European Commission

European Research Area

Progress Report 2016

ERA Monitoring Handbook

EUR 28430 EN

Research and Innovation

EUROPEAN COMMISSION

Directorate-General for Research and Innovation
Directorate B — Open Innovation and Open Science
Unit B.2 — Open Science and ERA policy

Contact: Fabienne Gautier

E-mail: Fabienne.gautier*@ec.europa.eu*
RTD-PUBLICATIONS@ec.europa.eu

European Commission
B-1049 Brussels

EUROPEAN COMMISSION

ERA PROGRESS REPORT 2016

ERA MONITORING HANDBOOK

Accompanying

SCIENCE-METRIX STUDY

'Data gathering and information for the 2016 ERA monitoring' – Technical Report

Produced and written by

Science-Metrix: David Campbell, Chantale Tippett, Guillaume Roberge, Bastien St-Louis, Rémi Lavoie, Grégoire Côté & Éric Archambault

Written by Science Metrix
September – 2016

Directorate-General for Research and Innovation
European Research Area

2017

EUR 28430

EUROPE DIRECT is a service to help you find answers to your questions about the European Union

Freephone number (*)
00 800 6 7 8 9 10 11

(*) The information given is free, as are most calls (though some operators, phone boxes or hotels may charge you)

More information on the European Union is available on the internet (http://europa.eu).

Luxembourg: Publications Office of the European Union, 2017.

PDF ISBN 978-92-79-65533-3 ISSN 1831-9424 doi: 10.2777/098457 KI-04-17-112-EN-N

Table of Contents

TABLES

ACRONYMS

AC	Associated Country
A&HCI	Arts & humanities citation index
BES	Business enterprise sector
CAGR	Compound annual growth rate
CWTS	Centre for science and technology studies, Leiden University
DG	Directorate-General
ERA	European Research Area
ERAC	European Research Area Committee
ERC	European Research Council
EMM	ERA monitoring mechanism
ESFRI	European Strategy Forum on Research Infrastructures
FTE	Full-time equivalent
GBARD	Government budget allocations for R&D
GDP	Gross domestic product
GDRC	Gender dimension in research content
GOV	Government sector
HEI	Higher education institution
HES	Higher education sector
HS	Health science
ISCED	International standard classification of education
JPI	Joint programming initiatives
JRC-COIN	Joint research centre, competence centre on composite indicators and scoreboards
MORE	Mobility and career paths of researchers in Europe
MS	Member State
NAP	National action plan
NSE	Natural sciences and engineering
OA	Open access
OECD	Organization for Economic Cooperation and Development
P2P	Public to public
PATSTAT	EPO Worldwide Patent Statistical Database
PCT	Patent cooperation treaty
PRI	Private research institute
RPBF	Research performance based funding
REI	Research excellence indicator
RI	Research infrastructure
RIO	Research and innovation observatory
R&D	Research and development
R&I	Research and innovation
RFOs	Research funding organisations
RPOs	Research performing organisations

SCI Expanded	Science citation index expanded
SI	Specialisation index
SII	Summary innovation index
SSCI	Social sciences citation index
SSH	Social sciences and humanities
S&T	Science and technology
WiS	Women in science
WIPO	World Intellectual Property Organization
WoS™	Web of Science™ database (by Thomson Reuters)

1 GENERAL INTRODUCTION

The European Research Area (ERA) Monitoring Handbook of Indicators has been developed to accompany the 2016 ERA Monitoring Report. It provides guidance on the theoretical underpinnings, the collection of data and the calculation of all indicators presented in the main report and the individual country profiles. Also included is a discussion of any pertinent considerations for the use of data on progress towards the common ERA at the organisational, Member State/Associated Country and/or European level.

Current version of the handbook - ERA Monitoring 2016

The 2016 ERA Monitoring Report gathers, systematises, and analyses internationally comparable data and indicators to monitor, at the pan-European level, progress towards a common ERA. It covers a wide range of themes, aligned with the five key priorities underpinning the achievement of the ERA (European Commission, 2012), including indicators on the effectiveness of national research systems, optimal transnational cooperation and competition, the openness of labour markets for researchers, gender equality and gender mainstreaming in research, and optimal access to and circulation and transfer of scientific knowledge, including via digital ERA. The ERA Progress Report provides a crucial evidence base for policies in these areas, while accounting for recent refinements in the key implementation priorities. For instance, the cross-cutting theme on the international dimension outside the ERA was transformed into a sixth priority for the 2015 ERA Roadmap, while Priorities 2 and 5 were split into two sub-priorities each (ERAC Secretariat, 2015).

Multiple lines of evidence have been used to triangulate the findings: the compilation of quantitative data, a desk research and document review, as well as interviews. This Handbook serves as a resource detailing the relevant guidelines for the collection of quantitative and qualitative data pertaining to relevant information in implementing the ERA Monitoring Mechanism. Upon future developments and new additions to the ERA Monitoring Mechanism (EMM), the data collection instruments included in the Handbook will be revised accordingly. As such, it is designed to reflect the state of the art in the mapping and monitoring of the ERA.

1.1 Aim and scope

Aim

This Handbook aims to provide specific guidelines and recommendations concerning the necessary data and indicators for monitoring progress around the six ERA priorities for each Member State/Associated Country at country and organisational level.

In particular, the Handbook promotes cross-country uniformity in terms of data collection, indicator computation and data-validation procedures. Furthermore, it provides interested stakeholders with detailed information on the data needed to examine multiple dimensions of progress in the six key priorities underpinning the ERA. It serves as a reference document and provides users with the methods needed to undertake the following:

- Calculate the indicators, so as to increase consistency of ERA monitoring indicators across countries and time periods
- Analyse and synthesise the collected quantitative data
- Assess and ensure the quality of the collected quantitative data
- Conduct the desk research and document reviews required to gather the qualitative information intended to clarify the quantitative findings
- Conduct interviews with relevant stakeholders to complement the information gathered in the desk research and document reviews

Scope

The Handbook is not intended to be specific to any version of the ERA Progress Reports. Rather, it is a live document intended to be used as the basis for the computation of indicators in current and future versions of the ERA Progress Report.

Current version of the Handbook

Although intended to act as a stand-alone document (i.e. untied to any of the specific versions of the ERA Progress Report), the current version of the Handbook was created to accompany the 2016 edition of the publication and thus includes some details specific to that edition. In the 2016 version, data are presented at the individual country level and the broader EU level for the current 28 EU Member States, plus associated countries (Iceland, Norway, Switzerland, Montenegro, the Former Yugoslav Republic of Macedonia, Albania, the Republic of Serbia, Turkey, Bosnia and Herzegovina, Israel, the Faroe Islands, the Republic of Moldova and Ukraine).

The Handbook has been thoroughly cross-referenced and contains an indexed list of key terms aimed at improving accessibility and readability (see Annex 2).

1.2 History and background of the ERA Progress Report

History

Since the ERA's conception in 2000, the EU Member States and Associated Countries have made substantial progress on the implementation of relevant policies and initiatives, and the conditions for the completion of the ERA are now in place. Yet continuous progress requires balancing efforts and speeding up the pace of implementation among the various actors. For example, a 2014 analysis of the state of play in each Member State and a selection of Associated Countries highlighted that only half of the Member States had implemented measures to at least a medium degree, and that progress was particularly low for priority 4 (Gender equality and gender mainstreaming in research). Additionally, regional differences in implementation were identified among Western European countries, which fared better than Central and Eastern European Member States. Notably, the distinction between Member States and Associated Countries did not appear to be relevant.

The European Council has declared that realising the ERA necessitates the monitoring of progress in close connection with the European Semester, and invited the European Commission to establish such a monitoring mechanism. Consequently, the EMM was developed by the European Commission in close collaboration with Member States, with the aim of assessing compliance to the ERA at the levels of national and regional policies, RFOs and RPOs.

In this context, in September 2013 the European Commission published the results of the first ERA Progress Report, which presented an overview of the political context, actions taken and recent progress towards achieving the ERA. The report was accompanied by the ERA Facts and Figures report, where the state of play in each of the ERA priorities in EU Member States and Associated Countries were presented, with more detail on the situation in each country presented in 'country fiches'. The 2014 ERA Progress Report followed a similar structure and approach to the 2013 version; however, it included some important adaptations and additions, such as the state of play of support provided by RFOs for the adoption of ERA measures. The ERA progress reports have been produced using qualitative and quantitative information from various sources, including, but not limited to, information contained in National Reform Programmes, results from ad hoc ERA surveys, official internationally comparable statistics from Eurostat, and measures identified by the Institute for Prospective and Technological Studies of the Joint Research Centre.

Data in the ERA Progress Report

A portion of the ERA Progress Report is dedicated to reporting back on a core set of indicators — elucidated throughout by qualitative information gathered through document reviews and interviews — which serve as the foundation for exposing progress at the organisational and national levels toward the common ERA. However, each year the ERA Progress Report builds on previous versions by improving on the definition of indicators, introducing new indicators where there is a need, or refining the scope of the methods and approaches used for data collection and analysis.

1.3 Structure of the Handbook

The Handbook of Indicators on ERA monitoring is made up of six sections and three annexes:

- The first (current) section provides a brief overview of the aim and scope of the Handbook, as well as a background to the ERA priorities and progress.

- The second section describes all indicators used in the ERA Monitoring publication, including definitions, rationale and computation method (with the necessary data, data source, formulas and any calculation specifications or comments that may be of relevance).
- The third section describes the general approach used to analyse and present quantitative indicators, in particular the analysis of progress towards achieving the ERA.
- The fourth section details the general quality plan of the ERA Monitoring publication, focusing on the methodological principles employed in the verification and validation of data.
- The fifth section presents the approach taken to conduct the desk research and document reviews.
- The sixth section presents the approach taken to conduct interviews with relevant ERA stakeholders.
- There are three annexes. The first synthesises recent changes to international classification standards that were taken into account. The second provides an overview of how key terms are defined. Finally, the third provides lists of the indicators sorted by priority, type (Headline, EMM, etc.) and alphabetical order.

The sections and annexes are followed by the bibliography.

2 INDICATORS

In May 2015, the consulting group ICF International (henceforth referred to simply as ICF) performed an appraisal of available or potential indicators and proposed a core set of 22 indicators with which to monitor progress across ERA priorities (ICF International, 2015). Building on ICF's work, the European Research Area and Innovation Committee (ERAC) selected eight core high level indicators (one per priority, or sub-priority for Priorities 2 and 5) that are regarded as being the most relevant in monitoring progress in achieving the ERA (ERAC Secretariat, 2015). In addition to these Headline indicators, the ERAC selected two complementary ERA Monitoring Mechanism (EMM) indicators per priority (including the sub-priorities for priorities 2 and 5; selected at an ad hoc workshop of the ERAC in March 2016) for a total of 24 EMM indicators (including the Headline indicators). Please refer to Table 1 for a list of Headline and EMM indicators. Additional indicators included in the 2016 Progress Report include indicators used in the 2014 ERA Facts and Figures report and its companion country fiches (European Commission & Directorate General for Research and Development, 2015), and indicators identified in discussion between Science-Metrix and the European Commission to further round out the quantitative assessment. Additionally, Science-Metrix computed composite indicators to facilitate integrated assessments, including assessments across indicators within a given priority, as well as assessments across priorities. The following sections present the indicators according to the data source used for their computation. Refer to Annex 3 for a complete list of the indicators covered in this Handbook, sorted by priority, type (Headline, EMM, etc.) and alphabetical order.

Table 1 Matrix of Headline and complementary EMM indicators

Priority	Input Indicator	Output Indicator	Outcome/Impact Indicator
Priority 1: More effective national research systems	GBARD as percentage of GDP (Eurostat)	Adjusted Research Excellence Indicator (REI) (source: JRC)	European Innovation Scoreboard Summary Innovation Index (SII) (source: EIS)
Sub-priority 2a: Optimal transnational cooperation	Participation in public-to-public partnerships per researcher in the public sector (ERA-Learn 2020 report on P2P)	GBARD allocated to Europe-wide transnational, as well as bilateral or multilateral, public R&D programmes per FTE researcher in the public sector (Eurostat)	International co-publications with ERA partners per 1 000 researchers in the public sector (WoS and Eurostat)
Sub-priority 2b: European Strategy Forum on Research Infrastructures (ESFRI)	Share of developing ESFRI projects in which a Member State or an Associated Country participates (ESFRI)	Availability of national roadmaps with identified ESFRI projects and corresponding investment needs (ESFRI)	Share of operational ESFRI landmarks in which a Member State or an Associated Country is a partner (ESFRI)
Priority 3: Open Labour Market for Researchers	Share of doctoral candidates with a citizenship of another EU Member State	Researcher's posts advertised through the EURAXESS job portal per 1 000 researchers in the public sector (EURAXESS and Eurostat)	Share of researchers expressing satisfaction that the hiring procedures in their institution are open, transparent and merit-based (MORE2 Survey)
Priority 4: Gender equality and gender mainstreaming in research	Share of female PhD graduates (Eurostat)	Gender dimension in research content (WoS)	Share of women in grade A positions in HES (WiS—Women in Science database)
Sub-priority 5a: Knowledge circulation	Share of product and/or process innovative firms cooperating with higher education institutions or public/private research institutions (Eurostat)	Share of public research financed by the private sector (Eurostat)	Number of public-private co-publications per million population (CWTS and Eurostat)
Sub-priority 5b: Open access	Share of RFOs that provide funds to cover the costs of making publications available in OA and share of RPOs making their research data available in OA (data unavailable)	Share of publications available in open access (green and gold) (1Science, WoS)	Presence or absence of national OA policies in RIO policy repositories (JRC Research and Innovation Observatory (RIO) policy repositories)
International dimension outside ERA (Priority 6)	International co-publications with non-ERA partners per 1 000 researchers in the public sector (WoS and Eurostat)	Non-EU doctorate students as a share of all doctorate students (EIS)	Licence and patent revenues from abroad as a share of GDP (Eurostat)

Note: The cells in light green represent Headline indicators, while the cells in light grey hold EMM complementary indicators.

Source: Assembled by Science-Metrix from ERAC documentation

2.1 ESFRI

National roadmaps form part of the European Strategy Forum on Research Infrastructures (ESFRI) and are the blueprints for the setting of national priorities and funding strategies for pan-European research infrastructure activities. The ESFRI is a forum comprising EU Member States (MS) and Associated Countries (AC) that supports and guides policy relating to research infrastructures in Europe. It also aims to encourage the effective use of research infrastructures through collaboration between EU countries and internationally. In 2012, the mandate of the ESFRI was expanded to include support for the implementation of projects and to maintain Europe's leadership role in research and innovation. ESFRI Member States have been encouraged to link national roadmaps to the European ESFRI roadmap in an effort to better allocate resources and efforts.

2.1.1 P2b – Headline indicator – Availability of national roadmaps with identified ESFRI projects and corresponding investment needs

Definition of indicator

This indicator presents the availability of national roadmaps for research infrastructures for each Member State and assesses if the national roadmap contains identified ESFRI projects with corresponding investment needs.

Rationale

This indicator pertains to sub-priority 2b (ESFRI), which aims to improve the effectiveness with which ESFRI regional and national research infrastructures (RIs) of pan-European interest are financed and shared across all MS/AC. It serves as a measure of the presence and comprehensiveness of national roadmaps and compliance to the request made by the European Commission for Member States to link their roadmaps to the ESFRI roadmap (European Commission, 2012). This relates to the key priority of increasing the effectiveness of investments in and use of RIs identified by the European Commission for the reinforcement of the ERA (European Commission, 2012).

Computation method

Data needed

(NRM)	National roadmap document
(RM)	ESFRI Roadmap

Source of data

For national roadmaps: *ESFRI website*

(https://ec.europa.eu/research/infrastructures/index_en.cfm?pg=esfri-national-roadmaps)

Specifications

This binary indicator is given the value of 1 if the MS/AC has a roadmap available on the ESFRI website and if it contains identified ESFRI project and investment needs. If one of the aforementioned elements is missing the value zero is attributed. The year reported is the year for which the roadmap was published. The year indicates the publication date of the latest available roadmap.

Comments/critical issues

Note that not all MS/AC have submitted a national roadmap to ESFRI (BE, CY, LV, LU, MT, SK, IS, ME, MK, AL, RD, TR MD and UA are missing) and the last update varies from 2007 to 2014.

2.1.2 P2b – EMM indicator – Share of developing ESFRI Projects in which a Member State or an Associated Country participates

Definition of indicator

This indicator is the proportion of ESFRI projects in which a given country participates.

Rationale

This indicator pertains to sub-priority 2b (ESFRI) and relates to financial commitments for the construction and operation of ESFRI, national and regional research infrastructures (RIs) to improve access to RIs of pan-European interest. In an increasingly competitive, globalised and knowledge-based economy, the goal of ESFRI is to fully exploit the Member States' potential for scientific and technological innovation by structuring their research objectives, developing common protocols, sharing expertise, fostering multidisciplinarity and maintaining competitiveness (ESFRI, 2016). Member States are therefore encouraged to take part in ESFRI projects — that is, early development phase projects aiming to establish RIs.

Computation method

Data needed

(PROJ) Number of ESFRI projects in which the country is participating in: **Unit=Total**;

($PROJ_{TOT}$) Total number of ESFRI projects: **Unit=Total**.

Source of data

ESFRI

Specifications

$$Share\ (\%)\ of\ ESFRI\ projects\ in\ which\ the\ country\ participates = \frac{PROJ}{PROJ_{Tot}} \times 100$$

Comments/critical issues

None identified.

2.1.3 P2b – EMM indicator – Share of operational ESFRI Landmarks in which a Member State or an Associated Country is a partner

Definition of indicator

This indicator is the proportion of ESFRI landmarks in which a given country is a partner.

Rationale

This indicator pertains to sub-priority 2b (ESFRI) and relates to financial commitments for the construction and operation of ESFRI, national and regional research infrastructures (RIs) to improve access to RIs of pan-European interest. In line with the previous indicator on ESFRI projects, ESFRI landmarks are successfully implemented ESFRI projects that are delivering science services or effectively advancing in their construction (ESFRI, 2016).

Computation method

Data needed

(LAND) Number of ESFRI landmarks in which the country is a partner: **Unit=Total**;

($LAND_{TOT}$) Total number of ESFRI landmarks: **Unit=Total**.

Source of data

ESFRI

Specifications

$$Share\ (\%)\ of\ ESFRI\ landmarks\ in\ which\ the\ country\ is\ a\ partner = \frac{LAND}{LAND_{Tot}} \times 100$$

Comments/critical issues

None identified.

2.2 EURAXESS Portal

EURAXESS is a pan-European initiative providing a diversity of support services to assist researchers in their career development in Europe or in maintaining their connection to European research. As part of the EURAXESS initiative, a job portal provides easy access to all open job offerings throughout the ERA ([1]).

2.2.1 P3 – Headline indicator – Number of researcher postings advertised through the EURAXESS job portal, per 1 000 researchers in the public sector (2012-2014)

Definition of indicator

This indicator is the ratio of the number of researcher posts advertised through the EURAXESS job portal to the number of researchers in the public sector.

Rationale

This indicator pertains to priority 3 (open labour market for researchers) and measures active international recruitment efforts by a given country's institutions. This relates to the goal of creating an open labour market for researchers established by the Commission for reinforcing the European Research Area (European Commission, 2012). Evidence suggests that researchers who have moved internationally have a greater research impact than those who have not and that countries with more open research systems perform better in terms of innovation (DG Research and Innovation, 2014). It is therefore interesting to monitor the effort made by a MS/AC to recruit international researchers. The indicator is normalised by 1 000 researchers in the public sector in order to allow for a better comparison between MS/AC.

Computation method

Data needed

(RPA) Number of researcher's posts advertised through the EURAXESS job portal: **Unit=Total**;

(RES_{HES}) Researchers in the higher education sector: **Unit=Full time equivalent (FTE)**;

(RES_{GOV}) Researchers in the government sector: **Unit=Full time equivalent (FTE)**.

Source of data

For *(RPA)*: *the European Commission provided historical data from the EURAXESS portal;*

For *($RES_{HES\ or\ GOV}$)*: *Eurostat—Statistics on research and development (online data code rd_p_persocc).*

Filters applied

For *(RES_{HES})*:
SECTPERF set to "Higher education sector"
OCCUP set to "Researchers"
SEX set to "Total"
UNIT set to "Full-time equivalent (FTE)"

For *(RES_{GOV})*:
SECTPERF set to "Government sector"
OCCUP set to "Researchers"

[1] http://ec.europa.eu/euraxess/

SEX set to "Total"
UNIT set to "Full-time equivalent (FTE)"

Specifications

$$Researcher's\ posts\ advertised\ through\ the\ EURAXESS\ job\ portal\ per\ 1\ 000\ researchers\ in\ the\ public\ sector = \frac{RPA}{(RES_{HES} + RES_{GOV})/1000}$$

Comments/critical issues

None identified.

2.3 European Commission, Joint Research Centre, Competence Centre on Composite Indicators and Scoreboards (JRC-COIN)

The JRC-COIN of the European Commission develops and implements various methodologies to produce composite indices summarising multi-dimensional phenomenon into simplified pictures. These simplified pictures convey key messages to decision makers on key European issues, thereby assisting the development of policies and the monitoring of progress towards key objectives ([2]).

2.3.1 P1 – Headline indicator – Adjusted Research Excellence Indicator (REI)

Definition of indicator

This indicator defines the research excellence of a country through a composite indicator integrating four components: share of top 10% most highly cited publications per total publications (data source: CWTS); PCT patent applications per population (OECD); European Research Council (ERC) grants per public R&D (DG-RTD, Eurostat, OECD) and participation in Marie Skłodowska-Curie fellowships (DG-EAC). Dates refer to actual data years, except for Marie Skłodowska-Curie fellowships. It was calculated using the latest available data as of April 2016 (i.e. 2013), taking into consideration the presence of a citation window for the highly cited publications indicator.

Rationale

This indicator pertains to priority 1 — that is, to more effective national systems. As one of the key types of actions promoted under the ERA to achieve this priority, it relates to the establishment of Research Performance Based Funding (RPBF) systems (i.e. systems applying the core principles of international peer review in grant competitions), and it becomes highly relevant to monitor the establishment of such systems and their impact on research excellence across ERA countries. The adjusted REI does this by integrating four dimensions of high relevance to monitor progress towards more effective national R&I systems, looking at both the funding mechanisms and the resulting R&I outputs (Vértesy, 2015). It covers ERC grants per public R&D, which is a good proxy to appreciate the success of countries in securing ERA-wide project-based competitive funding. It covers participation in Marie Skłodowska-Curie fellowships, which is a good proxy to appreciate the extent of researcher exchanges across national, sectoral and disciplinary boundaries (regardless of career stage), which are themselves expected to foster more integrated and efficient R&I ecosystems. It covers PCT patent applications per population, which is a good output indicator to capture the inventiveness of national R&I systems. Finally, it covers the share of top 10% most highly cited publications per total publications, which is a good proxy of the excellence of the research output of a nation.

Computation method

Data needed

For details on the methodology, please refer to Vértesy (2015).

[2] https://ec.europa.eu/jrc/en/coin

Source of data

Calculations by European Commission, DG Joint Research Centre, Competence Centre on Composite Indicators and Scoreboards (JRC-COIN).

Specifications

For details on the methodology, please refer to Vértesy (2015).

Comments/critical issues

See the 'Limitations' sub-section under the 'Composite indicators' section for a discussion of the limits of composite indicators such as the adjusted REI.

