




Scientific Output
and Collaboration of

Companies Publishing the Most in the ERA



*Research and
Innovation*



EUROPEAN COMMISSION
Directorate-General for Research and Innovation
Directorate C — Research and Innovation
Unit C.6 — Economic analysis and indicators

E-mail: rtd-innovation-papers-studies@ec.europa.eu
RTD-PUBLICATIONS@ec.europa.eu

Contact: Carmen Marcus, Matthieu Delescluse and Pierre Vigier (Head of unit)

European Commission
B-1049 Brussels

Scientific Output and Collaboration of Companies Publishing the Most in the ERA

Authors of the study

David Campbell, Andréa Ventimiglia and Éric Archambault
Science Metrix Inc, Canada

This report is part of the study *Analysis and Regular Update of Bibliometric Indicators* carried out by Science Metrix-Canada under the coordination and guidance of the European Commission, Directorate-General for Research and Innovation, Directorate Research and Innovation, Economic analysis and indicators Unit.

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

(*) Certain mobile telephone operators do not allow access to 00 800 numbers or these calls may be billed

LEGAL NOTICE

Neither the European Commission nor any person acting on behalf of the Commission is responsible for the use which might be made of the following information.

The views expressed in this publication, as well as the information included in it, do not necessarily reflect the opinion or position of the European Commission and in no way commit the institution.

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

Cataloguing data can be found at the end of this publication.

Luxembourg: Publications Office of the European Union, 2013

ISBN 978-92-79-33014-8

doi:10.2777/37043

© European Union, 2013

Reproduction is authorised provided the source is acknowledged.

Cover images: earth, © #2520287, 2011. *Source:* Shutterstock.com; bottom globe, © PaulPaladin #11389806, 2012. *Source:* Fotolia.com

Table of Contents

Executive Summary	ii
Tables.....	vii
Figures	vii
Acronyms	viii
1 Introduction.....	1
2 Publication and Collaboration Profiles of Companies	3
2.1 Geographic Distribution of the Scientific Production of the Most Publishing Companies within the ERA.....	3
2.1.1 Company Level.....	3
2.1.2 Industry Level.....	7
2.2 Collaboration Patterns of Companies Having the Largest Scientific Production within the ERA.....	9
2.2.1 Company Level.....	9
2.2.2 Comparative Analysis of the Propensity of the 100 Companies Publishing the Most within the ERA to Collaborate Internally or Externally across Industrial Sectors.....	15
2.2.3 Comparative Analysis of the Propensity of the 100 Companies Publishing the Most within the ERA to Collaborate within and Outside the ERA across Industrial Sectors	17
3 Conclusion.....	19
3.1 Geographic Distribution of the Scientific Production of the Most Publishing Companies within the ERA.....	19
3.1.1 Company Level.....	19
3.1.2 Industry Level.....	20
3.2 Collaboration Patterns of Companies Having the Largest Scientific Production within the ERA.....	20
3.2.1 Company Level.....	20
3.2.2 Comparative Analysis of the Propensity of the 100 Companies Publishing the Most within the ERA to Collaborate Internally or Externally across Industrial Sectors.....	22
3.2.3 Comparative Analysis of the Propensity of the 100 Companies Publishing the Most within the ERA to Collaborate within and Outside the ERA across Industrial Sectors	22
4 Methods.....	24
4.1 Standardisation of Author Addresses for Firms and Subsidiaries.....	24
4.1.1 Delineating the Dataset.....	24
4.1.2 List Matching	25
4.1.3 Frequency-Based Extraction	25
4.1.4 Retrieval, Cleaning, and Matching of Subsidiaries.....	26
4.1.5 Identifying the Main Affiliates of Each Company within the ERA.....	26
4.2 Bibliometric Indicators.....	27
4.3 Graphic Representation of Data	27

Executive Summary

Background

Science-Metrix has been selected as the provider of bibliometric indicators for the European Commission's Directorate-General for Research & Innovation (DG Research) starting in September 2010 and extending to September 2014. This work involves the collection, analysis, and updating of bibliometric data that will be integrated into the European Commission's evidence-based monitoring of progress towards the objectives set forth in the Lisbon framework and the post-Lisbon Strategy for the European Research Area (ERA). The bibliometric component of this monitoring system is part of a package of six complementary studies reporting on the dynamics of research activities along the whole route of knowledge, from R&D investments to publications, patents, and licensing.

The analyses provided by Science-Metrix to the European Commission focus on the scientific performance—including impact and collaboration patterns—of countries, regions, and research performers (such as universities, public research institutes, and companies). These analyses rely on a series of indicators designed to take into account national and sector specificities, as well as to allow for a comprehensive analysis of the evolution, interconnectivity, performance, and impact of national research and innovation systems in Europe. The statistics also provide an overall view of Europe's strengths and weaknesses in knowledge production across fields and subfields of science. In measuring progress towards past and current objectives, this information aims to support the coherent development of research policies for the ERA.

The Present Report

More specifically, this report aims to measure the scientific production of research-intensive companies within the ERA zone, portray the geographic distribution of their scientific activities, and characterise their collaboration profiles. Data is presented for the top 100 firms which published the most within the ERA zone from 2007 to 2011; this includes multinational companies with activities outside the ERA, although only their production within this zone is considered. The production and collaboration profiles are based on the following set of bibliometric indicators:

Number of publications: Number of peer-reviewed scientific publications written by authors located in a given organisational entity (e.g., a company) or its affiliates within the ERA. Both full (FULL) and fractional (FRAC) counting is used.

Number of exclusively intra-company publications: Only those publications produced by the company without any contribution from authors in external organisations or from its affiliates located outside the ERA.

Number of exclusively intra-company publications with a single affiliate: Only those publications produced exclusively by authors from a single affiliate (i.e., city) within the ERA of the considered company.

Number of exclusively intra-company publications with multiple affiliates: Only those publications produced by authors from at least two affiliates (i.e., cities) within the ERA of the considered company without any external partner. (External partners of a company include its affiliates of the company located outside the ERA.)

Number of co-publications: Those publications of the considered company that were co-authored with at least one external partner. (External partners of a company include its affiliates located outside the ERA.)

Number of co-publications (Acad or RPO): Those publications of the considered company that were co-authored with at least one external partner in the academic or RPO sector.

Number of co-publications (Firm): Those publications of the considered company that were co-authored with at least one external partner in the private sector. (External partners of a company include its affiliates located outside the ERA.)

Number of co-publications (Unclassified): Those publications of the considered company that were co-authored with at least one external partner for which the sector is unknown.

The report is primarily descriptive, focusing on the salient points relevant to the report's two main parts: geographic distribution of the scientific production of the most publishing companies within the ERA (Section 2.1) and their collaboration patterns (Section 2.2).

