International Audit of Research, Development & Innovation in the Czech Republic

International Cooperation in R&D: Bibliometric Analysis

Annex 7 to the Second Interim Report

Manchester Institute of Innovation Research
Manchester Business School, University of Manchester
Manchester, UK

Philip Shapira
Abdullah Gok

esf evropský sociální fond v ČR
EVROPSKÁ UNIE
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY
OP Vzdělávání pro konkurenceschopnost
INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Table of Contents

Table of Figures

Executive Summary

This document presents preliminary findings from the bibliometric analysis of Czech international R&D collaboration. The analysis draws on a dataset of more than 85,600 science, social sciences, and arts and humanities journal article records from the Web of Science published between 1980 and mid-2010 with at least one author from the Czech Republic or the Czech part of the former Czechoslovakia.

We find that about two-fifths of all Czech research publications are co-authored and that Czech international R&D collaboration is strongly European and has grown more during the transitions of recent decades. Four-fifths of the Czech Republic's international collaboration papers are with European countries. The US is also a major collaborator. Czech scientists co-publish with German, American, French and British scientist in every subject group very extensively while they collaborate with other countries extensively only on some of the subjects. At present, collaboration with Asia is limited.

Internationally collaborated Czech publications are generally of higher citation quality than purely domestic papers. Czech international R&D collaboration is greatest (by absolute numbers of papers) in the fields of physics and material science, chemistry and chemical engineering, basic life sciences, clinical medicine, biomedical sciences and biological sciences. When international science papers are compared with purely domestic papers, Czech international R&D collaboration is relatively higher in physics and materials science and in basic life sciences, but relatively lower in biomedical sciences and clinical medicine. International collaboration in several social science disciplines is weak compared to this subject group's national significance. In particular, in economics and business and in politics and public administration, there are low levels of international collaboration relative to purely domestic research outputs.

International research collaboration through co-authorship is dominated by two institutions, the Academy of Sciences of the Czech Republic (ASCR) and Charles University (UK). These institutions are also powerful nationally in research, but there are other Czech research institutions that collaborate less internationally than their national ranking would suggest. Three Czech organizations lead in sponsoring internationally collaborated Czech research – the Ministry of Education Youth and Sport of the Czech Republic, the Czech Science Foundation (GACR), and the Academy of Science of the Czech Republic. While international institutions are also important sponsors (especially the EU), this suggests that there is capability (real or latent) within the Czech research system to influence the direction and nature of future Czech international R&D collaboration.

Introduction

This document is an interim report which reviews the progress achieved so far in the bibliometric analysis for Wp(g): International Cooperation in R&D. The next section outlines the objectives of Wp(g) and the issues to be covered in the bibliometric analysis in this work package as presented in the proposal document. This is followed by a analysis of interim results. The report concludes with a discussion of planned activities for the final part of this work package.

The analysis presented in this report draws on a dataset of more than 85,600 journal article records from the Web of Science published between 1980 and mid-2010 with at least one author from the Czech Republic or the Czech part of the former Czechoslovakia. Our analysis covers science, social sciences, and arts and humanities. About two-fifths of the records (about 34,000 papers) comprise internationally-collaborated publications with at least one Czech co-author. This provides a large evidence base for the bibliometric exploration of Czech international R&D cooperation patterns.

Details of the bibliometric search strategy and database development are provided in the Appendix.

Scope of the Bibliometric Analysis

Under work package Wp(g), the bibliometric analysis of scientific publication records seeks to analyse:

1. Patterns of international co-publication of CZ researchers - who publishes with whom and in what fields;
2. Changing patterns of international co-authorship over time (since 1980s to present; key break points will be to probe: international co-publication in the COMECON period (pre 1989) period of transformation pre-EU membership (1990-2003); post EU membership 2004 onwards); and current era of knowledge globalisation, including changes in distributions of countries and engagement with "rising" countries (e.g. in Asia) as well as traditional N America and European partners; and by seniority (years in publishing) profile of researchers
3. Leading institutions and research groups in CZ engaged in international collaborative research
4. Rising and falling technical fields (by ISI subject categories) engaged in international collaborative research
5. Quality and impact of CZ international collaborative research (using citations, journal impact placement, other measures)
6. Recent funding trends, using funding source data where available in publication records (which should give an indication of the varied domestic and international funding sources accessed by CZ researchers engaged in international collaborations)
7. Selected comparison of CZ international co-research publication profiles with those of benchmark countries (for example: Austria; Sweden)

This interim report covers tasks 1 through 6 above.[1] As yet, we have not undertaken comparisons with benchmark countries. This can best occur once we have received feedback on our approach to date and on the preferred set of benchmark countries.

[1] We judge that it is not feasible at this point to proceed with an earlier idea to assess the role of the Czech research 'diaspora' in international collaborative research. The name-matching algorithms that would be required would be complex and exceed the resources available.

Interim Results

1.1 Patterns of international co-publication of CZ researchers

For the period 1980 through to mid-2010, we identified 85,635 Web of Science papers with at least one Czech author. Of these, 34,030 papers - or about 40% - were internationally collaborated with authors from other countries, with 51,605 papers – about 60% - domestically co-authored (with all Czech authors).

Around 80% of the internationally collaborated publications (ICP) have at least one wider European co-author while almost 90% of these European co-authors are mainly from EU27 countries other than Czech Republic (72.1% of the total) as depicted in Figure 1 authors represents one-fifth of the total. North American (mostly US) researchers are co-authors for about one-quarter of ICP papers while only one-fifth has Asian co-authors. Less than 4% of the Czech ICP papers had co-authors from the Middle East, Africa, and South and Central America and the Caribbean.

