVaI in a mashup.⁷⁴

This calls for a lightweight data format that is close to the presentation layer. JSON⁷⁵ seems to be the obvious candidate, followed by XML. To anchor the structure and the semantics of the data, CERIF is the source of inspiration to turn to.

Linked Open Data (LOD)

The Linked Open Data movement, in particular the Government Open Data, is definitely gaining momentum. Research information is a natural candidate for open exposition. At the same time, it is structurally rather complex and in need of very precise metadata descriptions. The main risk is misinterpretation of the data by its users; the relationship between the producers and the consumers of the information is a very detached one.

EuroCRIS activities in the development of the Linked Open Data presentation of CERIF data are concentrated in the LOD Task Group. We recommend that the IS VaVaI tap on the outputs of that task group in this aspect.

4.3 Conclusions

We recommend that information as rich as possible should be used for the evaluation – and any other function of the IS VaVaI. However, this information has to be based on one data source that is relied on: that primary source should be interlinked and enriched with data from other sources. The basic requirements for such a **primary data source** are coverage, authoritativeness, and control. The natural choice here is the IS VaVaI: It is the authoritative source covering specifically the Czech RD&I, and it is under the control of the Czech Republic.

We recommend that the information in the IS VaVaI be connected with data in citation and patent databases and other available sources to the maximum extent that is feasible and affordable. In particular, interlinking with a citation database provides the international context and the citation information of the publications while maintaining reliable relationships to the Czech institutions and the authors who are/were affiliated with them.

Combining information from several sources is a very demanding task, if the resulting data are required to be usable in an administrative process that will affect the future of research organisations. We rather recommend picking one source, an authoritative one, and enriching it with data from the other sources. We therefore prefer to speak of additional, rather than alternative sources of information.

Where there are several additional sources of information, it is inevitable that contradictory information will be seen. One should be prepared for such situations

The Web of Science and Scopus citation databases are the principal candidates for integration into the IS VaVaI data. Partially this is already being done within the current Metodika. Patent databases are also partially integrated, which is mainly used for the verification of the reported IP outputs.

The source of information that is gaining coverage – since it is intimately connected with the booming world of electronic publishing – is the DOI. The metadata that is accessible (without fee or registration) to anyone seeing the identifier is sufficient for

⁷⁴ See http://en.wikipedia.org/wiki/Mashup_%28web_application_hybrid%29

⁷⁵ The JavaScript Object Notation, see <http://en.wikipedia.org/wiki/JSON>

most purposes. This source will be very useful when combined with the IS VaVaI, especially for verifications.

Field-specific databases, such as PubMed for medical sciences or DBLP for computer science, can also be used to enrich the information in the IS VaVaI. However, the real value of these databases is in indexing the journals and other sources than the citation ones. Enrolling each such database has also its costs, so such decisions shall be based on a cost/benefit analysis.

It is known that the coverage of Social Sciences and of the Humanities in standard citation databases is not satisfactory for an evaluation of research performance of institutions in these fields. For this purpose national integrating Current Research Information Systems (CRIS) are being built: authoritativeness and full representation is achieved through the researchers and their institutions reporting about all the research outputs they achieved. For the purposes of evaluation, publishing channels are usually categorized. That is what we also suggest for the Czech Republic.

Throughout Europe and in many countries internationally the need for effective RD&I strategic management at the national level is recognized by the decisions to build specialized information systems that would comply with the system of RD&I governance in the country, reflect the specific legislative and organisational context, and track the agenda of the State in the RD&I domain. Such information system is an integrating node of all other partial and discipline-specific information sources and tools. The Czech Republic has the great advantage of having a comprehensive CRIS in operation: it only needs to be extended to implement the new Evaluation Methodology, which will become one of the users of the national research information infrastructure.

5. Proposal for the use of the IS VaVaI as an information tool for evaluation

In this chapter we cover

- The potential role of the IS VaVaI as an information tool for evaluation (Section 5.1), and
- The options for an enhancement of the IS VaVaI user-friendliness, effectiveness and efficiency (Section 5.2).

We propose two categories of extensions: those that would be needed for a maximised use of the IS VaVaI in the context of the EM and the institutional funding principles, and those that will improve the user-friendliness, effectiveness and efficiency aspects of the IS VaVaI in general, as well as the reliability of the information.

5.1 The use of the IS VaVaI for the Evaluation Methodology implementation

5.1.1 Introduction

The Evaluation Methodology needs quality information so as to achieve a smooth and efficient process, and a higher degree of trust in its outcomes. The IS VaVaI, if extended as proposed in this report, will provide integrated, verified, authoritative data about EvUs' research activities and research outputs. That will give the evaluation panels reliable and comparable input data in a unified form.

We identified several extensions that would be useful for maximising the use of the IS VaVaI for the proposed evaluation. To this end, the IS VaVaI shall collect the most relevant parts of the information, with continuous updates. The self-evaluation forms of the EvUs and RUs shall, to the maximum extent possible, either come pre-filled with the appropriate summaries of the information from the IS VaVaI, or offer the RUs to pick the most important *N* items from a list.

At the same time, an evaluation following the proposed EM is a large-scale administrative process. The right information has to reach the right participants in the process at the right times. Individual steps and their outcomes shall be recorded and archived. The outcomes of the evaluation will be published using the IS VaVaI, tailored for the various audiences: for the EvUs and RUs, for strategic planning, and for the general public.

The IS VaVaI shall also record the Performance Agreements, which will allow for their interim and final evaluation. The funding that is connected with performance agreements shall be tracked in the IS VaVaI, together with the basic data about the performance agreement: Its identification, dates of entering in force and the dates of funding start and end, the parties, and the performance criteria. A reference to the full text of the agreement, in case it is published, shall be provided.

The Evaluation Methodology will be the most important use of the information in the IS VaVaI, but it is there for other purposes, too: most notably for analytical processing and strategic management at all levels of the RD&I system (research organisations, funders, the RD&I Council).

Already now the IS VaVaI contains data about the basic entities in the research information domain - research projects and research outputs - that are needed for the evaluation. The information is collected through a standard process that has been in operation for many years now. The collection process has several levels of verification, with clear roles and responsibilities.

- Data in the RIV is collected by the research organisations, the funders verify that the research outputs indeed were created in the research activities they reference.

- Data in the CEP, the CEA and the VES are provided primarily by the funders who are the natural guarantors for the agenda.

This division of roles and responsibilities is the fundamental precondition for **validity** of the data.

In the framework of the current evaluation system, the IS VaVaI data are integrated with external sources: citation databases, patent databases, and ERIH. This integration (with a necessary provision for the added costs) should be moved into the IS VaVaI core to make it more systematic. That would support additional verifications at all levels.

5.1.2 Providing data for the self-assessment reports

The EM foresees that the evaluation Evaluated Units (EvUs) will be responsible for the transfer of information on their activities in the context of their self-assessment reports. Extension of the information registered in the IS VaVaI will allow for the self-assessment report to be more extensively and directly based on IS VaVaI data. This will save the EvUs considerable amounts of workload, compared to an ad-hoc approach. The evaluation will be also based on higher quality data, thanks to the standardized, continuously running system of data collection. The data is accessible for everyone who is interested. The fewer data is collected specifically for the evaluation, the better. The data quality will be assured, and the workload for EvU/RU and the overall cost will be lower.

The data will be immediately re-usable for various strategic analyses, funders, research organisations, possible partial benchmarking, etc. The new data will be available in the IS VaVaI for other purposes as well, not only the evaluation. This has the advantage that if there are multiple purposes for a single piece of information, gaming becomes much more difficult.

In order to act as an authoritative and comprehensive source of information that needs to be provided for the EM, we recommend **extending existing components** of the RD&I information system as well as adding **new ones**. The yearly frequency of data updates seems to be sufficient for most uses, given the pace at which the underlying research funding processes take place. Data collection with a frequency lower than once per year would, however, result in a worse data quality due to personnel migrations and both the human and the institutional memories fading out. To stress the importance of regular reporting and increase the motivation of the research organisations, we recommend the introduction of a constraint that only up to 10–20 % of research outputs of an organisation can be reported later than in the year that follows the realisation (publication) year of the output.

Below we list the questions from the self-evaluation forms⁷⁶ that can be entirely or partially supported by the IS VaVaI at its present state or with the proposed extensions implemented.

The following questions are entirely backed by data from the existing IS VaVaI:

- National competitive funding ('targeted' funding) -- aggregated CEP data (Q019)
- Research Outputs -- aggregated RIV data (Q024)
- International conferences -- RIV data (type=M - Organizing a conference), the RU shall pick the 5 most prominent ones (Q042)

⁷⁶ As used in the Small Pilot Evaluation, see also the report The Small Pilot Evaluation: Feedback and Results (Background report 10)

- Funding from contract research -- data about contract research was collected for Pillar III of the current PRFS. It is not publicly available, but with proper clearance it could be used to pre-fill the form (Q022)

The RIV, when interlinked with data from citation databases (see Section 4.2.1), also provides the underlying data for the bibliometric reports to inform the panels.

The following questions can be partially answered using data from the IS VaVaI at its present state:

- Institutional funding for research - the current CEA tracks institutional funding at the RO level; it should be refined to the EvU level i.e., universities should disclose the distribution to the faculty/institute level (Q013)
- Participation in Centres (Q031)
- Participation in incubators and clusters -- data from the CEP in the case the centres, incubators or clusters have the form of projects with national/regional funding (Q034)
- National collaborations and partnerships -- collaborations on projects tracked in the CEP, but other forms of partnerships and collaborations are possible too (Q032)
- Non-academic collaborations -- projects with partners from outside academia (e.g. TA CR), contract research (Q033)

We propose to start the collection of **new types of data** in the IS VaVaI providing information for the following questions:

- Scientific prizes (Q035)
- Membership in editorial boards (Q041)
- Total R&D personnel of EvU (Q008, Q009), Number of researchers of EvU (Q010), Number of researchers of RU (Q014) - aggregations of researcher-level micro-data
- Academic advisory boards and Non-academic advisory boards (memberships in) (Q036, Q037)
- Experimental equipment (Q026), Shared use of equipment (Q027), Other national research infrastructures (non-competitive access) (Q028), Other research infrastructures (competitive access) (Q029)
- PhDs awarded per year (Q017)

This will regard in particular the following items and topics:

Researcher Esteem Indicators and Profiles

Currently the IS VaVaI registers researchers only indirectly, through their participation in projects and through their authorships of publications and other research outputs. Organizing conferences or other events is also included. However, this provides only a partial picture of the researchers' activities and achievements. Most of the RU self-evaluation points out the indicators of esteem the units consider the most prominent or important. The evaluation could be supported to a much larger

extent if the relevant information was present and continuously updated in the information system, and the RUs able to point at it, instead of having to collect it.⁷⁷

Such information has many uses beyond the evaluation: most notably, it can support **on-line researchers CVs**, which is a very useful for the researchers.

Apart from conference organization, the IS VaVaI is currently ignorant of the researchers achievements: prizes or other awards, invited lectures, memberships in scientific and professional societies and associations, in programme committees of conferences, in evaluation panels, positions in policy bodies, offices in the academic sphere, memberships in editorial board of journals - none of these are currently represented.