Key Findings—Geographic Distribution of the Scientific Production of the Most Publishing Companies within the ERA (see Section 2.1)

Company Level

- Of the 100 companies publishing the most within the ERA, Siemens—a German multinational conglomerate specialising in electronic and electrical equipment—leads, having published a quarter more scientific papers (i.e., about 5,000 publications in FULL) than the second most publishing firm in ERA countries from 2007 to 2011.
- Following is a group of other Europe-based multinational firms whose number of publications (FULL) ranges from 3,000 to 4,000. In the Pharmaceutical & Biotechnology Sector, they include GlaxoSmithKline (headquartered in the UK; nearly 4,000 papers), Novartis International (Switzerland; about 3,900), and AstraZeneca (UK; about 3,600). They also include Philips Electronics (Netherlands; about 3,500) in the Leisure Goods sector, Thales (France; about 3,400) in Aerospace & Defence, and STMicroelectronics (Switzerland; nearly 3,000) in Technology Hardware & Equipment.
- Although 50% of the selected companies published their scientific contributions with affiliates distributed in at least six countries and 15 cities of the ERA, it should be noted that the research activities of private firms are generally highly concentrated in a few sites in Europe.
- Indeed, for 80% of the selected companies, at least 65% of the company's scientific output was produced in only two countries and at least 40% of the company's scientific output was produced in only two cities.
- Also, the city in which the firm's affiliates published the most is responsible for 40% or more of the company's papers for 43% of selected companies, and it published at least twice as many papers as the second most important city for 52% of selected companies.
- The concentration of output is noticeable even when the scientific output is distributed over a wide range of sites. For instance, Pfizer, Philips Electronics, Novartis, BASF, Microsoft, Roche Holding, Hewlett-Packard, Fiat, Nestlé, Deutsche Telekom, Vodafone, Rolls-Royce, Takeda Pharmaceutical, and Boehringer Ingelheim each have affiliates in at least 10 cities while at least 50% of their output is produced in only one city.
- Although there is a correlation between companies' numbers of publications and the number of countries or cities in which they are distributed, the correlation coefficients are moderately high with many outliers. As such, those companies with the most publications are not necessarily those with affiliates dispersed in the largest number of countries or cities.
- When ranked by the number of cities in which a company's affiliates are located, the top ten companies are Siemens (106), ABB Group (56), and Thales (55), followed by General Electric, Merck, Pfizer, EADS, Ericsson, Robert Bosch, and Sanofi (all between 36 and 51 affiliates).

- Alternatively, those with the lowest number of affiliates include Thomson Reuters, BT, and Soitec (4 affiliates each); NTT Docomo (3 affiliates); Yahoo! (2 affiliates); and NIZO Food Research, ASML, and Honda (1 affiliate each).
- The countries in which the affiliates of the 100 companies publishing the most (based on the leading two cities) within the ERA are most often located include, in descending order, Germany, the UK, France, Italy, Switzerland, the Netherlands, and Sweden.

Industry Level

- Among the 100 companies having published the most within the ERA from 2007 to 2011, the most prevalent industry is Health Care (ICB 4), which accounted for about 24% of all publications (based on fractional counting) produced by the leading firms (i.e., the top 100). This industry is followed closely by Industrials (ICB 2; 22%).
- The Technology (ICB 9) and the Consumer Goods (ICB 3) industries also appear to be highly research intensive with, respectively, a 20% and 11% share of publications produced by leading firms.
- Other industries appearing among the 100 leading firms include Oil & Gas, Basic Materials, Consumer Services, Telecommunications, and Utilities.
- At the sectorial level, Pharmaceuticals & Biotechnology (3-digit ICB: 457) clearly stands out with a 23% share of publications produced by leading firms, a 17% share of leading firms, and an average production of 610 papers per firm.
- Other sectors that stand out among the 100 firms publishing the most within the ERA include Technology Hardware & Equipment, Aerospace & Defence, and Electronic & Electrical Equipment.
- Industries and their sectors do not differ profoundly in terms of the propensity of companies to disperse their research activities over multiple sites across countries and cities. Dispersion only appears slightly lower than the average across industries in Consumer Goods, Telecommunications, and Utilities. However, in the latter two, the population of companies is less than 10.

Key Findings—Collaboration Patterns of Companies Having the Largest Scientific Production within the ERA (see Section 2.2)

Company Level

- The scientific collaboration network of the 100 companies that have published the most within the ERA allowed the identification of seven coherent sub-networks (or communities), each corresponding to organisations that are active in similar industrial sectors/markets. These clusters represent research-intensive industrial sectors in Europe and appear to cover nine of the 17 thematic priorities of the European Commission under the FP7. These groupings are as follows:

Aerospace & Defence: Aerospace & Defence (ICB 271; light purple cluster);

Automobiles & Parts: Automobiles & Parts (ICB 335; light blue cluster);

Basic Materials & Food Producers: Chemicals (ICB 135; mostly dark blue and red clusters), Industrial Metals & Mining (ICB 175; dark blue and green cluster), Food Producers (ICB 357; red cluster), and Personal Goods (376; red cluster);

Health Care: Health Care Equipment & Services (ICB 453; various clusters) and Pharmaceuticals & Biotechnology (ICB 457; almost exclusively dark blue cluster);

Oil, Gas & Utilities: Oil & Gas (ICB 0001; green cluster), Electricity (ICB 753; green cluster), and Gas, Water & Multi-Utilities (ICB 757; green cluster);

Other Industrials: Other sectors under the Industrials category (ICB 235, 272, 273, 275, and 279; various clusters);

Technology: Software & Computer Services (ICB 953; orange cluster), Technology Hardware & Equipment (ICB 957; mainly orange cluster), and Leisure Goods (374; orange cluster); and

Telecommunications & Media: Fixed Line and Mobile Telecommunications (ICB 653 and 657; mostly light purple and orange clusters) and Media (ICB 555; only two companies, one in the green cluster and the other in the orange cluster).

- This clustering pattern in the network indicates that the research fields of highest relevance to companies in a given industrial sector/market are generally similar.
- Furthermore, the proximity between various clusters in the network allowed the study to highlight the cross-applicability of research findings from one sector to the next. For example, the Aerospace & Defence grouping is positioned near the Telecommunications & Media grouping. In fact it even includes, at its periphery, a major player in the Telecommunications industry (France Telecom). This is not surprising since research progress in ICT is known to contribute importantly to the developments taking place in Aerospace & Defence.
- Siemens is the most central player in the collaboration network of the most-publishing firms in the ERA, with connections to all clusters. This is not surprising since it is the most publishing firm within the ERA and because it is a multinational company specialising in the Electronic & Electrical Equipment sector for the ICT, Energy, Health Care, and Industry markets. It is actually an important bridge between the two poles in the network: companies performing research in the life sciences and companies performing research in the physical, material, and engineering sciences.
- Thomson Reuters is by far the company with the highest proportion of exclusively intra-company publications (limited to the ERA). Note that the external partners of a company include its affiliates located outside the ERA.
- It is followed by Vodafone (44%), then ThyssenKrupp, Technicolor, BAE Systems, ABB Group, RWE, and Deutsche Telekom (all between 30% and 37%).
- Turning to publications co-authored with external partners, the affiliates of Thomson Reuters in the ERA zone are the least opened to co-authorship with external partners, with only 9% of its papers being co-authored with at least one external partner.
- Eighteen companies published 90% or more of their papers with at least one external partner, those at the top of the list being a diversified group including Medtronic, Danone, Renault, Eli Lilly and Company, Bruker Corp, ASML, and NIZO Food Research, all with 93% or more.
- Companies with the highest proportion of co-publications with at least one external partner in the academic or RPO sector are Danone (95%), Medtronic (93%), Novozymes (88%), Bruker Corp (86%), NIZO Food Research (85%), Fiat (85%), and Renault (84%).
- Companies with the highest proportion of co-publications with at least one external partner in the private sector include ASML (51%), Eli Lilly and Company (45%), IMS Health (44%), Johnson & Johnson (38%), Vattenfall (37%), UCB (37%), and Boehringer Ingelheim (37%).

