A National CRIS Infrastructure as the Cornerstone of Transparency in the Research Domain

Tomáš Chudlarský, Jan Dvořák

Institute of Information Studies and Librarianship
Faculty of Arts
Charles University in Prague
Czech Republic

Outline

- Motivation
- Czech R&D&I Information System
- Usage, Users, Statistics

Hypothesis

A national CRIS is useful, and is used massively.

Really?

Why do it?

To build trust!

The International Audit of R&D&I in the Czech Republic:

Building trust in government and among members of the R&D&I community is a precondition for successful performance of the National Research and Innovation System. Such trust is based on fair principles transparently and impartially implemented combined with conspicuous punishment for those at any level who break the rules.

(Synthesis Report, recommendation #1)

To build trust: How? Why?

Openness
Accountability
Transparency

Trust

Less lobbying
Equality of opportunities
Better achievements

National CRIS

- For the government
  - to monitor state-budget-funded research
  - to support strategic planning

- For everyone else in the country
  - to get research information

- For the rest of the world
  - to see what's going on in the country

National CRIS in the Czech Republic

The Czech R&D&I Information System

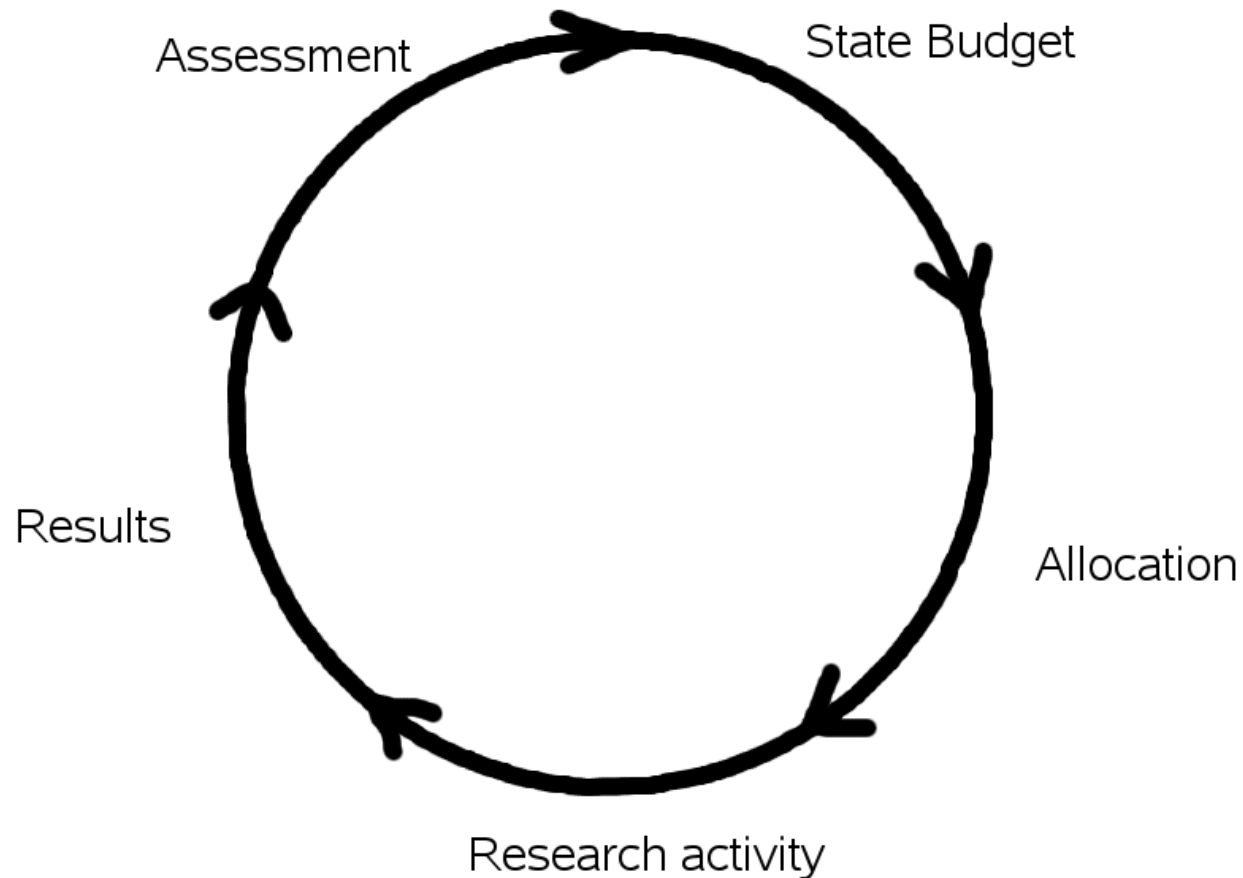   – monitors budget ~ 1 billion EUR / year