We recommend to introduce a separate component in the IS VaVaI to manage the information about researchers. The natural chain of reporting and responsibility for the data is:

1. Researcher
2. Research organisation
3. Institutional funding provider
4. IS VaVaI

With such an extension in place, a large proportion of evaluation forms may be prepared in a semiautomatic way.

Data for the calculation of FTE researchers

The single most laborious task on the SPE self-evaluation form (aside from the ones for which the data can be obtained from the IS VaVaI already) was the question on the **researcher headcounts and FTEs**. This information also plays an important role in translating scores into funding⁷⁸ where the FTE sizes of the EvUs and RUs need to be known.

The Czech Statistical Office collects the data with the granularity of the whole Research Organisations, but it is bound by the law to only make the data available in an anonymized form. The statistical methodology also does not capture the effect of multiple concurrent employments of researchers in organisations from different sectors. As a result, the head counts of the sectors and at the country level may be slightly overestimated since it is often the same researcher employed at several organisations.⁷⁹

We therefore suggest establishing a central register of researchers in organisations that have the Research Organisations status. The ideal variant includes the following information:

1. The identity of the researcher
2. The FTE (Full Time Equivalent) of each researcher
3. Information on the extent to which the employee conducts research (an attribute with categories)

⁷⁷ In line with [IPn-KA1.II.9]

⁷⁸ See the Final report 2 – The Institutional Funding Principles

⁷⁹ In line with [IPn-KA1.II.9]

4. The internal subdivision of the EvU to departments
5. The association of researchers with departments and positions

The dynamic nature of this information calls for a lightweight update mode to avoid staleness. This will allow for the panels to see the researchers of the RU in the context of their organisation.

Research infrastructures⁸⁰

We suggest the extension of the information in the RD&I IS with an **Infrastructure Catalogue** in three main categories: Facilities, Services and Equipment. Use of the information will allow efficient equipment sharing and advertisement (including international) of research-oriented services.

Tracking usage would be done in three dimensions: by whom the infrastructure is used, in what project and to which outputs it contributed. The data will be available during assessment of the (mainly large national) infrastructures and will help with strategic management of infrastructure.

5.1.3 Handling the evaluation process

Another core functionality will be a direct support of the evaluation procedure: The assistance with the draft of the structure of the subject panels, registration of EvUs into the evaluation, definition of RUs, generation and maintenance of the list of researchers, eligibility checks, preparation of data summaries and presentations for the panels, systematic collection of the self-assessment reports (with the EvUs/RUs re-using the data already collected in the IS VaVaI), realizing the workflow and recording the conclusions drawn up by the panels, the audit of the performance agreements and preparation and implementation of the funding principles of EM.

This support of the evaluation process can be implemented either as a module of the IS VaVaI, or as a separate system that takes data out from the IS VaVaI and eventually feeds back the evaluation outcome.

The evaluation for the whole country will be a large exercise, also from an information management perspective. It will involve several hundreds of referees and panel members, several thousand users in the EvUs and RUs, as well as several tens of users in auxiliary bodies of the evaluation (such as the helpdesk). The IS VaVaI will register these users and manage their access rights. The authorizations of the users will depend on their capacities - on the positions they will have in the evaluation bodies, on their appointments by the EvUs, and on the stages of the relevant sub-processes. Substitutability needs to be ensured, as well as support for the people in positions changing.

Specifically, at both the EvU and RU levels two basic kinds of users will exist:

1. Users who prepare the submissions
2. Users who approve the submissions

An approval can be done through:

- An on-line action of the properly authenticated approving user;
- Downloading a digital document, placing a digital signature on it, and uploading it back in the system; or

⁸⁰ In line with [IPn-KA1.III]

- Downloading a digital document, printing it, undersigning it, and sending it in via surface mail (optionally also complemented by uploading a scan of the document).

The choice of the form of approval depends on the type of statement that is to be made.

Identification of the subject panels and panel membership management

The evaluation panels as well as the auxiliary bodies of the evaluation will be defined in the IS VaVaI together with their relationships.

The RIV data at the national level will support the process of structuring the panels and identifying the subject panels. If the goal is to have at most 25 panels that should see approximately balanced numbers of research outputs, a non-trivial phase of merging related disciplines has to take place.

The tentative and the appointed members of the panels will be registered in the system together with their positions (panel member, panel chair, panel secretary). For each position the current holder as well as the history shall be recorded, with the exact dates of the changes taking effect.

Managing the contracts with the panel members would be outside the scope of the extended IS VaVaI, but it will provide useful underlying information to a human resources system where the contracts will be managed.

EvU and RU registration including eligibility checks

The EvU registrations will be supported. Underlying data about the EvU will already exist in the IS VaVaI, so no new identification data shall be collected. The registration will then basically consist of making the contact person(s) known to the system.

The IS VaVaI will keep the list of researchers in the EvU and their FTEs (see Section O). Alternatively, reviewing the list of researchers over the evaluation period will be the first large-scale task for the EvU. The basis of the list would be created from the RIV (list of output creators who were affiliated with the EvU), the CEP (list of researchers who participated in RD&I projects on behalf of the EvU) and the CEZ (list of researchers who participated in the EvU's institutional research plans). Any missing researchers shall be added to the list. The information about the researcher FTEs shall be input to support the self-evaluation report creation.

When the list of researchers is consolidated the EvU can start defining its RUs and assign the researchers into them. The formal eligibility criteria for the RUs can be checked on-line thanks to the tight integration with the underlying data in the IS VaVaI.

EvU and RU level self-assessment reports, selection of the outputs for peer review

The IS VaVaI will provide support for both the EvUs and the RUs in filling out the self-assessment reports. As a matter of fact, most of the information will be possible to fill in - at least partly - from the IS VaVaI information. Section 5.1.2 has more detail on the specific questions and the degree of support from the side of the IS VaVaI their answering will have.

Generally, the questions on the self-evaluation forms are of three types:

1. Questions asking for summary data, typically over the years of the evaluation period.
2. Questions asking for a certain number of the most prominent items.
3. Questions asking for a free form view, comment or justification.

IS VaVaI as an authoritative, continuously updated information service is expected to contain complete, accurate and up-to-date information. Therefore questions of the first type can be pre-filled without burdening the user. If the user wants to check the list of underlying items (the micro-data), they will be able to do so. If they spot an

inaccuracy, they can have the data in the IS VaVaI corrected through the normal data collection mechanisms. Also the panels can consult the lists if they want to delve into the characteristics of the items that add up to the summary figures.

Answering questions of the second type can be also supported by the IS VaVaI: the system contains all (or most of) the eligible items. It's up to the user to pick the most relevant ones and to comment on their importance. At the same time, automated checks can prevent situations such as an EvU listing a project they haven't participated in. However, the system will not preclude the users from adding an item that is not represented in the IS VaVaI.

Selection of the research outputs for the peer review is essentially a task of this type, too.

Optionally, the user who is tasked with selecting the N most prominent items of a kind may want to ask other users (e.g. heads of departments) to provide a few suggestions from which (s)he could pick. The information system will provide support for this kind of interaction as well.

The tight integration between the information system and the reporting process will result in a higher quality of the information provided, and yet at just a fraction of the workload compared to what would be needed for collecting the information ad hoc. The added benefit is that the information should ideally remain available also outside of the evaluation.

Peer review and panel decision support

The IS VaVaI will also support the process of the panel evaluation. This involves the following steps:

1. The panels select the referees for the submitted outputs
2. The referees read and review the outputs that are assigned to them
3. The panels make their judgements (based on the reviews, the self-assessment reports, the bibliometric reports, and on their discussion)

Recording and disseminating the results of the evaluation

The whole body of information from the evaluation is to be preserved. Parts of it shall be made accessible to the RUs, and parts shall be made public. All these access modes will be handled by the IS VaVaI.

Alternative sources of information

Combining information from several sources is a very demanding task, if the resulting data are required to be usable in an administrative process that will affect the future of research organisations. We rather recommend picking one source, an authoritative one, and enriching it with data from the other sources. We therefore prefer to speak of additional, rather than alternative sources of information.

As discussed already in Section 4.2.1, the Web of Science and Scopus citation databases are the principal candidates for integration into the IS VaVaI data. Partially this is already being done within the current evaluation methodology. Patent databases are also partially integrated, which is mainly used for the verification of the reported IP outputs.

The source of information that is gaining coverage – since it is intimately connected with the booming world of electronic publishing – is the DOI. The metadata that is accessible (without fee or registration) to anyone seeing the identifier is sufficient for most purposes. This source will be very useful when combined with the IS VaVaI, especially for verifications.

Field-specific databases, such as PubMed for medical sciences or DBLP for computer science, can also be used to enrich the information in the IS VaVaI. However, the real

value of these databases is in indexing the journals and other sources than the citation ones. Enrolling each such database has also its costs, so such decisions shall be based on a cost/benefit analysis.

Where there are several additional sources of information, it is inevitable that contradictory information will be seen. One should be prepared for such situations.

5.2 Options for the enhancement of the IS VaVaI

A set of extensions to the IS VaVaI can be envisaged – and would be useful – in order to optimise its user-friendliness, efficiency, effectiveness, and most important, reliability. In the sections below, we list these possible extensions, starting with those that we consider as most relevant and important.

5.2.1 Subject classification

The IS VaVaI uses a classification consisting of 123 fields. The classification was introduced in 1997 and has not been modified since. The classification uses two-letter codes.

Throughout the classification's lifetime various groupings of these fields were promoted. The current PRFS (Metodika 2013-2015), which introduced a peer-review component into the funding system, uses 11 groups of fields.

Any research project in the CEP (and any institutional research plan in the CEZ) is classified using one, two or three fields.

Any output in the RIV is classified in one field (by the submitting institution). This makes for a much finer-grained classification than the one found in citation databases where the document typically inherits its classification from the source it appears in. Also, the sources are often classified in several categories, which leaves a trace of fuzziness about the classification of the individual documents. This should be borne in mind when interpreting analytical outputs from bibliographic databases.

The proposed Evaluation Methodology is based on the internationally recognized OECD Fields of Science 2007 classification (see Appendix B to this report). It has 6 major areas with 36 fields and 6 “other” categories. There is a third level of the classification consisting of approximately 190 sub-fields. The use of this classification gives the EM the advantageous property of international compatibility.

However, the subject classification system is the most sensitive spot of any research information system. Disruptive changes have the potential to distort any discipline-based time series that is derived from the data. Subject classifications have to be modified or changed with utmost care. It is also important to emphasize that virtually no CRIS or bibliographic database use the OECD classification natively. Rather, they employ their own schemes or schemes that are the most relevant for them, and when asked for OECD-compatible analyses, they use a mapping from their native classification into the target one. Few of such mappings can be described as straightforward, smooth, and free of issues.

When comparing information from two research information systems that both use different native subject classification schemes, the user has to be made aware of the inherent inconsistencies that are bound to occur.