Comparative Analysis of the Propensity of the 100 Companies Publishing the Most within the ERA to Collaborate Internally or Externally across Industrial Sectors

Using the above seven groupings of industrial sectors, the propensity of the most publishing firms to collaborate internally (i.e., between affiliates of a company, excluding affiliates outside the ERA) as well as externally was compared across industrial sectors.

- Generally, the most publishing ERA companies publish a great majority of their papers with external partners; when all industrial sectors are combined, companies co-authored, on average, about 80% of their publications with at least one external partner and this average score never drops below 66% regardless of the industrial sector.
- A far greater majority of papers co-published with at least one external partner involved the academic/RPO rather than the private sector (69% versus 22% when all industrial sectors are combined). Again, this tendency is the same irrespective of the industrial sector, but with some level of variation across industrial groupings.
- The Telecommunications & Media industry is generally less open to collaboration with external partners compared to other industrial groupings. Also interesting is the fact that the papers produced by companies in the Telecommunications & Media sectors do not involve multiple affiliates as often as in other sectors; however, this does not come as a surprise since the

geographical spread of companies over multiple affiliates in this area is not as pronounced as it is in other sectors.

- The most opposite collaboration pattern was observed for the Health Care sector, which is mostly made up of pharmaceutical firms. It is in this sector that the average percentages of exclusively intra-company publications and exclusively intra-company publications with a single affiliate are the smallest. Alternatively, the average percentages of co-publications with at least one external partner, with at least one academic or RPO partner, or with at least one external firm, are highest in this sector.

Comparative Analysis of the Propensity of the 100 Companies Publishing the Most within the ERA to Collaborate Within and Outside the ERA across Industrial Sectors

Using the above seven groupings of industrial sectors, the propensity of the most publishing firms to collaborate externally within and outside the ERA with organisations in the academic or RPO sector, or alternatively with organisations in the private sector, was compared across industrial sectors.

- The 100 most publishing companies within the ERA generally collaborate more extensively with external partners located within the ERA than with those outside the ERA, regardless of whether these partners are classified in the academic/RPO or private sectors.
- In two groupings of industrial sectors, namely Aerospace & Defence and Automobiles & Parts, ERA companies tend to collaborate more with external partners (either academic/RPO or private partners) located within the ERA than with those outside the ERA, to a greater extent than is generally the case (i.e., when all sectors are combined).
- Although collaboration in Health Care remains more pronounced within than outside the ERA, the propensity to collaborate with external partners (either academic/RPO or private partners) outside the ERA is more pronounced in this grouping of industrial sectors than is generally the case (i.e., when all sectors are combined).

Tables

Table I	Distribution of publication activity of the 100 most publishing companies in the ERA by geographic location of the affiliates at the country and city level within the ERA (2007–2011)	5
Table II	Distribution of the 100 most publishing companies in the ERA by industry and industrial sector based on the Industry Classification Benchmark (ICB) for the 2007–2011 period.....	8
Table III	Co-publication patterns of the 100 companies publishing the most within the ERA (2007–2011)	13
Table IV	Co-publication patterns of the 100 companies publishing the most within the ERA by grouping of industrial sectors (2007–2011)	16
Table V	Co-publication patterns within and outside the ERA of the 100 companies publishing the most within the ERA by grouping of industrial sectors (2007–2011).....	18

Figures

Figure 1	Collaboration network of the 100 most publishing companies within the ERA (2007–2011).....	10
----------	--	----

Acronyms

DG Research	Research Directorate-General
EFTA	European Free Trade Association
ERA	European Research Area
EU	European Union
EU-27	The 27 member countries of the European Union
FP7	Seventh Framework Programme of the European Community for Research, Technological Development (2007 to 2013)
FRAC	Fractional Counting
FULL	Full Counting
GERD	Gross Expenditures on R&D
ICB	Industry Classification Benchmark
NACE	Nomenclature generale des Activites economiques dans les Communautés europeennes (Industrial Sector Classification)
NSE	Natural Sciences and Engineering
NSF	United States National Science Foundation
NUTS2	Eurostat Nomenclature of Territorial Units for Statistics (Level 2)
R&D	Research and Development
RFP	Request for Proposal
RPO	Non-university Research Performing Organisations
RTD	Research and Technological Development
S&T	Science and Technology
SSH	Social Sciences and Humanities
STC	Science, Technology, and Competitiveness
STI	Science, Technology, and Innovation

1 INTRODUCTION

This report analyses the scientific production profiles of companies performing research within the European Research Area (ERA). The report is part of a suite of six bibliometric reports aimed at supporting the European Commission's evidence-based monitoring of progress towards the objectives set forth in the Lisbon framework and the post-Lisbon Strategy for the European Research Area (ERA), particularly the Seventh Framework Programme for Research and Technological Development (FP7). The suite of reports provides a descriptive analysis of the scientific performance and collaboration patterns of countries, NUTS2 regions, universities, Research Performing Organisations (RPOs), and companies with an emphasis on Europe.¹ It also provides a deeper analysis of the linkages between R&D inputs and outputs from an econometric perspective aimed at expanding the knowledge base on the factors driving the scientific productivity (i.e., the efficiency with which research inputs are converted into research outputs) of nations and regions.²

The broad analysis presented in this report covers bibliometric indicators related to the size and geographic distribution of scientific output (i.e., number of publications) as well as to the collaboration patterns of 100 firms which were selected with a view to providing information on companies publishing the greatest number of peer-reviewed scientific publications within the ERA. This report covers the period from 2007 to 2011. A final update will follow, covering the 2007–2012 period. The content to be included in this update can be expected to differ from that covered in the current edition in order to better meet some of the punctuality needs of the Commission at the time of production. Note that additional data and maps on companies have been produced as part of the project in addition to those which are presented in this report and its companion Excel data book; the goal of the present report is to synthesize the information of relevance to the Commission's officials in performing their duty. Indeed, presenting all of the collected information would overwhelm the reader.