2.4 European Innovation Scoreboard

Formerly called the Innovation Union Scoreboard, the European Innovation Scoreboard provides an international benchmark of the innovation performance of ERA countries, taking account of the multi-faceted nature of innovation (http://ec.europa.eu/growth/industry/innovation/facts-figures/scoreboards_en).

2.4.1 P1 – EMM indicator – European Innovation Scoreboard Summary Innovation Index (SII)

Definition of indicator

Formerly called the Innovation Union Scoreboard indicator, this composite indicator is produced every year by the European Commission to benchmark MS/AC, accounting for a wide spectrum of innovation indicators.

Rationale

This indicator pertains to priority 1 (More effective national research systems). It integrates a multitude of indicators distributed across eight dimensions covering R&I enablers (Human resources; Open, excellent and attractive research systems; and Finance and support), firm activities (Firm investments; Linkages & entrepreneurship; and Intellectual assets) and R&I outputs (Innovators and Economic effects). It thus presents a comprehensive picture of the state of a country's R&I system along the full path from inputs, through outputs, and on to outcomes/impacts. It is therefore highly relevant to monitor progress towards more effective national research systems under priority 1.

Computation method

This indicator includes a variety of innovation variables from different sources. The reader is referred to the latest Innovation Union Scoreboard report (DG Internal Market, Industry, Entrepreneurship and SMEs, 2016) for more details.

Source of data

DG Internal Market, Industry, Entrepreneurship and SMEs.

Comments/critical issues

The very broad set of indicators (25 in total) covered by this indicator mean that it covers a broader set of issues than those specific to priority 1; in fact, some of the indicators included in the SII are also EMM indicators in other ERA priorities (e.g. public-private co-publications per million population (Priority 4) and non-EU doctorate students as a percentage of all doctorate students (Priority 6)). See the 'Limitations' sub-section under the 'Composite indicators' section for a discussion of the limits of composite indicators such as the SII.

2.4.2 P5a – EMM indicator – Number of public-private co-publications per million population

Definition of indicator

This indicator is the number of publications co-authored by at least one researcher from the public sector and one researcher from the private sector per capita, according to their affiliation address and by full counting (refer to Annex 2 for a definition of full counting).

Rationale

This indicator pertains to sub-priority 5a (optimal access to and circulation and transfer of scientific knowledge) and relates to open innovation and knowledge transfer between the public and private sectors. In order to align needed skills with training, Member States have put in place measures, such as joint programmes and research training in private companies, to stimulate the partnership between universities, research institutions and the private sector (DG Research and Innovation, 2015). This indicator may serve as a proxy to measure the level of sharing of scientific knowledge and the level of collaboration between research institutions, scientists and businesses.

Computation method

Data needed

($PAP_{PUB\text{-}PRIV.COLLAB}$) Number of co-publications between the public and the private sectors: **Unit=Total (full counting)**;

(POP) Total population: **Unit=Total**.

Source of data

Centre for Science and Technology Studies at Leiden University (CWTS).

Comments/critical issues

The data were provided by CWTS for the period 2009-2013 only.

2.5 Eurostat

Eurostat is the official statistical office of the European Union. It provides reliable and objective statistics on European Member States, Associated Countries and candidate countries, allowing for comparison at the country and regional level. Most statistics are freely available online through the Eurostat website. The statistics produced at Eurostat cover a wide range of topics divided among nine primary categories: General & regional statistics, Economy & finance, Population & social conditions, Industry, trade & services, Agriculture and fisheries, External trade, Transport, Environment & energy, and Science & technology.

2.5.1 P3 – EMM indicator – Share of doctoral candidates with a citizenship of another EU Member State

Definition of indicator

This indicator is the proportion of doctoral candidates with a citizenship of another Member State to the total number of doctoral candidates in a given country.

Rationale

This indicator pertains to priority 3 (open labour market for researchers). This priority seeks to improve framework conditions for researcher mobility across Europe in order to retain highly skilled Europeans rather than have them pursue career goals in other competitive economies (Science-Metrix, 2016). Actions are expected from Member States to expand structured doctoral training programmes and remove barriers for cross-border mobility. This indicator can act as a proxy to monitor the extent to which a country's academic system is open to other European doctoral candidates (the openness may be in the portability of a national grant or other mechanism that may facilitate the switch to a new country academic institution). By promoting an

open academic system, a MS/AC can attract and retain skilled students who will eventually contribute to the R&I workforce either in academia or the industrial sector.

Computation method

Data needed

($PhD_{F.EUR}$) Number of foreign European doctoral candidates: **Unit= Number**;

(PhD) Number of doctoral candidates: **Unit= Number**.

Source of data

For *($PhD_{F.EUR}$)*: *Eurostat—Learning mobility (online data code: educ_uoe_mobs02);*

For *(PhD)*: *Eurostat—Participation in education and training (online data code: educ_uoe_enrt01).*

Filters applied

For *($PhD_{F.EUR}$)*:
UNIT set to "Number"
PARTNER set to "Europe"
SEX set to "Total"
ISCED11 set to "Doctoral or equivalent level"

For *(PhD)*:
UNIT set to "Number"
WORKTIME set to "Total"
ISCED11 set to "Doctoral or equivalent level"
SEX set to "Total"
SECTOR set to "Total"

Specifications

$$Share\ (\%)\ of\ doctoral\ candidates\ with\ a\ citizenship\ of\ another\ EU\ Member\ State = \frac{PhD_{F.Eur}}{PhD} \times 100$$

Comments/critical issues

Data for this indicator is only available for 2013 (academic year 2012-2013) and 2014 (academic year 2013-2014).

ISCED

The International Standard Classification of Education (ISCED11) categorises education programmes by level. ISCED level 8 covers:

- Programmes leading to advanced research qualifications devoted to advanced study and original research. Includes academic and professional Doctoral programmes and can be referred to as 'Doctoral or equivalent' for international comparison purposes.

2.5.2 P6 – EMM indicator – Non-EU doctorate students as a share of all doctorate students

Definition of indicator

This indicator is the proportion of non-European doctoral students to the total number of doctoral students in a given country.

Rationale

This indicator pertains to priority 6 (international cooperation) identified by the Commission for reinforcing the European Research Area. By attracting outstanding researchers from international locations, the EU will improve its capacity to address grand challenges and increase its competitiveness. Enrolling international students represents the first step toward this goal. However, approaches to increasing international collaboration vary from MS to MS and are

uncoordinated. As such, it is interesting to monitor the openness and attractiveness of each country's education system and research institutions with this indicator.

Computation method

Data needed

(PhD_{NON-EU}) Non-EU doctorate students: **Unit= Number**;

(PhD_{TOT}) Total EU doctorate student: **Unit= Number**.

Source of data

For *(PhD_{NON-EU})*: *European Innovation Scoreboard 2015 – Database (indicator 1.2.3)*

For *(PhD_{TOT})*: *European Innovation Scoreboard 2015 – Database (indicator 1.2.3)*

Specifications

$$Non-EU\ doctorate\ students\ as\ a\ share\ (\%)\ of\ all\ doctorate\ students = \frac{PhD_{NON\_EU}}{PhD_{TOT}} \times 100$$

Students and graduates

The International Standard Classification of Education (ISCED-97) categorises education programmes by level. ISCED-97 Level 6 (also referred to as ISCED 6) covers:

- The second stage, which leads to the award of an advanced research qualification (e.g. PhD, non-PhD programmes with an advanced research component, etc.). The programmes are devoted to advanced study and original research **(ISCED 6)**.

Eurostat also makes use of a 'direct PhD' code, which includes only those graduates pursuing PhD programmes (excluding those pursuing non-PhD programmes with an advanced research component). This indicator is calculated using either the direct PhD code or the more general ISCED Level 6. In most countries, the number of graduates at ISCED 6 and PhD level is the same.

Comments/critical issues

There were some changes to the ISCED in 2011. For more explanation, see Annex 1, which presents recent changes to international classification standards. At the time of producing the ERA Monitoring report, the European Innovation Scoreboard 2016 was not yet released, so the values for this indicator were taken from the 2015 edition. The time series in the 2015 edition covers the 2005-2012 period and uses ISCED-97. Eurostat now uses ISCED 2011 for reporting educational data; however, the new Eurostat time series covers a very limited number of recent years.

2.5.3 P1 – EMM indicator – GBARD as a percentage of GDP

Definition of indicator

This indicator is the government budget allocations for R&D (GBARD) divided by the gross domestic product (GDP) of a given country. GBARD represents budget provisions and not actual spending.

Rationale

This indicator pertains to priority 1 (More effective national research systems). Public funding for R&D is a key driver, if not *the* key driver, of the strength and international competitiveness of domestic R&I systems. It is therefore highly relevant to consider input indicators of R&D investments to contextualise the strength of national R&I systems as measured by output indicators. The GBARD covers all government financed R&D (including government financed R&D performed in business enterprise, private non-profit or HES sectors), giving a complete representation of government investment in R&D. Normalising by GDP accounts for the relative size of a country's economy and allows comparing the scores across MS/AC.

Computation method

Data needed

(GBARD) Government budget allocations for R&D: **Unit=Euro**;

(GDP) Gross domestic product at market price: **Unit=Euro**.

Source of data

For *(GBARD)*: *Eurostat— Statistics on research and development (online data code: gba_nabsfin07);*

For *(GDP)*: *Eurostat—Annual national accounts (online data code: nama_10_gdp).*

Filters applied

For *(GBARD)*: *NABS07 set to "Total R&D appropriations"*
UNIT set to "Million euro"

For *(GDP)*: *UNIT set to "Current prices, million euro"*
NA_ITEM set to "Gross domestic product at market prices"

Specifications

$$\text{GBARD as a percentage of GDP} = \frac{GBAORD}{GDP} \times 100$$

Comments/critical issues

None identified.

2.5.4 P1 – Additional indicator – GBARD as a percentage of government expenditures

Definition of indicator

This indicator is the proportion of the government budget allocations for R&D (GBARD) to total government expenditures. GBARD represents budget provisions and not actual spending.

Rationale

In line with GBARD as a percentage of GDP, this indicator pertains to priority 1 (More effective national research systems). The same rationale holds. The main difference is in the way this indicator accounts for the relative size of a country. It does so by normalising a country's GBARD to its total governmental expenditures.

Computation method

Data needed

(GBARD) Government budget allocations for R&D: **Unit=Euro**;

(GOV_{EXP}) Total government expenditure: **Unit=Euro**.

Source of data

For *(GBARD)*: *Eurostat— Statistics on research and development (online data code: gba_nabsfin07);*

For *(GOV_{EXP})*: *Eurostat—Annual government finance statistics (online data code: gov_10a_main).*

Filters applied

For *(GBARD)*: *NABS07 set to "Total R&D appropriations"*
UNIT set to "Million euro"

For *(GOV_{EXP})*: *UNIT set to "Million euro"*
SECTOR set to "General government"
NA_ITEM set to "Total general government expenditure"

Specifications

$$GBARD\ as\ a\ percentage\ of\ government\ expenditures = \frac{GBAORD}{GOV_{Exp}} \times 100$$

Comments/critical issues

None identified.

2.5.5 P1 – Additional indicator – Percentage of GBARD allocated as project based funding

Definition of indicator

The indicator is the proportion of government budget allocations for R&D (GBARD) allocated as project-based funding to the total GBARD of a given country. GBARD represents budget provisions and not actual spending.

Rationale

In line with the goal of improving the effectiveness of national research systems (priority 1), the European Commission has called on Member States to emphasise open and competitive funding of research through calls for proposals applying the core principles of international peer review (European Commission, 2012). This indicator represents the proportion of a country's GBARD devoted to the funding of specific projects. Projects are typically conducted by individuals or groups performing an R&D activity limited in scope, budget and time, normally requiring the submission of a project proposal outlining the research activity to be conducted. A higher share of GBARD allocated to project-based R&D activities would indicate a higher degree of commitment to a competitive funding process, which in turn should lead to increased value from public money invested in R&D.

Computation method

Data needed

($GBARD_{PF}$) Amount of GBARD allocated as project based funding: **Unit=Euro**;

(GBARD) Total amount of GBARD: **Unit=Euro**.

Source of data

For *($GBARD_{PF}$)*: *Eurostat—Statistics on research and development (online data code: gba_fundmod);*

For *(GBARD)*: *Eurostat—Statistics on research and development (online data code: gba_nabsfin07).*

Filters applied

For *($GBARD_{PF}$)*: *NABS07 set to "GBARD - project funding"*
UNIT set to "Million euro"

For *(GBARD)*: *NABS07 set to "Total R&D appropriations"*
UNIT set to "Million euro"

Specifications

$$Share\ (\%)\ of\ GBOARD\ allocated\ as\ project\ based\ funding\ = \frac{GBARD_{PF}}{GBARD} \times 100$$

Comments/critical issues

($GBARD_{PF}$) is only available for 14 countries (Czech Republic, Germany, Ireland, Greece, France, Luxembourg, Netherlands, Austria (2011-2012 only), Portugal, Slovakia (2012-2013 only), Iceland, Norway, Switzerland (2010 and 2012 only) and Turkey).

2.5.6 P2a – Headline indicator – GBARD allocated to transnational cooperation per researcher in the public sector.

Definition of indicator

This indicator is the government budget allocations for R&D (GBARD) allocated to transnational cooperation normalised by the number of researchers from the public sector. Transnational coordinated R&D contains GBARD allocated to Europe-wide, bilateral or multilateral transnational public R&D programmes and GBARD allocated to transnational public R&D performers. However, for this indicator, only the GBARD allocated to Europe-wide transnational public R&D programmes and the GBARD allocated to bilateral or multilateral public R&D programmes are taken into account. This is because these two address cooperation through programmes, while the third sub-category (GBARD allocated to transnational public R&D performers) does not involve joint programming and therefore does not contribute to ERA sub-priority 2a (implementing joint research agendas).

Rationale

This indicator pertains to sub-priority 2a — that is, implementing joint research agendas to address grand challenges of high importance to Europeans. It reflects a given country's emphasis on collaboration and sharing of experiences in R&D across borders, whether national, regional or organisational. Europe-wide transnational public R&D programmes include R&D programmes that involve the flow of funds across borders for research purposes, as well as those that include transnational cooperation. Bilateral or multilateral public R&D programmes comprise non-European Commission funded R&D research conducted jointly by at least two Member State governments, involving either the flow of funds or transnational cooperation. Thus, this indicator is a good proxy to measure government support to transnational collaborations across the ERA.

Computation method

Data needed

($GBARD_{TRANS(EU\text{-}WIDE)}$)	Amount of GBARD allocated to Europe-wide transnational public R&D programmes: **Unit=Euro**;
($GBARD_{TRANS(BI\text{-}MULTI)}$)	Amount of GBARD allocated to bilateral or multilateral public R&D programmes: **Unit=Euro**;
(RES_{HES})	Researchers in the higher education sector: **Unit=Full time equivalent (FTE)**.
(RES_{GOV})	Researchers in the government sector: **Unit=Full time equivalent (FTE)**.

Source of data

For *($GBARD_{TRANS}$)*:	*Eurostat—Statistics on research and development (online data code: gba_tncoor);*
For *($RES_{HES\ or\ GOV}$)*:	*Eurostat—Statistics on research and development (online data code rd_p_persocc).*

Filters applied

For *($GBARD_{TRANS(EU\text{-}WIDE)}$)*:	*NABS07 set to "National contributions to Europe-wide transnational public R&D programmes"* *UNIT set to "Million euro"*

For *($GBARD_{TRANS(BI\text{-}MULTI)}$)*: *NABS07 set to* "National contributions to bilateral or multilateral public R&D programmes"
UNIT set to "Million euro"

For *(RES_{HES})*: *SECTPERF set to "Higher education sector"*
OCCUP set to "Researchers"
SEX set to "Total"
UNIT set to "Full-time equivalent (FTE)"

For *(RES_{GOV})*: *SECTPERF set to "Government sector"*
OCCUP set to "Researchers"
SEX set to "Total"
UNIT set to "Full-time equivalent (FTE)"

Specifications

$$GBARD\ allocated\ to\ transnational\ cooperation\ per\ researcher\ in\ the\ public\ sector$$

$$= \frac{GBOARD_{TRANS(EU-WIDE)} + GBOARD_{TRANS(BI-MULTI)}}{RES_{GOV} + RES_{HES}}$$

Comments/critical issues

None identified.

2.5.7 P5a – Headline indicator – Share of product and/or process innovative firms cooperating with higher education institutions or public/private research institutions

Due to the nature of the data source, this indicator was divided into two indicators: (a) Percentage of product or process innovative firms cooperating with public or private research institutes, and (b) Percentage of product or process innovative firms cooperating with universities or other higher education institutions. Because pre-aggregated data are provided separately for each of these two indicators (i.e. there is no pre-aggregated data combining both types of cooperation), and because the microdata are not available, it is impossible to determine how many firms are involved in both types of partnerships. In turn, summing the number of firms across these two types of partnerships would result in multiple double-ups of those companies collaborating with both public/private research institutes and with universities or other higher education institutions for their innovation activities.

Definition of indicator

(a) The indicator is the proportion of product and/or process innovative firms cooperating with government, public or private research institutes (PRIs) to the total number of product and/or process innovative firms.

(b) The indicator is the proportion of product and/or process innovative firms cooperating with universities or higher education institutes (HEIs) to the total number of product and/or process innovative firms.

Rationale

This indicator pertains to priority 5 (knowledge circulation) and helps to assess the potential for knowledge transfer and open innovation between the public and private sectors within a given country. A higher rate of private firm engagement with HEIs or PRIs should better facilitate the transfer of research results to the market in line with the goal of optimising circulation of, access to and transfer of scientific knowledge established by the European Commission (2012). This indicator represents the degree of cooperation between private industry and other sectors and can be used as a proxy for the willingness of private firms to collaborate with higher education and/or public/private research institutes and the potential for knowledge transfer.

Computation method

Data needed

($COOP_{HEI\ OR\ PRI}$) Product and/or process innovative firms cooperating with HEIs or PRIs: **Unit=Total**;

(FIRM) Total number of product and/or process innovative firms: **Unit=Total**.

Source of data

For *($COOP_{HEI\ OR\ PRI}$)*: *Eurostat—Community innovation survey (online data codes: inn_cis8_coop, inn_cis7_coop, inn_cis6_coop);*

For *(FIRM)*: *Eurostat—Community innovation survey (online data codes: inn_cis8_type, inn_cis7_type, inn_cis6_type).*

Filters applied

For *($COOP_{HEI\ OR\ PRI}$)*:
SIZECLAS set to "Total"
NACE_R2 set to "Innovation core activities (Com.Reg. 1450/2004)"
TYPE_INN set to "Product and/or process innovative enterprises, regardless of organisational or marketing innovation (including enterprises with abandoned/suspended or on-going innovation activities"
INDIC_IN set to "Enterprises co-operating with universities or other higher education institutions" or "Enterprises co-operating with Government, public or private research institutes"
UNIT set to "Number"
TIME set to "2012", "2010" or "2008" (depending on which table is used)

For *(FIRM)*:
SIZECLAS set to "Total"
NACE_R2 set to "Innovation core activities (Com.Reg. 1450/2004)"
TYPE_INN set to "Product and/or process innovative enterprises, regardless of organisational or marketing innovation (including enterprises with abandoned/suspended or on-going innovation activities"
INDIC_IN set to "Total number of enterprises in the population in 2012" (change 2012 for 2010 or 2008 depending on the table)
UNIT set to "Number"
TIME set to "2012", "2010" or "2008" (depending on which table is used)

Specifications

$$Share\ (\%)\ of\ innovative\ firms\ cooperating\ with\ HEI = \frac{COOP_{HEI}}{FIRM} \times 100$$

$$Share\ (\%) of\ innovative\ firms\ cooperating\ with\ PRI = \frac{COOP_{PRI}}{FIRM} \times 100$$

Comments/critical issues

To maintain consistency across time, the filter NACE_R2 was set to "Innovation core activities (Com.Reg. 1450/2004)" for all years (we are aware that the data from cis8 in 2012 contains a more up-to-date version of the NACE_R2 innovation category but it was not available in the previous editions).

Note that the definition of the type of collaboration (INDIC_IN) slightly changed from cis7 to cis8. The former version was 'Enterprises co-operating with Government or public research institutes', while the newest definition is 'Enterprises co-operating with Government, public or private research institutes'. A flag (i.e. definition differs) was added to the cis8 data to make sure the reader is aware of this change.

2.5.8 P5a – EMM indicator – Share of public research financed by the private sector

Definition of indicator

This indicator is the share of the total amount of research funds allocated to the public sector from all sources coming from the private sector.

Rationale

This indicator pertains to priority 5 (Knowledge circulation) and relates to the open innovation and knowledge transfer between the public and private sectors. In Europe, the private sector employs relatively few researchers. Young graduates have little experience outside academic circles and often lack the skills to pursue a career in the private sector (DG Research and Innovation, 2015). Enterprises are encouraged to fund research in the public sector to align their needs with academic training and facilitate the transition of young graduates to the job market. This indicator can be used as a proxy for cooperation and knowledge transfer between the public and private sectors.

Computation method

Data needed

($F_{HEI-BES}$) Amount of research funds allocated to the higher education sector by the business enterprise sector: **Unit=Euro**;

($F_{HEI-ABES}$) Amount of research funds allocated to the higher education sector by the abroad business enterprise sector: **Unit=Euro**;

($F_{GOV-BES}$) Amount of research funds allocated to the government sector by the business enterprise sector: **Unit=Euro**;

($F_{GOV-ABES}$) Amount of research funds allocated to the government sector by the abroad business enterprise sector: **Unit=Euro**;

(F_{HEI}) Amount of research funds allocated to the higher education sector by all sectors: **Unit=Euro**.

(F_{GOV}) Amount of research funds allocated to the government sector by all sectors: **Unit=Euro**.

Source of data

Eurostat—Statistics on research and development (online data code rd_e_gerdfund).

Filters applied

($F_{HEI-BES}$) *SECTPERF set to "Higher education sector"*
SECTFUND set to "Business enterprise sector"
UNIT set to "Million euro"

($F_{HEI-ABES}$) *SECTPERF set to "Higher education sector"*
SECTFUND set to "Abroad - Business enterprise sector"
UNIT set to "Million euro"

($F_{GOV-BES}$) *SECTPERF set to "Government sector"*
SECTFUND set to "Business enterprise sector"
UNIT set to "Million euro"

($F_{GOV-ABES}$) *SECTPERF set to "Government sector"*
SECTFUND set to "Abroad - Business enterprise sector"
UNIT set to "Million euro"

($F_{HEI-TOT}$) *SECTPERF set to "Higher education sector"*
SECTFUND set to "All sector"
UNIT set to "Million euro"

($F_{GOV-TOT}$) *SECTPERF set to "Government sector"*
SECTFUND set to "All sector"
UNIT set to "Million euro"

Specifications

$$Share\ (\%)\ of\ public\ research\ financed\ by\ the\ private\ sector$$

$$= \frac{F_{HEI-BES} + F_{HEI-ABES} + F_{GOV-BES} + F_{GOV-ABES}}{F_{HEI-TOT} + F_{GOV-TOT}} \times 100$$

Comments/critical issues

None identified.

2.5.9 P1 – Additional indicator – Researchers per 1 000 active population

Definition of indicator

This indicator measures the total number of researchers of a country in all sectors relative to the total size of its active population.