For these reasons we strongly warn against the classification schemes to be simply switched. Below we formulate a proposal that leads to the implementation of the OECD classification while the inherent disruptive impact is minimized. We aim for a **seamless map-ability of a refined IS VaVaI classification scheme** into the OECD FOS.⁸¹

The modifications suggested below will make the IS VaVaI compatible with the OECD FOS classification. This will considerably improve the general level of international comparability of research information from the Czech Republic, and the international compatibility of the EM results in particular.

The matching of the IS VaVaI classification with the OECD FOS is based on equivalences or overlaps at the most detailed levels: the 123 RD&I IS categories vs. the 190 OECD sub-fields. Matching of higher levels in the hierarchies sometimes gives misleading results, partly because the OECD field labels give only a rough idea about the range of sub-fields that are under it.

Below we list the types of cases that we encountered when doing the classification schemes matching for the Small Pilot Evaluation. For cases that pose difficulties we present the pragmatic approach taken for this study and a recommendation for a systematic resolution of the issue (which may take longer to implement).

In Appendix B to this report we provide an overview of IS VaVaI fields matched into the OECD FOS ones.

Cases where seamless mapping was possible

When doing the mapping for the purpose of the bibliometric analyses in the Small Pilot Evaluation, we were able to seamlessly map 106 out of the 123 IS VaVaI categories. There were basically two types of correspondence:

1. One-to-one correspondences (such as *AN – Psychology* → *5.01 Psychology*)
2. Several-to-one mappings (such as *BK – Fluid mechanics* → *1.03 Physical sciences*)

Composite IS VaVaI field

Some categories in the IS VaVaI **combine sub-fields from several OECD fields** together. E.g. the category *FN – Epidemiology, infectious diseases and clinical immunology* combines sub-fields from *3.03 Health sciences* (namely, *epidemiology* and *infectious diseases*) on the one hand and sub-fields from *3.02 Clinical medicine* (namely, *clinical immunology*) on the other hand.

In two cases this spans across two OECD major fields, in both cases between the Social Sciences and the Humanities:

1. *AC – Archaeology, anthropology, ethnology*: here *archaeology* goes under *6.01 History and archaeology*, whereas *anthropology* and *ethnology* are found under *5.04 Sociology*.
2. *AJ – Literature, mass media, audio-visual activities*: here *literature* goes under *6.02 Languages and literature*, while *mass media* and *audio-visual activities* match against *5.08 Media and communications*.

⁸¹ This is a proposed solution for [IPn-KA1.II.2]

Here we recommend to refine the IS VaVaI classification by splitting these composite categories into the constituent ones. We suggest that existing projects and research outputs having these classifications shall be classified finer. If this is not possible in its entirety then it should at least be made for data from several most recent years. A combination of semi-automatic and manual methods shall be used.

Field of approach vs. field of application

In some cases one encounters differences in emphasis between the two classifications. E.g. *GB – Agricultural machines and constructions* in the RD&I IS classification stresses the fact that the machines and constructions are used for agriculture. The OECD classification does not consider this aspect to warrant a specific sub-field: the closest matches are *2.03 Mechanical engineering for agricultural machines* and *2.01 Civil engineering for agricultural constructions*. Similarly, *GA – Agricultural economics* is mapped into *5.02 Economics and business*, simply because in such cases the primary aspect to determine the OECD classification field is the field of the primary scientific approach, method or artefact, rather than the field of its application.

Recommendation for future development of the IS VaVaI is to leave the categorization as it is, for the sake of compatibility, except for cases described in the section ‘IS VaVaI categories on the edge of research’, below.

The specific case of Applied Statistics

The OECD classification explicitly states that applied statistics should be classified under the field of application.⁸² The IS VaVaI classification has the category *BB – Applied statistics, operational research*. There are over 2,000 publications in this category in the years 2008-2012. An inspection of a sample of the titles of the works and of the institutions reveals that almost all fields of application are covered.

We recommend that the category be discontinued, thus forcing new research projects/outputs to be classified under the field of application. Existing research projects/outputs will remain classified as they were.

IS VaVaI categories on the edge of research

Certain categories in the IS VaVaI classification are arguably more fitting into a broader classification of human activities rather than a classification of research fields. Examples include categories such as the following:

1. *AK – Sport and leisure time activities* (of which the scientific parts are likely to go to one of: 3.03 Health sciences, 5.04 Sociology, or 5.06 Political science)
2. *JY – Firearms, ammunition, explosives, combat vehicles* (where the development and innovation aspects would probably be more appropriately classified under 2.03 Mechanical engineering, 2.04 Chemical engineering, or 2.05 Materials engineering, the effects of them under 3.02 Clinical medicine or 1.06 Biological sciences and the social aspects in 5.04 Sociology or elsewhere in the Social Sciences major area)

⁸² A possible justification is the following one: The knowledge from the field of statistics has been successfully packaged into statistical software; sometimes it has even been integrated into scientific instruments. Statistics has become a fundamental supportive pillar of the scientific method. Whenever research has a quantitative component, statistics is likely to be used. Increasingly often, this happens entirely behind the scenes for the researchers, without them having to tackle the complexity that is involved. Whether statistics was or was not used in a research has essentially become non-information.

We recommend that these categories be discontinued, thus forcing new research projects/outputs to be classified under the field that better describes the approach taken. Existing research projects/outputs will remain classified as they were.

OECD fields with no matching IS VaVaI categories

Four OECD fields were not matched by any category of the IS VaVaI classification: *2.08 Environmental biotechnology*, *2.10 Nano-technology*, *3.04 Health biotechnology*, and *4.04 Agricultural biotechnology*. There are the following reasons for this:

- The scope of the fields in question is relatively narrow, especially when compared to whole disciplines such as Physics or Biological sciences (which are also at the 2nd level).
- The practical applicability of the fields in question has only recently grown to justify their inclusion as fields in the OECD classification. The IS VaVaI classification has not reflected these developments (yet).

(We also did not map any IS VaVaI category into *1.07 Other natural sciences* and *6.05 Other humanities*, but that is not a problem.)

We recommend to consider an extension of the IS VaVaI classification with four categories that would correspond to these fields.

5.2.2 Improve RIV compatibility with established bibliographic conventions

While the RIV is not a bibliographic database in itself, it very much benefits from linking to bibliographic databases. To that end, it is advantageous that it actually has its content structured in a similar way. Among other goals, the RIV should be able to generate the standard bibliographic record for a publication, since that is what its users and consumers of the publication metadata are used to work with. We put forward a series of recommendations for extensions with the aim of achieving this higher degree of compatibility.

First author. The name of the first author is a very relevant piece of bibliographic information. Currently it is not necessarily recorded in the RIV: every reporting organisation has to specify its affiliated authors only. Unless there is just a single organisation reporting the research output and all the authors are affiliated with that organisation, it's possible that the first author is not listed, and it's certain that (s)he is not flagged as the first author. We recommend the entry of the first author to be made mandatory.

Full list of authors if their number ≤ 15. Full lists of creators are required for results of applied research. We recommend that this be mandated also of publications, unless they have more than 15 authors.

Editors for type C, D. Currently chapters in books and papers in proceedings record the ISBN and the title of the encompassing source, but not its authors or editors. However, the need to search for the documents in external information sources, as well as the need to generate standard bibliographic records, requires that the secondary responsibilities for the publication be recorded as well.

Separate name of event for type D. Current RIV does not really have a field for the name of the conference/event. It is therefore typically put into the R30 field (the title of the book of proceedings in the original language), usually as some form of a subtitle. This makes the information about the event – which is otherwise well structured – look very unbalanced. It's very difficult to search on conference names, for instance.

ISSN of book series for type B, C. Some book series (for instance the Lecture Notes in ... series by Springer) are regular continuing sources that are assigned ISSNs. It makes sense to record these ISSNs, too.

Multiple ISSNs. We are finding ourselves in a period of transition from printed scholarly communication to a fully on-line one. Journals and other continuing sources typically use two ISSNs, one for the on-line form and the other for the legacy form. Using the generic mechanism of recording arbitrary identifiers (see Section 4.1) it is possible to record both ISSNs. Recording **multiple ISBNs** is possible using this approach too, with the limitation induced by the fact that there are other attributes that are functionally dependent on the ISBN, such as the publisher name and the place of publication. For full generality, this should be reflected in a more general data structure.

Separate title of book series for types B, C, D. Book series title is also an important attribute that should not be confused with the title of the book/book of proceedings itself. It should be recorded separately.

Possible other improvements. There are possible other minor improvements that would raise the level of regularity of the information in the RIV.

5.2.3 Introduction of a RIV master list of publishing channels

The RIV presently collects information about journals as simple attributes (one ISSN, the title, the publication year). This form of incoming data, especially given the volumes of around 40,000 journal article records per year, requires tedious checking and unification of the journal title spelling and conventions. A similar situation is with books, book chapters and conference papers. All together these publication outputs make up 80 % of all research outputs per year.

The remedy is well known: use a controlled list of publishing channels: journals, series, and book publishers. The model here is similar to the introduction of the controlled list of patent offices in the 2012 RIV data collection campaign. As a result, the rate of imprecise information entered about the patent issuer dropped dramatically.

However, the experience cannot be translated directly: the list of publishing channels is about a hundred times larger and much more dynamic. It can be created from the existing RIV data, most of which had been verified already. The list will have to be maintained to reflect additions and changes in the journals' attributes, which can be done with the aid of relevant sources – the ISSN portal, the Ullrich's database, the ISSN agency for the Czech Republic (the National Technical Library), the National Library catalogue, the Journal Citation Report by Thomson Reuters, the Scopus title list by Elsevier, the ERIH PLUS database, the WorldCat service by OCLC. The full history of the list shall be preserved.

In any case, the possibility of reporting a research output in a channel that is not (yet) on the master list shall be retained. It should, however, be encouraged that institutions use entries from the master list.

The existence of the master list of publishing channels would allow for a simpler matching between incoming RIV data and external bibliographic sources the data is to be verified against. This would result in an improved linking into citation databases. It would also lower the burden on the maintainers of the institutional CRISs, as they would be able to re-use the master list instead of compiling and maintaining their own.

5.2.4 Revision of the RIV output types classification

The IS VaVaI definitions of the output types are stable since 2009 and should be regarded as primary. The definitions should allow for all research outputs to be reported. It is up to the evaluation to make a selection of outputs it takes into account:

such selection is based on the research output types, years and possibly other criteria. In this way, the evaluation of research institutions, an evaluation of research infrastructures, an evaluation of applied research, and an evaluation of funding programmes (with their specific needs) will make use of different selections of research outputs from the RIV. This is a general approach catering for the needs [IPn-KA1.II.8] and [IPn-KA1.II.10].

The definitions of the publication output types (types J, B, C, D) are in line with international practice and satisfy the requirements of the EM. Regarding other output types we make the following two suggestions.

Better characterise Software

The Software output type (code R) is a broad category. Its present definition is rather general (which is not a disadvantage), but getting further details about each piece of software would be useful. Namely, the following information items would be useful as additional characteristics of software:

- Type of software (packaged software, a tool, a software library, a service accessible on-line, ...);
- Type of contribution of the institution's staff to the software (created from scratch, contributed a new module, bug fixing, project leadership, ...);
- Approximate scale of the contribution;
- The user base.