According to Figure 2 which shows the top 20 countries where the co-authors of Czech authored publications belong to, almost a quarter of the publication have a co-author who has a Germany address. The second largest country in this figure is the US which makes up 22% of all publications. They are followed by France and the UK with 15% and 14% of the publications respectively. Slovakian and Italian addressed co-authors are also found about 10% of the internationally co-authored publications. Russia, Poland, Netherlands and Sweden are also amongst the top 10 countries with whom Czech authored scientists co-published.

Figure 1: Czech International Co-authorship Pattern by Regions

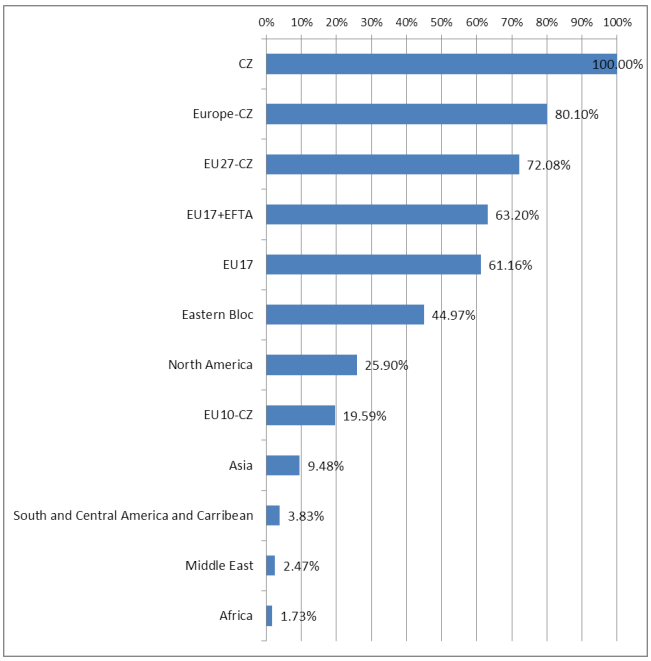


Figure 2: International Co-authorship Pattern by Top 20 Countries

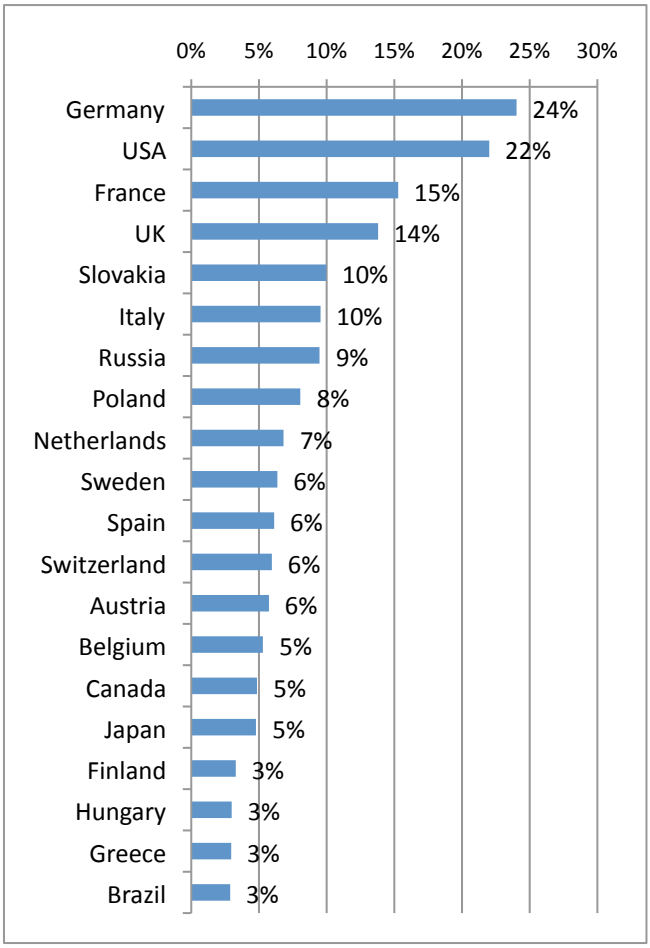
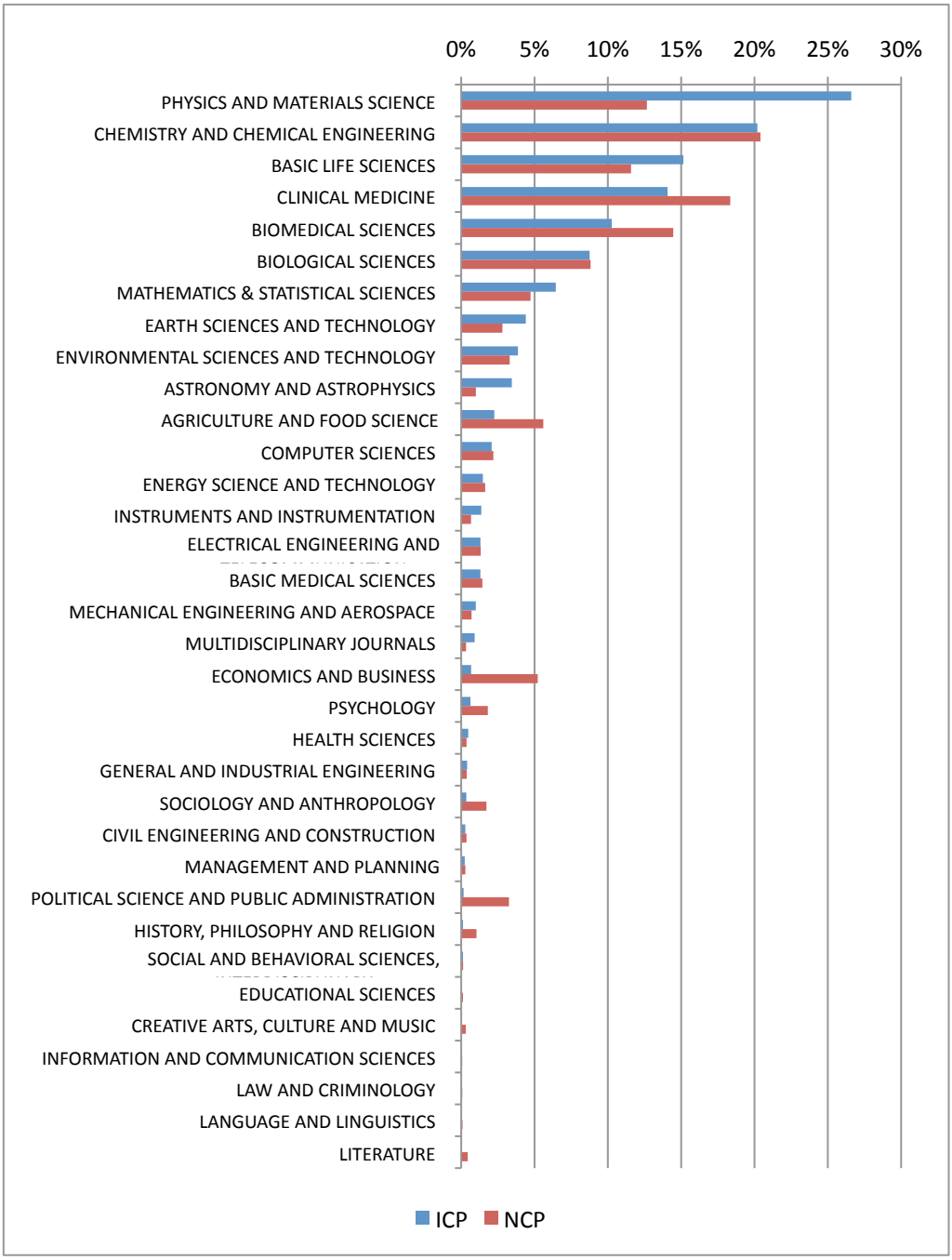