Organisation of This Report

This report contains one section (Section 2) briefly presenting the following information for the leading 100 companies for the 2007–2011 period in the sciences in general (overall in Scopus):³

- Total number of publications by the considered company;
- Share of each affiliate of the considered company in all publications by the company (based on fractional counting);

¹ Delivered output at this time: Campbell, D., Picard-Aitken, M., Côté, G., Trépanier, É., Ventimiglia, A., and Archambault, É. (2011). *Analysis and Regular Update of Bibliometric Indicators: Country and Regional Scientific Production Profiles (Analytical Report 2.3.1)*. Report prepared by Science-Metrix for the European Commission Directorate-General for Research, 99 pages.

² Delivered outputs at this time: Campbell, D., Caruso, J., and Archambault, É. (2012). *Analysis and Regular Update of Bibliometric Indicators: Cross-Cutting Analysis of Scientific Output vs. Other STI Indicators (Analytical Report 2.3.2)*. Report prepared by Science-Metrix for the European Commission Directorate-General for Research, 67 pages.

Campbell, D., Caruso, J., and Archambault, É. (2012). *Cross-Cutting Analysis of Scientific Output versus Other STI Indicators. Conference Proceedings of the 17th International Conference on Science and Technology Indicators (STI)*, 5-8 September 2012, Montréal, Québec, Canada.

³ The dataset does not cover a sufficient number of years to allow trend analysis.

- Percentage of exclusively intra-affiliate publications (one or more authors in one affiliate only of the considered company);
- Percentage of exclusively intra-company publications (one or more authors in the considered company, possibly in different affiliates);
- Percentage of co-publications with at least one external partner;
- Percentage of co-publications with at least one external company partner; and
- Percentage of co-publications with at least one external academic and/or RPO partner.

This is followed by summary conclusions presented in Section 3. The methods used in this study are presented in Section 4.

2 PUBLICATION AND COLLABORATION PROFILES OF COMPANIES

Section 2.1 starts by providing a ranking of companies performing research within the ERA based on their number of peer-reviewed scientific papers published between 2007 and 2011. It subsequently provides information on the geographic distribution, within Europe, of these companies' most publishing affiliates. Section 2.2 provides a general overview of the structure of the knowledge-intensive private sector in the ERA. This is achieved by detecting underlying communities of firms through the analysis of scientific partnerships among the most publishing companies in the ERA. In turn, this provides insights on the most research-intensive industrial sectors in Europe. It also characterises the co-publication patterns of these companies (e.g., collaboration with external partners [i.e., universities, RPOs, firms] versus intra-company linkages between affiliates within and outside the ERA).

2.1 GEOGRAPHIC DISTRIBUTION OF THE SCIENTIFIC PRODUCTION OF THE MOST PUBLISHING COMPANIES WITHIN THE ERA

This section starts with an overview at the level of individual companies (Section 2.1.1) and ends with an overview at the level of industrial sectors, aggregating company data only for the most publishing companies within the ERA (Section 2.1.2).

2.1.1 Company Level

Table I presents the 100 companies publishing the most within the ERA, along with their number of publications in full (FULL) and fractional (FRAC) counting. (Refer to Section 4 for explanations on the counting methods.) This table also provides the number of countries/cities in which the companies' affiliates are publishing as well as the names of the two countries/cities in which these affiliates are publishing the most, with their respective shares of the companies' outputs indicated in parentheses. Please note that even in the case of multinational firms, only those papers produced by affiliates located within the ERA are considered. Furthermore, only cities in which a company's affiliates have published at least one paper (based on FRAC) are considered. All remaining cities and the countries in which they are located are classified as "Unknown". They were left out since it does not really make sense to consider the scientific production of cities in which a company published less than one paper (based on FRAC) in the past four years to appreciate its geographical spread. For instance, cities with less than one paper (based on FRAC) could result from erroneous entries in the database or from noise in a company's distribution pattern (e.g., one or a few punctual publication(s) without further continuation of research activity in the corresponding city).

Of the 100 companies publishing the most within the ERA, Siemens—a German multinational conglomerate company specialising in electronics and electrical engineering (Industry Classification Benchmark [ICB] sector 273; i.e., Electronic & Electrical Equipment) for the information and communication technologies (ICT), energy, health care, and industry markets—leads the pack, having published a quarter more scientific papers (i.e., about 5,000 publications in FULL) than the second most publishing firm from 2007 to 2011 in ERA countries. Coming next is a group of other Europe-based multinational firms whose number of publications (FULL) from 2007 to 2011 ranges from about 3,000 to nearly 4,000; half of them are in the Pharmaceutical & Biotechnology Sector (i.e., ICB 457) and the other half are in various industrial sectors mostly

related to the physical, material, and engineering sciences (i.e., Leisure Goods, Aerospace & Defence, and Technology Hardware & Equipment).

In the Pharmaceutical & Biotechnology sector, they include GlaxoSmithKline, headquartered in the UK, with nearly 4,000 papers (mostly active in the over-the-counter drugs, prescription drugs, biologics, vaccines, and consumer health care markets); Novartis International, headquartered in Switzerland, with nearly 3,900 papers (mostly active in the over-the-counter and prescription drugs markets); and AstraZeneca, headquartered in the UK, with about 3,600 papers (mostly active in the over-the-counter drugs, prescription drugs, and biologics markets). Philips Electronics—which is headquartered in the Netherlands and classified in the Leisure Goods sector with most of its activities in the consumer, health care, and lighting markets—ranks 5th, slightly behind AstraZeneca, with nearly 3,500 papers. It is followed by Thales, which is headquartered in France and is classified in the Aerospace & Defence sector, with about 3,400 papers (mostly active in the aerospace, defence, transportation, and security markets) and by STMicroelectronics, which is headquartered in Switzerland and is classified in the Technology Hardware & Equipment sector, with about 3,000 papers (mostly active in the semiconductor chips market). Among the remaining companies in the top 100, the number of publications in the 2007–2011 period declines gradually from about 2,700 to about 270 without any obvious breaks in the distribution. As such, no additional sub-group clearly stands out in the tail of the distribution.

Although there is a strong correlation between the number of publications of the top 100 companies obtained using FULL and FRAC counting methods (Pearson's $r = 0.98$, Spearman rank's $r = 0.95$; data not shown), some companies exhibit outlier-like behaviour in this correlation. For example, Vodafone, Thomson Reuters, and Honda each gain about 25 places in the ranking when using FRAC instead of FULL counting. This provides a preliminary indication that these firms do not co-author as many of their publications with as many researchers from other organisations as the average company among those selected. Conversely, Medtronic, Danone, and Bruker Corp lose between 20 and 26 places in the ranking when using FRAC instead of FULL counting, indicating that they co-author more of their publications with, on average, more researchers from other organisations than the average company among those selected. Section 2.2 investigates in greater detail such differences in the collaboration pattern of the most publishing companies in the ERA.