All of these items can conveniently be collected using pre-defined vocabularies with an option of free-text comments.

Add Research Datasets

The current trend towards (a) preserving research datasets, and (b) making them shareable and re-usable, which is most visibly represented by the Research Data Alliance⁸³, is going to gain momentum also in the Czech Republic. We recommend that research datasets be included in the RIV for the following reasons:

1. Make research data exist in the national research information landscape, make it linkable and available in its context (e.g. of the relevant projects)
2. Support the sharing and re-use of research data by giving credit to their contributors and the institutions

5.2.5 More comprehensive support for funders

Funders in the current IS VaVaI receive only a basic support. To enhance the effectiveness and efficiency of their reporting obligations, a few extensions are suggested.

Notifications about the funders' obligations⁸⁴

The current IS VaVaI produces a report on the past, the current and the upcoming reporting obligations (called the *Data expected in the IS VaVaI* report). This report is available as:

⁸³ <https://rd-alliance.org/>

⁸⁴ Related to [IPn-KA1.I.1]

1. The entire report that combines all funders - targeted at the information system personnel;
2. Separate reports by individual funders.

The latter should be made systematically available to the funders. The current concept document suggests the Vklap application as the information carrier. This has a certain weakness that the application is not in use by all funders. Also, the nature of a locally running application (where all the resources need to get downloaded to the user's computer before the application can run) does not guarantee the desired level of isolation between the funders. There may be a more effective way in simply creating funder-specific pages on the IS VaVaI web portal with proper authentication. Automated email notifications of changes can be an additional benefit.

Support for funders in the RIV data collection process

A significant proportion (22 %, from Table 5) of research outputs are funded from several streams by several funders. In the current RIV data collection process these are separate research output records that travel from the research organisation (or its organisational unit) to the central information system via the distinct funders. The funders are expected to do the following:

1. Verify the basic, objective information about the result output record (e.g. the bibliographic record for a publication). This includes the list of creators.
2. Verify that the research output was created within the research activities the submitting organisation references in the record (in the structured funding acknowledgements section).

This has the following consequences:

- There is no guarantee that the base records of the research output are made identical by the reporting institution. Since the records are in different submissions, there is no way to check that before they reach the central database.⁸⁵
- If a deficiency is spotted in the base record, the funder can either return the record to the reporting institution (asking for it to be corrected and re-submitted), or correct it by themselves and propagate further, or they can decide to drop the record from the submission. The research organisation need not be notified about the funder's action.

This leaves room for possible differences in the base records of the same research output reported in the RIV by the same research organisation. Some records may reach the RIV with delay, or not at all. There is only little the research organisations themselves can do to prevent the issue; they are not in a position to harmonize the policies and practices of different funders.

⁸⁵ Note that this situation is rare, as both the Vklap and any decent institutional CRIS handle this fan-out to funders by copying the same base record. Nevertheless there is a potential for confusion if the data collection is not coordinated strictly.

Table 5. Frequency of funders reporting outputs (research outputs of all types, years 2009-2013)

Funders per output	# outputs	% outputs
1	227,517	78.239 %
2	53,115	18.265 %
3	9,415	3.238 %
4	728	0.250 %
5	20	0.007 %
6	2	0.001 %
Total	290,797	100.000 %

We believe the IS VaVaI has a potential to make the RIV verifications by the funders both more effective and more efficient. This requires a slight modification of the data flow and a specialized tool to support the funders.

Here is a sketch of the modified process:

1. The research organisation makes a single RIV submission (not specific by funders) and sends this directly to the Office of the Government into the RIV.
2. The incoming records in the RIV are put in the initial “to-be-verified” state. They will not be displayed as a part of the official RIV yet.
3. The funders will be invited to check the records that are relevant to them. Additionally, other verifications can be carried out on the records as well. All these verification activities can be carried out concurrently.
4. If an issue is found in the base record, it is visible to all the funders who are relevant for the record, and to the reporting organisation.
5. It is the responsibility of the reporting organisation to resolve the issue by either submitting an updated record, or withdrawing the record (in extreme cases).
6. The acts of verification are recorded. Where the verifications use external information sources, the pointers to these sources are recorded too. This will contribute to improving the connectivity of the IS VaVaI data.
7. When a record passes all verifications, it is marked “verified” and allowed in the official RIV metadata set.

This requires a shared working area for the funders and the research organisations to support this workflow. The funders shall introduce their processes and train their staff to perform these verifications. Alternatively, they can contract such a service. The verifications should be complemented with a mechanism for appeal.

This modification does not require a change in the Act on public support of RD&I.

The following are the expected effects of this re-arrangement:

- A yet more thorough verification by funders resulting in a higher data quality

- Contribution to an improved connectivity where the verifications use external information sources, such as citation or patent databases
- A more uniform spread of the workload throughout the year. The research organisations would be able to set their own reporting frequency without generating excessive workload.⁸⁶
- Funders get the room to actually use their evaluation committees to assess the outputs of projects.
- Funders can have more up-to-date information about projects that have not yet reported any result.

5.2.6 Improved presentation of the IS VaVaI data and analytical outputs

The current web portal (see Section 2.2.4) that presents the IS VaVaI data is functioning and is being used by most, if not all players in the Czech RD&I domain. That said, the portal (launched 2007) is already showing its age and the options for improvement reach far beyond rolling out a new graphical design.

The search system seems not to be up-to-date with today's user expectations. It should be redesigned to give a more straightforward user experience. In the interviews users named the most problematic aspect of the current portal being the need to enter precise names of institutions whose outputs/activities the users want to see. Today users expect a simple query interface based on text search,⁸⁷ with an auto-complete functionality.⁸⁸

However, at the same time users want to search in clearly defined subsets of the information that is available (e.g. based on a specific year range, funder, output types, etc.). So integrating a generic web search engine, such as Google, would not satisfy all requirements. The right balance appears to be with the **faceted search technology**: it can combine discrete constraints applying to structured attributes with text search.

When implementing these changes, the **graphical design** should go through a refresh as well. Ideally, it should be visually integrated across the whole range of IS VaVaI tools (the Vklap, the Checking web service). An appropriate, visually appealing style would aid the reception of the system both in the Czech Republic and internationally.

There is a pronounced need for **analytical outputs**⁸⁹, especially among strategic managers and their support staff. The micro-data should be accompanied with the most commonly expected and used aggregations at the meso (organisations, funding programmes, subject fields) and the macro (funding providers, organisation types, disciplinary areas, regions) levels. When doing the aggregations, it is imperative to keep the link to the underlying micro-data.

The primary form used to represent the aggregations is with charts, possibly interactive, with an option to get the underlying values. Network visualizations can

⁸⁶ Frequencies below once per year would threaten an organisation's ability to report the outputs of their research.

⁸⁷ Full-text search in the traditional sense cannot be offered, as the information system only contains metadata, not the research outputs as such.

⁸⁸ See also [IPn-KA1.I.3]

⁸⁹ In line with [IPn-KA1.I.4]

help understanding collaboration patterns. The web interface should provide for customisable data outputs that would be amenable for subsequent off-line processing.

Analytical outputs from the IS VaVaI cannot be considered a one-time task. Rather, it's a whole development path that is going to be driven by user requirements.

5.2.7 Tracking research output lifecycles and impacts

Ideally, one would like to follow the exploitation of the results in RIV. This is partially done for patents where a category of financial income is recorded.

This could be extended to any recorded research output. Beside the IP outputs this information could be useful for the evaluation of the applicability and the actual benefits of applied research⁹⁰, where it is highly desirable to distinguish the outputs that were used (preferably by another party) from those never getting into the application sphere. It could be possible to track financial yields stemming from the sale of IP, leasing etc. For publication results in SSH it is desirable to follow the most important responses, such as notable reviews. Other items to follow may be prizes and awards.⁹¹

IS VaVaI could be extended by a program support for importing and monitoring this class of information. The selection of the monitored types of results should be set up on the basis of a consensus within the Council for Research, Development and Innovation. This support would also allow for subsequent addition of external identifiers (see Section 4.1) that become available only after a research output has been reported into the RIV.⁹²

While the IS VaVaI provides the technical opportunity duly to manage information related to research outputs and their subsequent exploitation, it should be considered that keeping track of this information - over time and for all research outputs - represents a high burden on the involved organisations. Also the quality of the information provided typically is very hard to assess. Finally, there are also more methodological issues that inhibit the use of micro-data for the assessment of research impacts. These include, e.g., the concept of project fallacy and the issue of attributing specific impacts to 1 single research output or public intervention. For these reasons, impacts of research activities are typically assessed at a higher level of aggregation.

5.2.8 Improve support for projects with postponed ex-post evaluation

Grant projects can report their results up to two years after the project ended. In many cases the principal investigators exercise the option to postpone the evaluation of the project by the ex-post evaluation committee until the time when the publications from the project actually appear (by one and half years maximum). That way, no uncertainty as to the publication status of the project results is left and the evaluation committee gets a much fuller picture of the project.

However, this does not play well with the regular yearly beat for reporting about projects. The funder (the Czech Science Foundation) is forced to introduce artificial

⁹⁰ See also [IPn-KA1.II.7]

⁹¹ In line with [IPn-KA1.II.12]

⁹² The current practice requires a re-submission of the research output record with the additional information added. The original record need not be erased.

states of “temporarily suspended” with zero funding, and the information about the real end of the project gets distorted.

A possible improvement of this situation (in compliance with the requirements of Act no. 130/2002 Coll.) would be to extend the list of possible states with the state “finished, waiting for evaluation” and collect the date by which the evaluation is expected to take place. The financial statement would still be due by the usual schedule (i.e., July 1st of the year following the last year of the project), only the evaluation verdict would be postponed.

This extension would greatly improve the quality of the data from the Czech Science Foundation (and possibly from other funders as well, if they start supporting grant projects in the future).

5.2.9 Collect more detailed information about the cost structure of projects

The CEP finance structure has traditionally been centred around the different funding sources (the state budget funding and the total eligible costs). Only recently (from 2014) a particular detail was added: the amount of investment covered by the state budget funding. This is driven by an information need of the current Metodika PRFS that makes use of this information in its Pillar III (evaluation of applied research).

It turns out that collecting many more details would be useful. The following cost types are typical on projects and should be considered:

- Staff costs,
- External service costs,
- Travel expenses,
- Cost of material,
- Investment costs,
- Upkeep and maintenance costs,
- Other running costs,
- Overhead costs.

The refined cost planning/reporting structure would allow for a comparison and benchmarking of funders.

5.3 Cost estimates

In this section we give estimates of the costs of the adaptations of the IS VaVaI to support the EM, as well as the other modifications that are proposed here. At this stage the estimates can only be preliminary and rough, since IT support is a service component in the whole EM implementation framework.

The evaluation of institutions at the scale of the Czech Republic is a big exercise in all aspects. The information management aspect will not be an exception.

Information technology has the demonstrated potential of considerably increasing efficiency and effectiveness of processes. Equally so, information technology has a demonstrated potential for failures, usually with large impact. The most commonly quoted reasons for IT project failures are: missing or insufficient support from the part of superior management, insufficient knowledge of the problem domain by the implementers, and unrealistic estimates of the effort that needs to be invested.