Rationale

This indicator relates to the goal of encouraging more effective national research systems (priority 1), identified by the European Commission for the reinforcement of the European Research Area (European Commission, 2012). The number of researchers in a given country proportional to the total active population is a relative measure of a given country's involvement at an early phase of the R&D cycle (i.e. input stage) and can be used as a proxy for comparing the strength of national R&I systems. Researchers are defined as 'professionals engaged in the conception or creation of new knowledge, products, processes, methods and systems and also in the management of the projects concerned' (OECD, 2002). Because it is normalised to the total active population of a country, it facilitates cross-country comparisons.

Computation method

Data needed

(RES) Total number of researchers: **Unit=Full time equivalent (FTE)**;

(POP_{ACT}) Total active population: **Unit=Total**.

Source of data

For *(RES)*: *Eurostat—Statistics on research and development (online data code rd_p_persocc);*

For *(POP_{ACT})*: *Eurostat—Employment and unemployment (Labour Force Survey) (online data code lfsi_emp_a).*

Filters applied

For *(RES)*: *SECTPERF set to "All sectors"*
OCCUP set to "Researchers"
SEX set to "Total"
UNIT set to "Full-time equivalent (FTE)"

For *(POP_{ACT})*: *AGE set to "From 15 to 64 years"*
UNIT set to "Thousand persons"
SEX set to "Total"
INDIC_EM set to "Active population"

Specifications

$$Number\ of\ researchers\ per\ 1\ 000\ active\ population = \frac{RES}{POP_{Act}}$$

Comments/critical issues

None identified.

2.5.10 P4 – Additional indicator – Share of women researchers

Definition of indicator

This indicator is the proportion of women researchers to the total number of researchers in all sectors of the economy. Some of the text below has been taken directly from the She Figures Handbook 2015 (DG Research and Innovation, 2016a).

Rationale

In the context of reinforcing the European Research Area, the European Commission has made it a priority to address the gender imbalance among researchers in Europe (European Commission, 2012). This indicator pertains to priority 4 (and relates to gender balance in career progression) by characterising the rate of participation and progression of women in public research systems. Women, while outnumbering men among tertiary education graduates, are still underrepresented in the research community, particularly among researchers in leadership or decision-making positions. The Commission has recognised that the better inclusion of women in the research and innovation workforce would lead to a more efficient use of highly skilled individuals and increase economic growth (DG Research and Innovation, 2012; European Commission, 2012).

Computation method

Data needed

(F) Number of women researchers in all sectors of the economy: **Unit=Head count;**

(T) Total number of researchers in all sectors of the economy: **Unit=Head count**.

Source of data

Eurostat—Statistics on research and development (online data code rd_p_persocc). The computation of this indicator is as specified in the *She Figures Handbook 2015* (DG Research and Innovation, 2016a).

Filters applied

For *(F)*: *SECTPERF set to "All sectors"*
OCCUP set to "Researchers"
SEX set to "Females"
UNIT set to "Head count"

For *(T)*: *SECTPERF set to "All sectors"*
OCCUP set to "Researchers"
SEX set to "Total"
UNIT set to "Head count"

Specifications

$$Share\ of\ women\ researchers = \frac{F}{T}\ \mathrm{x}\ 100$$

Researchers

The OECD's *Proposed Standard Practice for Surveys on Research and Experimental Development* (Frascati Manual, 2002) provides an international definition for researchers: 'Researchers are professionals engaged in the conception or creation of new knowledge, products, processes,

methods and systems and also in the management of the projects concerned' (§301, Frascati Manual, OECD, 2002).

Comments/critical issues

None identified.

2.5.11 P4 – EMM indicator – Share of female PhD graduates

Definition of indicator

This indicator is the proportion of women PhD graduates to the total number of PhD graduates. Some of the text below has been taken directly from the *She Figures Handbook 2015* (DG Research and Innovation, 2016a).

Rationale

This indicator pertains to priority 4 (and relates to gender balance in career progression) through measuring the rate of graduation of women from the highest level of tertiary education. The European Commission has noted that '[t]he persistence of gender bias in careers, of gender imbalance in decision-making roles, and the lack of a gender dimension in research programmes remain common challenges' (DG Research and Innovation, 2014). In light of this, a key priority for reinforcing the European Research Area is emphasising gender equality and gender mainstreaming in research (DG Education and Culture, 2011). This indicator aims to characterise the rate and progress of women's graduation from doctoral programmes.

Computation method

Data needed

(F) Number of women PhD graduates pursuing direct PhD programmes: **Unit=Total**;

(T) Total number of PhD graduates pursuing direct PhD programmes: **Unit=Total**.

Source of data

For *F* and *T*: *Eurostat – Education Statistics (online data code: educ_grad5).* The computation of this indicator is as specified in the She Figures Handbook 2015 (DG Research and Innovation, 2016a).

Specifications

$$Share\ of\ women\ PhD\ graduates = \frac{F}{T} \text{ x } 100$$

Students and graduates

The International Standard Classification of Education (ISCED-97) categorises education programmes by level. ISCED-97 Level 6 (also referred to as ISCED 6) covers:

- The second stage, which leads to the award of an advanced research qualification (e.g. PhD, non-PhD programmes with an advanced research component etc.). The programmes are devoted to advanced study and original research **(ISCED 6)**.

Eurostat also makes use of a 'direct PhD' code, which includes only those graduates pursuing PhD programmes (excluding those pursuing non-PhD programmes with an advanced research component). This indicator is calculated using either the direct PhD code or the more general ISCED Level 6. In most countries, the number of graduates at ISCED 6 and PhD level is the same.

The number of graduates refers to those graduating in the reference year and not to the number of graduates in the population.

The number of graduates also refers to non-nationals graduating in the country, but does not include nationals graduating abroad.

Comments/critical issues

There were some changes to the ISCED in 2011, although the She Figures continues to use the 1997 classifications. For more explanation, see Annex 1, which presents recent changes to international classification standards.

2.5.12 P6 – EMM indicator – Licence and patent revenues from abroad as a share of GDP

Definition of indicator

This indicator is the proportion of the amount of foreign revenue generated from licensing and patenting to the GDP.

Rationale

This indicator pertains to priority 6 (Openness of Member State/Associated Country (MS/AC) for international cooperation). Patents provide formal protection under the law for original inventions. Owners of patents are therefore entitled to royalties when their inventions are commercially marketed by other parties. As such, patents are a source of revenue originating from funding investments in research and development. The royalties and licences paid by foreign actors indicate that those actors use the technology developed by a European country. This can be seen as a proxy to monitor how much a country makes its technology available to the rest of the world and — indirectly — how open it is to international cooperation. It can also be viewed as a proxy for the strength of European R&I systems.

Computation method

Data needed

($FREV_{PAT}$) Foreign revenues generated from licensing and patenting activities: **Unit=1 000 million Euro**;

(GDP) Gross domestic product: **Unit=million Euro**.

Source of data

For *($FREV_{PAT}$)*: *Eurostat—International trade in services (online data code bop_its_ybk);*

For *(GDP)*: *Eurostat—Annual national accounts (online data code nama_10_gdp).*

Filters applied

For *($FREV_{PAT}$)*:
INDIC_BP set to "Exports (in 1 000 million ECU/EUR)"
PARTNER set to "Rest of the world"
POST set to "Current account, Services, Other services, Royalties and license fees"

For *(GDP)*:
UNIT set to "Current prices, million euro"
NA_ITEM set to "Gross domestic product at market prices"

Specifications

$$Licence\ and\ patent\ revenues\ from\ abroad\ per\ GDP\ (\%) = \frac{FREV_{Pat} \times 1000}{GDP} \times 100$$

Comments/critical issues

None identified.

2.6 First ERA-Learn 2020 Annual Report on P2P Partnerships

The mission of the ERA-Learn 2020 project — a new initiative started in January 2015 that builds upon previous ERA-NET projects — is to provide an integrated framework that will strengthen the community of P2P (public-to-public) partnerships and support national funding organisations (ERA-Learn, 2015).

2.6.1 P2a – EMM indicator – Member States participation in Public-to-public partnerships per researcher in the public sector

Definition of indicator

This indicator is the amount of combined funds committed to ERA-NET, Joint Programming Initiatives (JPI) and Article 185 initiatives relative to the number of researchers in the public sector.

Rationale

This indicator pertains to sub-priority 2a (transnational cooperation) and relates to the implementation of joint research agendas. ERA-NET projects, JPIs and Article 185 initiatives are all P2P partnerships. Article 185 is a reference to the Treaty on the Functioning of the European Union. Data on the amount of funds committed by Member States to these joint research programming efforts are captured in the 1st ERA-Learn 2020 Annual Report on P2P Partnerships (ERA-Learn, 2015) and can therefore be used to assess the state of play in regard to P2P partnerships in all Member States within the ERA.

Computation method

Data needed

(CBUDG) Committed budget to ERA-NET, JPIs and Article 185 initiatives: **Unit=euro**;

(RES_{HES}) Researchers in the higher education sector: **Unit=Full time equivalent (FTE)**;

(RES_{GOV}) Researchers in the government sector: **Unit=Full time equivalent (FTE)**.

Source of data

For *(*CBUDG*)*: *1st ERA-Learn 2020 Annual Report on P2P Partnerships (ERA-Learn, 2015).*

For *(*RES*)*: *Eurostat—Statistics on research and development (online data code: rd_p_persocc).*

Specifications

$$Participation\ in\ Public-to-public\ partnerships\ per\ researcher\ in\ the\ public\ sector = \frac{(CBUDG)}{(RES_{HES}) + (RES_{GOV})}$$

Comments/critical issues

Data are only available for Member States.

2.7 MORE2 survey

The MORE2 Higher Education Institutions (HEI) survey, conducted in spring 2012, collected data on mobility patterns, career paths and working conditions of researchers working in HEIs. It reached more than 10 000 institutions located in what are now the EU-28 and the Associated Countries (but were then the EU-27, the Associated Countries and the candidate countries). The survey was designed and implemented in order to offer maximum accuracy at the EU and individual country levels.

2.7.1 P3 – EMM indicator – Share of researchers expressing satisfaction that the hiring procedures in their institution are open, transparent and merit-based

Definition of indicator

This indicator represents the proportion of researchers having answered positively to the three following questions from the MORE2 survey:

(1) Are you satisfied with the extent to which research job vacancies are publicly advertised and made known by your institution?
(2) Do you think that the recruitment process at your home institution is sufficiently transparent?

(3) Do you think that recruitment at your institution is sufficiently merit-based?

Rationale

This indicator pertains to priority 3 (Open labour market for researchers) and relates to the perception that researchers have regarding their respective institution. Institutions with recruitment processes that are open, transparent and merit based are considered more attractive for researchers and may thus provide a better and more open labour market. Most attractive institutions also have the potential to interest researchers from abroad and contribute to international and inter-sectoral mobility, which is thought to boost the competitiveness of research systems.

Computation method

Data needed

(Np_i) Respondent having answered positively to the three questions on openness, transparency and merit-based recruitment procedures. **Unit=Head count**;

(Nn_i) Respondent having answered negatively to ***at least*** one of the three questions on openness, transparency and merit-based recruitment procedures. **Unit=Head count**;

(W_i) Sampling weight at the country level (provided in the raw data of the MORE2 survey).

Source of data

MORE2 Survey raw dataset

Specifications

$$Share\ (\%)\ of\ researchers\ \text{expressing satisfaction that the hiring procedures in their institution are open, transparent and merit based} = \frac{\sum(Np_i \times W_i)}{\sum(Nn_i \times W_i + Np_i \times W_i)}$$

Where the sums are taken over all the respondents *i* of a given Member State or Associated Country.

Comments/critical issues

Weighting procedures are described in the document 'Guidelines for the data analysis of the EU HEI survey data' (IDEA Consult, 2013). The sampling weight used was the one to be applied for the computation of results at country level (i.e. weight in the raw dataset). It serves to increase accuracy when aggregating results at this level.

Note that the respondents who did not answer to all three questions were left out of this analysis as it is impossible to know to which category they belong.

2.8 OECD—Science, Technology and Industry Scoreboard

The OECD Science, Technology and Industry Scoreboard is a report published by the Organization for Economic Cooperation and Development every two years. It provides a variety of indicators at country level used to assess the state of science, technology, innovation and industry (OECD, 2015b).

2.8.1 P1 – Additional indicator – R&D tax incentives as a proportion GBARD

Definition of indicator

This indicator represents the amount of R&D tax incentives relative to the government budget allocations for R&D (GBARD). GBARD represents budget provisions and not actual spending.

Rationale

Public funding for R&D activities is an essential element in sustaining and improving research output leading to more effective national research systems, a key priority (i.e. priority 1) identified by the European Commission (2012) for reinforcing the European Research Area. This input indicator is linked to the creation of more effective national research systems by fostering/boosting private investments in R&D activities. Tax rebates for R&D are a market-based means to incentivise private companies to engage in R&D and represent a government's willingness to forgo near-term revenue for future growth. This indicator measures the value of R&D tax incentives as a proportion of total GBARD.

Computation method

Data needed

(TAX) Indirect government support through R&D tax incentive as a percentage of GDP: **Unit=% GDP**;

(GDP) Gross domestic product at market price: **Unit=Euro**;

(GBARD) Government budget allocations for R&D: **Unit=Euro**.

Source of data

For *(TAX)*: *OECD—Science, Technology and Industry Scoreboard 2015 (section on R&D tax incentives link);*

For *(GDP)*: *Eurostat—Annual national accounts (online data code nama_10_gdp);*

For *(GBARD)*: *Eurostat—Government budget allocations for R&D (online data code gba_nabsfin07).*

Specifications

$$R\&D\ tax\ incentives\ as\ a\ proportion\ of\ GBARD = \frac{TAX\ \times GDP}{GBARD} \times 100$$

Comments/critical issues

The GDP and GBARD have to be matched on the reference year of the tax incentive, which varies from 2011 to 2013.

Notes on tax incentive data (OECD, 2015):

For France, Norway, Portugal, Spain and the United Kingdom, preliminary R&D tax incentive estimates are reported for 2013 (or closest year). Figures are rounded to the second decimal unless rounding would result in a value of zero.

For Belgium, Ireland, Israel, Spain and Switzerland figures refer to 2012. For Iceland figures refer to 2011.

Estimates of direct funding for Belgium, France, Italy and Portugal are based on imputing the share of direct government-funded BERD in the previous year to the current ratio of BERD to GDP. For Austria, the 2011 share is used for 2013.

In Austria, R&D tax incentive support is included in official estimates of direct government funding of business R&D. It is removed from direct funding estimates to avoid double counting.

Estonia, Germany, Luxembourg, Sweden and Switzerland did not provide information on expenditure-based R&D tax incentives for 2013. For Israel, the R&D component of incentives cannot be identified separately at present. No data on the cost of expenditure-based R&D tax incentive support are available for Poland.

Estimates do not cover sub-national and income-based R&D tax incentives and are limited to the business sector (excluding tax incentive support to individuals). Data refer to estimated initial revenue loss (foregone revenues) unless otherwise specified.

Estimates refer to the cost of incentives for business expenditures on R&D, both intramural and extramural, unless otherwise specified. Direct support figures refer only to intramural R&D expenditures.

Country specific notes:

Austria: Estimates, on a cash basis, refer to the refundable research premium.

Belgium: Estimates, on an accrual basis, include the R&D tax credit (for R&D capital) and payroll withholding tax credit for young innovative companies, private companies and partnership agreements with universities. They exclude the investment deduction for environmental projects as the R&D component cannot be identified.

Czech Republic: Estimates are on a cash basis.

Denmark: Estimates refer to the cost of accelerated depreciation of R&D capital and the R&D tax credit for deficit-related R&D expenditure, and exclude personal income incentives for research and key personnel.

Finland: Estimates refer to the R&D tax allowance for R&D labour costs, provided on an experimental basis over the tax years 2013 and 2014.

France: Estimates, on an accrual basis, refer to the crédit d'impôt recherche and special provisions for social security contributions by young and innovative firms (JEIs) and young university enterprises (JEU), but exclude the cost of accelerated depreciation incentives for capital R&D.

Greece: Estimates, on an accrual basis, refer to the R&D tax allowance.

Hungary: Estimates refer to the R&D tax allowance and the special provision for social security and vocational training contributions for researchers but exclude the local business tax allowance. No figures are available for the R&D component of the tax incentive for capital development.

Iceland: Estimates refer to the R&D tax credit providing a deduction of eligible R&D expenses from the income tax at an enhanced rate of 20 percent. No further details were provided.

Ireland: Estimates, on a cash basis, refer to the R&D tax credit on current, machinery and buildings expenditures.

Italy: The cash-based estimate, referring to fiscal year 2013, is based on corporate tax return data. The estimate refers to the R&D tax credit for SMEs providing a fixed payment for newly hired, qualified researchers and a volume-based credit for R&D collaborations with universities and public research consortia (Law 449/1997).

Netherlands: Estimates, on a cash basis, refer to the WBSO payroll tax credit for R&D labour and the R&D tax allowance (RDA) for non-labour related R&D expenditures.

Norway: Estimates for the fully refundable SKATTEFUNN R&D tax incentive cover current and machinery costs.

Poland: Estimates for the cost of accelerated depreciation provisions and tax deductions for R&D Centres are not available. New Technology Tax Relief scheme for the acquisition of intangible assets is excluded as it does not necessarily apply to R&D.

Portugal: Estimates, on an accrual basis, for the SIFIDE-II R&D tax credit which includes current and R&D-related capital expenditures.

Slovakia: Estimates, on an accrual basis, refer to the R&D tax allowance scheme, which is restricted to grant recipients (Tax relief for subsidy recipients: Income Tax Act S30b).

Slovenia: Estimates, on an accrual basis, refer to the R&D tax allowance scheme.

Spain: Estimates, based on the tax authorities' data on claims, refer to the R&D and innovation tax credit. Estimates include support for technological innovation. According to data from a non-random subset of firms (Informes Motivados), this accounts for more than 45% of all qualifying expenditures and nearly 20% of all deductions. Estimates do not include the cost of the accelerated depreciation provision for R&D capital and allowances for employers' social security contributions which was less than EUR 1 million when introduced in 2007.

Turkey: Estimates, on a cash basis, refer to deductions for current R&D and machinery expenditures in eligible R&D centres and companies (Law 5746) and to partial relief on social security contributions both in these and in firms based in Technoparks. Figures may include the cost of standard deductions for current R&D expenditures and may therefore overstate tax support in relation to other countries. Estimates for the cost of accelerated depreciation provisions are not available.

United Kingdom: Estimates for fiscal year 2013, on an accrual basis, refer to the Research & Development Relief for Corporation Tax and the Research and Development Expenditure Credit (RDEC) Scheme for large companies, introduced for expenditure incurred on or after 1 April 2013. Estimates for the cost of accelerated depreciation provisions are not available.

Israel: 'The statistical data for Israel are supplied by and under the responsibility of the relevant Israeli authorities or third party. The use of such data by the OECD is without prejudice to the status of the Golan Heights, East Jerusalem and Israeli settlements in the West Bank under the terms of international law'.

2.9 PATSTAT

As with scientific publications, which contain useful information to measure the scientific output of a nation, the bibliographic data on patents can be used to measure a country's level of technological innovation. The measurement of patent data, which is commonly referred to as technometrics, is often regarded as a good proxy for measuring the innovative capabilities of an entity — be it an individual, an institution or a country — as patents provide formal protection for new technologies, new processes or new products, which often result from research activities.

To ease the patent application process for inventors who wish to protect their inventions globally, the Patent Cooperation Treaty (PCT) allows them to file a single application for protection in the 148 countries of their choice that ratified the treaty. It is worth mentioning that the World Intellectual Property Organization (WIPO), responsible for managing PCT applications, does not grant patents. Through their PCT application process, inventors choose under which jurisdictions they wish to seek protection. Patents are ultimately delivered by individual offices. PCT applications have grown in popularity over recent decades given their inclusiveness and flexibility. From a statistical standpoint, data derived from PCT applications present a clear advantage over data derived from specific national offices. PCT data remove the 'home advantage' effect, whereby locals usually perform better than foreigners in their home market. Thus, the statistics produced in this study rely exclusively on PCT data.

A patent is usually assigned a group of 'assignees' and a group of 'inventors'. Assignees are individuals and/or legal institutions who filed for protection and legally own the patent. Inventors, which are strictly individuals, produced the invention but do not have ownership rights to the patent. An individual can be both an assignee and an inventor on the same patent. Upon review, by the patent examiner(s), of the patent applications, a patent may, or may not, be issued. Two types of indicators are usually produced with patent data. Indicators of Intellectual Property (IP) measure the ownership of inventions, and therefore typically rely on issued patents looking at the assignee field. Indicators on inventorships typically measure inventions themselves, and therefore usually rely on patent applications looking at the inventor field. Statistics on inventorships using patent applications have a clear advantage over issued patents for the analysis of timely data. Because a patent is usually issued a few years following the filing of an application, issued patents are running behind becoming visible only years after the innovative activity has taken place (DG Research and Innovation, 2016a).

The current study's patent indicators are relying exclusively on PCT applications (to minimise the 'home advantage') and the inventor field to monitor and investigate trends in the inventiveness of ERA countries. Since the European Patent Office (EPO) Worldwide Statistical Database (PATSTAT) covers, in addition to data from over 150 patenting offices worldwide, PCT data, it was selected for producing the following patent indicators. The following version of PATSTAT was used: Edition

2016 – Spring ([3]). To limit the analysis to the proper documents in PATSTAT, the following filter was applied throughout: appln_kind = 'W' (in table 201_appln).

2.9.1 P1 – Additional indicator – Number of patent applications per 1 000 researchers

Definition of indicator

This indicator is the ratio of the number of PCT patent applications, based on the inventor field, to the number of researchers (in thousands) in all sectors using fractional counting (refer to Annex 2 for a definition of fractional counting).

Rationale

This indicator pertains to priority 1 — that is, to more effective national systems. Because one of the key types of actions promoted under the ERA to achieve this priority relates to the establishment of Research Performance Based Funding (RPBF) systems (i.e. systems applying the core principles of international peer review in grant competitions), it becomes highly relevant to monitor the impacts such changes have generated, or will generate, on the inventiveness of countries. One such impact relates to the size of a country's technological production (i.e. the number of inventions it produced). However, such a metric is not internationally comparable due to differences in the size of countries. Perhaps more important is the impact of RPBF systems on the technological productivity of countries, which is a key dimension in measuring the efficiency with which national R&I systems convert R&D inputs into R&D outputs. Additionally, technological productivity metrics have the advantage of accounting for differences in country size. In this study, the number of patent applications per 1 000 researchers in all sectors is used as a proxy for measuring the technological productivity of countries.

Computation method

Data needed

(PAT) Number of PCT patent applications: **Unit=Total (fractional count)**;

(RES) Researchers in all sectors: **Unit=Full time equivalent (FTE)**.

Source of data

For *(PAT)*: *PATSTAT (Edition 2016 – Spring);*

For *(RES)*: *Eurostat—Statistics on research and development (online data code rd_p_persocc).*

Filters applied

For *(RES):*

SECTPERF set to "All sectors"
OCCUP set to "Researchers"
SEX set to "Total"
UNIT set to "Full-time equivalent (FTE)"

Specifications

$$Patent\ applications\ per\ 1\,000\ researchers = \frac{PAT}{\left(\frac{RES}{1\,000}\right)}$$

Comments/critical issues

The yearly data are provided based on the priority date of PCT applications. Although the data only go up to 2013 using this approach, instead of 2014 with the application date, the trend is more robust with the former. For instance, using the application date there is a quasi-systematic drop in the number of PCT applications in 2014. Note that like many other productivity metrics,

[3] https://forms.epo.org/service-support/ordering/patstat-order-form.html

this indicator comes with limitations in that the typical production size of researchers varies across technology domains. As such, differences in the specialisation patterns of countries reduce its cross-country comparability when used as a proxy for productivity.