Located at the Office of the Government

- Reflects the needs and importance

- Neutral grounds

Running

- SW: by InfoScience Praha

- HW & infrastructure: Czech Technical University, Computing & Information Centre
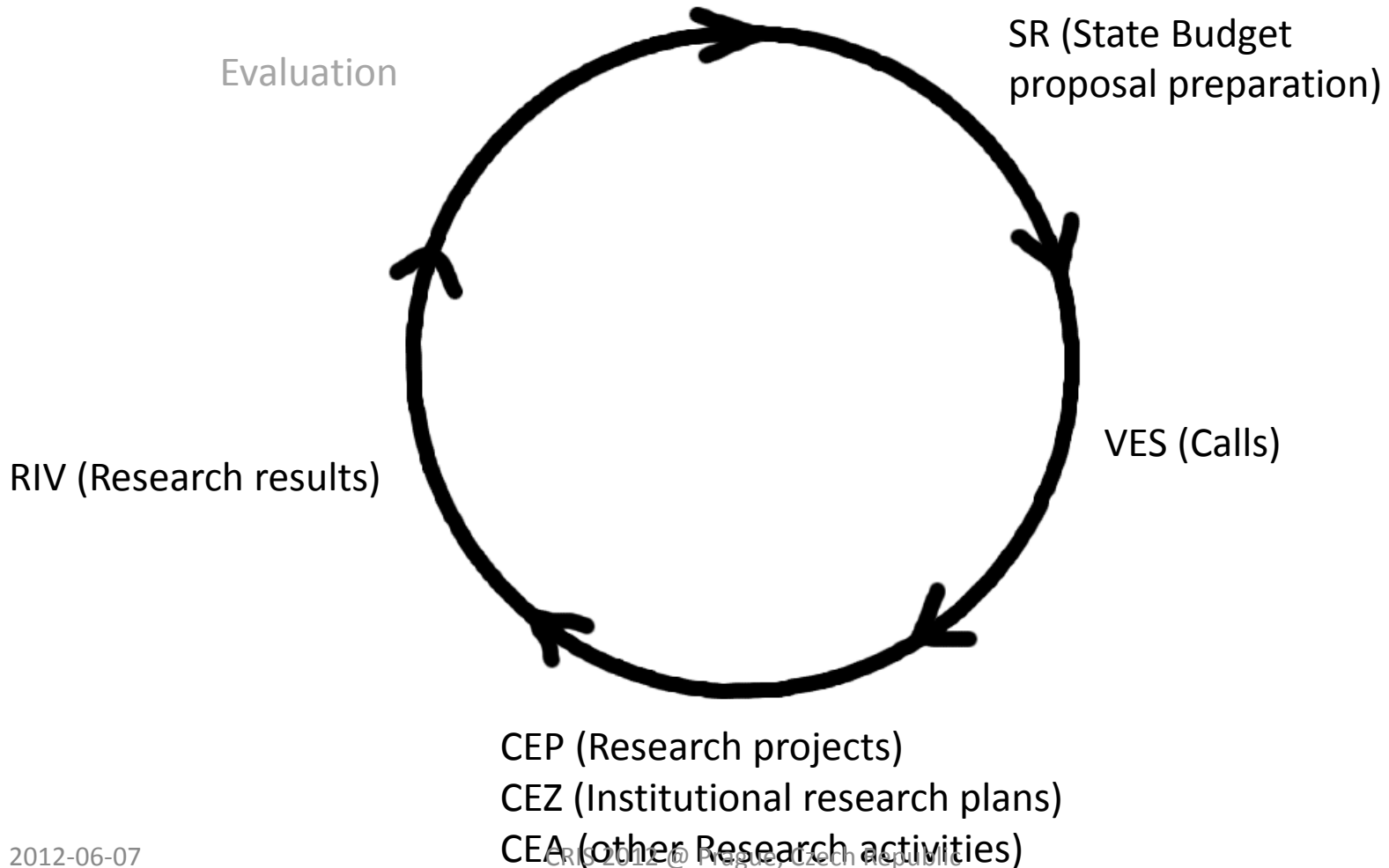
R&D management life cycle