Table I Distribution of publication activity of the 100 most publishing companies in the ERA by geographic location of the affiliates at the country and city level within the ERA (2007–2011)

Company	Industry (3-digit ICB)	Total		Countries		Cities	
		Pubs (FULL)	Pubs (FRAC)	#	Two most publishing	#	Two most publishing
Siemens	273	5,006	2,315	21	DE (64%), AT (5%)	106	München (DE: 25%), Erlangen (DE: 13%)
GlaxoSmithKline	457	3,965	1,585	12	UK (62%), IT (13%)	34	Harlow (UK: 22%), Stevenage (UK: 19%)
Novartis International AG	457	3,888	1,543	12	CH (58%), IT (12%)	31	Basel (CH: 56%), Siena (IT: 11%)
AstraZeneca	457	3,561	1,468	6	UK (48%), SE (42%)	19	Macclesfield (UK: 31%), Mölndal (SE: 24%)
Philips Electronics	374	3,483	1,611	10	NL (66%), DE (22%)	31	Eindhoven (NL: 60%), Aachen (DE: 11%)
Thales	271	3,362	1,540	8	FR (69%), IT (9%)	55	Paris (FR: 27%), Marcoussis (FR: 8%)
STMicroelectronics	957	3,067	1,363	9	FR (62%), IT (26%)	27	Crolles (FR: 48%), Milan (IT: 14%)
EADS	271	2,742	1,348	5	DE (42%), FR (27%)	41	München (DE: 17%), Toulouse (FR: 12%)
Merck	457	2,461	926	13	DE (26%), UK (19%)	48	Darmstadt (DE: 20%), Oss (NL: 14%)
Nokia	957	2,454	1,248	13	FI (56%), DE (16%)	31	Helsinki (FI: 25%), Tampere (FI: 21%)
Pfizer Incorporated	457	2,286	950	15	UK (68%), DE (6%)	47	Sandwich (UK: 60%), Madrid (ES: 4%)
Bayer	135	2,196	836	11	DE (84%), FR (5%)	33	Berlin (DE: 31%), Wuppertal (DE: 21%)
International Business Machines	953	2,127	1,140	15	CH (41%), IL (23%)	33	Riischlikon (CH: 41%), Haifa (IL: 23%)
France Telecom	653	2,046	1,166	5	FR (83%), PL (0%)	15	Issy-Les-Moulineaux (FR: 26%), Lannion (FR: 26%)
ABB Group	275	1,957	1,166	14	CH (34%), SE (22%)	56	Baden (CH: 23%), Västerås (SE: 18%)
Alcatel-Lucent	957	1,929	964	11	FR (50%), DE (15%)	28	Villarsceaux (FR: 19%), Marcoussis (FR: 13%)
Electricite de France	753	1,833	861	4	FR (78%), UK (6%)	22	Chatou (FR: 23%), Clamart (FR: 22%)
Roche Holding AG	457	1,772	625	8	CH (66%), DE (19%)	18	Basel (CH: 64%), Penzberg (DE: 14%)
Infineon Technologies	957	1,686	849	10	DE (63%), AT (21%)	19	Neubiberg (DE: 29%), München (DE: 24%)
Novo Nordisk	457	1,623	602	6	DK (90%), CH (1%)	15	Gentof (DK: 35%), Måløv (DK: 28%)
Sanofi	457	1,566	522	8	FR (59%), DE (24%)	36	Frankfurt (DE: 18%), Lyon (FR: 17%)
BASF	135	1,497	605	9	DE (77%), CH (9%)	26	Ludwigshafen (DE: 65%), Basel (CH: 9%)
Unilever	357	1,490	641	4	UK (61%), NL (33%)	9	Sharnbrook (UK: 34%), Vlaardingen (NL: 33%)
Johnson & Johnson	457	1,368	505	11	BE (63%), NL (7%)	27	Beerse (BE: 47%), Mechelen (BE: 16%)
Microsoft Corporation	953	1,319	633	10	UK (59%), IT (9%)	19	Cambridge (UK: 58%), Povo (IT: 8%)
Ericsson	957	1,299	659	14	SE (40%), HR (8%)	37	Stockholm (SE: 24%), Zagreb (HR: 6%)
NXP	957	1,251	622	7	NL (61%), FR (13%)	12	Eindhoven (NL: 38%), Nijmegen (NL: 23%)
Deutsche Telekom	653	1,191	683	7	DE (65%), IL (11%)	14	Berlin (DE: 51%), Beer-Sheva (IL: 11%)
General Electric Co	272	1,167	501	14	UK (22%), DE (18%)	51	München (DE: 11%), Firenze (IT: 10%)
Statoil ASA	053	1,140	473	3	NO (64%), UK (1%)	16	Trondheim (NO: 25%), Stavanger (NO: 18%)
Nestlé S.A.	357	1,054	521	4	CH (83%), FR (7%)	15	Lausanne (CH: 76%), Sophia Antipolis (FR: 4%)
Boehringer Ingelheim	457	997	390	9	DE (80%), AT (5%)	11	Biberach (DE: 54%), Ingelheim (DE: 26%)
SAP	953	987	561	6	DE (60%), FR (7%)	15	Karlsruhe (DE: 28%), Dresden (DE: 11%)
Royal Dutch Shell	053	966	421	6	NL (51%), UK (14%)	15	Rijswijk (NL: 30%), Amsterdam (NL: 14%)
TOTAL	053	929	414	7	FR (41%), DE (8%)	20	Pau (FR: 22%), Paris (FR: 10%)
Robert Bosch	335	862	457	6	DE (79%), ES (1%)	37	Stuttgart (DE: 30%), Reutlingen (DE: 10%)
DSM	135	832	299	6	NL (61%), CH (25%)	16	Geleen (NL: 38%), Basel (CH: 20%)
Eli Lilly and Company	457	823	259	9	UK (39%), ES (16%)	14	Windlesham (UK: 33%), Alcobendas (ES: 13%)
Schneider	273	778	328	8	FR (34%), UK (29%)	22	Stafford (UK: 29%), Grenoble (FR: 18%)
CinetIQ	271	773	413	2	UK (84%), BE (2%)	11	Malvern (UK: 46%), Farnborough (UK: 29%)
AREVA	753	743	331	3	FR (50%), DE (36%)	23	Erlangen (DE: 30%), Paris (FR: 20%)
Rolls-Royce	271	743	311	4	UK (67%), DE (14%)	13	Derby (UK: 54%), Dahlewitz (DE: 13%)
Daimler	335	715	397	2	DE (85%), TR (0%)	10	Ulm (DE: 28%), Sindelfingen (DE: 23%)
Volkswagen	335	684	338	6	DE (58%), CZ (10%)	17	Wolfsburg (DE: 32%), Ingolstadt (DE: 16%)
ArcelorMittal	175	659	265	6	FR (61%), BE (16%)	26	Maizières-les-Metz (FR: 41%), Zelzate (BE: 13%)
ALSTOM	275	656	299	9	CH (39%), FR (20%)	25	Baden (CH: 30%), Birr (CH: 9%)
Schlumberger Limited	057	652	356	7	UK (31%), FR (11%)	19	Cambridge (UK: 21%), Clamart (FR: 8%)
Telefonica	653	608	270	2	ES (70%), PL (0%)	6	Madrid (ES: 37%), Barcelona (ES: 24%)
Eni	053	606	317	7	IT (51%), FR (3%)	16	San Donato Milanese (IT: 31%), Novara (IT: 8%)
SAFRAN	271	605	264	3	FR (79%), BE (2%)	24	Pessac (FR: 20%), Moissy Cramayel (FR: 15%)

Note: Only cities (within the ERA) in which a company's affiliates have published at least one paper (based on FRAC) are considered. All remaining cities and the countries in which they are located are classified as "Unknown". The two most publishing countries/cities are listed with their respective shares of the company's output indicated in parentheses.