2.9.2 P2a – Additional indicator – International co-invention rate with ERA partners

Definition of indicator

The rate is defined as the number of PCT patent applications co-invented by at least one inventor from a given ERA country (or region within the ERA) with at least another co-inventor from another ERA country, proportional to the total number of PCT patent applications in the given country (or region). Full counting is used (refer to Annex 2 for a definition of full counting). Note that the data are produced based on the priority date of patent applications.

Rationale

This indicator pertains to sub-priority 2a, which relates to the implementation of joint research agendas within the ERA. As with the international co-publication rate with ERA partners, this indicator measures the extent of transnational cooperation in innovation within the ERA. It is therefore a good proxy to measure the outcomes resulting from transnationally allocated research funding promoted under sub-priority 2a.

Computation method

Data needed

($PAT_{ERA.COLLAB}$) Number of PCT co-applications with another ERA country based on the inventor field: **Unit=Total (full counting)**;

(PAT) Number of PCT applications based on the inventor field: **Unit=Total (full counting)**.

Source of data

Computed using PATSTAT (Edition 2016 – Spring).

Specifications

$$International\ co-invention\ rate\ with\ ERA\ partners = \frac{PAT_{ERA.Collab}}{PAT}$$

Comments/critical issues

The yearly data are provided based on the priority date of PCT applications. Although the data only go up to 2013 using this approach, instead of 2014 with the application date, the trend is more robust with the former. For instance, using the application date there is a quasi-systematic drop in the number of PCT applications in 2014. Also note that this indicator is characterised by strong yearly fluctuations, especially for the smaller countries that have few PCT patent applications. This makes the analysis of trends difficult in the short term. To circumvent this issue in the analysis of growth, as well as to maximise the coverage of countries, a three-year rolling window (or three-year moving average of the scores) was applied in presenting the data.

2.9.3 P6 – Additional indicator – International co-invention rate with non-ERA partners

Definition of indicator

The rate is defined as the number of PCT patent applications co-invented by at least one inventor from a given ERA country (or region within the ERA) with at least another co-inventor from a non-ERA country, to the total number of PCT patent applications in the given country (or region). Full counting is used (refer to Annex 2 for a definition of full counting). Note that the data are produced based on the priority date of patent applications.

Rationale

This indicator pertains to priority 6, which promotes the openness of MS/AC for international cooperation beyond the ERA. As with the international co-publication rate with non-ERA partners, this indicator measures the extent of international cooperation in innovation beyond the ERA. It is therefore a good proxy to measure the outcomes resulting from actions designed to achieve this priority.

Computation method

Data needed

($PAT_{NON\text{-}ERA.COLLAB}$) Number of PCT co-applications with a non-ERA country based on the inventor field: **Unit=Total (full counting)**;

(PAT) Number of PCT applications based on the inventor field: **Unit=Total (full counting)**.

Source of data

Computed using PATSTAT (Edition 2016 – Spring).

Specifications

$$International\ co-invention\ rate\ with\ non-ERA\ partners = \frac{PAT_{NON-ERA.Collab}}{PAT}$$

Comments/critical issues

The yearly data are provided based on the priority date of PCT applications. Although the data only go up to 2013 using this approach, instead of 2014 with the application date, the trend is more robust with the former. For instance, using the application date there is a quasi-systematic drop in the number of PCT applications in 2014. Also note that this indicator is characterised by strong yearly fluctuations, especially for the smaller countries that have few PCT patent applications. This makes the analysis of trends difficult in the short term. To circumvent this issue in the analysis of growth, as well as to maximise the coverage of countries, a three-year rolling window (or three-year moving average of the scores) was applied in presenting the data.

2.10 JRC Research and Innovation Observatory Policy Repositories

The European Commission recently launched the Joint Research Centre (JRC) Research and Innovation Observatory (RIO) initiative, which aims to 'monitor and analyse research and innovation developments at Country and EU levels to support better policy making in Europe' ([4]).

2.10.1 P5b – EMM indicator – Open access policies in national action plans

Definition of indicator

Originally, this indicator was designed to identify the presence or absence of a national open access (OA) policy or policies in National Action Plans (NAPs). Following a preliminary analysis of NAPs, it appeared difficult to adequately characterise these policies — that is, to count the number of such policies for OA to research data on the one hand, and to research publications on the other hand, as well as to identify the years of adoption of these policies. Additionally, the structure of NAPs varies substantially across countries and the absence of a common reporting structure for an OA policy or policies in these documents makes it such that the indicator, if it relied on this source, would have limited cross-country comparability. Consequently, this indicator was constructed relying on the identification of national policies on OA in the RIO policy repositories.

[4] https://rio.jrc.ec.europa.eu/

Rationale

This output indicator pertains to sub-priority 5b (optimal access to and circulation and transfer of scientific knowledge) and relates to the open access of publications and data resulting from publicly funded research. Open access articles are accessible online without restrictions to the public (DG Research and Innovation, 2015). By hosting their research results (articles or simply data) on an open access domain, researchers facilitate the dissemination of knowledge by allowing anyone, anywhere, to benefit from their research. This can ultimately lead to more efficient science. Moreover, it has been argued that research financed by public funds should be made available in open access so that it benefits the population. This indicator investigates the presence of national policies designed to foster greater accessibility to scientific data and/or publications via open access.

Computation method

Data needed

Qualitative information contained in the RIO policy repositories.

Source of data

The Joint Research Centre (JRC) Research and Innovation Observatory (RIO) policy repositories (https://rio.jrc.ec.europa.eu/)

Specifications

The qualitative information on OA policies contained in the RIO policy repositories was codified according to two categories: OA policies for research data and OA policies for research publications. For each category, the number of policies is determined, as is the year range over which these policies were adopted.

Comments/critical issues

Data are only available for Member States.

2.11 She Figures

The She Figures provide statistics on the state and the progression of gender equality in science and technology (S&T) across Europe. Women are still well under-represented in S&T, but even more so in decision-making positions in research institutions. In an effort to have a more balanced gender representation in scientific research, such statistics provide information on the progress to correct this gender bias. Some of the text below has been taken directly from the *She Figures Handbook 2015* (DG Research and Innovation, 2016a).

2.11.1 P4 – Additional indicator – Share of women heads of institutions in the Higher Education Sector

Definition of indicator

This indicator is the proportion of women heads of institutions in the Higher Education Sector (HES) to the total number of heads of institutions in the HES.

Rationale

This indicator pertains to priority 4 (gender equality in research) and relates to gender balance in decision-making processes. A well-balanced gender representation in heads of higher education institutions ensures the relevance of research activities and maximises the full potential of human resources. Specific laws and national strategies on gender equality in public research have been adopted by a number of Member States, but progress is slow and uneven between countries. While gender balance was mostly achieved among PhD students and graduates at the EU-28 level in 2013, women continue to be under-represented in the highest grades of a typical academic career (DG Research and Innovation, 2016b). Leaders of universities and research institutions also tend to be dominated by male figures, who are instrumental in guiding decision-making in research across Europe. According to the *She Figures 2015* publication, on average among Member States in 2014, women occupied 20 % of organisation heads in the HES (DG Research and Innovation, 2016b). This indicator therefore acts as a proxy for the extent to which national

public research systems manage to ensure career progression for women — up to the decision-making sphere (DG Research and Innovation, 2016b) — as well as to track the progress made over time to correct a gender bias in decision-making processes.

Computation method

Data needed

($HEAD_{FEM}$) Number of women heads of institutions in the higher education sector: **Unit=Head count**;

($HEAD_{MALE}$) Number of men heads of institutions: **Unit=Head count**.

Source of data

DG Research and Innovation—WiS—Women in Science database.

Specifications

$$Share\ of\ women\ heads\ of\ institution\ in\ the\ HES = \frac{HEAD_{Fem}}{HEAD_{Male} + HEAD_{Fem}} \text{ x } 100$$

Comments/critical issues

For more information, the reader is referred to the *She Figures Handbook 2015* (DG Research and Innovation, 2016a).

2.11.2 P4 – Headline indicator – Share of women in grade A positions in HES

Definition of indicator

This indicator presents the proportion of women occupying the highest-level research positions (Grade A) in HES to the total of Grade A positions.

Rationale

This indicator pertains to priority 4 (gender equality in research) and relates to gender balance in career progression. This indicator enables tracking the progress made with regard to women's presence at the highest level of academia by analysing its trend through time. According to the DG Research and Innovation (2016b), women represented a majority of university graduates in the first stage of tertiary education (~60 %), while still representing close to half of them in the second stage of tertiary education (~47 %) in the EU-28 in 2013. Despite this figure, women represent a small minority of Grade A professors (21 %, 2013), heads of higher education institutions (20 %, 2014) and board members (including leaders) in research decision-making (28 %, 2014) (DG Research and Innovation, 2016b). Therefore, it is relevant to monitor the proportion of women present at each level of academia in order to observe whether there is progress toward reducing vertical segregation, defined as the under- or over-representation of a clearly identifiable group of workers in occupations or sectors at the top of an ordering based on 'desirable' attributes (EGGE, 2009).

Computation method

Data needed

($GRADE.A_{FEM}$) Number of women in grade A academic position: **Unit=Head count**;

($GRADE.A_{MALE}$) Number of men in grade A academic position: **Unit=Head count**.

Source of data

DG Research and Innovation—WiS—Women in Science database.

Specifications

$$Proportion\ of\ women\ in\ grade\ A\ academic\ positions = \frac{GRADE.A_{Fem}}{GRADE.A_{Male} + GRADE.A_{Fem}} \text{ x } 100$$

Senior grades/Academic staff

The grades presented in ERA monitoring are based upon national mappings according to the following definitions:

(A) The single highest grade/post at which research is normally conducted.

Comments/critical issues

The classification of academic positions into grades may vary across countries. This should be taken into account when comparing or aggregating statistics.

It is important to note that these data are not always completely cross-country comparable as the seniority of grades is not yet part of a formal international classification. Furthermore, it is not always possible to distinguish research staff from teaching staff, although the target population for 'academic staff' is researchers in higher education institutions (excluding staff involved in teaching or administration only and not at all in research).

For more information, the reader is referred to the *She Figures Handbook 2015* (DG Research and Innovation, 2016a).

2.12 Web of Science™ (WoS™)

The Web of Science (WoS™), produced by Thomson Reuters, was used as the main data source for computing the indicators presented in this section. The version of the WoS™ used in this monitoring exercise includes three databases: the Science Citation Index Expanded (SCI Expanded), the Social Sciences Citation Index (SSCI), and the Arts & Humanities Citation Index (A&HCI). Together these databases index some 12 000 journals whose publications are peer reviewed and cover all major fields of scientific research in the natural sciences and engineering (NSE), health sciences (HS) and social sciences and humanities (SSH). The WoS™ includes comprehensive bibliographic information on peer-reviewed scientific publications, such as their titles, abstracts, authors, author affiliations and references. This information can be analysed and tracked to measure an entity's (e.g. a country, an institution, a researcher) contribution to the scientific literature and its collaboration behaviour with other entities. For the purpose of this project, only high-quality and original contributions to scientific knowledge are considered. This covers two types of peer-reviewed documents: research articles and reviews, which are collectively referred throughout as 'publications' (or 'papers'). Note that a licence from Thomson Reuters is required to access WoS™ for the purpose of producing large-scale bibliometric datasets.

2.12.1 P1 – Additional indicator – Number of publications per 1 000 researchers in the public sector

Definition of indicator

This indicator is the ratio of the number of publications to the number of researchers (in thousands) in the public sector using fractional counting (refer to Annex 2 for a definition of fractional counting).

Rationale

This indicator pertains to priority 1 — that is, to more effective national systems. As one of the key types of actions promoted under the ERA to achieve this priority relates to the establishment of Research Performance Based Funding (RPBF) systems (i.e. systems applying the core principles of international peer review in grant competitions), it becomes highly relevant to monitor the impacts such changes have generated, or will generate, on the scientific production capacity of countries. One such impact relates to the size of a country's scientific production (i.e. the number of papers it published). However, such a metric is not internationally comparable due to differences in the size of countries. Perhaps more important is the impact of RPBF systems on the research productivity of countries, which is a key dimension in measuring the efficiency with which national R&I systems convert R&D inputs into R&D outputs. Additionally, scientific productivity metrics have the advantage of accounting for differences in country size. In this study, the number of publications per 1 000 researchers in the public sector is used as a proxy for measuring the scientific productivity of countries.

Computation method

Data needed

(PAP) Number of peer-reviewed scientific publications: **Unit=Total (fractional count)**;

(RES_{HES}) Researchers in the higher education sector: **Unit=Full time equivalent (FTE)**;

(RES_{GOV}) Researchers in the government sector: **Unit=Full time equivalent (FTE)**.

Source of data

For (PAP): *WoS™ (Thomson Reuters);*

For ($RES_{HES\ or\ GOV}$): *Eurostat—Statistics on research and development (online data code rd_p_persocc).*

Filters applied

For *(RES_{HES})*:
SECTPERF set to "Higher education sector"
OCCUP set to "Researchers"
SEX set to "Total"
UNIT set to "Full-time equivalent (FTE)"

For *(RES_{GOV})*:
SECTPERF set to "Government sector"
OCCUP set to "Researchers"
SEX set to "Total"
UNIT set to "Full-time equivalent (FTE)"

Specifications

$$Publications\ per\ 1\,000\ researchers\ in\ the\ public\ sector = \frac{PAP}{\left(\frac{RES_{GOV} + RES_{HES}}{1\,000}\right)}$$

Comments/critical issues

As most peer-reviewed scientific publications involve an actor from the public sector (at least 97 % in recent years; Science-Metrix, unpublished data), the denominator for this indicator was limited to the government and higher education sectors. Note that like many other productivity metrics, this indicator comes with limitations in that the typical production size of researchers varies across scientific subfields. As such, differences in the specialisation patterns of countries reduce its cross-country comparability when used as a proxy for productivity.

2.12.2 P2a – EMM indicator – International co-publications with ERA partners per 1 000 researchers in the public sector

Definition of indicator

This indicator measures, using fractional counting (refer to Annex 2 for a definition of fractional counting), the number of publications of an ERA country (or region within the ERA) involving at least one co-author from another ERA country. The number is presented relative to the given country's (or region's) researcher population size.

Rationale

This indicator pertains to sub-priority 2a, which relates to the implementation of joint research agendas within the ERA. The number of international co-publications with ERA partners per 1 000 researchers in the public sector is a good proxy to measure the outcomes resulting from the transnationally allocated research funding that is promoted under sub-priority 2a. The normalisation by the number of researchers accounts for size differences across countries, allowing their direct comparison.

Computation method

Data needed

($PAP_{ERA.COLLAB}$) Number of co-publications with another ERA country: **Unit=Total (fractional count)**;

(RES_{HES}) Researchers in the higher education sector: **Unit=Full time equivalent (FTE)**;

(RES_{GOV}) Researchers in the government sector: **Unit=Full time equivalent (FTE)**.

Source of data

For *($PAP_{ERA.COLLAB}$)*: *WoS™ (Thomson Reuters);*

For *($RES_{HES\ or\ GOV}$)*: *Eurostat—Statistics on research and development (online data code rd_p_persocc).*

Filters applied

For *(RES_{HES})*:
SECTPERF set to "Higher education sector"
OCCUP set to "Researchers"
SEX set to "Total"
UNIT set to "Full-time equivalent (FTE)"

For *(RES_{GOV})*:
SECTPERF set to "Government sector"
OCCUP set to "Researchers"
SEX set to "Total"
UNIT set to "Full-time equivalent (FTE)"

Specifications

$$Intl.co-publications\ with\ ERA\ partners\ per\ 1\ 000\ researchers = \frac{PAP_{ERA.Collab}}{\left(\frac{RES_{GOV}+RES_{HES}}{1000}\right)}$$

Comments/critical issues

As most peer-reviewed scientific publications involve an actor from the public sector (at least 97 % in recent years; Science-Metrix, unpublished data), the denominator for this indicator was limited to the government and higher education sectors.

2.12.3 P6 – Headline indicator – International co-publications with non-ERA partners per 1 000 researchers in the public sector

Definition of indicator

Using fractional counting (refer to Annex 2 for a definition of fractional counting), this indicator measures the number of publications of an ERA country (or region within the ERA) involving at least another co-author from a non-ERA country. The number is presented relative to the given country's (or region's) researcher population size.

Rationale

This indicator pertains to priority 6, which promotes the openness of MS/AC for international cooperation beyond the ERA. The number of international co-publications with non-ERA partners per 1 000 researchers in the public sector is a good proxy to measure the outcomes resulting from actions designed to achieve this priority. The normalisation by the number of researchers accounts for size differences across countries, allowing their direct comparison.

Computation method

Data needed

($PAP_{NON\text{-}ERA.COLLAB}$) Number of co-publications with non-ERA countries: **Unit=Total (fractional count)**;

(RES_{HES}) Researchers in the higher education sector: **Unit=Full time equivalent (FTE)**;

(RES_{GOV}) Researchers in the government sector: **Unit=Full time equivalent (FTE)**.

Source of data

For *($PAP_{NON\text{-}ERA.COLLAB}$)*: *WoS™ (Thomson Reuters);*

For *($RES_{HES\ or\ GOV}$)*: *Eurostat—Statistics on research and development (online data code rd_p_persocc).*

Filters applied

For *(RES_{HES})*:
SECTPERF set to "Higher education sector"
OCCUP set to "Researchers"
SEX set to "Total"
UNIT set to "Full-time equivalent (FTE)"

For *(RES_{GOV})*:
SECTPERF set to "Government sector"
OCCUP set to "Researchers"
SEX set to "Total"
UNIT set to "Full-time equivalent (FTE)"

Specifications

$$Intl.co-publications\ with\ non-ERA\ partners\ per\ 1\ 000\ researchers = \frac{PAP_{NON-RA.Collab}}{\left(\frac{RES_{GOV}+RES_{HES}}{1000}\right)}$$

Comments/critical issues

As most peer-reviewed scientific publications involve an actor from the public sector (at least 97 % in recent years; Science-Metrix, unpublished data), the denominator for this indicator was limited to the government and higher education sectors.

2.12.4 P2a – Additional indicator – International co-publication rate with ERA partners

Definition of indicator

The rate is defined as the number of publications co-authored by at least one researcher from a given ERA country (or region within the ERA) with at least one co-author from another ERA country, proportional to the total number of publications in the given country (or region). Full counting is used (refer to Annex 2 for a definition of full counting).

Rationale

This indicator measures the same dimension as is captured by the number of international co-publications with ERA partners per 1 000 researchers in the public sector. As such, the same rationale applies. The main difference between the two resides in the approach used to normalise the number of international co-publications with ERA partners. In the former case the denominator is the total number of publications of the country of interest, while in the latter case it is the number of researchers in the given country. The main benefit of the former over the latter is that it is dependent on a single data source, which allows producing data for a larger number of ERA countries and up to a more recent year (i.e. 2015 instead of 2014). Additionally, because the data come from a single high-quality source, the former indicator is believed to be more robust in performing international comparisons. It was added for the purpose of increasing country coverage in the composite indicator for priority 2. The number of international co-

publications with ERA partners per 1 000 researchers in the public sector had to be kept for the presentation of individual indicators since it is the formal one selected by the ERAC in the EMM indicators.

Computation method

Data needed

($PAP_{ERA.COLLAB}$) Number of co-publications with another ERA country: **Unit=Total (full counting)**;

(PAP) Number of peer-reviewed scientific publications: **Unit=Total (full counting)**.

Source of data

Computed using WoS™ (Thomson Reuters).

Specifications

$$International\ collaboration\ rate\ with\ ERA\ partners = \frac{PAP_{ERA.Collab}}{PAP}$$

Comments/critical issues

None identified.

2.12.5 P6 – Additional indicator – International co-publication rate with non-ERA partners

Definition of indicator

The rate is defined as the number of publications co-authored by at least one researcher from a given ERA country (or region within the ERA) with at least one co-author from a non-ERA country, proportional to the total number of publications in the given country (or region). Full counting is used (refer to Annex 2 for a definition of full counting).

Rationale

This indicator measures the same dimension as is captured by the number of international co-publications with non-ERA partners per 1 000 researchers in the public sector. As such, the same rationale applies. The main difference between the two resides in the approach used to normalise the number of international co-publications with non-ERA partners. In the former case the denominator is the total number of publications of the country of interest, while in the latter case it is the number of researchers in the given country. The main benefit of the former over the latter is that it is dependent on a single data source, which allows producing data for a larger number of ERA countries and up to a more recent year (i.e. 2015 instead of 2014). Additionally, because the data come from a single high-quality source, the former indicator is believed to be more robust in performing international comparisons. It was added for the purpose of increasing country coverage in the composite indicator for priority 6. The number of international co-publications with ERA partners per 1 000 researchers in the public sector had to be kept for the presentation of individual indicators since it is the formal one selected by the ERAC in the Headline indicators.

Computation method

Data needed

($PAP_{NON\text{-}ERA.COLLAB}$) Number of co-publications with non-ERA countries: **Unit=Total (full counting)**;

(PAP) Number of peer-reviewed scientific publications: **Unit=Total (full counting)**.

Source of data

Computed using WoS^{TM} (Thomson Reuters).

Specifications

$$International\ collaboration\ rate\ with\ non-ERA\ partners = \frac{PAP_{NON-ERA.Collab}}{PAP}$$

Comments/critical issues

None identified.

2.12.6 P4 – EMM indicator – Gender dimension in research content

Definition of indicator

This indicator relates to the proportion of a given country's scientific production (measured by the number of peer-reviewed scientific publications by full counting, see Annex 2 for more details) in which a gender dimension has been identified in the research content relative to the same proportion at world level. The resulting indicator is a specialisation index (SI), whereby a score above 1 means that a country is specialised — i.e. it puts more emphasis on the gender dimension in its research output — relative to the world, while a score below 1 means that it is not specialised relative to the world.

The concept of the gender dimension in research covers both the biological characteristics (i.e. the sex) and social/cultural aspects (i.e. the gender) of men and women. Scientific publications that involve a gender dimension are extracted by performing keyword-based queries in the titles, abstracts and author keywords of scientific publications. The selected keywords focus on well-defined gender topics (e.g. feminism, gender pay gap, gender equality, LGBT), as well as research content in which a distinction, or a comparison, is made between men and women (e.g. publications reporting sex-disaggregated data). Excluded from the gender dimension are studies pertaining to the animal kingdom (e.g. feminisation of fish populations) and other non-human biological entities, such as plants. Papers investigating specific medical conditions (e.g. menopause, erectile dysfunction) were also specifically excluded as they would return a very large number of scientific publications in the medical fields.

In *She Figures 2015* (DG Research and Innovation, 2016b), this indicator was simply presented as the proportion of a country's research output integrating a gender dimension in its research content (GDRC). This was adequate since the data were reported by main field of science. In the context of the current monitoring exercise, the data are only presented for all fields combined. Since the gender dimension in research content is more frequently observed in particular subfields (e.g. Nursing, Cultural Studies, Clinical Medicine) relative to others (e.g. Acoustics, Civil Engineering, Mining & Metallurgy), it is important that the GDRC indicator accounts for the distribution of a country's publication output across subfields so as to optimise cross-country comparability of the scores. For example, if one country publishes most of its output in the medical sciences, and another country publishes most of its output in the physical sciences, it is obvious that the former country will have a greater proportion of its total output integrating the GDRC than the latter. In the context of the ERA Monitoring Mechanism, the GDRC takes care of this issue by comparing the proportion of a given country's output integrating the GDRC to the world reference by subfield, and subsequently aggregates the subfield scores accounting for how prevalent each subfield is in the corresponding country's total output.