Figure 3 shows the relative importance of the subject categories of the ICP and NCP. It is evident that Physics and Material Science publications are almost two times more important in international publications than non-international publications as their shares are around 26% and 13%, respectively. Other similar areas whose importance are higher in ICP include basic life sciences, mathematics and statistical sciences, earth sciences and technology, and astronomy and astrophysics. On the hand, the share of clinical medicine, biomedical sciences, agriculture and food sciences and a group of social sciences including economics and business, political science, sociology and psychology have higher share in NCP than in ICP. The difference in economics and business publications is remarkable as this subject's relative importance changed by almost ten-fold.

Figure 3: Comparison of ICP and NCP according to Subject Categories



1.2 Changing patterns of international co-authorship over time

Figure 4 shows that international scientific publications have increased steadily by time from 1980 to 2010. Increase in the number of co-publications with all different country groups is steady until early1990s while after 1990, there are three spikes in the pace of increase. Especially the spike in the pace starting around 1997 is noteworthy as in three years period number of publications increased almost by 50%. The increase experienced in 1990s partially saturated in the first half of 2000s where there is a low climbing plateau until 2004. Finally, 2004 marks the beginning of the unprecedented increase in the number of international co-publications in the Czech Republic. One potential bias in this picture against earlier years is that it is highly likely that some of the journals located in the Czech Republic/Czechoslovakia might have not been included in the ISI-WoS database and some of the increase can be attributed to the increase in the coverage of ISI-WoS by time. However, in any case, the trends discussed above are strong and the main conclusions laid in should be true.

Figure 5 and Figure 6 are other representations of the changing patterns over time. In this case, publications are grouped into the three main timeframes: COMECON Era (1980-1989), Transformation Era (1990-2003) and Post-EU Membership Era (2004-2010). The figures are also normalised by taking yearly average within the respective era to allow an easy comparison. These figures depict the accelerating increase in the number of publications as the eras advance. There is a relative increase in the number of international collaborations with the regions that have less collaboration with Czech scientists (i.e. Asia, South and Central America and Caribbean, Middle East, Africa) is more than the other regions that have already had stronger links. This means that the breadth of international scientific cooperation is also increasing.