Our estimates are based on the experience both with the IS VaVaI and with the Small Pilot Evaluation. We give estimates for two out of the following three levels:

1. Essential: Implementing this level is necessary for the evaluation as it is foreseen.

2. **Recommended:** We recommend implementing extensions at this level as well, they will bring important benefits for the quality of the underlying data and/or the smoothness of the evaluation process.
3. **Optional:** We suggest the extensions at this level be also implemented, but not implementing them will not have major consequences for the evaluation.

The Recommended level includes the extensions and enhancements of the Essential level; likewise, the Optional level includes all of the Recommended level.

Other uses of the IS VaVaI may find they need the same or very similar extensions, but would assign different priorities. Also, the enhancements serve the evaluation as well as other uses of the IS VaVaI. We calculate the cost against the evaluation as a proportion of the total cost of each enhancement. Our estimates include all the associated project management, quality assurance, and other overhead at a standard rate of 500 €/man-day. Where extensive manual processing is required, the rate is lowered to 300 €/man-day.

The following three types of costs are calculated:

- Initial set-up costs,
- Costs per evaluation campaign,
- Running costs per year.

These types of costs are combined into an aggregated cost figure that includes one half of the initial set-up costs (counting on the set-up costs to amortize over the first two evaluation campaigns), the per-evaluation costs, and the running costs for the period between the evaluations (six years in our estimates).

In the sections below, we give reference under each section heading to the previous sections in the report where the task/activity/area is described.

5.3.1 Direct support of the EM implementation: Handling the evaluation process

See Section 5.1.3

This section lists the cost estimates of tasks/activities/areas that are directly connected with handling the EM process. The direct EM costs therefore make 100% of the total costs.

This part can be either a module of the IS VaVaI, or a separate system that takes data out from the IS VaVaI and eventually feeds back the evaluation outcome. The overheads connected with either approach are approximately equal, so we do not see a need to make a distinction in the cost estimates.

Panel structure set-up and membership management

This activity will require the customization and maintenance of a workflow system with user access control, including the reviewers and members of auxiliary and supportive bodies of the EM. At the “essential” level the start dates and end dates of each membership or other relationship are tracked. A basic set of database reports for the purposes of the help desk operations, of employment contract management and of website publication. The “recommended” level adds support for substitutability in specific roles, as well as a more comprehensive set of database reports for the various purposes. The added value of the “recommended” level lies in a smoother handling of situations of temporary unavailability, which are bound to occur frequently in an undertaking of the size of the national evaluation of institutions.

Estimated costs		The "essential" level	The "recommended" level
Initial set-up	Man-days	50	100
	Rate/Man-day	500 €	500 €
	Cost	25 000 €	50 000 €
Per evaluation campaign	Man-days	12	20
	Rate/Man-day	500 €	500 €
	Cost	6 000 €	10 000 €
Total per evaluation period		18 500 €	35 000 €
EM cost intensity factor		100%	100%
Direct EM costs		18 500 €	35 000 €

EvU and RU registration including eligibility checks

This is the first contact of EvUs with the evaluation infrastructure. User accounts need to be set up for both the users who enter information and the users who give confirmations. The capacities of the users need to be tracked and once again, the appropriate access control structures need to be set up. Whereas the “essential” variant basically allows a single user of each type (enter/confirm) per EvU/RU to be present in the system, in the “recommended” variant more users and more detailed structures (up to the EvU/RU to set up) will be supported by the system. In this way the evaluation infrastructure will support the internal processes on the EvU side, thus reducing the indirect costs of the evaluation. Specifically the RU registration step, which involves the distribution of the EvU’s researchers into the RUs, can be very communication-intensive. In particular, the eligibility checks will give early indication about the feasibility of the RU portfolio an EvU is setting up.

Estimated costs		The "essential" level	The "recommended" level
Initial set-up	Man-days	15	40
	Rate/Man-day	500 €	500 €
	Cost	7 500 €	20 000 €
Per evaluation campaign	Man-days	6	16
	Rate/Man-day	500 €	500 €
	Cost	3 000 €	8 000 €
Total per evaluation period		6 750 €	18 000 €
EM cost intensity factor		100%	100%
Direct EM costs		6 750 €	18 000 €

EvU and RU level self-assessment reports, selection of the outputs for peer review

The “essential” level supports filling in by a single user, whereas the “recommended” level allows multiple users per EvU and RU to work on the self-assessment report in a shared editing mode. The process of selection of the outputs for peer review will also be better supported in the “recommended” level, where a process of nomination by department heads and selection by a responsible would be supported inside the evaluation infrastructure. This would again lower the administrative burden on the

RUs, while leaving full control of the core of the selection agenda with the RU appointed representatives.

Estimated costs		The "essential" level	The "recommended" level
Initial set-up	Man-days	65	160
	Rate/Man-day	500 €	500 €
	Cost	32 500 €	80 000 €
Per evaluation campaign	Man-days	15	30
	Rate/Man-day	500 €	500 €
	Cost	7 500 €	15 000 €
Total per evaluation period		23 750 €	55 000 €
EM cost intensity factor		100%	100%
Direct EM costs		23 750 €	55 000 €

Peer review and panel decision support

This supports the routing of tasks as well as of the associated forms and/or documents. While the “essential” level offers a single workflow definition and a single set of form templates for all panels, the “optional” level gives per-panel customizability of the workflows and form templates. This gives the panels more adaptability to the specificities of their fields.

Estimated costs		The "essential" level	The "optional" level
Initial set-up	Man-days	90	200
	Rate/Man-day	500 €	500 €
	Cost	45 000 €	100 000 €
Per evaluation campaign	Man-days	40	100
	Rate/Man-day	500 €	500 €
	Cost	20 000 €	50 000 €
Total per evaluation period		42 500 €	100 000 €
EM cost intensity factor		100%	100%
Direct EM costs		42 500 €	100 000 €

Recording and disseminating the results of the evaluation

This supports the information publishing and preservation aspect of the EM. The “essential” level offers a single model of information recording and dissemination. In the “optional” one, panels can define what needs to be preserved and what shall be published. Again, this is the mechanism to account for the specificities of certain fields. This step also involves the feedback by the EvUs on the panel reports.

Estimated costs		The "essential" level	The "optional" level
Initial set-up	Man-days	10	15
	Rate/Man-day	500 €	500 €
	Cost	5 000 €	7 500 €
Per evaluation campaign	Man-days	40	100
	Rate/Man-day	500 €	500 €
	Cost	20 000 €	50 000 €
Total per evaluation period		22 500 €	53 750 €
EM cost intensity factor		100%	100%
Direct EM costs		22 500 €	53 750 €

Infrastructure for the direct EM implementation support

The necessary IT infrastructure stack, such as the server hardware, operating systems, databases and application servers, including backups and monitoring, to run the EM support system outlined in Section 5.1.3. In the "essential" variant a dedicated PC-class server is envisioned for handling the additional data that is specific to the evaluation. In the "recommended" variant one proposes enterprise-grade hardware for the task.

Estimated costs	The "essential" level	The "recommended" level
Per evaluation campaign	45 000 €	80 000 €
Running costs / year	2 000 €	4 000 €
Total per evaluation period	57 000 €	104 000 €
EM cost intensity factor	100%	100%
Direct EM costs	57 000 €	104 000 €

Totals for the support to the EM implementation

The tasks of the EM implementation support that were outlined in this sub-section 5.3.1 have the following totals:

Estimated costs for Direct support of the evaluation process	The "essential" level	The "recommended" level	The "optional" level
Initial set-up	115 000 €	200 000 €	247 500 €
Per evaluation campaign	101 500 €	153 000 €	213 000 €
Running costs / year	2 000 €	4 000 €	4 000 €
Total per evaluation period	171 000 €	277 000 €	360 750 €
Direct EM costs	171 000 €	277 000 €	360 750 €

5.3.2 Enriched IS VaVaI: EM compatibility, new components, data quality enhancements

This sub-section concentrates the enhancements of the IS VaVaI to support its connectivity and data quality, as well as the two new components: the Researchers Registry and the Research Infrastructure Catalogue.

Providing data for the self-assessment reports and into the direct evaluation support

See Section 5.1.2

The IS VaVaI data needs to be exported and transformed into the specific form needed for the evaluation support infrastructure. This involves the full database of objects (research organisations, their units, research projects, research outputs, researchers) that are relevant for the evaluation.

Estimated costs		The "essential" level
Per evaluation campaign	Man-days	25
	Rate/Man-day	500 €
	Cost	12 500 €
Total per evaluation period		12 500 €
EM cost intensity factor		100%
Direct EM costs		12 500 €

Generic support for interconnections using pointers and identifiers

See Section 4.2.1

This task results in a generic structure inside the IS VaVaI that allows any type of pointer or identifier to be attached to any object in the system. In the “recommended” level, a corresponding structure would be added to drive the integrity checks for consistent presence of such pointers and identifiers with objects of the desired type and possibly other configurable criteria.

Estimated costs		The "essential" level	The "recommended" level
Initial set-up	Man-days	10	20
	Rate/Man-day	500 €	500 €
	Cost	5 000 €	10 000 €
Running costs / year	Cost	1 000 €	2 000 €
Total per evaluation period		8 500 €	17 000 €
EM cost intensity factor		50%	50%
Direct EM costs		4 250 €	8 500 €

The Web of Science data connector for the IS VaVaI

See Section 4.2.1

The connector shall use the web service API of the Web of Science database. At the “essential” level just the basic bibliographic data is pulled into the IS VaVaI; at the “recommended” level a more complete set of data is integrated into the IS VaVaI, which would allow for more comprehensive validations to take place.

Estimated costs		The "essential" level	The "recommended" level
Initial set-up	Man-days	12	20
	Rate/Man-day	500 €	500 €
	Cost	6 000 €	10 000 €
Running costs / year		1 200 €	2 000 €
Total per evaluation period		10 200 €	17 000 €
EM cost intensity factor		70%	70%
Direct EM costs		7 140 €	11 900 €

The Scopus data connector for the IS VaVaI, addition of the EID identifier to the RIV data structure

See Section 4.2.1

The connector shall use the web service API of the Scopus database. At the “essential” level just the basic bibliographic data is pulled into the IS VaVaI; at the “recommended” level a more complete set of data is integrated into the IS VaVaI, which would allow for more comprehensive validations to take place. In either case, the EID identifier needs to be added as a data item to the RIV data structure.

Estimated costs		The "essential" level	The "recommended" level
Initial set-up	Man-days	16	24
	Rate/Man-day	500 €	500 €
	Cost	8 000 €	12 000 €
Running costs / year	Cost	1 600 €	2 400 €
Total per evaluation period		13 600 €	20 400 €
EM cost intensity factor		70%	70%
Direct EM costs		9 520 €	14 280 €

Extended validation of the incoming RIV data

See Section 3

Here the two levels represent different levels of thoroughness of the validations. The initial set-up phase involves defining the method and the workflows.