ICB codes: 053 = Oil & gas producers, 057 = Oil equipment, services, & distribution, 135 = Chemicals, 175 = Industrial metals & mining, 235 = Construction & materials, 271 = Aerospace & defence, 272 = General industrials, 273 = Electronic & electrical equipment, 275 = Industrial engineering, 279 = Support services, 335 = Automobiles & parts, 357 = Food producers, 374 = Leisure goods, 376 = Personal goods, 453 = Health care equipment & services, 457 = Pharmaceuticals & biotechnology, 555 = Media, 653 = Fixed line telecommunications, 657 = Mobile telecommunications, 753 = Electricity, 757 = Gas, water, & multi-utilities, 953 = Software & computer services, 957 = Technology hardware & equipment.

Source: Computed by Science-Metrix using Scopus

Table I (Cont'd) Distribution of publication activity of the 100 most publishing companies in the ERA by geographic location of the affiliates at the country and city level within the ERA (2007–2011)

Company	Industry (3-digit ICB)	Total		Countries		Cities	
		Pubs (FULL)	Pubs (FRAC)	#	Two most publishing	#	Two most publishing
Hewlett-Packard Company	957	583	311	10	UK (55%), IL (14%)	17	Bristol (UK; 55%), Haifa (IL; 11%)
Abbott	457	580	189	9	DE (39%), NL (34%)	17	Weesp (NL; 29%), Ludwigshafen (DE; 23%)
Syngenta AG	135	570	189	5	UK (56%), CH (29%)	12	Bracknell (UK; 43%), Stein (CH; 18%)
Toshiba Corporation	272	550	296	7	UK (75%), SE (10%)	8	Bristol (UK; 43%), Cambridge (UK; 32%)
Bruker Corp	453	522	156	7	DE (75%), LV (6%)	13	Bremen (DE; 28%), Rheinstetten (DE; 20%)
Les Laboratoires Servier	457	519	214	4	FR (94%), PL (1%)	9	Paris (FR; 38%), Suresnes (FR; 27%)
Evonik Industries	272	502	200	2	DE (90%), UK (1%)	11	Hanau (DE; 34%), Marl (DE; 23%)
Thomson Reuters	555	501	472	4	UK (50%), ES (49%)	4	London (UK; 50%), Barcelona (ES; 49%)
BT	653	499	251	1	UK (68%)	4	Ipswich (UK; 65%), London (UK; 2%)
Intel Corporation	957	497	220	9	IL (48%), ES (10%)	21	Haifa (IL; 31%), Jerusalem (IL; 11%)
Thermo Fisher Scientific	453	496	156	12	DE (33%), UK (25%)	27	Berlin (DE; 17%), Uppsala (SE; 11%)
BP	053	485	219	5	UK (53%), NO (4%)	12	Sunbury on Thames (UK; 33%), Aberdeen (UK; 7%)
BMW	335	443	242	4	DE (81%), AT (1%)	7	München (DE; 79%), Dingolfing (DE; 1%)
UCB	457	440	176	4	BE (41%), UK (35%)	7	Braine-l'Alleud (BE; 36%), Slough (UK; 24%)
Finmeccanica	271	431	223	3	IT (69%), UK (12%)	10	Rome (IT; 48%), Giugliano in Campania (IT; 9%)
Carl Zeiss	453	414	182	3	DE (90%), IL (2%)	9	Oberkochen (DE; 53%), Jena (DE; 31%)
Takeda Pharmaceutical	457	390	155	7	CH (64%), DE (13%)	11	Konstanz (CH; 64%), Roskilde (DK; 5%)
Lundbeck	457	385	153	5	DK (80%), FR (8%)	6	Copenhagen (DK; 80%), Paris (FR; 6%)
Solvay	135	382	154	6	FR (49%), IT (24%)	13	Pessac (FR; 20%), Bollate (IT; 20%)
NTT Docomo, Inc.	653	380	187	3	DE (84%), FR/UK (1%)	3	München (DE; 84%), London (UK; 1%), Paris (FR; 1%)
Sumitomo Corporation	272	376	195	3	DE (74%), UK (3%)	5	Heidelberg (DE; 58%), Augustin (DE; 14%)
DuPont (mostly Danisco)	135	363	131	12	UK (23%), FI (17%)	25	Kantvik (FI; 17%), Brabrand (DK; 13%)
Danone	357	360	105	4	NL (46%), FR (33%)	7	Wageningen (NL; 45%), Palaiseau (FR; 32%)
Medtronic	453	350	80	7	NL (38%), IT (22%)	10	Maastricht (NL; 35%), Rome (IT; 13%)
Renault	335	349	136	2	FR (77%), RO (4%)	7	Guyancourt (FR; 57%), Rueil-Malmaison (FR; 11%)
ThyssenKrupp	175	344	197	5	DE (82%), IT (3%)	30	Duisburg (DE; 26%), Dortmund (DE; 15%)
Technicolor	555	342	211	4	FR (65%), DE (14%)	9	Cesson-Sévigné (FR; 44%), Paris (FR; 15%)
GDF SUEZ	757	339	154	2	FR (38%), BE (22%)	6	Saint Denis (FR; 22%), Bruxelles (BE; 21%)
E.ON	757	330	168	4	DE (62%), UK (9%)	12	Aachen (DE; 32%), Essen (DE; 10%)
Johnson Matthey	135	329	157	2	UK (86%), DE (1%)	9	Reading (UK; 48%), Cleveland (UK; 19%)
NIZO Food Research B.V.	357	326	105	1	NL (95%)	1	BA Ede (NL; 95%)
Acree AB	957	316	157	2	SE (93%), FR (1%)	5	Kista (SE; 78%), Hudiksvall (SE; 7%)
Novozymes	457	315	106	3	DK (85%), UK (5%)	5	Bagsvaerd (DK; 84%), Nottingham (UK; 4%)
Volvo	275	315	122	2	SE (80%), FR (2%)	5	Göteborg (SE; 40%), Trolhättan (SE; 33%)
BAE Systems	271	310	178	2	UK (68%), SE (5%)	14	Bristol (UK; 28%), Chelmsford (UK; 23%)
IMS Health	279	308	122	5	DE (33%), UK (29%)	7	London (UK; 29%), Frankfurt (DE; 23%)
Vodafone	657	308	207	7	DE (61%), ES (5%)	14	Dresden (DE; 52%), Madrid (ES; 5%)
Saint-Gobain	235	307	133	4	FR (80%), DE (3%)	10	Aubervilliers (FR; 44%), Cavailon (FR; 26%)
Vattenfall	753	307	128	4	SE (54%), DE (16%)	14	Stockholm (SE; 24%), Ivkarleby (SE; 9%)
Fiat	335	303	111	5	IT (69%), DE (5%)	15	Orbassano (IT; 55%), Maranello (IT; 4%)
ASML	957	302	118	1	NL (92%)	1	Veldhoven (NL; 92%)
Veolia Environnement	757	300	137	3	FR (75%), DE (2%)	9	Maison Laffitte (FR; 34%), Limay (FR; 14%)
RWE	757	295	144	3	DE (62%), UK (15%)	17	Essen (DE; 15%), Swindon (UK; 14%)
Yahoo! Inc.	953	292	143	2	ES (80%), IL (11%)	2	Barcelona (ES; 80%), Haifa (IL; 11%)
L'Oréal	376	286	131	2	FR (92%), DE (2%)	9	Clichy (FR; 38%), Aulnay sous Bois (FR; 20%)
Honda	335	284	166	1	DE (93%)	1	Offenbach (DE; 93%)
Agilent Technologies	273	283	99	9	DE (33%), UK (20%)	17	Waldbronn (DE; 22%), Glostrup (DK; 13%)
Telecom Italia	653	282	135	2	IT (77%), CH (1%)	5	Turin (IT; 70%), Rome (IT; 4%)
Soitec	957	278	85	2	FR (86%), DE (9%)	4	Crolles (FR; 76%), Courtaboeuf (FR; 9%)
OHB Group	271	271	128	4	DE (63%), IT (15%)	8	Bremen (DE; 28%), München (DE; 25%)