Figure 4: Change of international Co-authorship by Year

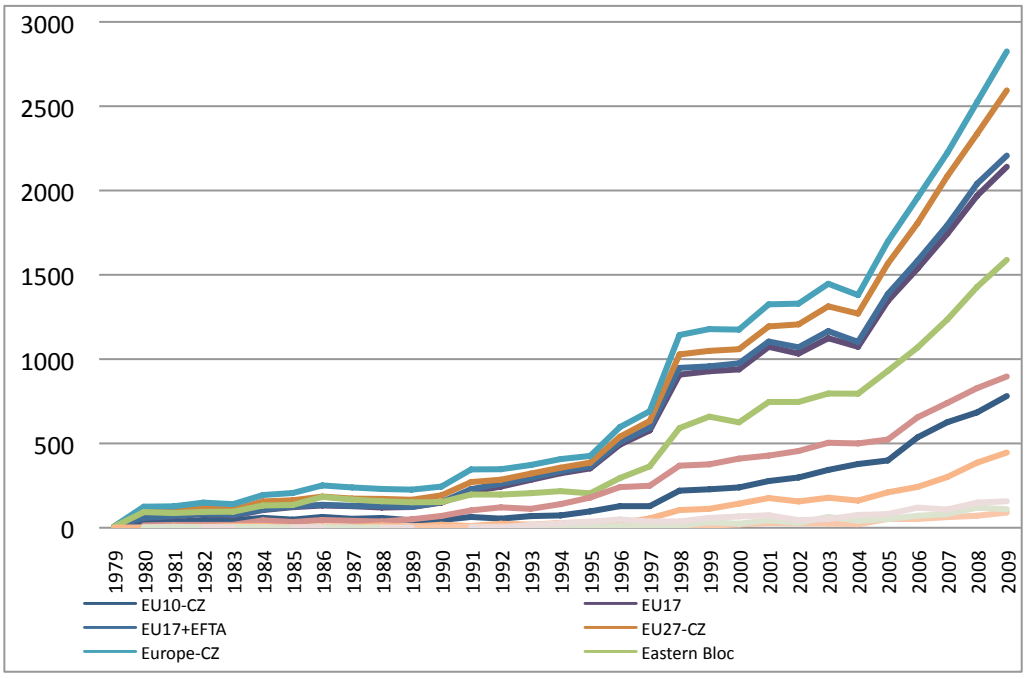


Figure 5: Change of international Co-authorship by Era (Normalised to Yearly Average)

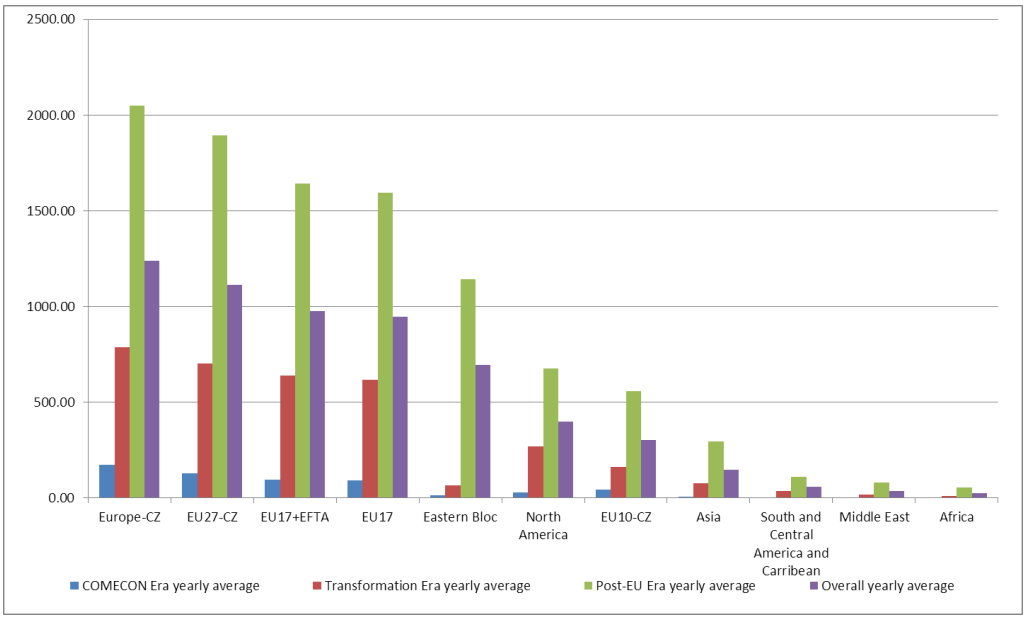
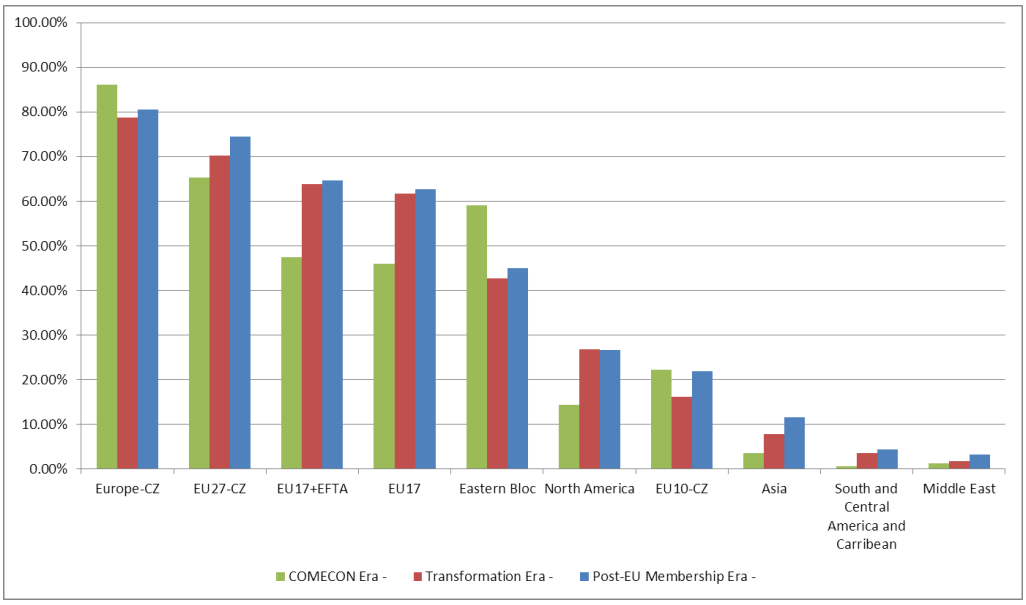