Assessment — State Budget — Allocation — Research activity — Results (circular diagram)

Czech R&D&I IS: scope

Objects	Acronym	Object count	Inrease per year	Collected since
Research projects	CEP	37,000	1,800	1994
Institutional research plans	CEZ	889	0	1998
R&D result records	RIV	870,000	100,000	1998
Cleansed R&D results		630,000	60,000	
Calls	VES	458	30	2000
Funding schemes	CEA	235	10	2010

Czech R&D&I IS: scope

Evaluation

SR (State Budget proposal preparation)

VES (Calls)

RIV (Research results)

CEP (Research projects)
CEZ (Institutional research plans)
CEA (other Research activities)

Users and Usage

- The Research, Development and Innovation Council
  - primary user
  - strategic planning
  - allocation of the R&D budget to the funders
- Funders (ministries and the 2 funding agencies)
  - responsibility for the data
  - evaluation of project proposals
  - project outputs can be confronted with project inputs

Users and Usage

- The research-performing organizations
  - supply the metadata about their outputs
    - basis of the evaluation
  - search potential collaborators
  - track competition
- Innovative businesses
  - research outputs are concentrated in one system
- The public
  - look for information about the state-funded R&D
  - justification of the expenditures to the tax payer

National CRIS in the Czech Republic

- Comprehensive
- Has many kinds of potential users
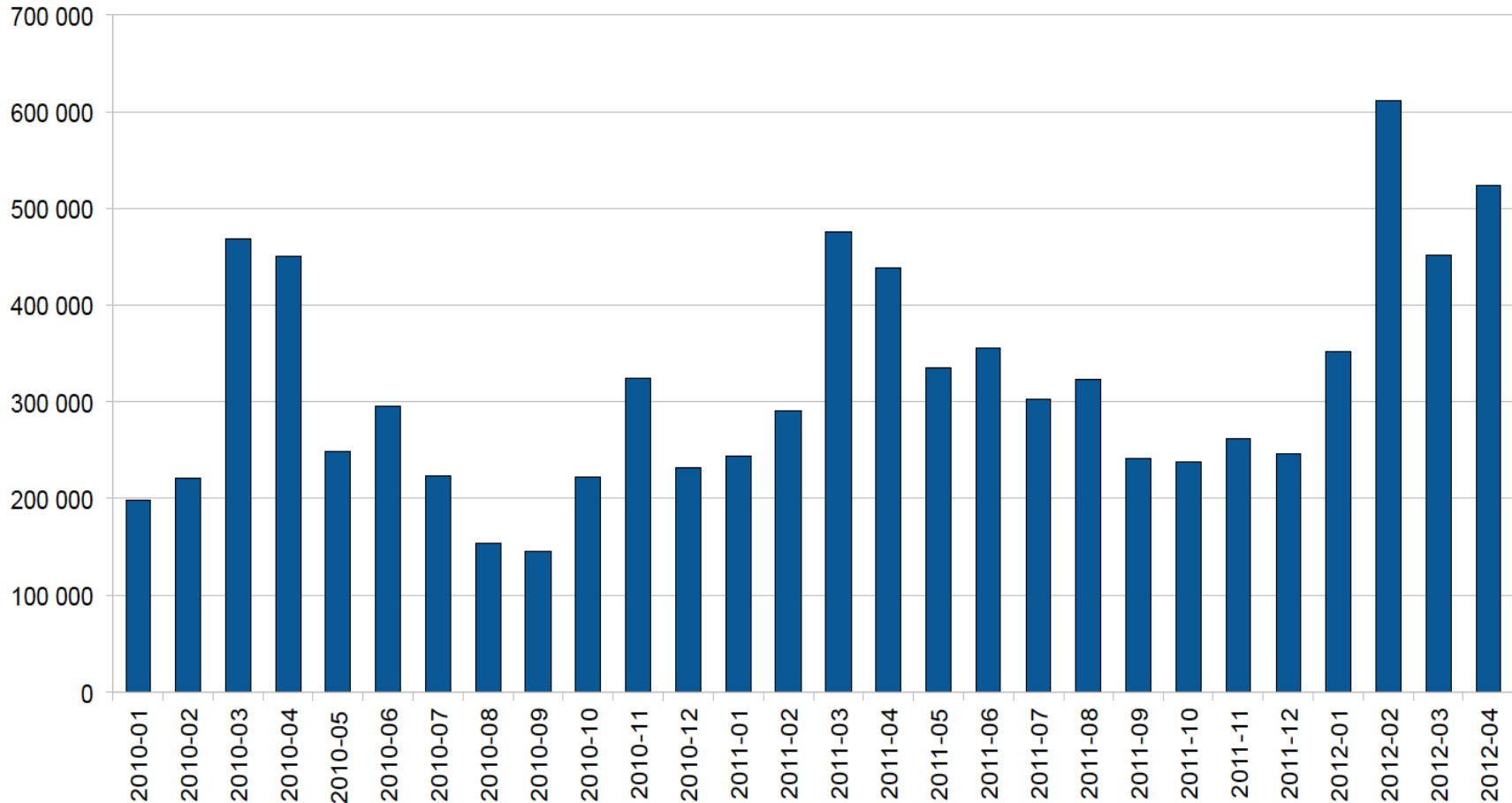

- Is it actually being used?

Usage statistics: Questions

- What is the CRIS usage over time?

- Who uses the CRIS?

- Which parts of the CRIS are used?

Usage statistics: the data

- January 2010 - April 2012 (28 months)
  152 million raw requests
  - 78 mil. requests for data or action

- Cleaning
  - exclude general search engines => 8.9 mil. dataset

- Domain detection
  - ".cz" domain => 6.7 mil. dataset
  - 2/3 ~ browser (humans)
  - 1/3 ~ specialized agents by universities and academy

Usage over time

Peaks ~ evaluation data release

Average 300,000 requests/month, growing   (but total load 20x larger)

Who uses it?