Estimated costs		The "essential" level	The "recommended" level
Initial set-up	Man-days	8	10
	Rate/Man-day	500 €	500 €
	Cost	4 000 €	5 000 €
Running costs / year	Man-days	130	250
	Rate/Man-day	300 €	300 €
	Cost	39 000 €	75 000 €
Total per evaluation period		236 000 €	452 500 €

Estimated costs	The "essential" level	The "recommended" level
EM cost intensity factor	70%	70%
Direct EM costs	165 200 €	316 750 €

The Researcher Registry

See Section 5.1.2

At the “essential” level this is a basic registry with data collection from research organisations. The “recommended” level adds an extensive matching to link each researcher with his/her previous occurrences in the IS VaVaI, and also the research organisations’ internal structure will be mapped closely.

Estimated costs		The "essential" level	The "recommended" level
Initial set-up	Man-days	50	150
	Rate/Man-day	500 €	500 €
	Cost	25 000 €	75 000 €
Running costs / year	Cost	5 000 €	15 000 €
Total per evaluation period		42 500 €	127 500 €
EM cost intensity factor		70%	70%
Direct EM costs		29 750 €	89 250 €

The Research Infrastructure Catalogue

See Section 5.1.2

The “essential” level represents a basic catalogue of research facilities: the large research infrastructures, laboratories, e-Science and computer network facilities, academic libraries and other components of research infrastructure of the Czech Republic. The “recommended” level adds a catalogue of notable equipment and research services.

Estimated costs		The "essential" level	The "recommended" level
Initial set-up	Man-days	80	200
	Rate/Man-day	500 €	500 €
	Cost	40 000 €	100 000 €
Running costs / year		8 000 €	20 000 €
Total per evaluation period		68 000 €	170 000 €
EM cost intensity factor		60%	60%
Direct EM costs		40 800 €	102 000 €

Closer integration with the Industrial Property Office Information

See Section 4.2.2

This task builds support for obtaining the list of expired documents and integrating that into the RIV. At the “recommended” level one envisions a display of the information on the IS VaVaI web portal.

Estimated costs		The "essential" level	The "recommended" level
Initial set-up	Man-days	5	8
	Rate/Man-day	500 €	500 €
	Cost	2 500 €	4 000 €
Running costs / year		500 €	800 €
Total per evaluation period		4 250 €	6 800 €
EM cost intensity factor		40%	40%
Direct EM costs		1 700 €	2 720 €

Integration with additional sources of data on IP outputs

See Section 4.2.2

The “recommended” level represents identifying and integrating one such source. At the “optional” level, integration of two to three source of data on IP outputs is foreseen.

Estimated costs		The "recommended" level	The "optional" level
Initial set-up	Man-days	5	15
	Rate/Man-day	500 €	500 €
	Cost	2 500 €	7 500 €
Running costs / year		500 €	1 500 €
Total per evaluation period		4 250 €	12 750 €
EM cost intensity factor		40%	40%
Direct EM costs		1 700 €	5 100 €

Additional demand for computing resources due to the extensions

The aforementioned extensions will cause an increased demand for computing resources (processing power, storage). The two levels represent a range of costs of provision for this demand.

Estimated costs		The "essential" level	The "recommended" level
Running costs / year		10 000 €	15 000 €
Total per evaluation period		60 000 €	90 000 €
EM cost intensity factor		60%	60%
Direct EM costs		36 000 €	54 000 €

Totals for the Enriched IS VaVaI

Estimated costs for Enriched IS VaVaI	The “essential” level	The “recommended” level	The “optional” level
Initial set-up	90 500 €	218 500 €	223 500 €
Per evaluation campaign	12 500 €	12 500 €	12 500 €
Running costs / year	66 300 €	132 700 €	133 700 €
Total per evaluation period	455 550 €	917 950 €	926 450 €
Direct EM costs	306 860 €	613 600 €	617 000 €

5.3.3 Improved accessibility of the IS VaVaI data

This sub-section summarizes the tasks from Sections 4.2.6 and Section 5.2.6 of this report.

CERIF compatibility layer

See Section 4.2.6

This area includes the conversions of the IS VaVaI data into the CERIF format. The difference of between the “recommended” and the “optional” levels consists in different levels of detail in the representation of past data of the IS VaVaI.

Estimated costs		The “recommended” level	The “optional” level
Initial set-up	Man-days	70	90
	Rate/Man-day	500 €	500 €
	Cost	35 000 €	45 000 €
Running costs / year		7 000 €	9 000 €
Total per evaluation period		59 500 €	76 500 €
EM cost intensity factor		0%	0%
Direct EM costs		0 €	0 €

Simple presentation API

See Section 4.2.6

This area of activities consists of a simple API (at the “recommended” level) that gives a dynamic listing of research outputs and research projects of a researcher. At the “optional” level this is extended with support of groups of researchers with time constraints and constraints by organisations.

Estimated costs		The “recommended” level	The “optional” level
Initial set-up	Man-days	10	20
	Rate/Man-day	500 €	500 €
	Cost	5 000 €	10 000 €
Running costs / year		1 000 €	2 000 €
Total per evaluation period		8 500 €	17 000 €
EM cost intensity factor		0%	0%
Direct EM costs		0 €	0 €

Linked Open Data (LOD) presentation of the IS VaVaI data

See Section 4.2.6

The Linked Open Data presentation of the data needs a careful and detailed definition of semantics of the data that is to be published, since this cannot be modified later. At the “recommended” level a basic support using just the CERIF vocabularies is included. At the “optional” level one envisions the inclusion of other LOD vocabularies, too. This makes the data more re-usable also outside of the RD&I information domain.

Estimated costs		The "recommended" level	The "optional" level
Initial set-up	Man-days	40	70
	Rate/Man-day	500 €	500 €
	Cost	20 000 €	35 000 €
Running costs / year		4 000 €	7 000 €
Total per evaluation period		34 000 €	59 500 €
EM cost intensity factor		0%	0%
Direct EM costs		0 €	0 €

Improved presentation of the IS VaVaI data and analytical outputs

See Section 5.2.6

At the “essential” level this enhancement represents developing a new, modern application for presentation of the IS VaVaI data on the WWW, which would use a combination of full-text and faceted search. The “recommended” level allows for displaying of various summary indicators of the working selection of objects, such as funding amounts (total vs. per year) or counts of research outputs, with appropriate visualizations.

Estimated costs		The "essential" level	The "recommended" level
Initial set-up	Man-days	120	200
	Rate/Man-day	500 €	500 €
	Cost	60 000 €	100 000 €
Running costs / year		12 000 €	20 000 €
Total per evaluation period		102 000 €	170 000 €
EM cost intensity factor		20%	20%
Direct EM costs		20 400 €	34 000 €

Additional infrastructure for the improved accessibility of the IS VaVaI data

The additional consumers of the IS VaVaI data as well as the availability of analytical reports will result in an increasing demand for computing resources to service the requests. We estimate two levels of such additional infrastructure demands in the two variants.

Estimated costs	The "essential" level	The "recommended" level
Running costs / year	10 000 €	15 000 €
Total per evaluation period	60 000 €	90 000 €
EM cost intensity factor	50%	50%
Direct EM costs	30 000 €	45 000 €

Totals for improved accessibility of the IS VaVaI data

Estimated costs for Improved accessibility of the IS VaVaI data	The "essential" level	The "recommended" level	The "optional" level
Initial set-up	60 000 €	160 000 €	190 000 €
Running costs / year	22 000 €	25 000 €	53 000 €
Total per evaluation period	162 000 €	200 000 €	413 000 €
Direct EM costs	50 400 €	79 000 €	79 000 €

5.3.4 the enhancements of the IS VaVaI

This sub-section lists the estimates for all the other enhancements of the IS VaVaI that were proposed in this report and not listed in the previous sub-sections.

Subject classification refinement and extension – the modification

See Section 5.2.1

This task is essential to the EM.

Estimated costs		The "essential" level
Initial set-up	Man-days	8
	Rate/Man-day	500 €
	Cost	4 000 €
Total per evaluation period		2 000 €
EM cost intensity factor		60%
Direct EM costs		1 200 €

Subject classification refinement and extension – re-classification of existing data

See Section 5.2.1

This task represents different levels of depth in re-classifying the existing data in connection with the refinement of the subject classification. At both levels a combination of semi-automatic and manual processing is used.

Estimated costs		The "essential" level	The "recommended" level
Initial set-up	Man-days	20	120
	Rate/Man-day	300 €	300 €
	Cost	6 000 €	36 000 €
Total per evaluation period		3 000 €	18 000 €
EM cost intensity factor		60%	60%
Direct EM costs		1 800 €	10 800 €

Improve RIV compatibility with bibliographic conventions

See Section 5.2.2

The “essential” level represents a basic support for the improvements. The “recommended” level adds an support on the side of the IS VaVaI portal to present the RIV research outputs in a format that more closely resembles bibliographic citations (using a single style).

Estimated costs		The "essential" level	The "recommended" level
Initial set-up	Man-days	24	36
	Rate/Man-day	500 €	500 €
	Cost	12 000 €	18 000 €
Total per evaluation period		6 000 €	9 000 €
EM cost intensity factor		50%	50%
Direct EM costs		3 000 €	4 500 €

Revision of the RIV output types classification

See Section 5.2.4

At the “essential” variant a basic level of support is estimated. The “recommended” level includes more metadata about both the research software and research datasets.

Estimated costs		The "essential" level	The "recommended" level
Initial set-up	Man-days	30	60
	Rate/Man-day	500 €	500 €
	Cost	15 000 €	30 000 €
Total per evaluation period		7 500 €	15 000 €
EM cost intensity factor		50%	50%
Direct EM costs		3 750 €	7 500 €

RIV master list of publishing channels – generic support for the master list

See Section 5.2.3

The “essential” level counts with journals and book publishers; the “recommended” variant adds support for book series.

Estimated costs		The "essential" level	The "recommended" level
Initial set-up	Man-days	10	15
	Rate/Man-day	500 €	500 €
	Cost	5 000 €	7 500 €
Total per evaluation period		2 500 €	3 750 €
EM cost intensity factor		50%	50%
Direct EM costs		1 250 €	1 875 €

RIV master list of publishing channels – list creation and maintenance

See Section 5.2.3

The “essential” level counts with journals and book publishers; the “recommended” level adds support for book series.

Estimated costs		The "essential" level	The "recommended" level
Initial set-up	Man-days	70	120
	Rate/Man-day	300 €	300 €
	Cost	21 000 €	36 000 €
Running costs / year	Man-days	30	50
	Rate/Man-day	300 €	300 €
	Cost	9 000 €	15 000 €
Total per evaluation period		64 500 €	108 000 €
EM cost intensity factor		50%	50%
Direct EM costs		32 250 €	54 000 €

More comprehensive support for funders

See Section 5.2.5

The “essential” level features the basic support and workflow for the funders. At the “recommended” level, additional support for coordination of several verifying users per funder and the management of the verification process are introduced.