Figure 6: Change in the Pattern of Collaboration by Regions and Eras (as percentage of total number of publications)



1.3 Leading Czech and international institutions for collaborative research

The Academy of Science of the Czech Republic (ASCR) leads the NCP list (39% or all NCP), but is even more dominant (51%) in international collaborated publications (Table 1). The second and third institutions in ICP, Charles University (UK) and Masaryk University (MU), are also in the same position in NCP and their relative shares are roughly equal in both sub-datasets. Similarly, CVUT, UPOL, VSCHT and JCU appear in both top 10 lists with similar shares. CZU and UP appears in NCP top 10 list, but rank less highly in ICP. The two leading institutions, ASCR and UK, contribute to nearly four-fifths of Czech internationally contributed publications. (Note: percentage totals exceed 100% since some publications may have authors from two or more difference Czech institutions).

The leading foreign institutions collaborating with the Czech Republic include the Slovak Academy of Science, the Joint Institute of Nuclear Research (Russia) and CNRS (France). (See Table 2.)

Table 1: Top 10 Czech institutions for NCP and ICP

Internationally Collaborated Publications (ICP)			**Non-Internationally Collaborated Publications (NCP)**		
Author Affiliations	# Records	Percentage of ICP	Author Affiliations	# Records	Percentage of NCP
ASCR	17242	51%	ASCR	19965	39%
UK	9267	27%	UK	13425	26%
MU	2139	6%	MU	3546	7%
CVUT	1354	4%	VŠCHT	3174	6%
UPOL	1184	3%	UPOL	2171	4%
VŠCHT	1087	3%	CVUT	1421	3%
JCU	1004	3%	JCU	1407	3%
UP	477	1%	VFU	1134	2%
SZÚ	420	1%	CZU	1133	2%
IKEM	376	1%	UP	1047	2%

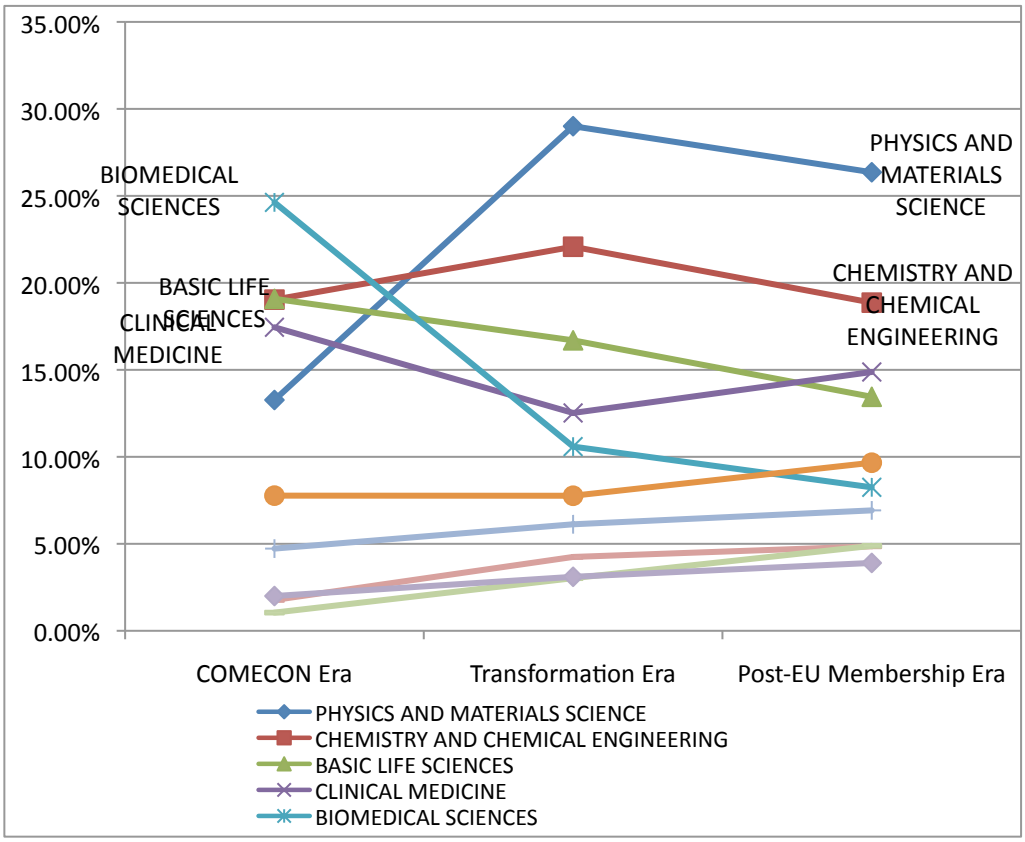
Table 2: Top 10 Non-Czech institutions for ICP

Author Affiliations	# Records	Percentage of ICP
SLOVAK ACAD SCI	1388	4%
JOINT INST NUCL RES	1061	3%
CNRS	976	3%
UNIV PARIS 06	931	3%
LUND UNIV	843	2%
UNIV PARIS 07	823	2%
COMENIUS UNIV	729	2%
UNIV LANCASTER	715	2%
UNIV PARIS 11	697	2%
UNIV GRENOBLE 1	678	2%

Rising and falling technical fields engaged in international collaborative research

Figure 7 shows the relative importance of top 10 subject categories of international collaborative publications in different eras. The most visible result in Figure 7 is that the relative importance of physics and material science increased from about 13% in the COMECON era to around 26% in the Post-EU Era. While areas like basic life sciences, clinical medicine, biomedical sciences have declined in relative importance and chemistry and chemical engineering have not increased its relative importance, all the other fields' relative importance have risen steadily at the expense of these areas that declined.