More specifically, the ratio of publications including a gender dimension to the total number of publications is first computed at the subfield level (according to the Science-Metrix classification; Archambault, Caruso and Beauchesne, 2011) for each MS/AC as well as for the world (all countries combined). The ratios are then normalised by the world ratios (for each subfield) to obtain a SI for each subfield. The SIs of each country across subfields are then multiplied by the corresponding subfield proportion in the given country's total output; if the SI of a country for GDRC in the Social Sciences Methods subfield equals 1.14 and this subfield represent 3 % of the country's total output, then the weighted SI for this country and subfield will equal 0.0342 (i.e. 1.14 * 0.03). Subsequently, the weighted SI scores of a given country are summed across subfields to obtain an aggregated SI score reflecting the country's emphasis on GDRC research relative to the world, while accounting for differences in the specialisation patterns of countries across scientific subfields.

Rationale

This indicator pertains to priority 4 (gender equality in research) and relates to the promotion of cultural and institutional change on gender. Since 2014, applicants to Horizon 2020, the latest EU Research and Innovation funding programme, are required to specify how they intend to integrate a gender dimension in their research content. This new requirement makes it relevant to start monitoring the extent to which researchers in different countries incorporate this aspect in their research content to provide baseline figures against which to measure progress in the future.

Computation method

Data needed

($PAP_{CO\text{-}GD\text{-}SUB}$) Number of papers with a gender dimension in a given subfield for a given country (i.e. a MS or AC): **Unit=Total (full counting)**;

($PAP_{CO\text{-}SUB}$) Total number of papers for a given country (i.e. a MS or AC) in a given subfield: **Unit=Total (full counting)**;

($PAP_{W\text{-}GD\text{-}SUB}$) Number of papers with a gender dimension in a given subfield for the world (all countries combined): **Unit=Total (full counting)**;

($PAP_{W\text{-}SUB}$) Total number of papers in a given subfield for the world: **Unit=Total (full counting)**;

($PAP_{CO\text{-}TOT}$) Total number of papers for a given country (i.e. a MS or AC): **Unit=Total (full counting)**.

Source of data

Computed using WoS™ (Thomson Reuters).

Specifications

$$Gender\ dimension\ in\ research\ content = \sum_{SUB}^{TOT} \frac{\frac{PAP_{CO-GD-SUB}}{PAP_{CO-SUB}}}{\frac{PAP_{W-GD-SUB}}{PAP_{W-SUB}}} \times \frac{PAP_{CO-SUB}}{PAP_{CO-TOT}}$$

Comments/critical issues

Note that full counting is used (refer to Annex 2 for a definition of full counting). Also note that this indicator is characterised by strong yearly fluctuations, especially for the smaller countries, which make it difficult to analyse trends in the short term. To circumvent this issue in the analysis of growth, as well as to maximise the coverage of countries, a four-year rolling window (or four-year moving average of the scores) was applied in presenting the data.

2.12.7 P5b – Headline indicator – Share of publications available in open access (green and gold)

Definition of indicator

This indicator is the proportion of a country's publications that are available in open access (OA) as per Peter Suber's definition ([5]) of gratis OA, which includes libre OA. In addition to the proportion of total OA, the indicator is also produced for two sub-types of OA: gold and green. The former refers to

> papers made available for free by the publishers themselves, be it on their website (e.g., in fully gold OA journals on Springer Open and BioMedCentral, or as hybrid OA, that is, OA papers from otherwise paywalled journals on, for example, Springer's website) or on the site of an aggregator

[5] http://sparcopen.org/our-work/gratis-and-libre-open-access/

(e.g., Scielo, and also PubMedCentral, on which the majority of papers are archived by the publishers themselves) (Archambault et al., 2016).

The latter refers to 'papers made available for free by parties other than publishers, usually the authors themselves, who archive papers in institutional repositories, subject repositories such as arXiv, or commercial repositories such as ResearchGate' (Archambault et al., 2016).

Briefly, 1science provided an index list of peer-reviewed publications available in OA through their oaFindr database. The URLs pointing to the OA version of these publications were harvested and codified by OA type. Note that the sum of the proportions for gold and green OA add up to more than the total proportions of OA since both types of OA overlap; some papers can be accessible through different URLs, some of which may be coded as gold OA, while others may be coded as green OA. The OA papers, thus coded as gold or green OA, were then matched to a large-scale bibliographic database of peer-reviewed literature — the WoS™ — to allow computing the proportion of all publications (i.e. OA papers plus non-OA papers) that is available in gold, green or gold/green (i.e. total) OA. The entirety of the WoS™ was used; in other words, OA availability was checked for each paper in the database through the 1science OA index. Note that the whole set of OA papers include uncategorised OA papers (i.e. the URL could not be classified as gold or green). For further details on the methodology, refer to Archambault and colleagues (2016).

Rationale

This indicator pertains to sub-priority 5b (optimal access to and circulation and transfer of scientific knowledge) and relates to the open access of publications and data resulting from publicly funded research. Open access articles are publicly accessible online without restrictions (DG Research and Innovation, 2015). Articles published in open access format can be freely read by anyone who can access the web. It therefore facilitates the mobility, transfer and circulation of knowledge between scientists, research institutions, the private sector and citizens who might lack the resources necessary to access the scientific literature. ERA members are expected to implement legal frameworks with the intention of making scientific research openly available (Science-Metrix, 2016), but open access may require more financial support from funders (DG Research and Innovation, 2015) since the publication costs related to open access publishing are covered by the authors instead of journal subscribers. This indicator enables assessing the state of play as regards the extent to which ERA countries disseminate the results of their research via OA channels.

Computation method

Data needed

($PAP_{OA\text{-}Total}$) Number of publications in gold or green OA: **Unit=Total (full counting)**;

($PAP_{OA\text{-}Gold}$) Number of publications in gold OA: **Unit=Total (full counting)**;

($PAP_{OA\text{-}Green}$) Number of publications in green OA: **Unit=Total (full counting)**;

(PAP) Total number of publications: **Unit=Total (full counting)**.

Source of data

For *(PAP_{OA}, $PAP_{OA\text{-}Gold}$, $PAP_{OA\text{-}Green}$)*: *1science (http://www.1science.com/index.html) matched to the WoS™;*

For *(PAP)*: *WoS™ (Thomson Reuters).*

Specifications

$$Share\ (\%)\ of\ Total\ OA\ publications = \frac{PAP_{OA}}{PAP} \times 100$$

$$Share\ (\%)\ of\ Gold\ OA\ publications = \frac{PAP_{OA-Gold}}{PAP} \times 100$$

$$Share\ (\%)\ of\ Green\ OA\ publications = \frac{PAP_{OA-Green}}{PAP} \times 100$$

Comments/critical issues

The share of publications available in OA can be provided by the publication year of the papers. However, a 2005 publication might only become available in OA years after its original publication date. This phenomenon, referred to as 'delayed OA', makes it impossible to study the growth in the share of OA publications using a single snapshot (e.g. Spring 2016 in the case of this study) of those papers in the WoS^{TM} that are available in OA. Although an analysis of the trend in the share of papers available in OA based on their publication year shows a strong increase based on this study's 2016 snapshot, the yearly shares (even those of earlier publication years) will continue to change with future snapshots; it is also normal for older papers to be less accessible via OA. To adequately study the growth of OA availability, it would be necessary to use trends based on the production year (or date) of the snapshots instead of the publication year of the papers. This will only become possible as new snapshots of OA publications become available.

Nevertheless, in this study the analysis of trends based on the publication year revealed a striking drop in the share of OA papers in the most recent year (i.e. 2015, refer to the main study's report for the data). This drop is particularly pronounced for green OA and appears to be due to short-term delayed OA, which is mostly attributable to embargo periods. These embargoes are a period following publication, after which publishers release the copyright of traditional subscription-based journals, thereby either making their full content directly available to the public, or making the content partially available by allowing researchers to post their papers online on various archives or personal websites. Because researchers might be busy doing other things, however, it might take longer still before their papers become accessible to all in the latter case (i.e. via green OA).

Finally, note that the proportions of OA papers computed in this study are slightly underestimated since, like any other harvesters, the 1science harvester does not capture 100 % of all OA papers. It is estimated that the recall achieved in this monitoring exercise is roughly equal to 75 %; in other words, about a quarter of WoS^{TM} papers have erroneously been classified as being paywalled (Archambault et al., 2016). Thus, to obtain adjusted proportions, one simply has to multiply the proportions reported in this study by a correction factor (i.e. by 1/0.75 = 1.33); in this study, only the unadjusted figures are provided. Note that the most accurate adjustment might be one that varies across countries.

2.13 Composite indicators

Science-Metrix designed two types of composite indicators as experimental tools to synthesise progress towards achieving the ERA both within and across priorities. The first type of composite — the Headline composite — aims to give a balanced reflection of performance across the eight headline indicators selected by ERAC as being the most relevant in monitoring progress in achieving the ERA. Thus, the sub-priorities 2a and 2b are represented separately, as are sub-priorities 5a and 5b. The second type of composite — the Meta-composite — aims to provide a comprehensive overview of performance towards achieving each of the six ERA priorities' relevant dimensions by integrating multiple indicators within each priority. The Meta-composite has been constructed using a bottom-up approach, whereby intermediate priority composites were first constructed to synthesise performance within each priority. Since the number of relevant dimensions, and of indicators available to measure them, varies across priorities (and sub-priorities), this approach carries two benefits: it provides a synthetic view of progress towards achieving the ERA both within (the intermediate priority composites) and across (the Meta-composite) priorities, and equalises the contribution of each priority to the Meta-composite (i.e. each priority is represented by a single intermediate composite) ([6]). In short, the Headline composite integrates only the indicators identified as the most salient by the ERAC, while the Meta-composite integrates a broader evidential base for each of the six priorities and overall (which includes, where possible, the Headline indicators but also a considerable number of others as well).

[6] For priority 2, a differential weighting approach was used to ensure that the sub-priorities 2a and 2b, although they differ in number of indicators, contribute, in as much as is possible, equally to the composite for priority 2. The same applies for the sub-priorities 5a and 5b of priority 5.

Indicator and country selection

For the composite indicators designed by Science-Metrix, computing changes in composite performance over time was not undertaken because of data limitations. In short, a minimum data coverage threshold was established to ensure the quality of the composites, and too many countries and indicators (even priorities) would have been excluded for passing this threshold had longitudinal computations been undertaken. The applied threshold consisted in a minimum coverage of 75 % of time series (or data points in the static approach), both across indicators for countries and across countries for indicators. The time series that were originally considered for building the composite in a dynamic fashion consisted of only two data points to measure progress between two reference years (e.g. between 2011 and 2015).

The threshold used in determining which indicator and country could be included was applied following data imputation. Imputation of missing data was performed by replacing missing data points by more recent or older data points with a maximum gap of two years between the reference and imputation year. For the dynamic option, it was requested that the imputation did not shorten or lengthen the time series (e.g. 2011-2015) by more than one year. In exceptional circumstances, outliers have been replaced using the same approach (refer to the 2016 ERA Monitoring Handbook for further details on outlier detection).

At the outset of this initial exploratory phase, it was decided that the composite indicators would be computed using a static approach only. For the Headline composite, the static approach was the only possible option since the Headline indicator for sub-priorities 2b ([7]) and 5b ([8]) could not be reported in a dynamic fashion. For the Meta-composite, the use of a dynamic approach would have resulted in a dataset of 21 indicators covering six out of eight priorities/sub-priorities (2b and 5b would have been omitted) and covering 27 Member States (Malta would have been omitted) as well as one Associated Country (Norway). In comparison, the static approach resulted in a dataset of 27 indicators covering all priorities/sub-priorities and covering all 28 Member States, plus 5 Associated Countries (Iceland, Norway, Switzerland, Serbia and Turkey) ([9]). The list of indicators included in the Headline composite and the Meta-composite is provided in Table 2.

In the static approach, because the most recent year of available data across selected indicators varied (i.e. the reference year of individual indicators), the reference year for the composites was set in relation to the year of the ERA Monitoring Mechanism to which this report refers to. Thus, the reference year of the composites was set to '2016' although the data correspond to the actual performance of countries in different years.

[7] The Headline indicator for sub-priority 2b (i.e. availability of national roadmaps with identified ESFRI projects and corresponding investment needs) has been substituted with the complementary EMM indicator on ESFRI landmarks since it could not be included in this study's composite (it is a qualitative indicator). No time series is available on ESFRI landmarks. ESFRI landmarks were chosen over ESFRI projects since they represent successful ESFRI projects (i.e. operational).

[8] Growth in the share of publications available in open access cannot be reported accurately due to delayed open access issues (see Section 2.12.7 for more details).

[9] In the country selection phase for building the Meta-composite, the 75 % threshold was applied once to all indicators across all priorities. In the resulting selection, in very few instances, some countries had fewer than 75 % of the selected indicators in one, or many, of the intermediate composite indicators embedded within the Meta-composite. These instances have been clearly flagged throughout this report.

Table 2 Indicators incorporated in the Headline composite and the Meta-composite

Priority	Headline composite	Meta-composite ***
1	Adjusted Research Excellence Indicator (DG Joint Research Centre, Competence Centre on Composite Indicators)	Adjusted Research Excellence Indicator (DG Joint Research Centre, Competence Centre on Composite Indicators)
		Government budget appropriations or outlays for R&D as a percentage of GDP (Eurostat)
		Number of researchers per 1 000 population (Eurostat)
		Number of papers published per 1 000 researchers (Science-Metrix using Web of Science).
2a	National GBARD (EUR) allocated to Europe-wide, bilateral or multilateral transnational public R&D programmes per FTE researcher in the public sector (Eurostat)	National GBARD (EUR) allocated to Europe-wide, bilateral or multilateral transnational public R&D programmes per FTE researcher in the public sector (Eurostat)
		Member State participation (EUR) in Public-to-Public collaborations per FTE researchers in the public sector (Eurostat and 1st ERA-Learn 2020 Annual Report on P2P Partnerships)
		International co-publication rate with ERA partners (Science-Metrix using Web of Science)
		International co-invention rate with ERA partners (Science-Metrix using PATSTAT data on PCT applications)
2b	Percentage of ESFRI Landmarks in which a Member State/Associate Country is a partner (ESFRI data) *	Percentage of ESFRI Landmarks in which a Member State/Associate Country is a partner (ESFRI data)
		Percentage of ESFRI Projects in which a Member State/Associate Country participates (ESFRI data)
3	Number of researcher postings advertised through the EURAXESS job portal per thousand researchers in the public sector (EURAXESS historical data and Eurostat)	Number of researcher postings advertised through the EURAXESS job portal per thousand researchers in the public sector (EURAXESS historical data and Eurostat)
		Share of doctoral candidates with a citizenship of another EU Member State (Eurostat)
		Share of researchers expressing satisfaction that the hiring procedures in their institution are Open, Transparent and Merit-based (MORE2 Survey)
4	Share of women in Grade A academic positions in the Higher Education Sector (Women in Science database, DG Research and Innovation)	Share of women in Grade A academic positions in the Higher Education Sector (Women in Science database, DG Research and Innovation)
		Gender dimension in research content (Science-Metrix using Web of Science)
		Share of women heads of institutions in the higher education sector (Women in Science database, DG Research and Innovation)
		Proportion of female PhD graduates (Eurostat)
		Share of women researchers (Eurostat)
5a	Share of product or process innovative firms cooperating with public or private research institutions (1) and with higher education institutions (2) (Eurostat) **	Share of product or process innovative firms cooperating with public or private research institutions (Eurostat)
		Share of product or process innovative firms cooperating with higher education institutions (Eurostat)
		Share of public research financed by the private sector (Eurostat)
		Number of public–private co-publications per million population (CWTS)
5b	Share of publications available in (Green and/or Gold) Open Access (1Science & Science-Metrix).	Share of publications available in (Green and/or Gold) Open Access (Science-Metrix using 1Science data)
6	Co-publications with non-ERA partners per 1 000 researchers in the public sector	International co-publication rate with non-ERA partners (Science-Metrix using Web of Science)
		Non-EU doctorate students as a share of all doctorate students (Eurostat)
		Licence and patent revenue from abroad as a share of GDP (Eurostat)
		International co-invention rate with non-ERA partners (Science-Metrix using PATSTAT data on PCT applications)

Note: * The Headline indicator for sub-priority 2b (i.e. availability of national roadmaps with identified ESFRI projects and corresponding investment needs) has been substituted since it is a qualitative indicator not suitable for the composite. ** The Headline indicator for sub-priority 5a aims to capture the share of innovative firms cooperating with public/private and higher education sector partners; as consolidated data were not available for these two sectors, the Headline was split into two indicators. Nine indicators are therefore integrated to cover the eight Headline indicators in the Headline composite. The two indicators for sub-priority 5a individually carry less weight than any other indicator in the composite as they are highly correlated. They each received a weight of about 0.5, approximately half the weight of the other indicators. In the Meta-composite, sub-priorities 2a and 2b, as well as sub-priorities 5a and 5b, have each been treated as a single priority (i.e. 2a&b and 5a&b) to ensure that each intermediate composite indicator includes a minimum of two indicators, as well as to allow for a more balanced distribution of indicators across priorities. There are therefore three indicators for priority 3, four for priorities 1 and 6, five for priorities 4 and 5, and six for priority 2. For priority 2, a differential weighting approach was used to ensure that the sub-priorities 2a and 2b, although they differ in number of indicators, contribute as equally as possible to the composite for priority 2. The same applies for the sub-priorities 5a and 5b of priority 5. The data source is provided in parentheses next to the indicator name.

Source: Compiled by Science-Metrix

Normalisation, standardisation & weighting of indicators in computing the composite indicators

The method for computing the composite indicators was designed to promote equal representation of all the indicators (or intermediate priority composites in the Meta-composite) integrated into a given composite, or giving them as equal a weight as possible. This method included four facets. First, for indicators where the distribution of performance scores at the country level were very skewed, logarithmic transformations were applied to the scores before computation, to yield a more normal distribution. Second, country scores on an indicator (whether logarithmically transformed or not) were standardised between 0 and 1 using the minimum and maximum scores across countries for the given indicator (i.e. [country score – min score across countries] / [min score across countries – min score across countries]). This standardisation of the scores is intended to facilitate integration of indicators on different scales — that is, when the range of scores is very wide for some indicators, while it is narrower for others. Both the normalisation and standardisation of the indicators serve to optimise the uniformity with which each indicator will contribute to the resulting composite; it ensures that none of the individual indicators will exert a disproportionate effect on the composite measure at the expense of the other indicators. Third, a weighting algorithm was applied to the individual indicators to address their inter-correlations. In short, if three indicators are integrated, and two of them are highly correlated, it implies that the resulting composite will integrate fewer than three distinct dimensions due to the redundancy in the latter two indicators. As such, the dimensions (less than two) captured by each of the two inter-related indicators will have a much greater influence on the resulting score than the dimension captured by the third indicator if each indicator is equally weighted in the composite indicator. The amount of redundancy in the dimensions captured by different indicators was captured by their correlation matrix, which served to algorithmically assess and compensate for existing redundancy across indicators. Briefly, redundant indicators receive a smaller weight reflecting the extent to which the dimension they intend to measure is captured by other indicators. Finally, a sensitivity analysis was performed to ensure that the weighting did improve the representation of each indicator in the composite relative to a uniform weighting scheme. This was achieved by comparing the variance, with and without the differential weighting scheme applied, in the coefficient of determination (R^2) between each component of a composite and the composite itself. For all composites, the application of a differential weighting scheme improved the resulting composite by reducing the observed variance in R^2; in other words, the influence of each component on the composite was more evenly distributed among them.

Limitations

Clustering sets of entities (e.g. cars, computers, species, countries) based on a variable number of characteristics (i.e. there can be few or many variables) can be achieved using various statistical procedures (e.g. exploratory factor analysis, multidimensional scaling, *k*-means clustering). These procedures aim to reduce the complexity (i.e. dimensionality) of a dataset towards producing groups of entities sharing similar patterns across the original set of measured characteristics. The synthetic information resulting from such procedures can then prove to be very useful in supporting decision-making in varied contexts since it focuses the analysis on the most discriminant composite dimensions (or composite indicators). For example, just as the clustering of fish populations based on their pattern of genetic variation across multiple regions of their genome assists conservation biologists in setting sensible fishing quotas to preserve biodiversity (Bradbury et al., 2015), the clustering of consumers according to varied characteristics such as their purchasing behaviour has allowed companies to evolve their offers towards more personalised services/products (Einav and Levin, 2013). In fact, many authors recently argued that the analysis of very large datasets using novel data mining techniques aimed at clustering diverse entities on the basis of heterogeneous sets of variables will revolutionise the delivery of services in governments as well as the way in which governments operate, just as it did in the private sector (Yiu, 2012).

Yet just as with any other analytical methods, clustering approaches are not without drawbacks. For instance, there is a risk of oversimplification hiding important information on the individual characteristics of entities, possibly leading to oversights on the part of decision-makers developing and implementing policies using such information. For this reason, ICF recently recommended not to use clustering approaches — or equivalently composite indicators — in their appraisal of available or potential indicators with which to monitor progress across ERA priorities (ICF International, 2015).

Although the study team fully understand these limitations in the use of composite indicators to monitor progress towards achieving the ERA, it also sees a value in their use to group countries according to their performance level in complying with the set of actions implemented under each ERA priority and globally across all six priorities. When multiple indicators are used to characterise the performance of countries it is often difficult to highlight general trends without a well-structured ranking mechanism. Note, however, that the clustering approach and composite indicators produced in this study will not overshadow any relevant information at a lower aggregation level (i.e. individual indicators measuring progress in relation to specific actions). This synthetic analysis only aims at supplementing the presentation of data at the indicator level.

3 GENERAL APPROACH TO THE ANALYSIS AND PRESENTATION OF QUANTITATIVE INDICATORS

The general time frame to be assessed was the 2005-2015 period, with each results table providing an assessment of static performance in the most recent year for which high-quality data were available across countries, as well as a longitudinal assessment of evolving performance, where the length of this assessment period was again determined by quality of available data. As very up-to-date data were often unavailable to compute a given indicator for certain countries, the selection processes for performance snapshots required balancing country coverage with the timeliness of assessment, to ensure that the need for a very timely snapshot did not exclude the coverage of too many countries, and that the need for exhaustive coverage across countries did not lead to the assessment of outdated results.

The quantitative results tables present growth over the period assessed for each indicator, displayed as a compound annual growth rate (CAGR), which shows the average year-over-year change in a country's performance, taking compounding effects into account. The CAGR assumes an exponential growth between the starting and ending year of a reference period, which is rarely the case across all countries, especially for the smaller ones. Additionally, there is some temporal heterogeneity among the selected indicators; some measure the structural aspects of a nation that change in the long term, whereas others show high short-term fluctuations in many countries. Since the CAGR measures growth using the longest available period for each indicator (from 2005 onward), it might indicate an upward or downward trend that no longer holds in the most recent years, especially for the smaller countries and indicators subjected to short-term fluctuations. In the tables created to report the results of this study, a micro bar chart showing the actual trend for each country is presented next to the CAGR to help detect both long-term and short-term progress towards realising the ERA.