Note: *Ibid.* Takeda Pharmaceutical was formerly Nycomed.

Source: Computed by Science-Metrix using Scopus

To investigate the geographical distribution of the research activities performed by the 100 companies publishing the most within the ERA, the spread of a company's scientific publications by country and city in which its affiliates are located was characterised (Table I). Although 50% of the selected companies published their scientific contributions with affiliates located in at least six countries and 15 cities of the ERA, it should be noted that the research activities of private firms are generally highly concentrated in a few sites in Europe. Indeed, for 80% of the selected companies at least 65% of a company's scientific output was produced in only two countries and at least 40% of a company's scientific output was produced in only two cities. In addition, the city in which the firm's affiliates published the most is responsible for 40% or more of the company's papers for 43% of selected companies, and it published at least twice as many papers as the

second most important city for 52% of selected companies. The concentration of output is noticeable even when the scientific output is distributed over a wide range of sites; for instance, Pfizer, Philips Electronics, Novartis, BASF, Microsoft, Roche Holding, Hewlett-Packard, Fiat, Nestlé, Deutsche Telekom, Vodafone, Rolls-Royce, Takeda Pharmaceutical, and Boehringer Ingelheim each have affiliates in at least 10 cities while at least 50% of their output is produced in only one city (ordered in descending order of the number of cities in which they published their output).

Note that those companies with the most publications are not necessarily those with affiliates dispersed in the largest number of countries or cities; although there is a correlation between the number of publications (fractional counting) and the number of countries as well as between the number of publications and the number of cities, the correlation coefficients (i.e., Pearson's r of 0.63 and 0.74, respectively; very similar coefficients using full counting as well as using Spearman rank's r) are moderately high with many outliers (data not shown). One such outlier among the companies with a smaller production is Agilent Technologies, which is ranked 98th for the size of its output (fractional counting; 97th in full) but is ranked 20th and 37th for the number of countries and cities, respectively, in which its affiliates are located. Among the companies with the largest production, AstraZeneca is also an outlier, being ranked 6th in output size (fractional counting; 4th in full) but ranked 45th and 32nd for the number of countries and cities, respectively, in which its affiliates are located.

When ranked by the number of cities in which a company's affiliates are located, the top ten companies are Siemens (106), ABB Group (56), and Thales (55), followed by General Electric, Merck, Pfizer, EADS, Ericsson, Robert Bosch, and Sanofi (all between 36 and 51 affiliates). Alternatively, those with the smallest number of affiliates include Thomson Reuters, BT, and Soitec (4 affiliates each); NTT Docomo (3 affiliates); Yahoo! (2 affiliates); and NIZO Food Research, ASML, and Honda (1 affiliate each). Besides the top 10 companies noted above, there are also a few companies that are noticeable for their geographic spread, with affiliates in almost twice the average number of countries and cities, namely International Business Machines, GlaxoSmithKline, Nokia, Novartis International, and Bayer. Finally, the countries in which the affiliates of the 100 companies publishing the most (based on the leading two cities) within the ERA are most often located include, in descending order: Germany, the UK, France, Italy, Switzerland, the Netherlands, and Sweden.

2.1.2 Industry Level

Among the 100 companies publishing the most within the ERA from 2007 to 2011, the most prevalent industry is Health Care, which accounted for about 24% of all publications (based on fractional counting) produced by the leading firms (i.e., the top 100) (Table II). This industry is followed closely by Industrials (22%). On average, Health Care firms published slightly more than Industrial firms as the former account for 21% of leading firms (521 papers per firm on average) compared to 20% (509 papers per firm on average) for the latter; on average, firms in both industries published more papers than the 100 leading firms did (457 papers). The Technology industry also appears to be highly research-intensive, with a 20% share of publications produced by leading firms, a 15% share of leading firms, and the highest average production among leading firms (605 papers per firm on average). The Consumer Goods industry is also important with an 11% share of publications produced by leading firms, a 13% share of leading firms, and the fifth largest average production among leading firms (382 papers per firm on average). Finally, although firms in the Telecommunications industry are less predominant, they rank 4th in terms of average production by firm (414 papers per firm on average).