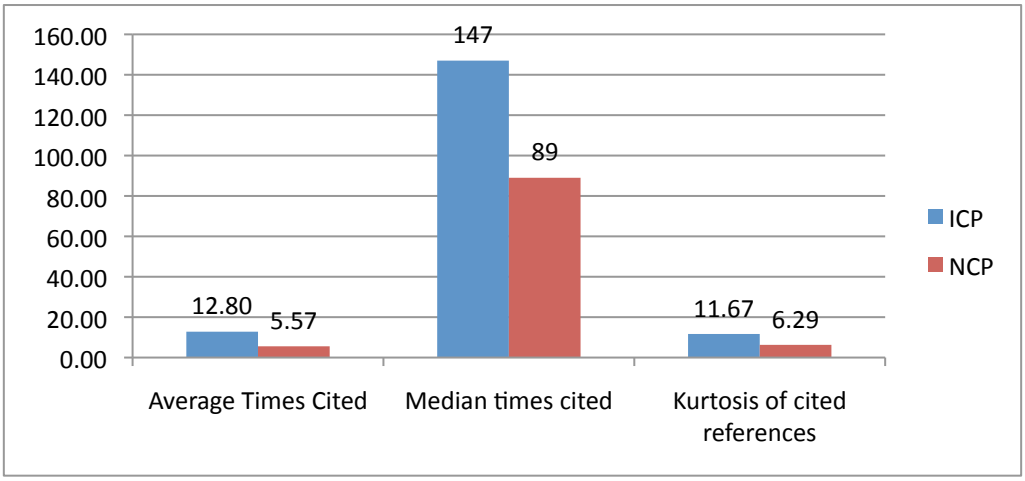
Figure 7: Change in the Top 10 Subject Categories by Eras (as percentage of all publications in this era)



Quality and impact of CZ international collaborative research

The impact of ICPs in terms of the average times cited is more than double of the impact of NCPs as depicted in Figure 8. However, it is also worth mentioning that kurtosis of the cited references for the former is significantly higher than the latter which indicates that not necessarily all of the ICPs are higher impact than NCPs.

Figure 8: Impact of ICPs and NCPs



Recent funding trends

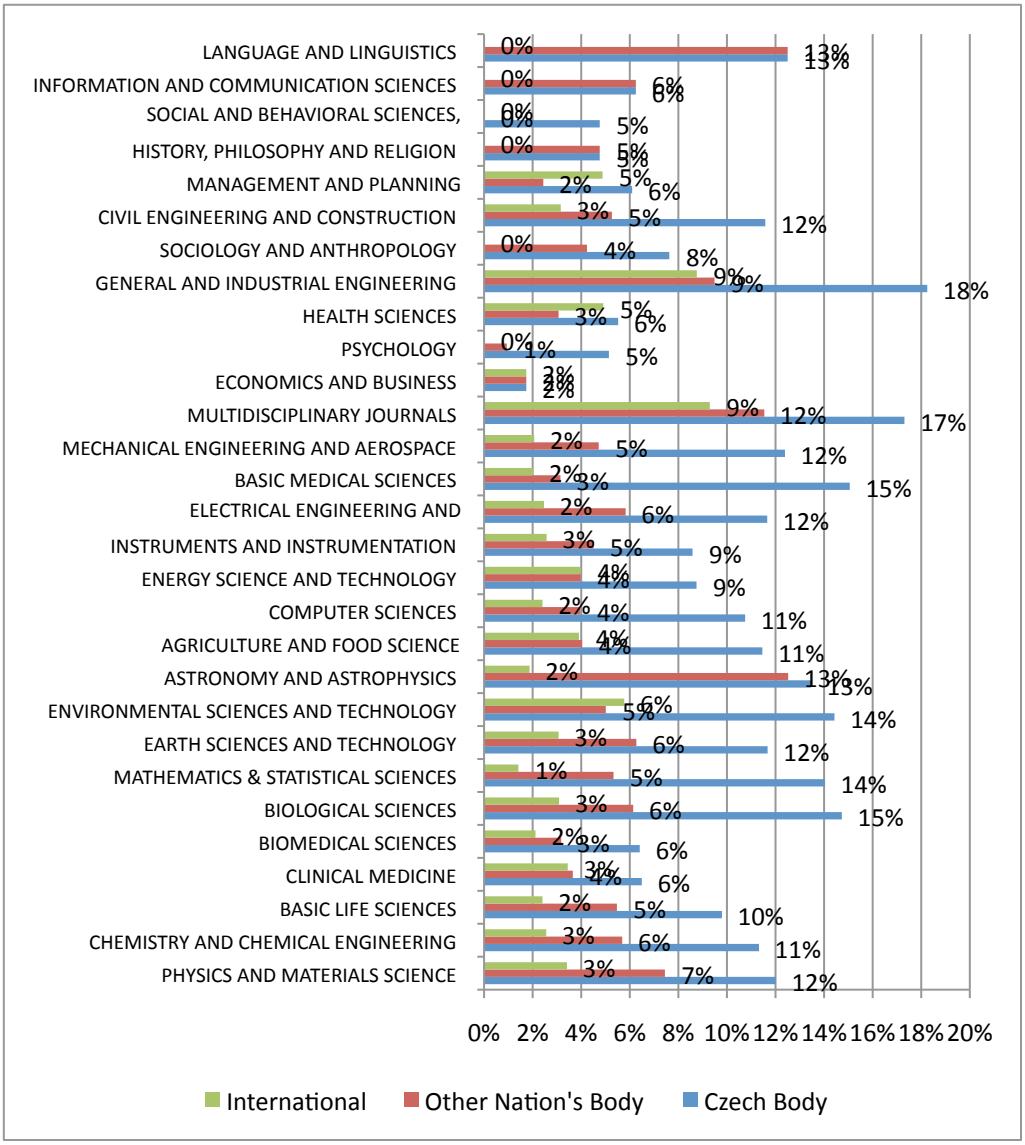
Table 3 shows the top funding organisations for NCP and ICP. While it is evident that Czech funders such as Ministry of Education Youth and Sport of the Czech Republic, GACR, Academy of Science of the Czech Republic and GAAV are dominant in both the funding of ICP and NCP, other nation's funding organisations including DFG, US NSF, CNRS, VEGA, BMBF and US DOE are funding Czech authored ICPs. Finally, EU appears in both lists although its relative importance is higher in ICP.