For example, one can see from Table 3 that the CAGR score for the Netherlands shows strong growth from 2006 to 2012 (CAGR of 17.5 %) in spite of a drastic decline in the most recent year (i.e. 2013). While the reference period for CAGR was 2006-2013 for this indicator (i.e. licence and patent revenue from abroad as a share of GDP), the presence of the trend column allowed the research team to detect the 2013 outlier to make an exception to the reference period for the Netherlands (i.e. 2006-2012). This is justified by the fact that the latest point in the time series is an obvious outlier that might hide an issue with the quality of the data for 2013 in particular (see the 'Quality plan' section for details on the detection of outliers). This way, the CAGR does not consider this potentially faulty data point, and yet the full trend is shown so that people can see the important drop on 2013, be it true or false. Care is taken in the interpretation of the data to highlight this issue, noting that new data for more recent years will enable determining whether the drastic decrease observed in 2013 is real. For France, the CAGR shows an average 6.7 % annual increase from 2006 to 2013. It is nevertheless interesting to note the recent decline starting in 2011 as revealed by the trend. Additionally, note that for a few indicators where short-term fluctuations were particularly pronounced, moving averages have been used to measure performance and growth (e.g. average scores across 2005-2007, 2006-2008… 2012-2014). In such cases, the CAGR measures the year-on-year percent change in the rolling average of an indicator between the starting and ending periods (e.g. between 2005-2007 and 2012-2014).

Table 3 Sample table showing the presentation layout used to report the data for each indicator — licence and patent revenue from abroad as a share of GDP (2006-2013)

Country	Weight in GDP	Score (2013)	CAGR (2006-13)	Lead/Gap to EU-28 CAGR	Trendline (2006-13)
EU-28		**0.64%**	**9.6%**	**N/A**	
Cluster 1	9.3%	0.03	18.3%	12.5	
Cluster 2	29.5%	0.01	9.0%	3.3	
Cluster 3	61.2%	0.00	-17.6%	-23.4	
Cluster 4	N/A	N/A	N/A	N/A	
Cluster 1					
NL	4.5%	*3.72%*	*17.5%*	11.7	
CH	3.6%	*3.07%*	*8.6%*	2.8	
IE	1.2%	2.23%	28.8%	23.1	
Cluster 2					
FI	1.4%	1.38%	16.8%	11.1	
LU	0.3%	1.29%	5.3%	-0.5	
SE	3.0%	1.08%	1.7%	-4.0	
IS	0.1%	*0.90%*	:		
HU	0.7%	0.89%	10.6%	4.9	
DE	19.5%	0.77%	18.6%	12.8	
DK	1.8%	0.71%	1.8%	-3.9	
BE	2.7%	0.64%	8.2%	2.4	
Cluster 3					
UK	14.1%	0.46%	-3.0%	-8.7	
FR	14.6%	0.43%	6.7%	0.9	
AT	2.2%	0.25%	7.4%	1.7	
IT	11.1%	0.19%	18.2%	12.4	
CZ	1.1%	0.13%	*17.3%*	11.5	
NO	2.7%	0.08%	-11.7%	-17.4	
RO	1.0%	0.07%	*-13.2%*	-18.9	
ES	7.1%	0.07%	-0.3%	-6.1	
PL	2.7%	0.05%	8.0%	2.3	
EL	1.2%	0.00%	-100.0%	-105.8	
MT	0.1%	0.00%	-100.0%	-105.8	
PT	1.2%	0.00%	-100.0%	-105.8	
SK	0.5%	0.00%	-100.0%	-105.8	
BG	0.3%	0.00%	0.0%	-5.8	
EE	0.1%	0.00%	0.0%	-5.8	
HR	0.3%	0.00%	0.0%	-5.8	
CY	0.1%	0.00%	0.0%	-5.8	
LV	0.2%	0.00%	0.0%	-5.8	
LT	0.2%	0.00%	0.0%	-5.8	
SI	0.2%	0.00%	0.0%	-5.8	
ME	0.0%	0.00%	:		
MK	0.1%	0.00%	0.0%	-5.8	

Note: Provisional: EU-28 (2011-2013); 2013 (BE, BG, CZ, DK, DE, EE, IE, FR, HR, IT, CY, LV, LT, LU, HU, MT, NL, AT, PL, PT, RO, SI, SK, FI, SE, UK, NO, ME, MK); EL (2011-2013); ES (2012, 2013).
Potential outlier: 2013 (DE, NL).
Eurostat country flags have been retained in the EU-28 aggregate.
Exception to reference year: 2012 (NL, IS, CH).
Exception to reference period: 2006-2012 (NL, CH); CZ (2009-2013); RO (2008-2013).
Data unavailable: AL, RS, TR, BA, IL, FO, MD, UA.
(:) = missing data.

Source: Computed by Science-Metrix using Eurostat data (online data codes: bop_its_ybk and nama_10_gdp)

As no explicit, quantitative targets have been established as a definition of having 'achieved the ERA', the static assessment of performance in the present report cannot meaningfully speak of how well one country or region is standing relative to that target, nor how fast one country or region is progressing or regressing relative to that target in the longitudinal assessment. This issue stems from the fact that the goals to be reached in achieving the ERA constitute moving targets (e.g. ERA priorities and actions to achieve them are continuously evolving along with the needs of European societies). As such, it is difficult to establish reference values to be attained in relation to specific ERA policy actions; some of these targets could become obsolete in between each EMM round. Thus, both the performance and progress of countries are benchmarked against one another and against the EU-28 average ([10]), displayed as a lead or gap to that average (in percent for performance (not displayed in this report's tables), and in percentage point difference for the CAGR (displayed in this report's tables)). This lead/gap analysis has been colour-coded, from blue for the lowest scores to orange for the highest scores, to facilitate visual identification of patterns in performance (Table 3) ([11]). Additionally, performance in the most recent year is also benchmarked relative to performance across the ERA as a whole (i.e. relative to an unweighted average across the Member States and Associated Countries for which data are available for a given indicator) ([12]). This benchmarking is conveyed through the clustering approach implemented throughout ([13]). As mentioned above, because explicit targets are not defined for these indicators, the distance to such a target cannot be measured. Country-level performance is compared to the EU-28 (weighted) and ERA (unweighted) averages, but these should not be conflated with targets. For instance, the EU-28 (weighted) and ERA (unweighted) averages are close to 20 % for some gender parity indicators, while a reasonable target would likely be closer to 50 %, which would reflect absolute parity.

Computation of EU-28 aggregate

When reporting EU-28 aggregates, two methods have been used to compute the score for the EU-28, depending on the data composing the reported indicator. If an EU-28 aggregate was already available from the data source, then the already aggregated score was used. It is often the case for data issued from Eurostat tables that Eurostat provide pre-computed aggregates for the EU-28 GBARD and for the EU-28 GDP — for example, for the indicator *GBARD as a percentage of GDP*.

When a pre-aggregated score was not available, Science-Metrix computed it by summing the score of each MS separately for each part of the indicator. For example, when an indicator is a ratio between two values, the EU-28 numerator will be the sum of all MS numerators and the EU-

[10] Weighted averages are used to ensure representativeness of the whole (i.e. as if the EU-28 was a single country). Refer to the following sections for details on how EU-28 scores were computed.

[11] Assuming progress is reflected by increased scores over time for all indicators, except those characterised by a tipping point after which further increases lead to greater imbalance (e.g. share of women researchers).

[12] An unweighted average is used across countries to allow computing the standard deviation used in the clustering protocol (see footnote 13 below).

[13] The strongest performances are found in Cluster 1, which is more than one standard deviation above the ERA mean; the next strongest performances are found in Cluster 2, which is above the ERA mean, but within one standard deviation of it; performances listed in Cluster 3 are below the ERA mean, but within one standard deviation thereof; and finally the performances listed in Cluster 4 are the lowest, being more than one standard deviation below the ERA mean. Under this clustering approach, and assuming a normal distribution of the scores, 16 % of the countries should fall in each of Cluster 1 and 4, while there should be 34 % of countries in each of clusters 2 and 3. This approach therefore aims to highlight the few countries that really stand out above or below the ERA average (i.e. respectively those in Cluster 1 and 4). In some cases where results are highly skewed (i.e. with a few countries showing very high scores and the rest being concentrated in the low scores; in other words, the distribution is not normally distributed), it would be mathematically impossible to be more than one standard deviation below the mean, and in these cases there is no Cluster 4. In such cases, Cluster 3 can in fact be interpreted as a merge of Cluster 3 and 4. In exceptional circumstances, some data points (i.e. outliers) were presented and categorised, although they were not used in computing the ERA average (and the associated standard deviation) to determine the clusters' boundaries. Data points were considered as outliers if they were more than four standard deviations away from the ERA average. In a normal distribution, 100 % of data points must lie within four standard deviations of the average.

28 denominator will be the sum of all MS denominators. This is the case for the indicator *Share of doctoral candidates with a citizenship of another EU Member State* for the year 2013. Also, since a trend analysis is provided, special care was taken to ensure that the MS constituting the EU-28 score stayed the same throughout the presented period. Reporting periods were chosen such that a maximum of MS had data availability for the whole period. The footnote of each table describes which MS were excluded from the EU-28 score, if any. Note that while a country might have a score for several years, if it had one missing year in the period then it was not included in the EU-28 score for the whole period. Also note that for some indicators, the MS included in EU-28 score for performance (based on the latest available year) may differ from the MS included in EU-28 score for growth. This is explained by the fact that some MS may have data availability for the performance reporting year but not for all years included in the growth analysis, hence these countries will be excluded from the EU-28 in the growth values.

Some indicators may also include both methods — that is, a part is computed by summing each MS value and the other part comes from a pre-aggregated score.

4 ADDITIONAL RELEVANT DETAILS

Limits and possible biases in Headline and EMM indicators

Although the Headline and complementary EMM indicators were chosen as those most relevant in measuring progress towards the ERA, and as those most objective and impartial in performing cross-country comparisons among MS/AC, it is possible that some indicators present some biases favouring specific countries to the detriment of others. Among the possible biases, it is recognised that a country's size, geographic location and home language can exert a significant effect on, for example, the extent to which a country gets involved in various types of cross-country partnerships. Table 4 lists the biases that can most likely affect the selected EMM indicators (including the Headline indicators). In Table 5 the reader will find a matrix of the Headline and complementary EMM indicators organised by priority and type. The coloured capital letter or letters at the end of each indicator indicate which biases from those listed in Table 4 may apply to the indicator.

Table 4 List of possible biases applicable to the headline and EMM complementary indicators

Problem	Definition
Question of optimum **O**	What is the optimum? It is not necessarily 100 % for project-based funding or 100 % for the share of females.
Question of balance **B**	As regards gender, one should aim at gender balance, not at maximising the share of one gender. Is a perfect balance/an optimum 50 %? (Women are, however, more than 50 % of the population, but below 50 % in the younger cohorts.) One might need to develop an indicator that reaches its maximum at 50 %. For PhD graduates, the share of female graduates might soon exceed 50 %.
Country size bias (in favour of small countries) **C**	Small countries tend to be more international than larger countries, especially if they share a common language with a larger country.
Country location bias **L**	Countries at the geographic centre of the EU might have a higher share of intra-EU cooperation, while countries at the periphery of the EU or bordering non-EU countries might have a higher share of non-EU cooperation. There may be additional linguistic/historical biases: countries with international languages or countries that have been colonial powers might have a higher share of non-EU cooperation.
Economic structure bias **S**	Countries with a high share of manufacturing industry tend to have a higher propensity to patent inventions than countries with a lower share of manufacturing. Countries that host the headquarters of large companies tend to have a higher level of patenting than countries that do not. Countries with a high share of pharma, biotech, ICT, software and electrical machinery companies tend to have more patents than countries without such industries. The same may also apply to scientific publications issued from the private sector.
Country level data only in binary from **BI**	In some cases, indicators can take a binary form (e.g. 0/1 or yes/no), such as for the availability of national roadmaps with identified ESFRI projects and corresponding investment needs (ESFRI). This is too little information for benchmarking countries.
Potential bias from taxation differentials **T**	For indicators related to revenue from abroad (e.g. patents and licence fees) there might be a certain bias stemming from differentials in taxation (Luxembourg and Ireland, for example, scoring high).
Periodicity of data collection **P**	Data might not be updated annually (for example, results of the MORE study, She Figures is 3-yearly only, some of the She Figures data might, however, be available annually).
Historical factors **H**	Some data might be influenced by historical factors; e.g. the UK and France have many non-ERA students and PhDs from former colonies.

Note: The coloured capital letters are matched to headline and EMM indicators in Table 5.
Source: Assembled by Science-Metrix from ERAC documentation

Table 5 Matrix of Headline and complementary EMM indicators with potential bias(es) identified

Priority	Input Indicator	Output Indicator	Outcome/Impact Indicator
Priority 1: More effective national research systems	GBARD as percentage of GDP (Eurostat)	Adjusted Research Excellence Indicator (REI) (source: JRC)	European Innovation Scoreboard Summary Innovation Index (SII) (source: EIS)
Sub-priority 2a: Optimal transnational cooperation	Participation in public-to-public partnerships per researcher in the public sector (ERA-Learn 2020 report on P2P) **C**	GBARD allocated to Europe-wide transnational, as well as bilateral or multilateral, public R&D programmes per FTE researcher in the public sector (Eurostat) **C**	International co-publications with ERA partners per 1 000 researchers in the public sector (WoS and Eurostat) **C, L**
Sub-priority 2b: European Strategy Forum on Research Infrastructures (ESFRI)	Share of developing ESFRI Projects in which a Member State or an Associated Country participates (ESFRI)	Availability of national roadmaps with identified ESFRI projects and corresponding investment needs (ESFRI) **BI**	Share of operational ESFRI Landmarks in which a Member State or an Associated Country is a partner (ESFRI)
Priority 3: Open Labour Market for Researchers	Share of doctoral candidates with a citizenship of another EU Member State **C, L**	Researcher's posts advertised through the EURAXESS job portal per 1 000 researchers in the public sector (EURAXESS and Eurostat)	Share of researchers expressing satisfaction that the hiring procedures in their institution are open, transparent and merit based (MORE2 Survey) **P**
Priority 4: Gender equality and gender mainstreaming in research	Share of female PhD graduates (Eurostat) **B**	Gender dimension in research content (WoS) **O**	Share of women in grade A positions in HES (WiS—Women in Science database) **B, P**
Sub-priority 5a: Knowledge circulation	Share of product and/or process innovative firms cooperating with higher education institutions or public/private research institutions (Eurostat) **S**	Share of public research financed by the private sector (Eurostat) **S**	Number of public-private co-publications per million population (CWTS and Eurostat) **S**
Sub-priority 5b: Open access	Share of RFOs that provide funds to cover the costs of making publications available in OA and share of RPOs making their research data available in OA (data unavailable)	Share of publications available in open access (green and gold) (1science, WoS)	Presence or absence of national OA policies in RIO policy repositories (JRC Research and Innovation Observatory (RIO) policy repositories) **BI**

Priority	Input Indicator	Output Indicator	Outcome/Impact Indicator
International dimension outside ERA (Priority 6)	International co-publications with non-ERA partners per 1 000 researchers in the public sector (WoS and Eurostat) **C, L, H**	Non-EU doctorate students as a share of all doctorate students (EIS) **C, L, H**	Licence and patent revenues from abroad as a share of GDP (Eurostat) **S, T**

Note: The capital letters in colour refer to Table 4. The cells in light green represent Headline indicators while the cells in light grey hold EMM complementary indicators.
Source: Assembled by Science-Metrix from ERAC documentation

Aside from the biases, indicators may also contain drawbacks or limits that can be imposed by the data used in their construction or by their relevancy to effectively represent and monitor the progress of ERA members toward the achievement of an ERA priority. The following section aims to highlight the limits of each headline indicator and some of the EMM indicators.

The Headline indicator for priority 1, the adjusted research excellence, is a composite built of four components: highly cited publications, PCT patents, ERC grants and number of Marie Skłodowska-Curie Actions (MSCA) grants. Although the four components capture the main aspects of research effectiveness and excellence it can be argued that priority 1, more effective national research systems, needs more than these four components to be fully represented and evaluated. Moreover, the component highly cited publications has an intrinsic lag associated with it; a time window of two or three years is necessary to record the citations to scientific articles. Consequently, the latest available year for this indicator is 2013. One of the EMM indicators associated with priority 1 is the Summary Innovation Index (SII) from the European Innovation Scoreboard. This indicator provides an innovation performance score of the MS based on more than 25 components, which are divided across the following categories: Human resources, research systems, finance & support, firm investment, linkage & entrepreneurship, intellectual assets, innovators and economic effects. The principal concern about using this indicator to monitor priority 1 is that it includes many dimensions that are not covered by priority 1. Moreover, the primary goal of the SII is to measure the performance of both research and innovation systems rather than to measure the effectiveness of research systems alone.

Sub-priority 2a is represented by the Headline indicator GBARD allocated to Europe-wide transnational, as well as bilateral or multilateral, public R&D programmes per FTE researcher in the public sector. This indicator informs on the budgetary effort of governments toward joint programming processes and it reflects the transnational cooperation between governments. However, it might undervalue the real amount of transnational research budgets as many research programmes might include a transnational dimension while their associated funding might not be tagged with a transnational component. The funding of such research programmes will not be taken into account by the transnational GBARD. In addition, the indicator does not show how the transnational funding directly links with the European grand challenges. It also does not provide information on increases in government research expenditure. Lastly, this indicator may suffer from a potential country-size bias, as the lack of RIs or expert knowledge within the small countries may make them turn to larger countries in order to carry out research projects. Hence smaller countries tend to collaborate more internationally than larger ones, and thus their need for budget allocations to transnational R&D programmes may be different.

A similar country-size bias is applicable to the EMM indicator on internal co-publications with ERA partners and in this case another country bias may applied related to the location of a country. When a country is centrally located in Europe it tends to cooperate more with other European (or ERA) countries while a country located on the periphery of Europe will more evenly split its cooperation between European and non-European countries.

Sub-priority 2b focuses on a coordinated approach to transnational research infrastructures. In this context MS/AC are encouraged to actively participate in ESFRI projects and landmarks and reflect this participation in their RI national roadmap. The headline indicator of sub-priority 2b seeks to identify the MS/AC that have (a) a national roadmap, (b) identified ESFRI projects, and (c) corresponding investment needs. Although this indicator is very useful to identify MS/AC who effectively provided a roadmap regarding their RIs and who provide information on ESFRI projects in addition to their national RI, it does not cover the financial details of ESFRI investment

needs such as the quantity or the progress of financial investment. It is only a binary indicator. The availability of such roadmaps is a good starting point, but they are more a measure of communication effort than investment effort. Moreover, this indicator does not inform on the transnational accessibility to RIs, which is an important point when assessing optimal use of public investment in RIs.

As the Headline indicator for priority 3, the ERA committee selected Researcher's posts advertised through the EURAXESS job portal per 1 000 researchers in the public sector. This indicator fits very well with the priority, as EURAXESS is an open and transparent recruitment system. Moreover, it can serve to directly measure a country's institutions' willingness to be open about recruitment. However, since the portal only displays the job vacancies, there is no information to assess if the recruiting procedures are really merit based. In addition, some MS institutions may prefer to use national job portals and these vacancies will not be reported in EURAXESS. The jobs posted by private companies will also not show up on the portal. One of the accompanying EMM indicators for this priority is the share of researchers expressing satisfaction that the hiring procedures in their institution are open, transparent and merit based. This indicator, issued from data from the MORE2 survey directly, measures the openness, transparency and merit-based recruitment mechanisms as perceived by the researchers from the HEI themselves. With more than 10 000 respondents, the MORE2 survey was designed to provide maximum accuracy at the EU and country levels. The downsides of using this data are that they only cover researchers from the HEI (GOV not covered), and the survey is not carried out periodically so data availability is limited. The second EMM complementary indicator under priority 3 is the Share of doctoral candidates with a citizenship of another EU Member State. It was pointed out by the ERAC that this indicator is loosely connected to the priority (Open labour market for researchers); it does not refer directly to open, transparent and merit-based recruitment procedures but rather to the training of students from other EU countries. In this sense, it is somewhat more of a mobility measure than an open, transparent and merit-based recruitment measure.

Moving to priority 4, gender equality and gender mainstreaming in research, the Headline indicator is Share of women in grade A positions in HES. This indicator illustrates well the priority action of addressing gender imbalances in research institutions and decision-making bodies. It focuses on senior-level positions, which are positions that women may have more difficulty accessing than men. Also, the methodology to gather the data behind this indicator has been refined for more than a decade so one can expect the data to be very accurate and representative of the reality. On the downside, this indicator only covers positions in HEI, and does not provide information on the government or business enterprise sectors.

In addition, the interpretability of this indicator may not be obvious, because a higher score does not necessarily translate to a better situation. For example, a score above 50% may indicate a bias in the recruitment process for men. This raises the question of balance: What is the right balance? 50 %/50 %? Or should women in grade A positions be represented in the same proportion as women in the population? The answer is still unclear. Also, this indicator does not take into account or act as a proxy for monitoring the inclusion of the gender dimension in research content, which is another part of the top action priority. However, this aspect is covered by the EMM indicator Gender dimension in research content. Note that this indicator uses research articles (from the WoS) in which a gender dimension has been identified by a keyword query. The keyword query is not flawless (as is always the case when identifying articles by keywords queries) and a vocabulary bias might in theory affect the results. For example, if a keyword related to gender dimension has been omitted in the query and this word is used more often in one country relative to another country then this might result in a small imbalance. In addition, although the WoS covers a wide range of scientific journals, it almost exclusively indexes articles written in English, which may lead to a language bias (resulting in underrepresentation) for countries publishing more heavily in local- and non-English-language journals. The next EMM indicator of priority 4 is the Share of female PhD graduates. This indicator is issued from She Figures and therefore the data periodicity is triannual. Moreover, it doesn't cover the full aspect of gender imbalance since it accounts only for students and not working positions.

The Headline indicator for sub-priority 5a (knowledge circulation) is Share of product and/or process innovative firms cooperating with higher education institutions or public/private research institutions. This indicator acts as a good proxy to measure the collaboration level between private firms and HEI or research institutions. Moreover, it is readily available through the biannual Community Innovation Survey (CIS). One of the issues with this indicator is that the data do not distinguish between large and small or medium firms; however, larger firms are

known to collaborate more with HEI or research institutions because of their larger R&D capacities. Hence countries with a higher proportion of SMEs are likely to observe a bias linked to this indicator. In addition, the data from CIS do not distinguish the level or extent of cooperation and record the cooperation in a binary form (yes/no). Also, a slight change in the questionnaire was made to the 2012 edition (and will apply to future iterations) such that private research organisations are now taken into account (prior to 2012, only HEI and public research institutes were included in the cooperation). However, trend analysis performed for the ERA Progress Report did not show any significant discrepancies between the old and new data. The two EMM indicators for this priority are Number of public-private co-publications per million population and Share of public research financed by the private sector. Although the first of these is interesting in terms of public-private cooperation, it covers only one type of knowledge transfer: that which ends up in a co-publication. It misses other types of knowledge transfer such as that used for the creation of a new product or new process. The second EMM indicator may convey a misleading picture as MS/AC may have different approaches to bring R&D to industry. For example, some countries have put in place an established system with private (or semi-private) research organisations to provide commissioned R&D to industry.