Estimated costs		The "essential" level	The "recommended" level
Initial set-up	Man-days	60	100
	Rate/Man-day	500 €	500 €
	Cost	30 000 €	50 000 €
Running costs / year		6 000 €	10 000 €
Total per evaluation period		51 000 €	85 000 €
EM cost intensity factor		50%	50%
Direct EM costs		25 500 €	42 500 €

Tracking research output lifecycles

At the “recommended” level, unstructured lifecycle events are supported. At the “optional” level, a taxonomy of admissible lifecycle events will be specified for each output type.

Estimated costs		The "recommended" level	The "optional" level
Initial set-up	Man-days	20	40
	Rate/Man-day	500 €	500 €
	Cost	10 000 €	20 000 €
Running costs / year		2 000 €	4 000 €
Total per evaluation period		17 000 €	34 000 €
EM cost intensity factor		50%	50%
Direct EM costs		8 500 €	17 000 €

Improved support for projects with postponed ex-post evaluation

The “essential” level contains the basic support for the extension. The “recommended” level adds support for identifying the situation in existing CEP data.

Estimated costs		The "essential" variant	The "recommended" variant
Initial set-up	Man-days	4	6
	Rate/Man-day	500 €	500 €
	Cost	2 000 €	3 000 €
Total per evaluation period		1 000 €	1 500 €
EM cost intensity factor		10%	10%
Direct EM costs		100 €	150 €

More detailed information about the cost structure of projects

The “recommended” level starts with a study of current practice of project funders in the CR and an effort to devise a harmonised taxonomy of cost categories. The “essential” level would not include such a study.

Estimated costs		The "essential" level	The "recommended" level
Initial set-up	Man-days	10	20
	Rate/Man-day	500 €	500 €
	Cost	5 000 €	10 000 €
Total per evaluation period		2 500 €	5 000 €
EM cost intensity factor		50%	50%
Direct EM costs		1 250 €	2 500 €

Totals for Other enhancements of the IS VaVaI

Estimated costs for the Other enhancements of the IS VaVaI	The "essential" level	The "recommended" level	The "optional" level
Initial set-up	100 000 €	204 500 €	214 500 €
Running costs / year	15 000 €	27 000 €	29 000 €
Total per evaluation period	140 000 €	264 250 €	281 250 €
Direct EM costs	70 100 €	133 525 €	142 025 €

5.3.5 Licences for citation databases

See Section 4.2.1

We consider it essential to interlink the RIV data with the raw data of at least one of the two leading citation databases (Web of Science by Thomson Reuters and Scopus by Elsevier). We actually recommend the use of the data of both of these databases, unless it turns out to be more expensive in which case we'd suggest the second database as optional only.

We do not state a preference for either the Web of Science or the Scopus database. It is up to the IS VaVaI responsible body to make this decision in cooperation with the evaluation implementers. It is important to maintain the ability to use both of these databases, which gives more flexibility.

We calculate with the prices of these licences in the range between 25 000 € and 100 000 € per year. Both citation database providers set the actual prices depending on the intended use and visibility of the data in a process of negotiation with the customer.

Estimated costs for the Licenses for citation databases	The "essential" level	The "recommended" level	The "optional" level
Running costs / year	100 000 €	125 000 €	200 000 €
Total per evaluation period	600 000 €	750 000 €	1 200 000 €
Direct EM costs	420 100 €	525 000 €	840 000 €

5.3.6 Estimates of the total costs of the adaptations and enhancements of the IS VaVaI outlined in this report

The cost estimates from the above sections have the following grand totals:

Estimated cost grand totals	The "essential" level	The "recommended" level	The "optional" level
Initial set-up	365 500 €	783 000 €	875 500 €
Per evaluation campaign	114 000 €	165 500 €	225 500 €
Running costs / year	205 300 €	335 700 €	419 700 €
Total per evaluation period	1 528 550 €	2 571 200 €	3 181 450 €
Direct EM costs	1 018 360 €	1 628 125 €	2 038 075 €

In CZK (using the exchange rate EUR/CZK = 27.500) this amounts to:

Estimated cost grand totals	The "essential" level	The "recommended" level	The "optional" level
Initial set-up	10 051 250 CZK	21 532 500 CZK	24 076 250 CZK
Per evaluation campaign	3 135 000 CZK	4 551 250 CZK	6 201 250 CZK
Running costs / year	5 645 750 CZK	9 231 750 CZK	11 541 750 CZK
Total per evaluation period	42 035 125 CZK	70 708 000 CZK	87 489 875 CZK
Direct EM costs	28 004 900 CZK	44 773 438 CZK	56 066 313 CZK

As described in the introduction to this section on the cost estimates, the areas covered include not only the direct support of the EM, but also various other adaptations and enhancements that will positively impact other uses of the IS VaVaI as well.

The “essential” level includes a basic, yet complete support of the tasks at hand. Specifically, one of the citation databases is integrated with the IS VaVaI data. The “recommended” level offers more features, flexibility and user comfort to all actors in the evaluation process, which translates into a larger efficiency, therefore lower direct and indirect costs of the evaluation. At the “optional” level a full support for the evaluation process, as well as all of the other proposed enhancements of the IS VaVaI, are covered.

The study team suggests to implement at least the “recommended” level of features. However, it is up to the EM implementers to select their own combination of variants of the extensions. Specifically, the handling of the evaluation process can be serviced either by a module of the IS VaVaI, or by a stand-alone application.

However, these estimates should be primarily regarded as indicative. The real costs of the IS VaVaI support of the evaluation will first and foremost depend on a detailed specification and analysis of the real EM implementation, and secondly, on the outcome of the negotiations with the citation database providers.

6. Conclusions

The Czech Research, Development and Innovation Information System (IS VaVaI) is the key element of research information infrastructure in the Czech Republic. It is an operating information system; all stakeholders have adapted to it. The system has been used to monitor and partially also to support the decisions on the allocation of public funding for research, development and innovation in the Czech Republic. It is an official administrative hub of information about R&D in the country. The administrator of the system is the supreme state administrative body for the domain - the Council for RD&I at the Office of the Government of the Czech Republic. The information system is established and operated on the grounds of Act No. 130/2002 Coll., on the public support of research, development, and innovation. IS VaVaI is an established, versatile instrument that assists all the players on the Czech RD&I scene. It maintains information that is necessary for qualified analyses at all levels. As a national research information hub it supports both the professional and the broader audiences and creates an environment of transparency.

The information in the IS VaVaI is authoritative, continuously updated, and publicly displayed. The system thus contains transparent and verifiable micro-data: data about individual projects, publications, other research outputs, etc. When aggregated to a meso- or macro-level, the information is much more reliable than ad-hoc surveys. The data can (and do) serve many purposes. It is one of the most comprehensive information systems of its kind in the world.

The main and core recommendation is to use the IS VaVaI to the maximum possible extent to support the EM. The suggested extensions of the IS VaVaI will allow to effectively and efficiently carry out the evaluation of the Czech Research Organisations, as well as to support re-use of the collected data for other purposes: evaluations of RD&I programmes, evaluations of applied research and innovations, and strategic analyses at the institutional, funders' and the national levels.

The synergy of the primary collected data with the external ones will extend the usability of data for all the above-mentioned purposes. The Czech national CRIS is an advanced one and has been running for a considerable time. All stakeholders are adapted and trained to using it in a routine way. All these factors will facilitate the extension and the use of the information system in the EM.

This report identifies the necessary extensions of the IS VaVaI. However, a real implementation of the EM with the inclusion of stakeholders (the Council for RD&I, the Ministry of Education, Youth and Sports, the Section for Science, Research and Innovation at the Office of the Government of the Czech Republic) is bound to bring additional or modified requirements. The implementation will need to start with a technical analysis working with a finalized set of requirements. This is natural: it reflects the supportive role of information technologies. It appears that the form of an implementation project would be the most appropriate. The cost estimates reflect the current, in some cases rather general, conceptual level of knowledge of the requirements. They shall not be used outside of this context.

Making use of the IS VaVaI with the extensions allows to lower the indirect cost estimate of the evaluation exercise, compared to similar evaluation exercises that are carried out without support of such an information system. Information re-use will save time of the researchers and of the EvU administrative staff.

Appendix A - Interview questions with answers

Question	The most frequent or important observations	Positive	Negative	Assuredly positive	Rather positive	Rather negative	Entirely negative
Components of the R&D Information System							
Do you use in your institution the R&D IS component CEA (Central registry of RD&I activities)				8	3	2	4
If yes, for what purpose?	Strategic decision/leadership, check how much organization spends from budget, preparation of financial overview and analysis, preparation of state budget, basis for R&D&I Council, basis for org. management, used by org. library, used by grant department, tracking of new calls/tenders in R&D&I, benchmarking against similar organization, search of potential partners, tracking of (project) budget, tracking of all kinds (CEA) of RD&I budget of the ministry						
If yes, are you satisfied with the web presentation of the information?				4	7	0	0
Do you use in your institution the R&D IS component VES (Database of calls for proposals)?				5	5	1	6
If yes, for what purpose?	Keeping track of management, control of public procurement, materials for RD&I Council, the state budget (contingent liabilities), success, evidence of announced and evaluated competitions, entering contests, search competitions, verification contests, processing financial reports, orientation in a particular financing, analytical studies						
If yes, are you satisfied with the web presentation of the information?				4	6	0	1
Do you use in your institution the R&D IS component CEP (Central catalogue of projects)?				17	0	0	0
If yes, for what purpose?	To support new applications for grants, materials for RVVI, state budget, analytical outputs, cooperating institutions, an overview of grants and other institutions, to maintain the image GACR, verification formalities projects, verifying the results of projects search projects, institutions, financing, analytical studies, cross-checks subsidies, reports success and external financing, check numbers of projects, import data into ASEP, VVVS, GaP, OBD, import providers, accounting and control a our project						
If yes, are you satisfied with the web presentation of the information?				9	7	0	0
Registry of R&D outputs (RIV)							
The purpose of the collection of information on R&D outputs							
Statement of results in RIV				16	0	0	0
Internal evaluation in the institution and or its units				12	1	2	0

The RD&I Information System as an information tool for evaluation

Question	The most frequent or important observations	Positive	Negative	Assuredly positive	Rather positive	Rather negative	Entirely negative
For the strategic decision-making				9	1	2	3
For the dissemination of the information in your institution				6	2	2	4
Information service for the professional sphere				8	6	1	1
Information service for the broader public				6	4	1	4
Support for the grant applications				7	2	0	3
Other (please specify):	Overall picture, effectivity of financing, keeping pace with competition, break-down picture, evaluation of subunits, departments, scientists; academic promotions to the ranks of associate or full professor; extensions of contracts, evaluation of projects						
Quality and reliability of the records in RIV (Register of R&D outputs)							
Are you aware of R&D outputs from your institution which are missing from RIV (and are not classified information)				1	7	2	6
If yes, the reason is:							
The absence of a proper type of result in RIV structure				1	7	2	3
The result was created within a R&D activity not financed from the state budget				2	3	5	3
The publication of the R&D output is not in the interest of your institution				1	2	8	4
Insufficient work capacity for the input of data				1	1	8	5
The collection of the information on RD&I results							
What sw tool is used in your institution for the collection of R&D data?	IS VaVaI (Vklap), ASEP, OBD, VVVS						
How is the collection of data on R&D results organized in your institution?	It follows the organisational hierarchy (university – faculty – department – researcher); with help of specialists - typically library, (department of) a dedicated vice-dean etc.						
To what extent a specialized unit (library) is involved in the collection of data?				11	0	1	1
Does your institution perceive the collection of information on R&D outputs as an important activity?				13	1	0	0
Does your institution keep a register of R&D outputs which are not included in RIV				1	9	0	2
Difficulties in the collection of data for the IS VaVaI				1	3	4	8
If yes, the cause is:							
Technical problems				2	1	7	4
Unclear methodology				1	5	4	5
Unclear responsibilities within the institution				0	1	12	2