Figure 9 shows the papers funded by different funders as percentage of all papers in a subject category. For instance, in the subject category of physics and material science, 12% of the all internationally collaborated papers are funded by a Czech body, 7% by other nation's body and 3% by an international body. Please note that they are not mutually exclusive as some of them might be co-funded. This also shows the areas which attract most funding by different funders. For example, 17% of the general and industrial engineering papers were funded by a Czech body.

Table 3. Funding Organisations for ICP and NCP

	Funding Organization for ICP	# Records	Funding Organization for NCP	# Records
1	Ministry of Education Youth and Sport of the Czech Republic	2155	Ministry of Education Youth and Sport of the Czech Republic	3014
2	GACR	1373	GACR	1946
3	Academy of Science of the Czech Republic	1126	Academy of Science of the Czech Republic	988
4	EU	1000	Ministry of Health Czech Republic	464
5	Deutsche Forschungsgemeinschaft	412	GAAV	383
6	US National Science Foundation	334	Ministry of Agriculture of the Czech Republic	352
7	GAAV	333	EU	258
8	CNRS	168	Charles University	181
9	VEGA grant agency	166	Ministry of Industry and Trade of the Czech Republic	93
10	BMBF	155	Ministry of Environment of the Czech Republic	59
11	US Department of Energy	149	Czech Government	53
12	Charles University	130	Ministry of Defence of the Czech Republic	52
13	Ministry of Health Czech Republic	119	AVOZ	15
14	Polish Ministry of Higher Education and Science	113	MZO	15
15	CONACYT-Mexico	96	VZ MSM	12
16	Alexander von Humboldt Foundation (Germany)	91	Gilead Sciences Inc	11
17	DST (India)	88	USB RIFCH	11
18	Chinese National Natural Science Foundation	84	Czech Geological Survey	10
19	Slovak Research and Development Agency	84	Institute of Organic Chemistry and Biochemistry	10

Figure 9: Papers funded by different funders as percentage of all papers in a subject category



Preliminary Findings

Among the series of observations and findings that this interim analysis presents, we note the following:

- About two-fifths of all Czech research publications are co-authored. This appears to be substantial (although we will compare with selected other EU countries when we undertake benchmarking).
- Czech international R&D collaboration is strongly European. Four-fifths of the Czech Republic's international collaboration papers are with European countries. The US is also a major collaborator. At present, collaboration with Asia is limited (again, our benchmarking will establish where the Czech Republic stands in terms of Asian collaboration).
- Czech international R&D collaboration is greatest (by absolute numbers of papers) in the fields of physics and material science, chemistry and chemical engineering, basic life sciences, clinical medicine, biomedical sciences and biological sciences.
- When international science papers are compared with purely domestic papers, Czech international R&D collaboration is relatively higher in physics and materials science and in basic life sciences, but relatively lower in biomedical sciences and clinical medicine. Whether this indicates a greater Czech propensity to internationally collaborate in pure rather than more applied sciences remains a question for further investigation. (We are aware that in certain fields, such as high energy physics, international collaboration is the main way to access large advanced research facilities.)
- International collaboration in several social science disciplines is weak compared to this subject group's national significance. In particular, in economics and business and in politics and public administration, there are low levels of international collaboration relative to purely domestic research outputs.
- In the Czech Republic's transitions since the 1980s, research collaborations with Western European countries appear to have grown significantly.
- Czech scientists co-publish with German, American, French and British scientist in every subject group very extensively while they collaborate with other countries extensively only on some of the subjects.
- International research collaboration through co-authorship is dominated by two institutions, the Academy of Sciences of the Czech Republic (ASCR) and Charles University (UK). These institutions are also powerful nationally in research, but there are other Czech research institutions that collaborate less internationally than their national ranking would suggest.
- Internationally collaborated Czech publications are generally of higher citation quality than purely domestic papers.
- Three Czech organizations lead in sponsoring internationally collaborated Czech research – the Ministry of Education Youth and Sport of the Czech Republic, the Czech Science Foundation (GACR), and the Academy of Science of the Czech Republic. While international institutions are also important sponsors (especially the EU), this suggests that there is capability (real or latent) within the Czech research system to influence the direction and nature of Czech international R&D collaboration.