Priority 5b, open access, is represented by the headline indicator Share of publications available in open access (green and gold). This indicator is produced by matching OA papers to the WoS database ; as a result, most limits of the WoS database are applicable here (e.g. coverage and language bias). In addition, a field of science bias may also apply; if OA is more predominant in a particular field of science. In this case, countries that publish in this field in greater proportion will benefit from a higher score. Although the data behind the indicator can easily be obtained annually (or even more frequently), there is a certain OA lag that obscures the most recent years. Indeed, some journals impose an embargo of a fixed time period (usually between 6 months and 2 years) on scientific publications before they officially become open. The maximal availability of OA paper is thereby shifted back by one or two years. The accompanying EMM indicator under this priority is the Presence or absence of national OA policies in RIO policy repositories. This indicator is reported in a binary form (yes/no), but when more than one policy is available it is indicated in the reporting table. However, countries are not benchmarked according to the number of policies present. This indicator is also based on the RIO policy repository, so it does not account for any policies not present there.

Lastly, priority 6, covered by the Headline indicator International co-publications with non-ERA partners per 1 000 researchers in the public sector, deals with the international dimension. This headline indicator is well suited to act as a proxy to assess the cooperation at the international level. Because it is built upon the WoS database, the WoS limitations must also be taken into consideration. Adding to these limitations, we can add the small-country bias (usually, smaller countries tend to be more international than larger ones), the location bias (countries situated at the periphery of Europe will have a larger propensity to collaborate with non-European countries) and the language bias (English-speaking countries may collaborate more internationally). Additionally, this indicator does not provide any information about the impact of the scientific publications. Finally, it was also pointed out by the ERAC that this indicator will react slowly to policy change since there is usually time lag between a policy being implemented, the funding mechanism engaging, and the publication of a research article. One of the EMM complementary indicators is Licence and patent revenues from abroad as a share of GDP. This indicator may suffer from an economical structure bias, whereas countries with a higher share of manufacturing industry will tend to have a higher patenting rate. The same applies to countries that host the headquarters of large companies, as most of the time the address on the patent will be that of the headquarters. Also, the sectors of operation can lead to patent imbalance; for instance, companies operating in pharma, biotech, ICT or software will tend to patent more than companies in other sectors.

5 QUALITY PLAN: VERIFICATION AND VALIDATION OF DATA

During the data gathering phase, the data quality was assured via a multi-faceted quality framework. The data quality framework involves various tests applicable on three dimensions: data relevance, data accuracy and data availability. Each indicator was evaluated by grading it for each dimension and by an overall assessment.

Relevance: the selection of indicators was highly influenced by the 2014 ERA Progress Report (as well as the Facts & Figures report). However, in the interests of reducing the survey burden,

some indicators were modified, merged together or added to the list of indicators to produce. The new indicators had to be relevant to at least one of the six ERA priorities to be added.

Accuracy: the accuracy of an indicator may be seen as the capacity of the indicator to adequately represent or describe the quantities it is designed to measure. We define two dimensions related to the accuracy: the data collection method and the degree of cross country standardisation. The former dimension was deemed fit if the data correctly estimated the quantities it was designed to represent. In other words, the accuracy of the data collection method evaluates how close the given values represent the (unknown) true values.

Since most of the data were collected from high-quality databases originating from international organisations, the European Commission and its agencies, or well-established bibliographic sources (i.e. WoS^{TM} and PATSTAT), one can expect the accuracy of the data collection method to be on par with the highest standard.

Next, data accuracy was also assessed for cross-country comparability. Data are said to be comparable across countries when the methods of data collection were the same or very similar across concerned countries.

To ensure data quality over the course of the project, additional validity checks were performed once the data were gathered and the indicators were computed. The two tests — namely, the detection of unreliable data point and coherence check — are explained below.

Availability: The availability of a particular dataset can be defined as the accessibility to data points for each country for a given time frame. Ideally, data would be available for each Member State at the benchmark year (2015).

Table 6 Dimensions of the data quality framework

	Depends on	Addressed by
RELEVANCE	• Relevance of indicator to the six ERA priorities in terms of content/policies perspective and the indicator's potential to adequately replace some of the ERA survey indicators	• Discussions with the Commission's officials • Identification of new indicators not relying on survey questions
ACCURACY OF DATA COLLECTION METHOD/COMPARABILITY	• Alignment between countries in reporting system, classifications used, etc., by data source	• Trying to rely as much as possible on existing official classifications and manuals for data collection (e.g. Frascati Manual); international standards; etc. • Validity/coherence checks after data gathering and computation of indicators
AVAILABILITY	• Availability of data up to benchmark year across ERA countries • Availability of secondary source databases	• Relying on international database offering EU coverage • Flagging system (to systematically register missing data)

Identification of unreliable data points through detection of outliers

Non-sampling errors (e.g. processing errors such as cleaning errors, wrong denominator in a share calculation, wrong units) could lead to inaccurate data points. In order to detect aberrations

in the time series, an automated test for detecting potential outliers was applied for each MS/AC time series for each indicator. A linear regression model was fitted in the time series of each country. Subsequently, a statistical procedure was applied to test the hypothesis that the studentised residual of each data point has been generated by the fitted model; when the *p*-value of a test was smaller than 0.05, the hypothesis was rejected implying that the data point is likely to be an outlier. Subsequently, the potential outliers were visually inspected by a seasoned analyst to assess the degree to which they may represent real variations; in other words, the data points that could represent real outliers (bad data or different definition) were differentiated from false outliers (data points likely representing real fluctuations). Note that this exercise is very complex, as data divergence may be caused by a precise political and/or economic condition unknown to the analyst performing this task. Therefore, actions were only taken on data points for which there was no ambiguity regarding their outlier status. When potential outliers were identified, the data source(s) used for the computation of the indicator was analysed to detect where the aberrant values might come from. Subsequently, the faulty data points were labelled with a flag. Note that only a small proportion of the outliers detected with the automated method were flagged as such.

Next, a method was applied to identify breaks in a time series and other possible outliers. A stepwise analysis was conducted via a script that compares the (y axis) difference between two successive points to the average (y axis) difference of the points before and after. This highlighted undocumented breaks in a time series or changes of regime that were not detected through the first method. Again, when the result was above a particular threshold, a manual validation was applied to each point before flagging them. Special care was taken when analysing data points flagged by the above validation procedures. For example, if a severe break in time series was detected, the reference period in the growth analysis could be shifted or reduced in order to avoid using data points that would lead to inaccurate results. Whenever a change was applied either in the reference period (for growth) or reference year (for performance), it is clearly indicated in the corresponding table notes.

Coherence checks

For data broken down in categories, totals were also available. When all of these data are available, coherence checks may be computed by summing each category and comparing the sum to the totals. For example, the sum of the GBARD funding modes should be equal to the total GBARD. Categories are defined regarding the following:

- Sex
- Institutional sectors
- Education levels (see the International Standard Classification of Education, UNESCO, 1997)
- R&D personnel categories
- Country aggregate (i.e. EU-28)

Table 7 lists the coherence checks that were applied to each dataset before the production of the indicators.

Table 7 List of coherence checks

Data	Verification Description	Verification Formula
Demographic data (table *demo_pjan and lfsi_emp_a*)	Verify population for both genders	{Men} + {Women} = {Total}
	Verify population for all countries	{AT} + {BE} + {BG} + {CY} + {CZ} + {HR} + {DE} + {DK} + {EL} + {EE} + {ES} + {FI} + {FR} + {HU} + {IE} + {IT} + {LT} + {LV} + {LU} + {MT} + {NL} + {PL} + {PT} + {RO} + {SE} + {SI} + {SK} + {UK} = {EU-28}
Employment data (table *rd_p_persocc*)	Verify R&D personnel for all sectors	{Business enterprise sector} + {Government sector} + {Higher education sector} + {Private non-profit sector} = {All sectors}
	Verify R&D personnel for all occupations	{Researchers} + {Technicians / equivalent staff} + {Other supporting staff} = {Total R&D personnel}
Education data (tables *educ_uoe_enrt01* and *educ_uoe_mobs02*)	Verify enrolment for all tertiary cycles	{Short cycle tertiary education} + {Bachelor's or equivalent level} + {Master's or equivalent level} + {Doctoral or equivalent level} = {Tertiary education (levels 5–8)}
	Verify enrolment for public and private sectors	{Public institutions} + {Private institutions} = {Total}
	Verify enrolment for both genders	{Men} + {Women} = {Total}
Economic data (table *nama_10_gdp*)	Verify GDP for all countries	{AT} + {BE} + {BG} + {CY} + {CZ} + {HR} + {DE} + {DK} + {EL} + {EE} + {ES} + {FI} + {FR} + {HU} + {IE} + {IT} + {LT} + {LV} + {LU} + {MT} + {NL} + {PL} + {PT} + {RO} + {SE} + {SI} + {SK} + {UK} = {EU-28}
R&D expenditure data (tables *gba_nabsfin07*, *gba_fundmod* and *gba_tncoor*)	Verify that GBARD is the same in both tables	{Total R&D appropriations} in *gba_nabsfin07* = {Total R&D appropriations} in *gba_fundmod*
	Verify GBARD for all appropriations	{Agriculture} + {Culture, recreation, religion and mass media} + {Defence} + {Education} + {Energy} + {Environment} + {Exploration and exploitation of space} + {Exploration and exploitation of the earth} + {R&D financed from General University Funds (GUF)} + {R&D financed from other sources than GUF} +

Data	Verification Description	Verification Formula
		{Health} + {Industrial production and technology} + {Political and social systems, structures and processes} + {Transport, telecommunication and other infrastructures} = {Total R&D appropriations}
	Verify GBARD for all countries	{AT} + {BE} + {BG} + {CY} + {CZ} + {HR} + {DE} + {DK} + {EL} + {EE} + {ES} + {FI} + {FR} + {HU} + {IE} + {IT} + {LT} + {LV} + {LU} + {MT} + {NL} + {PL} + {PT} + {RO} + {SE} + {SI} + {SK} + {UK} = {EU-28}
	Verify GBARD for all funding purposes	{Institutional funding} + {Project funding} = {Total R&D appropriations}
	Verify funding for all transnationally coordinated R&D	{National contributions to transnational public R&D performers} + {National contributions to Europe-wide transnational public R&D programmes} + {National contributions to bilateral or multilateral public R&D programmes} = {National public funding to transnationally coordinated R&D}
	Verify GBARD/capita	{GBARD per capita} = {GBARD}/{capita}
	Verify GBARD/GDP	{GBARD per GDP} = {GBARD}/{GDP}
Intramural R&D expenditure (table *rd_e_gerdfund)*	Verify funding for all sectors	{Business enterprise sector} + {Government sector} + {Higher education sector} + {Private non-profit sector} = {All sectors}
	Verify funding from all sectors	{Abroad} + {Business enterprise sector} + {Government sector} + {Higher education sector} + {Private non-profit sector} = {All sectors}
	Verify funding for all countries	{AT} + {BE} + {BG} + {CY} + {CZ} + {HR} + {DE} + {DK} + {EL} + {EE} + {ES} + {FI} + {FR} + {HU} + {IE} + {IT} + {LT} + {LV} + {LU} + {MT} + {NL} + {PL} + {PT} + {RO} + {SE} + {SI} + {SK} + {UK} = {EU-28}
Enterprises data (tables *inn_cis8_bas* and	Verify the number of enterprises for all countries	{AT} + {BE} + {BG} + {CY} + {CZ} + {HR} + {DE} + {DK} + {EL} + {EE} + {ES} + {FI} + {FR} + {HU} +

Data	Verification Description	Verification Formula
inn_cis8_coop)		{IE} + {IT} + {LT} + {LV} + {LU} + {MT} + {NL} + {PL} + {PT} + {RO} + {SE} + {SI} + {SK} + {UK} = {EU-28}

Most of the tables passed the coherence checks. Note that some very small discrepancies (0.999 < ratio between expected value and real value < 1.001) were not reported as the impact associated with them is minimal. The following section lists the table and the coherence checks applied to each of them with the results. For the values that did not pass the coherence checks, an explanation or the action taken to correct the table is given.

demo_pjan table

Sum male-female: ok

Sum MS: ok

Educ_uoe_enrt01

Sum ICSED categories: ok

Sum Private-public institutions: ok

Sum male-female: minor differences (see Table 8)

Table 8 Incoherence for total male + female (Educ_uoe_enrt01)

Geo	Time	ISCED11	Sector	Males	Females	Total (from table)	Male + female	Difference	Ratio
Austria	2013	Doctoral or equivalent level	Private institutions	155	202	356	357	1	1.003
Austria	2014	Doctoral or equivalent level	Private institutions	153	188	342	341	-1	0.997

Source: Science-Metrix

Action: No action taken, the difference is only 1 and the small numbers for the total makes the ratio fall within the decided boundaries but the effect still remains very small.

educ_uoe_mobs02

Sum ISCED categories: minor differences (see Table 9).

Table 9 Incoherence for total ISCED categories (educ_uoe_mobs02)

Geo	Time	Partner	Sex	Short-cycle tertiary education	Bachelor's or equivalent level	Master's or equivalent level	Doctoral or equivalent level	Tertiary education (levels 5-8)	Sum	Difference	Ratio
Austria	2013	Africa	Females	1	139	121	40	300	301	1	1.003
Austria	2013	Africa	Males	15	262	246	129	651	652	1	1.002
Austria	2013	Central and South America	Males	11	181	159	68	418	419	1	1.002
Austria	2013	Northern America	Females	8	199	221	71	500	499	-1	0.998
Austria	2013	Northern America	Males	10	172	164	75	420	421	1	1.002
Austria	2013	Oceania	Females	5	25	56	8	93	94	1	1.011
Austria	2014	Africa	Females	6	124	118	40	287	288	1	1.003
Austria	2014	Central and South America	Females	3	193	186	52	433	434	1	1.002
Austria	2014	Central and South America	Males	12	160	168	64	403	404	1	1.002
Austria	2014	Northern America	Males	8	70	186	72	337	336	-1	0.997

Source: Science-Metrix

Action: No action taken, the difference is only 1 and the small numbers for the total makes the ratio fall within the decided boundaries but it the effect still remains very small.

Sum male-female: minor differences (see Table 10**Error! Reference source not found.**)

Table 10 Incoherence for total male-female (educ_uoe_mobs02)

Geo	Time	Partner	ISCED11	Male	Female	Total	Male + female	Difference	Ratio
Austria	2013	Africa	Bachelor's or equivalent level	262	139	400	401	1	1.003
Austria	2013	Africa	Tertiary education (levels 5-8)	651	300	952	951	-1	0.999
Austria	2013	Central and South America	Bachelor's or equivalent level	181	203	383	384	1	1.003
Austria	2013	Central and South America	Master's or equivalent level	159	174	332	333	1	1.003
Austria	2013	Central and South America	Tertiary education (levels 5-8)	418	456	873	874	1	1.001
Austria	2013	Northern America	Tertiary education (levels 5-8)	420	500	919	920	1	1.001
Austria	2013	Oceania	Short-cycle tertiary education	2	5	6	7	1	1.167
Austria	2014	Africa	Doctoral or equivalent level	101	40	140	141	1	1.007
Austria	2014	Asia	Doctoral or equivalent level	584	403	986	987	1	1.001
Austria	2014	Central and South America	Bachelor's or equivalent level	160	193	352	353	1	1.003
Austria	2014	Central and South America	Doctoral or equivalent level	64	52	115	116	1	1.009
Austria	2014	Northern America	Master's or equivalent level	186	196	383	382	-1	0.997
Austria	2014	Northern America	Tertiary education (levels 5-8)	337	364	700	701	1	1.001

Source: Science-Metrix

Action: No action taken, the difference is only 1 and the small numbers for the total makes the ratio fall within the decided boundaries but it the effect still remains very small.

gba_fundmod

Sum institutional funding + project funding: ok except for some minor inconsistencies for Germany and Turkey (see Table 11).

Action: The inconsistencies for Germany might be explained by the fact that the flag "d" accompanies the values (meaning: definition differs) while the 2014 data points for Turkey are flagged with p (provisional). We decided to keep those values as is.

Table 11 Incoherence for sum institutional and project funding (gba_fundmod)

Geo	Time	Unit	GBAORD - institutional funding	GBAORD - project funding	Total R&D appropriations	Sum inst. + proj. funding	Difference	Ratio
Germany	2011	Million euro	15054.50	8844.60	23743.50	23899.10	155.5996	1.007
Germany	2012	Million euro	15591.40	8648.00	24070.20	24239.40	169.2012	1.007
Germany	2013	Million euro	16254.96	9355.20	25371.00	25610.16	239.1641	1.009
Germany	2014	Million euro	16508.96	9200.85	25363.50	25709.80	346.3027	1.014
Germany	2015	Million euro	17011.75	9294.52	25902.10	26306.27	404.166	1.016
Turkey	2014	Million euro	1498.99	403.09	1930.53	1902.09	-28.44604	0.985

Source: Science-Metrix

Sum of MS: ok

gba_nabsfin07

Comparison of between table gba_nabsfin07 and gba_fundmod: ok

Sum NABS07: ok except for the years and MS that do not have all the NABS07 values but have the total. Since we use only the total R&D appropriation we decided not to correct this table.

Sum MS: ok

R&D allocations / population = R&D appropriations per capita: ok

R&D allocations / GDP = Percentage of gross domestic product (GDP): ok

gba_tncoor

Sum of nabs07: minor differences (see Table 12)

Table 12 Incoherence for sum transnational and multilateral public R&D (gba_tncoor)

Geo	Time	Unit	National contributions to transnational public R&D performers	National contributions to Europe-wide transnational public R&D programmes	National contributions to bilateral or multilateral public R&D programmes	National public funding to transnationally coordinated R&D	Sum trans. and multi. public R&D prog.	Difference	Ratio
Latvia	2008	Million euro	0	0.665	0.04	0.704	0.705	0.001	1.001
Latvia	2009	Million euro	0	0.806	0.011	0.818	0.817	-0.001	0.999
Lithuania	2011	Million euro	0	0.765	0.041	0.805	0.806	0.001	1.001

Source: Science-Metrix

Action: No action taken, the differences are very small and will have no impact on the indicators.

Nabs as % total GBAORD: ok

inn_cis8_bas

sum of MS: ok

inn_cis8_coop

sum of MS: ok

lfsi_emp_a table

Sum male-female: minor differences (see Table 13).

Table 13 Incoherence for sum male plus female (lfsi_emp_a)

Geo	Year	Male	Female	Male + Female	Total (from table)	Difference	Ratio
Cyprus	2005	199	159	358	357	1	1.003
Cyprus	2006	202	164	366	365	1	1.003
Cyprus	2009	207	185	392	393	-1	0.997
Cyprus	2011	219	202	421	420	1	1.002
Cyprus	2012	223	204	427	426	1	1.002
Estonia	2014	336	313	649	648	1	1.002
Macedona	2008	556	355	911	912	-1	0.999
Macedona	2012	569	367	936	935	1	1.001
Macedona	2014	576	377	953	954	-1	0.999
Iceland	2007	95	79	174	173	1	1.006
Iceland	2009	92	81	173	172	1	1.006
Luxembourg	2007	118	94	212	211	1	1.005
Luxembourg	2008	120	92	212	213	-1	0.995
Luxembourg	2010	128	100	228	229	-1	0.996
Luxembourg	2012	137	110	247	246	1	1.004
Malta	2005	110	50	160	159	1	1.006
Malta	2006	111	51	162	161	1	1.006
Malta	2010	113	60	173	172	1	1.006
Malta	2012	113	67	180	179	1	1.006
Slovenia	2006	537	461	998	997	1	1.001
Slovenia	2011	540	459	999	998	1	1.001

Source: Science-Metrix

Action: No action taken, the difference is only 1 and the small numbers for the total make the ratio fall within the decided boundaries but the effect still remains very small.

Sum MS: For years 2005-2013, France is not present individually in the table but it is included in the total of EU-28. For year 2014-2015, France has provided data but there still is a small difference between the sum of MS and EU-28 (see Table 14)

Table 14Table 14 Incoherence for sum MS equal EU28 (lfsi_emp_a)

Year	Dataset	Sum EU28	EU28 (from table)	Difference	Ratio
2005	Females	89 871	102 638	-12767	0.876
2005	Males	112 177	126 603	-14426	0.886
2005	Total	202 046	229 241	-27195	0.881
2006	Females	91 226	104 115	-12889	0.876
2006	Males	113 118	127 595	-14477	0.887
2006	Total	204 338	231 709	-27371	0.882
2007	Females	92 056	105 107	-13051	0.876
2007	Males	113 777	128 304	-14527	0.887
2007	Total	205 834	233 410	-27576	0.882
2008	Females	93 200	106 361	-13161	0.876
2008	Males	114 571	129 169	-14598	0.887
2008	Total	207 772	235 531	-27759	0.882
2009	Females	93 920	107 234	-13314	0.876
2009	Males	114 306	128 992	-14686	0.886
2009	Total	208 232	236 226	-27994	0.881
2010	Females	93 488	106 859	-13371	0.875
2010	Males	113 199	127 912	-14713	0.885
2010	Total	206 689	234 770	-28081	0.880
2011	Females	94 024	107 397	-13373	0.875
2011	Males	112 859	127 534	-14675	0.885
2011	Total	206 879	234 931	-28052	0.881
2012	Females	94 968	108 433	-13465	0.876
2012	Males	113 159	127 927	-14768	0.885
2012	Total	208 122	236 360	-28238	0.881
2013	Females	95 400	108 971	-13571	0.875
2013	Males	113 037	127 811	-14774	0.884
2013	Total	208 437	236 782	-28345	0.880
2014	Females	109 836	109 465	371	1.003
2014	Males	128 248	127 865	383	1.003
2014	Total	238 081	237 330	751	1.003
2015	Females	110 017	109 644	373	1.003
2015	Males	128 350	127 970	380	1.003
2015	Total	238 367	237 614	753	1.003

Source: Science-Metrix

nama_10_gdp

Sum MS: ok

rd_p_persocc

sum of sector: minor differences (data not shown)

Action: No action taken, the difference is only 1 and the small numbers for the total make the ratio fall within the decided boundaries but the effect still remains very small.

Sum of occupation: some minor differences except for Sweden where we observed larger discrepancies between the total R&D personnel and the sum of occupations for the year 2007 (see Table 15).

Table 15 Incoherence for total occupation (rd_p_persocc)

Geo	Time	Sectperf	Sex	Unit	Researchers	Technicians / equivalent staff	Other supporting staff	Total R&D personnel	Verif	Diff	Ratio
Iceland	2013	All sectors	Total	Full-time equivalent (FTE)	1950	481	332	2766	2763	-3	0.999
Slovenia	2012	Government sector	Females	Full-time equivalent (FTE)	871	248	178	1295	1297	2	1.002
Sweden	2007	All sectors	Total	Full-time equivalent (FTE)	45812	17849	11277	75318	74938	-380	0.995
Sweden	2007	Government sector	Females	Full-time equivalent (FTE)	753	99	292	1172	1144	-28	0.976
Sweden	2007	Government sector	Total	Full-time equivalent (FTE)	1941	284	649	3253	2874	-379	0.883
Sweden	2009	Government sector	Females	Full-time equivalent (FTE)	517	183	198	896	898	2	1.002
Sweden	2011	Private non-profit sector	Total	Full-time equivalent (FTE)	194	0	68	260	262	2	1.008

Source: Science-Metrix

Action: Given that we only use the researcher portion of occupation and that the incoherencies are small we decided to keep the data as is.

rd_e_gerdfund

Sum of sectors (for SECTPERF): minor differences (see Table 16).