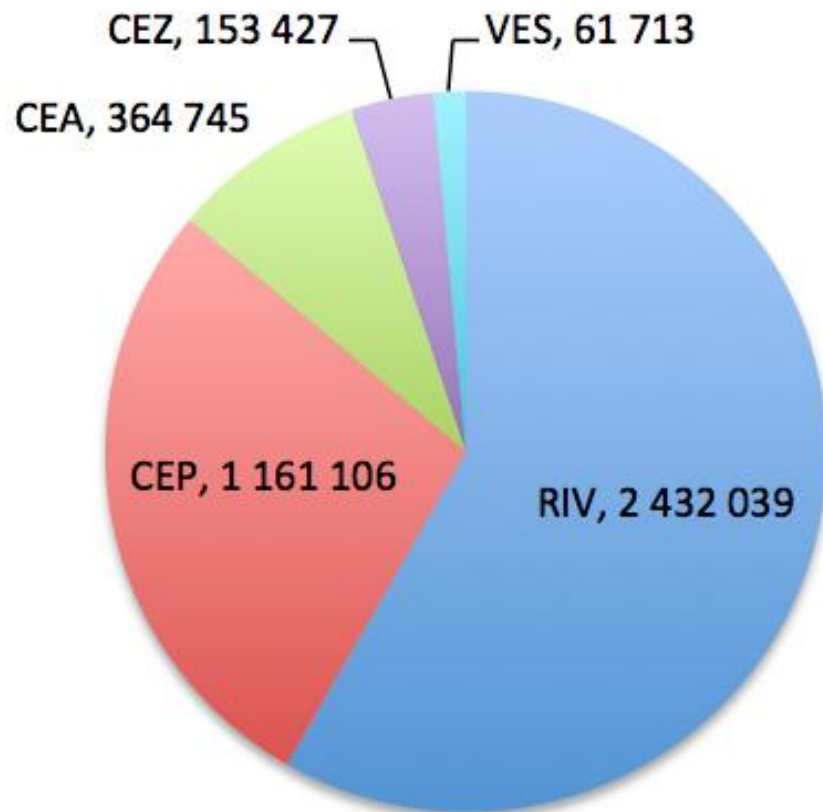| Rank	Domain	Hits	Hits (relative)	IP adresses	Institution
1	cz.vlada	952242	14,20%	2	Office of the Government
2	cz.cas	622134	9,28%	3448	Academy of Sciences
3	cz.iol	497092	7,41%	87685	a major internet provider for households
4	cz.cuni	472191	7,04%	5450	Charles University in Prague
5	cz.upcbroadband	372998	5,56%	67884	a major internet provider for households
6	cz.muni	345608	5,16%	4272	Masaryk University in Brno
7	cz.vsb	321163	4,79%	3639	Technical University Ostrava
8	cz.cvut	303298	4,52%	4260	Czech Technical University, Prague
9	cz.infoscience	133046	1,98%	4	IS maintainer
10	cz.o2isp	126982	1,89%	4268	a major internet provider for households
11	cz.tul	91930	1,37%	1916	Technical University Liberec
12	cz.vutbr	86902	1,30%	3831	Technical University Brno
13	cz.tmcz	74356	1,11%	11842	a major mobile internet provider
14	cz.zcu	71060	1,06%	1874	University of West Bohemia, Plzeň
15	cz.jcu	56083	0,84%	1822	University of South Bohemia
16	cz.osu	51325	0,77%	662	University of Ostrava
17	cz.eurotel	50788	0,76%	12663	a major mobile internet provider
18	cz.mendelu	48887	0,73%	1088	Mendel University Brno (agriculture)
19	cz.mkcr	47669	0,71%	1	Ministry of Culture
20	cz.env	46514	0,69%	2	Ministry of Environment
21	cz.upce	42968	0,64%	1690	University of Pardubice
22	cz.vscht	40946	0,61%	1144	Institute of Chemical Technology Prague
23	cz.ujep	35980	0,54%	452	University of Ústí nad Labem
24	cz.mvcr	35522	0,53%	2	Ministry of Interior

Do researchers use it?

- Users at universities and academy of sciences >= 39,000
  - Unique IP address ~ 1+ computer ~ 1+ person

  - 78,000 R&D personnel (43,000 researchers)
    - Source: Czech Statistics Office

Other observations

- Office of the Government ~ 14%
  - 11% internal CRIS tools
  - 3% ad-hoc queries

- Most funders use it frequently
  - Low usage by Technology Agency ~ new funder?
  - Low usage by Grant Agency ~ external reviewing?

Parts being used

The most frequently used:
-Research results
-Research projects

Data exports:
19,000 downloaded data sets

Validating service for input data:
74,000 executions
22,500 batch tickets

"Vklap" (a tool to prepare the data)
73,000 executions

Conclusions

The Czech R&D&I Information System is used massively and frequently. The usage is rising.

Such a national CRIS is a cornerstone for transparency in the research domain.

➜ www.isvav.cz