Next Steps

As this interim report illustrates, we have established a large-scale bibliometric database to analyse Czech international cooperation in R&D.

We have categorized the data into a series of regions (the various EU regional blocs overlap, but allow us to now easily focus on various European combinations). Publication years have been grouped into three different eras of the COMECON period, 1980 – 1989; the period of transformation (pre-EU membership), 1990-2003; and post EU membership, 2004 onwards. Funding organisations are categorised as Czech national funders, other national funders and international funders. Finally, publication subjects are grouped into 34 broader categories in line with the categorisation of the analysis being conducted in WP d, ii.

The analyses of this bibliometric data need to be incorporated and triangulated with evidence from other components of the work package on Czech international R&D collaboration (including survey, documentary and interview data) to develop robust conclusions and recommendations. However, it should be possible to discern trends and also aspects for further probing from the bibliometric data. This will be the key focus for the final stage of this work package: to develop a richer set of narrative insights and observations from the bibliometric evidence. Additionally, we will complete the final task of selected country benchmarking.

Appendix A - Bibliometric Database Development

1.1. Overview

This appendix summarizes the bibliometric search strategy used to identify, clean, and organize the publication records used in the analysis of Czech international cooperation in R&D.

1.2. Data Source and Specification

The data for the bibliometric analysis is derived from publication records listed in Thomson Reuters Web of Science (WoS). More than 10,000 journals published worldwide are indexed in the WoS, with coverage in the sciences, social sciences, arts, and humanities. The WoS indexes a variety of publication types, including journal articles, proceedings, reviews, abstracts and editorial materials.

In the analysis of Czech international cooperation in R&D, we focus on journal articles (which comprise the majority of all indexed records in the WoS). We accessed and downloaded all articles with a least one Czech Republic or Czechoslovak author listed in the three WoS databases SCI-EXPANDED, SSCI and A&HCI for the period 1980-2010. SCI-EXPANDED covers science; SSCI covers social science; and A&HCI covers arts and humanities. For 2010, there is a partial year (data was downloaded in August 2010/September 2010). The search terms "Czech Republic" and "Czechoslovakia" are used for publications after 1993 and before 1993 respectively, as depicted in Table 1. For the period 1993 – 1995, we also searched for "Czechoslovakia" since articles with this country address continued to appear in publications for a period after the dissolution of Czechoslovakia in January 1993.

Figure 10: Data Search Queries

Search Term		Years	Databases	Type
CU=(CZECH REPUBLIC)		1993-2010	SCI-EXPANDED + SSCI + A&HCI	Article Only
CU=(Czechoslovakia)		1993-1995	SCI-EXPANDED + SSCI + A&HCI	Article Only
CU=(Czechoslovakia) **CU=(Czech Republic)**	**OR**	1980-1992	SCI-EXPANDED + SSCI + A&HCI	Article Only

1.3. Data Cleaning

The WoS data records were downloaded and imported into VantagePoint text mining software. VantagePoint is used for data cleaning and analysis. We removed several non-vital record fields to reduce the file size. A critical step in the cleaning process was the consolidation of records published by authors within the boundaries of the present Czech Republic. This involved identifying and removing records *only* associated with authors from the Slovakian part of the former Czechoslovakia.

Initially the database consisted of 142,038 records of published articles. This initial data included records that did not have a Czech Republic address but were captured because they had authors from the Slovakian part of Czechoslovakia. To remove such records, all publications giving the country address as Czechoslovakia were checked

against the record's city address information. A list of major Czech and Slovakian cities within Czechoslovakia was compiled and used to identify Czech authors with Czechoslovakia address. After removing Slovakian publications (with no Czech Republic co-authors), the resulting dataset comprised 85,635 records.

We also undertook data cleaning to standardize institutional affiliations, funding organisations, and country names. Assistance with standardizing Czech Republic organisation names was provided by Lucie Vavrikova of Technology Centre. Several fields such as author country affiliations, publication years, funding organisations and subject categories are further processed to form groups within them. For publication years, as outlined in the proposal, we identified three major phases: COMECON Era (1980-1989), Transformation Era (1990-2003) and Post-EU Membership Era (2004-2010). For funding organisations, the major groups are international sponsors (such as the European Union), Czech sponsors, and other national sponsors. For subject categories, the 34 groupings used by the WP(d) are employed to ensure consistency. For author country affiliations, we identify these major groupings: Europe-CZ, EU10-CZ, EU17, EU17+EFTA, EU27-CZ, Eastern Bloc, North America, Asia, South and Central America and the Caribbean, Middle East, and Africa. Not all of these groupings are reported in the report.

In a final stage of data cleaning, the dataset is divided into two sub-datasets. The first sub-dataset is called "Non-Internationally Collaborated Publications (NCP)" which only includes the publications authored by Czech addressed authors while the second dataset, "Internationally Collaborated Publications (ICP)" includes publications that has at least one non-Czech addressed author. The former sub-dataset, NCP, makes up 51,605 or about three-fifths of the records, while the ICP comprises 34,030 or about two-fifths.

In Brighton, 15/02/2011

Erik Arnold

Technopolis Limited

Managing Director