Action: No action taken, the differences are very small and will have no impact on the indicators.

Table 16 Incoherence for sum sector (SECTPERF) (rd_e_gerdfund)

Geo	Time	Sectfund	Unit	Business enterprise sector	Government sector	Higher education sector	Private non-profit sector	All sectors	Sum sectors	Difference	Ratio
Romania	2008	Private non-profit sector	Million euro	0	0.082	0.136	0.013	0.217	0.231	0.014	1.065
Serbia	2014	Private non-profit sector	Million euro	0.013	0.026	0.001	0.002	0.04	0.042	0.002	1.050
Czech Republic	2012	Private non-profit sector	Million euro	0.002	0	0.032	0.032	0.065	0.066	0.001	1.015
Czech Republic	2005	Private non-profit sector	Million euro	0.007	0.001	0.008	0.057	0.072	0.073	0.001	1.014
Slovenia	2008	Private non-profit sector	Million euro	0.012	0.018	0.037	0.007	0.073	0.074	0.001	1.014
Czech Republic	2009	Abroad - Private non-profit sector	Million euro	0	0.004	0.044	0.031	0.078	0.079	0.001	1.013
Croatia	2005	Abroad - Private non-profit sector	Million euro	0	0.068	0.014	0	0.081	0.082	0.001	1.012
Serbia	2010	Private non-profit sector	Million euro	0.018	0.026	0.028	0.024	0.097	0.096	-0.001	0.990
Serbia	2013	Private non-profit sector	Million euro	0.023	0.031	0.019	0.001	0.075	0.074	-0.001	0.987
Malta	2010	Private non-profit sector	Million euro	0.032	0	0.019	0	0.052	0.051	-0.001	0.981
Slovakia	2009	Abroad - Private non-profit sector	Million euro	0	0.014	0.001	0	0.016	0.015	-0.001	0.938

Source: Science-Metrix

Sum of sectors (for SECTFUND): minor differences (data not shown).

Action: No action taken, the differences are very small and will have no impact on the indicators.

Share of government spending as percentage of GDP: ok (it also involves table nama_10_gdp)

Additional data considerations

Rounding error

In some cases, the row or column totals do not match the sum of the data. This may be due to rounding error.

Cut-off date

At the beginning of the project a cut-off date for each Eurostat table was established in collaboration with the Commission in order to maximise the chance of having the most up-to-date data while not delaying the project (see Table 17). All data were extracted at a time past the cut-off date. The project lasted for several months and therefore it is possible that some data source might have been updated between the cut-off date originally planned and the release of the ERA Progress Report.

Table 17 Eurostat table cut-off dates

Eurostat Table	Cut-off date	Extraction date
Community innovation survey		
online data code inn_cis8_coop	1-Mar-16	7-Jul-16
online data code inn_cis8_bas	1-Mar-16	7-Jul-16
Statistics on research and development		
online data code gba_fundmod	15-Mar-16	18-Mar-16
online data code gba_tncoor	15-Mar-16	12-May-16
Government budget appropriations or outlays on R&D		
online data code gba_nabsfin07	15-Mar-16	13-May-16
online data code gba_nabste	15-Mar-16	14-Apr-16
Annual national accounts		
online data code nama_10_gdp	15-Mar-16	14-Apr-16
Learning mobility		
online data code educ_uoe_mobs02	1-May-16	11-May-16
Statistics on research and development		
online data code rd_p_persocc	1-Apr-16	14-Apr-16
Participation in education and training		
online data code educ_uoe_enrt01	1-May-16	11-May-16
Annual government finance statistics		
online data code gov_10a_main	1-May-16	6-Jun-16
Population		
online data code demo_pjan	1-May-16	12-May-16
Employment and unemployment		
online data code lfsi_act_a	1-May-16	12-May-16

Source: Science-Metrix

6 DESK RESEARCH AND DOCUMENT REVIEWS

Desk research and document review provided the framework for the present project, situating the assessment exercise in the policy context of the movement towards an ever-more integrated European Research Area. The work included the analysis of documents at the level of stakeholder bodies representing several or more organisations and individual RPOs and RFOs, as well as at the level of regions and individual MS/AC. Efforts were made to identify and document examples of good practice, in particular for assessing institutional change at organisational level, as required for the completion of the ERA. Specifically, the study team conducted an initial review of documentation provided by the Commission with the intention of identifying preliminary evidence on progress towards the completion of the ERA. Additional information from sources other than the Commission were also considered. Refer to Table 18 for a list of the main sources used in the desk research and document review (additional documents are listed in this report's bibliography).

Table 18 Main sources used in the desk research and document review

Category	Number of documents
National Level	
ERAWATCH Country Reports	5
National Action Plans	27
OECD Policy reviews	1
Research and Innovation Observatory (RIO) Country Reports	18 (2014) 28 (2015)
Researchers' Report 2014: Country profiles	5
Organisation level	
Conference of European Schools for Advanced Engineering	7
ERAC documents	2
European Association of Research and Technological Organisations	2
European Commission – DG-RI	12
European Commission reports	3
European University Association	9
League of European Research Universities	7
Science Europe	6
Other sources (Reports from research organisations, academic/opinion papers)	2

Source: Compiled by Science-Metrix

Note that the document review established important contextual components for the subsequent interviews with key stakeholders, as well as the quantitative measurements of national- and ERA-level performance; one primary focus of this research is to deepen understanding of the ERA priorities, as these provide the primary structure for the assessment exercise at hand.

7 INTERVIEWS

Interviews (conducted by telephone) provided important findings from a variety of perspectives to facilitate interpretation of quantitative data, as well as the assessment of features of the ERA project that are not tracked by quantitative measures. Among other findings, these interviews provided insights into the benefits, difficulties and limitations that organisations are facing in implementing ERA initiatives and policies. In total, nearly 90 interviews were conducted with key members of stakeholder research funding organisations (RFOs) and research performing organisations (RPOs) from countries across the ERA, the chairs of the ERA-related groups and representatives of the ERA stakeholders' organisations and candidate organisations.

ANNEX 1: CHANGES TO INTERNATIONAL CLASSIFICATION STANDARDS

International Standard Classification of Education (ISCED)

The International Standard Classification of Education (ISCED) is the UN framework for classifying educational programmes at different levels. As part of revisions to the framework in 2011, new categories of tertiary education were introduced (OECD, European Union and UNESCO Institute for Statistics, 2015). However, the ISCED 1997 (ISCED-97) categories were used to ensure that data were available for a sufficiently long period to analyse trends; Eurostat data based on ISCED 2011 only cover data in 2013 and 2014. The ISCED-97 categories recognise two stages of tertiary education:

The first stage **(ISCED 5)** includes largely theory-based programmes to provide sufficient qualifications to gain entry to advanced research programmes and professions with high skills requirements (**ISCED 5A**) and programmes that are generally practically, technically or occupationally specific (**ISCED 5B**).

The second stage (**ISCED 6**) leads to the award of an advanced research qualification (e.g. PhD, non-PhD programmes with an advanced research component). The programmes are devoted to advanced study and original research.

Table 19 Correspondence between ISCED 2011 and ISCED 1997 levels

ISCED 2011	ISCED 1997
ISCED 01	
ISCED 02	ISCED 0
ISCED Level 1	ISCED Level 1
ISCED Level 2	ISCED Level 2
ISCED Level 3*	ISCED Level 3
ISCED Level 4*	ISCED Level 4
ISCED Level 5	ISCED Level 5
ISCED Level 6	
ISCED Level 7	
ISCED Level 8	ISCED Level 6

Note: * Content of category has been modified slightly.
Source: Reproduced from http://www.uis.unesco.org/Education/Documents/isced-2011-en.pdf

Change in GBAORD naming

Following the release of the 2015 Frascati manual (OECD, 2015) the government budget appropriations and outlays on R&D (GBAORD) was renamed to government budget allocations on R&D (GBARD). This change has yet to be implemented in all offices (for example, Eurostat still uses GBAORD), but the new naming is reflected through the ERA Monitoring Handbook and the ERA Monitoring Report.

ANNEX 2: LIST OF KEY TERMS

European Research Area (ERA)

Includes the 28 Member States of the European Union (EU-28) and 13 Associated Countries:

- Belgium (BE)
- Bulgaria (BG)
- Czech Republic (CZ)
- Denmark (DK)
- Germany (DE)
- Estonia (EE)
- Ireland (IE)
- Greece (EL)
- Spain (ES)
- France (FR)
- Croatia (HR)
- Italy (IT)
- Cyprus (CY)
- Latvia (LV)
- Lithuania (LT)
- Luxembourg (LU)
- Hungary (HU)
- Malta (MT)
- Netherlands (NL)
- Austria (AT)
- Poland (PL)
- Portugal (PT)
- Romania (RO)
- Slovenia (SI)
- Slovakia (SK)
- Finland (FI)
- Sweden (SE)
- United Kingdom (UK)
- Iceland (IS)
- Norway (NO)
- Switzerland (CH)
- Montenegro (ME)
- FYR Macedonia (MK)
- Albania (AL)
- Republic of Serbia (RS)
- Turkey (TR)
- Bosnia and Herzegovina (BA)
- Israel (IL)
- Faroe Islands (FO)
- Republic of Moldova (MD)
- Ukraine

Full counting of publications and patents

Each publication or patent application is counted once for each entity (e.g. country, institution, author) appearing in the publication's author affiliations, or in the patent application's inventor addresses. For example, if a publication is authored by one author from the US, two authors from the UK and one author from France, it would be counted once for each country even though the UK appears twice in the author affiliations. The same principal applies for the full counting of co-publications and co-inventions.

Fractional counting of publications and patents

Typically, publications are counted using full counting, whereby each publication is counted only once in each institution/country/world region regardless of the number of authors from that institution/country/world region. This means that a publication between a French, a German and a Canadian researcher would count once for France, once for Germany, and once for the EU-28 as an ERA publication. In some cases where the number of publications is normalised by another metric such as the number of researchers, full counting creates an asymmetry between the numerator and denominator when aggregating the data at the regional level (i.e. EU-28). In the above example, the publication by France and Germany would not add up together at the EU-28 level using full counting (it would be counted only once), while the number of French and German researchers would add up in the denominator. Summing the publications across countries would not work either since the sum across EU-28 countries would add up to more publications than there are in practice; in the above example, there would be two publication counts for the EU-28, although there is only one publication. To circumvent this issue, fractional counting of publications has been used where appropriate.

The fractional counting of publications prevents a single paper from being counted multiple times; the sum of fractions across all papers and countries will add up to the number of world papers in the reference database. A fraction of each publication is equally distributed among all author addresses, which can then be codified by author, institution or country depending on the aggregation level at which the data are produced. For example, if a publication is authored by one author from the US, two authors from the UK and one author from France, this publication would be counted 0.25 times for the US, 0.5 times for the UK and 0.25 times for France. At the EU-28

level, the fraction of the publication that would be counted would amount to 0.75 (the sum of fractions across Member States). The same principal applies to the fractional counting of PCT patent applications based on the inventor field.

Fractional counting of co-publications

Typically, co-publications are counted using full counting, whereby each co-publication is counted only once in each institution/country/world region regardless of the number of authors from that institution/country/world region. This means that a co-publication between a French, a German and a Canadian researcher would count once for France, once for Germany, and once for the EU-28 as an ERA co-publication, although the sum across EU-28 countries would amount to two ERA co-publications (i.e. the sum of France and Germany). Because such an asymmetry is not present for researchers — i.e. the sum of researchers across Member States is equal to the total number of EU-28 researchers — the number of co-publications with ERA partners per FTE researcher will be underestimated for the EU-28 as a whole relative to individual Member States when using full counting. Also note that counting co-publications involving at least two ERA countries by considering the whole EU-28 as one large country is conceptually problematic since the EU-28 is not a country but a region embedding multiple ERA countries. Thus, co-publications involving at least two ERA countries have been counted using fractional counting so that the sum of co-publication fractions across countries equals the total number of publications at the world level, making it possible to sum the number of ERA co-publications and researchers in a symmetrical fashion at any aggregation level.

For a co-publication between a French, a German and a Canadian researcher, there are six bilateral links to be taken into account (i.e. DE–FR, DE–CA, FR–DE, FR–CA, CA–DE and CA–FR) since the co-publication must be counted in the perspective of each country (i.e. each link must have its reciprocal link taken into account). Each link is attributed an equal fraction of the publication; in this case the fraction for each link equals 1/6. Of those links, four correspond to a co-publication between two ERA countries (i.e. DE–FR, DE–CA, FR–DE, FR–CA) such that two thirds (or 4/6) of this publication would count as an ERA co-publication when aggregating at the ERA level (i.e. pooled ERA countries). Since both Germany and France are Member States, the number of co-publications for the EU-28 would also amount to a fraction of two thirds. For individual countries, only half of the links including them must be counted; only the links corresponding to their perspective (i.e. the ones where a country appears first, although the ones where a country appears last would also work) should be counted. Thus, for Germany and France, one third (or 2/6) of this publication would count as an ERA co-publication.

ANNEX 3: INDEX LIST OF INDICATORS

Priority 1 – More effective national research systems

Headline indicator

— Adjusted Research Excellence Indicator (REI)

EMM indicators

— GBARD as a percentage of GDP

— European Innovation Scoreboard Summary Innovation Index (SII)

Additional priority 1 indicators

— GBARD as a percentage of government expenditures

— Percentage of GBARD allocated as project based funding

— Researchers per 1 000 active population

— R&D tax incentives as a proportion GBARDP1 – Additional indicator – Percentage of GBARD allocated

— Number of patent applications per 1 000 researchers

— Number of publications per 1 000 researchers in the public sector

Priority 2a – Transnational cooperation

Headline indicator

— GBARD allocated to transnational cooperation per researcher in the public sector

EMM indicators

— Participation in Public-to-public partnerships per researcher in the public sector

— International co-publications with ERA partners per 1 000 researchers in the public sector

Additional priority 2 indicator

— International co-invention rate with ERA partners

— International co-publication rate with ERA partners

Priority 2b – European Strategy Forum for Research Infrastructures (ESFRI)

Headline indicator:

— Availability of national roadmaps with identified ESFRI projects and corresponding investment needs

EMM indicators

— Share of developing ESFRI Projects in which a Member State or an Associated Country participates

— Share of operational ESFRI Landmarks in which a Member State or an Associated Country is a partnerP1 – EMM indicator – European Innovation Scoreboard

Priority 3 – Open labour market for researchers

Headline indicator:

— Number of researcher postings advertised through the EURAXESS job portal, per 1 000 researchers in the public sector

EMM indicators:

— Share of doctoral candidates with a citizenship of another EU Member State

— Share of researchers expressing satisfaction that the hiring procedures in their institution are open, transparent and merit-basedP1 – EMM indicator – European Innovation Scoreboard

Priority 4 – Gender equality and gender mainstreaming in research

Headline indicator

— Share of women in Grade A positions in HES

EMM indicators

— Gender dimension in research content

— Share of female PhD graduatesP1 – EMM indicator – European Innovation Scoreboard

Additional priority 4 indicators

— Share of women researchers

— Share of women heads of institutions in the Higher Education Sector

Priority 5a – Optimal circulation, access to and transfer of scientific knowledge

Headline indicator

— Share of product and/or process innovative firms cooperating with higher education institutions or public/private research institutions

EMM indicators

— Share of public research financed by the private sector

— Number of public-private co-publications per million population

Priority 5b – Optimal circulation, access to and transfer of scientific knowledge

Headline indicator

— Share of publications available in open access (green and gold)

EMM indicator

— Open access policies in national action plans

Priority 6 – International cooperation

Headline indicator

—International co-publications with ERA partners per 1 000 researchers in the public sector

EMM indicators

— Non-EU doctorate students as a share of all doctorate students

— Licence and patent revenues from abroad as a share of GDP

Additional priority 6 indicators

— International co-publication rate with non-ERA partnersP2a – Additional indicator – International co-publication rate with ERA partners

— International co-invention rate with non-ERA partners

REFERENCES

Archambault É., Caruso J., and Beauchesne O. (2011). Towards a Multilingual, Comprehensive and Open Scientific Journal Ontology, in Noyons, B., Ngulube, P. and Leta, J. *Proceedings of the 13th International Conference of the International Society for Scientometrics and Informetrics (ISSI)*, Durban, South Africa, pp 66-77.

Archambault, É., Côté, G., Struck, B., and Voorons, M. (2016). *Research Impact of Paywalled versus Open Access Papers*. http://www.1science.com/oanumbr.html.

Bradbury, I.R., et al. (2015). Genetic evidence of local exploitation of Atlantic salmon in a coastal subsistence fishery in the Northwest Atlantic. *Canadian Journal of Fisheries and Aquatic Sciences*, 72: 83-95.

DG Education and Culture (2011), *Supporting growth and jobs—An agenda for the modernization of Europe's higher education systems* [pdf], Publications Office of the European Union, Luxembourg, available at http://ec.europa.eu/education/library/policy/modernisation_en.pdf.

DG Internal Market, Industry, Entrepreneurship and SMEs (2016), *European Innovation Scoreboard 2016* [pdf], Publications Office of the European Union, Luxembourg, available at http://ec.europa.eu/DocsRoom/documents/17822/attachments/1/translations/en/renditions/native

DG Research (2008), *Mapping the maze: Getting more women to the top in research* [pdf], Office for Official Publications of the European Communities, Luxembourg, available at http://ec.europa.eu/research/science-society/document_library/pdf_06/mapping-the-maze-getting-more-women-to-the-top-in-research_en.pdf.

DG Research (2009a), *She Figures 2009* [pdf], Publications Office of the European Union, Luxembourg, available at http://ec.europa.eu/research/science-society/document_library/pdf_06/she_figures_2009_en.pdf.

DG Research (2009b), *Women in science and technology: Creating sustainable careers* [pdf], Office for Official Publications of the European Communities, Luxembourg, available at http://ec.europa.eu/research/science-society/document_library/pdf_06/wist2_sustainable-careers-report_en.pdf.

DG Research and Innovation (2012), *Structural change in research institutions: Enhancing excellence, gender equality and efficiency in research and innovation* [pdf], Publications Office of the European Union, Luxembourg, available at http://ec.europa.eu/research/science-society/document_library/pdf_06/structural-changes-final-report_en.pdf.

DG Research and Innovation (2014), *European Research Area—Progress Report 2014* [pdf], Publications Office of the European Union, Luxembourg, available at http://ec.europa.eu/research/era/pdf/era_progress_report2014/era_progress-report_150521.pdf.

DG Research and Innovation (2015), *European Research Area—Facts and Figures 2014* [pdf], Publications Office of the European Union, Luxembourg, available at http://ec.europa.eu/research/era/pdf/era_progress_report2014/era_facts&figures_2014.pdf.

DG Research and Innovation (2016a), *She Figures Handbook 2015* [pdf], Publications Office of the European Union, Luxembourg, available at http://ec.europa.eu/research/swafs/pdf/pub_gender_equality/she_figures_2015_Handbook_final.pdf.

DG Research and Innovation (2016b), *She Figures 2015* [pdf], Publications Office of the European Union, Luxembourg, available at https://ec.europa.eu/research/swafs/pdf/pub_gender_equality/she_figures_2015-final.pdf.

EGGE (European Commission's Expert Group on Gender and Employment) (2009), *Gender segregation in the labour market: Root causes, implications and policy responses in the EU* [pdf], Publications Office of the European Union, Luxembourg, available at http://ec.europa.eu/social/main.jsp?catId=738&langId=en&pubId=364&furtherPubs=yes.

Einav, L. and Levin, J. (2013). *The Data Revolution and Economic Analysis.* National Bureau of Economic Research (NBER) Working Paper No. 19035.

ERAC Secretariat. (2015). ERAC Opinion on the European Research Area Roadmap 2015-2020.

ERA-Learn (2015), *1st ERA-Learn 2020 annual report on P2P partnerships* [pdf], Deliverable D3.3, available at https://www.era-learn.eu/publications/other-publications/1st-annual-report-on-p2p-partnerships-2015.

ESFRI (European Strategy Forum on Research Infrastructures) (2016), *Strategy Report on Research Infrastructures* [pdf], Science and Technology Facilities Council, available at http://www.esfri.eu/sites/default/files/20160308_ROADMAP_single_page_LIGHT.pdf.

European Commission (2012), *Communication from the Commission to the European parliament, the council, the European economic and social committee and the committee of the regions: A reinforced European Research Area partnership for excellence and growth* [pdf], Brussels, available at http://ec.europa.eu/euraxess/pdf/research_policies/era-communication_en.pdf.

ICF International (2015), *Assessment of progress in achieving ERA in Member States and Associated Countries—Final Report to DG Research and Innovation*, Job Number 30260436.

IDEA Consult (2013), *Support for continued data collection and analysis concerning mobility patterns and career paths of researchers, Guidelines for the data analysis of the EU HEI survey data*, Brussels, available at http://ec.europa.eu/euraxess/pdf/research_policies/more2/Final%20report.pdf

OECD (2002), *Frascati Manual—Proposed standard practice for surveys on research and experimental development*, OECD Publishing, Paris, available at http://dx.doi.org/10.1787/9789264199040-en.

OECD (2015), Frascati Manual 2015: Guidelines for Collecting and Reporting Data on Research and Experimental Development, OECD Publishing, Paris, available at http://www.oecd.org/publications/frascati-manual-2015-9789264239012-en.htm

OECD (2015b), *OECD Science, technology, and industry scoreboard 2015—Innovation for growth and society*, OECD Publishing, Paris, available at http://www.oecd.org/science/oecd-science-technology-and-industry-scoreboard-20725345.htm.

OECD, European Union and UNESCO Institute for Statistics (2015), *ISCED 2011 Operational Manual—Guidelines for classifying national education programmes and related qualifications* [pdf], OECD Publishing, available at http://www.uis.unesco.org/Education/Documents/isced-2011-operational-manual.pdf.

Science-Metrix (2016), *Inception Report—Data gathering and information for the 2016 ERA monitoring*, Prepared for the European Commission, Contract 30-CE-0755994/00-15.

Vértesy, D. (2015) *Technical note on the robustness of a modified research excellence index. A Feasibility Study*. Publication Office of the EU: Luxembourg. DOI: 10.2760/81642.

Yiu, C. (2012). The Big Data Opportunity: Making government faster, smarter and more personal. *Policy Exchange*, ISBN: 978-1-907689-22-2, 34 pages.

How to obtain EU publications

Free publications:

- one copy:
  via EU Bookshop (http://bookshop.europa.eu);
- more than one copy or posters/maps:
  from the European Union's representations (http://ec.europa.eu/represent_en.htm);
  from the delegations in non-EU countries (http://eeas.europa.eu/delegations/index_en.htm);
  by contacting the Europe Direct service (http://europa.eu/europedirect/index_en.htm) or
  calling 00 800 6 7 8 9 10 11 (freephone number from anywhere in the EU) (*).

  (*) The information given is free, as are most calls (though some operators, phone boxes or hotels may charge you).

Priced publications:

- via EU Bookshop (http://bookshop.europa.eu)

The ERA Monitoring Handbook (2016) provides methodological guidance on the calculation of indicators included in the ERA PROGRESS REPORT 2016.

Organised by data source, information provided on each indicator includes a brief definition, rationale, computation method and any comments or critical issues for the reader to note. The handbook also includes a section on the verification and validation of data that outlines coherence checks and additional data considerations to be taken into consideration in the computation and interpretation of indicators. Finally, the annexes outline important information regarding international classification standards to which data for several of the indicators are tied, as well as key terminology and definitions.

Studies and reports

Publications Office