The RD&I Information System as an information tool for evaluation

Question	The most frequent or important observations	Positive	Negative	Assuredly positive	Rather positive	Rather negative	Entirely negative
Unwillingness of researchers to report outputs				0	2	8	4
Other (please specify):	Deadlines, changes that are announced too late, contract research, not keeping internal deadlines, problems with deadlines for panels, disinterest of commercial companies to reported errors, problems with the local information system of the organization						
Is a check of collected data on R&D results carried out in your institution?		15	1				
If yes, the following aspects are checked:							
Verification of the existence of the result		12	2				
Verification of the exactness of the reported information		13	2				
Verification of the assigned R&D type of the result		12	3				
Verification of the research field classification of the result		6	8				
Verification of the reported authorship		9	6				
Verification of the reported relationship with research activities		12	3				
Do you think the currently implemented mechanisms of the data verification in RIV are sufficient?				3	8	1	1
Formal machine checks		14	2				
Formal checks carried out by dedicated personnel		11	4				
Semantic checks carried out by professional panels		7	4				
Have you found instances of false positive/negative identification of erroneous records?		14	2				
If yes, how many cases per year?	Few examples, 1-2, 0-10, tens						
Are you able to estimate the cost associated with the supplying of the R&D outputs into RIV?		6	6				
	Exact numbers are not available, only few cases with direct estimation of 3-4 personweeks, 1-2 personmonth, 6 personmonths (big univ.), the workload with collections of outputs was also reported that it would be present independently of RIV (due to independent interest of the institution's management to collect it)						
Data forwarding and the communication with the superior body							
How do you transfer data to the body superior to your institution?	Official databox (<i>datové schránky</i>), mail, Vklap, OBD, through library, R&D department, through the rectorate						
If you use the VKLAP application, how are you satisfied with it?				4	6	2	1
What improvements would you suggest?	Improve handling of submission delivery notes (<i>průvodka</i>)						
Do you use the IS VaVal help desk?				13	1	1	2

The RD&I Information System as an information tool for evaluation

Question	The most frequent or important observations	Positive	Negative	Assuredly positive	Rather positive	Rather negative	Entirely negative
If yes, has the help-line assisted you in the solution of problems?				11	3	0	0
In the case the helpline did not solved the problem what was your subsequent course of actions?	in a such few cases they provided contact to relevant subject or person (funder, council)						
What aspect of the protocol of the data reporting do you consider the most difficult?	the process is set logically and clearly, but problems are with realistic estimation of the time needed, deadlines, summer vacations, etc.						
Synergy with other information systems							
Do you use in your institution in the process of data collection some import for external sources?				5	4	1	1
WoS				8	5	0	1
Scopus				5	4	1	2
Proprietary information system				5	3	0	0
Other information system:	INIS, Industrial Property Office, PubMed, the National Library Catalogue, the DOI metadata						
Presentation and the use of the data from the R&D IS							
Do you consider useful to have the records of R&D outputs of your institution (including subordinate units) visible in a centralized information system in a unified structure?				15	1	0	1
Do you consider useful to see the records of R&D outputs of other institution in a centralized information system in a unified structure?				14	2	0	0
Do you consider useful that the records of R&D outputs created in the CR are deposited in a central register?				14	3	0	0
Do you find useful to be able to export the data for a subsequent processing and analysis?				13	3	0	1
How important is for you the possibility of carrying out searches within RIV?				13	4	0	0
If you use the searching in the RIV, how satisfied are you with the current search options?				6	10	0	1
What improvements of the search options and parameters would you recommend?	Speed, full-text search, search by vedidk, managerial extension						
How useful do you find the public IS monitoring the distribution, use and generally the flow of public money spent on the R&D				16	0	0	0
The interviewee							
How frequently do you use the R&D IS	Very often 13x, often 4x, rarely 0x						
How skilled in the use of the R&D IS do you consider yourself:	Very experienced 14x, experienced 3x, unexperienced 0x, very unexperienced 0x						
Practical aspects of the use of the information on R&D outputs							

The RD&I Information System as an information tool for evaluation

Question	The most frequent or important observations	Positive	Negative	Assuredly positive	Rather positive	Rather negative	Entirely negative
Do you follow in your institution the exploitation of the results?		6	3				
In the academic sphere		6	6				
In applied research		7	5				
Practical exploitation of results		7	4				
Financial gains for your institution derived from the results		8	3				
Societal impacts of R&D outputs		5	7				

Appendix B - IS VaVaI fields mapped to the OECD FOS field categorisation

The mapping was used for the Small Pilot Evaluation. It has issues that are discussed and the remedy is outlined in Section 5.2.1.

OECD FOS field	IS VaVaI fields
1.01 Mathematics	BA – General mathematics
	BB – Applied statistics, operational research
1.02 Computer and information sciences	BC – Control theory, systems theory
	BD – Information theory
	IN – Informatics, computer science
1.03 Physical sciences	BE – Theoretical physics
	BF – Elementary particles and high energy physics
	BG – Nuclear, atomic and molecular physics, accelerators
	BH – Optics, masers and lasers
	BI – Acoustics and oscillation
	BJ – Thermodynamics
	BK – Fluid mechanics
	BL – Plasma physics and discharge through gases
	BM – Solid-state physics and magnetism
	BN – Astronomy and celestial mechanics, astrophysics
1.04 Chemical sciences	CA – Inorganic chemistry
	CB – Analytical chemistry, separation
	CC – Organic chemistry
	CD – Macromolecular chemistry
	CF – Physical chemistry and theoretical chemistry
	CG – Electrochemistry
	CH – Nuclear and quantum chemistry, photochemistry
1.05 Earth and related environmental sciences	DA – Hydrology and limnology
	DB – Geology and mineralogy
	DC – Seismology, volcanology and Earth structure
	DD – Geochemistry
	DE – Earth magnetism, geodesy, geography
	DG – Atmospheric sciences, meteorology
	DI – Air pollution and air quality control
	DJ – Water pollution and water quality control
	DK – Soil contamination and decontamination including

	pesticides
	DL – Nuclear waste, radioactive pollution and its control
	DM – Solid waste and its control, recycling
	DO – Land areas protection and conservation
1.06 Biological sciences	BO – Biophysics
	CE – Biochemistry
	EA – Morphologic disciplines and cytology
	EB – Genetics and molecular biology
	EE – Microbiology, virology
	EF – Botany
	EG – Zoology
	EH – Ecology - communities
2.01 Civil engineering	JM – Structural engineering
	JN – Civil engineering
	JO – Land transport systems and facilities
2.02 Electrical engineering, electronic engineering, information engineering	JA – Electronic engineering and optoelectronics, electrical engineering
	JB – Sensors, detecting elements, measurement and regulation
	JC – Computer hardware and software
	JD – Use of computers, robotics and its application
	JW – Navigation, connection, detection and countermeasure
2.03 Mechanical engineering	JF – Nuclear power engineering
	JQ – Machinery and tools
	JR – Other machinery industry
	JU – Aeronautics, aerodynamics, aeroplanes
	JV – Space technology
2.04 Chemical engineering	CI – Industrial chemistry and chemical engineering
2.05 Materials engineering	JG – Metallurgy, metal materials
	JH – Ceramics, fire-proof materials and glass
	JI – Composite materials
	JJ – Other materials
	JK – Corrosion and material surface treatment
	JL – Material fatigue and fracture mechanics
	JP – Industrial processes and processing
	JS – Reliability and quality management, industrial testing
2.06 Medical engineering	FS – Medical facilities, devices and equipment
2.07 Environmental engineering	DH – Mining industry including coal mining and processing
	JE – Non-nuclear power engineering, energy consumption

	and utilization
2.09 Industrial Biotechnology	EI – Biotechnology and bionics
2.11 Other engineering and technologies	GM – Food processing
	JT – Propulsion, engines and fuels
	JY – Firearms, ammunition, explosives, combat vehicles
	KA – Military science
3.01 Basic medicine	EC – Immunology
	ED – Physiology
	FR – Pharmacology and apothecary chemistry
3.02 Clinical medicine	FA – Cardiovascular diseases including cardiac surgery
	FB – Endocrinology, diabetology, metabolism, nutrition
	FC – Pulmonology
	FD – Oncology and hematology
	FE – Other disciplines of internal medicine
	FF – ENT (ear, nose, throat), ophthalmology, dentistry
	FG – Pediatrics
	FH – Neurology, neurosurgery, neurosciences
	FI – Traumatology and orthopedics
	FJ – Surgery including transplantology
	FK – Gynaecology and obstetrics
	FL – Psychiatry, sexuology
	FO – Dermatology and venereology
3.03 Health sciences	AQ – Safety and health protection, safety in operating machinery
	DN – Health impact of the environment
	FM – Hygiene
	FN – Epidemiology, infectious diseases and clinical immunology
	FQ – Public health care social medicine
3.05 Other medical sciences	FP – Other medical disciplines
4.01 Agriculture, forestry, and fisheries	DF – Pedology
	GC – Plant growing, crop rotation
	GD – Fertilization, irrigation, soil treatment
	GE – Plant breeding
	GF – Diseases, pests, weeds, and plant protection
	GK – Forestry
	GL – Fishery
4.02 Animal and dairy science	GG – Zootechnics
	GH – Nutrition of farm animals

	GI – Farm animal breeding and farm animal pedigree breeding
4.03 Veterinary science	GJ – Animal diseases and animal vermin, veterinary medicine
4.05 Other agricultural sciences	GB – Agricultural machines and constructions
5.01 Psychology	AN – Psychology
5.02 Economics and business	AE – Management, administration and clerical work
	AH – Economics
	GA – Agricultural economics
5.03 Educational sciences	AM – Pedagogy and education
5.04 Sociology	AO – Sociology, demography
5.05 Law	AG – Legal sciences
5.06 Political Science	AD – Political sciences
5.07 Social and economic geography	AP – Urban, regional and transportation planning
5.08 Media and communications	AF – Documentation, librarianship, work with information
5.09 Other social sciences	AK – Sport and leisure time activities
6.01 History and archaeology	AB – History
	AC – Archaeology, anthropology, ethnology
6.02 Languages and literature	AI – Linguistics
	AJ – Literature, mass media, audio-visual activities
6.03 Philosophy, ethics and religion	AA – Philosophy and religion
6.04 Art (arts, history of arts, performing arts, music)	AL – Art, architecture, cultural heritage

technopolis|group|

In collaboration with

NIFU

Nordic Institute for Studies in
Innovation, Research and Education