Table II Distribution of the 100 most publishing companies in the ERA by industry and industrial sector based on the Industry Classification Benchmark (ICB) for the 2007–2011 period

Industry (1/3-digit ICB)	Industry/Sector	Number/Percentage of firms among those which published the most within the ERA	Pubs (FRAC)	Average pubs per firm among those selected	Share of pubs produced by the 100 most publishing firms in the ERA	Avg No. of countries	Avg No. of cities
0001	Oil & Gas	6	2,200	367	4.8%	5.8	16.3
0530	Oil & Gas Producers	5	1,844	369	4.0%	5.6	15.8
0570	Oil Equipment, Services & Distribution	1	356	356	0.8%	7.0	19.0
1000	Basic Materials	9	2,834	315	6.2%	6.9	21.1
1350	Chemicals	7	2,372	339	5.2%	7.3	19.1
1750	Industrial Metals & Mining	2	462	231	1.0%	5.5	28.0
2000	Industrials	20	10,179	509	22.3%	6.5	25.0
2350	Construction & Materials	1	133	133	0.3%	4.0	10.0
2710	Aerospace & Defence	8	4,404	550	9.6%	3.9	22.0
2720	General Industrials	4	1,192	298	2.6%	6.5	18.8
2730	Electronic & Electrical Equipment	3	2,742	914	6.0%	12.7	48.3
2750	Industrial Engineering	3	1,587	529	3.5%	8.3	28.7
2790	Support Services	1	122	122	0.3%	5.0	7.0
3000	Consumer Goods	13	4,961	382	10.9%	3.9	12.8
3350	Automobiles & Parts	7	1,846	264	4.0%	3.7	13.4
3570	Food Producers	4	1,372	343	3.0%	3.3	8.0
3740	Leisure Goods	1	1,611	1,611	3.5%	10.0	31.0
3760	Personal Goods	1	131	131	0.3%	2.0	9.0
4000	Health Care	21	10,942	521	23.9%	8.1	19.7
4530	Health Care Equipment & Services	4	574	144	1.3%	7.3	14.8
4570	Pharmaceuticals & Biotechnology	17	10,368	610	22.7%	8.3	20.9
5000	Consumer Services	2	683	342	1.5%	4.0	6.5
5550	Media	2	683	342	1.5%	4.0	6.5
6000	Telecommunications	7	2,898	414	6.3%	3.9	8.7
6530	Fixed Line Telecommunications	6	2,691	449	5.9%	3.3	7.8
6570	Mobile Telecommunications	1	207	207	0.5%	7.0	14.0
7000	Utilities	7	1,923	275	4.2%	3.3	14.7
7530	Electricity	3	1,320	440	2.9%	3.7	19.7
7570	Gas, Water & Multi-Utilities	4	603	151	1.3%	3.0	11.0
9000	Technology	15	9,072	605	19.9%	8.1	18.1
9530	Software & Computer Services	4	2,477	619	5.4%	8.3	17.3
9570	Technology Hardware & Equipment	11	6,595	600	14.4%	8.0	18.4
All Sectors		100	45,692	457	100%	6.26	18.15

Note: Avg No. of countries = the average number of countries in which the affiliates of companies are distributed within the ERA. Avg No. of cities = the average number of cities (affiliates) of companies within the ERA. In performing meaningful comparisons across industries and sectors, one should take into account the population size for each category (i.e., number of companies in a given industry or sector). In some categories, the population size (within the 100 most publishing companies in the ERA) is so low that the results can hardly be regarded as providing a stable characterisation of the corresponding industry/sector.

Source: Computed by Science-Metrix using Scopos

At the sectorial level, Pharmaceuticals & Biotechnology clearly stands out with a 23% share of publications produced by leading firms, a 17% share of leading firms, and an average production of 610 papers per firm (Table II). Other sectors that stand out among the 100 firms publishing the most within the ERA include Technology Hardware & Equipment (Publications: 14.4% share; Firms: 11% share; Papers per firm: 600), Aerospace & Defence (Publications: 9.6% share; Firms: 8% share; Papers per firm: 550), and Electronic & Electrical Equipment (Publications: 6% share; Firms: 3% share; Papers per firm: 914). Note that in Electronic & Electrical Equipment, the publication share and high average production per firm are mostly attributable to Siemens.

Industries and their sectors do not differ profoundly in terms of the propensity of companies to disperse their research activities over multiple sites across countries and cities. In performing comparisons across industries and sectors, one should take into account the population size for each category (i.e., number of companies in a given industry or sector within the 100 most publishing firms); for instance, the largest departures from the averages for the whole set of 100 companies are strongest in sectors in which the population size is small. Consequently, findings for these industries and sectors can hardly be regarded as providing a stable characterisation of the corresponding industry/sector. The only industries in which dispersion appears slightly lower than the average for the selected companies are Consumer Goods, Telecommunications, and Utilities (Table II). However, in the latter two, the population is smaller than 10.

2.2 COLLABORATION PATTERNS OF COMPANIES HAVING THE LARGEST SCIENTIFIC PRODUCTION WITHIN THE ERA

This section starts with an overview at the level of individual companies (Section 2.2.1) and ends with Sections 2.2.2 and 2.2.3, which provide an overview at the level of industrial sectors aggregating companies' data only for the most publishing companies within the ERA.

2.2.1 Company Level

Figure 1 shows the collaboration network of the 100 companies publishing the most within the ERA. Using an algorithm for the fast unfolding of communities in large networks⁴, seven sub-networks (or communities; marked by node colour) were identified, each corresponding to organisations that are active in similar industrial sectors/markets. For example, the largest cluster, with 24 companies, is the dark blue cluster at the bottom of the network, for the most part composed of large multinational companies active in the Pharmaceuticals & Biotechnology sector (ICB 457; e.g., GlaxoSmithKline, Novartis International, AstraZeneca, Merck, and Pfizer, to name just a few). BASF, the largest chemical company in the world, is also part of this cluster, although one can see it is at the periphery, having important links to companies in the Oil & Gas and Industrials industries. Interestingly, linked to the dark blue cluster by several co-publications is a red cluster mainly composed of other large international firms active in the Chemicals (ICB 135; e.g., DSM, DuPont [mostly Danisco], and Syngenta) and Food Producers (ICB 357; e.g., Unilever, Nestlé, NIZO Food Research, and Danone) sectors. Considering that pharmaceuticals, chemicals, and food products all involve some similar aspects of laboratory R&D, the proximity of these two clusters indeed appears logical.

At the left of the network, there is another large cluster (green) which appears to be composed primarily of companies engaged in energy-related industries: there are BP, Eni, Royal Dutch Shell, Statoil ASA, TOTAL, and Schlumberger Limited in the Oil & Gas industry (ICB 0001); AREVA, Electricite de France, and Vattenfall in the Electricity sector (ICB 753); and E.ON, GDF SUEZ, RWE, and Veolia Environnement in the Gas, Water & Multi-Utilities sector (ICB 757). Furthermore, although these three groups are in the same cluster on the network map, they are easily distinguishable (i.e., energy-related companies are closer to other energy-related companies in their own ICB class than to other energy-related firms in other ICB classes). Not surprisingly, the green cluster is closely linked with a smaller cluster (light blue) mostly made up of companies in the automobiles and parts sector (ICB 335; e.g., BMW, Daimler, Fiat, Honda, Renault, and Volkswagen).

It is also closely related to an important light purple cluster of high-tech companies mostly engaged in the Aerospace & Defence sector (ICB 271; e.g., BAE Systems, EADS, OHB Group, QinetiQ, Rolls-Royce, SAFRAN, Thales). This cluster also includes, at its periphery, a major player in the Technology Hardware & Equipment sector (ICB 957; i.e., Alcatel-Lucent). This reflects the relevance of the Technology (mostly related to the computer sciences) industry to the developments taking place in the Aerospace & Defence Sector. This cluster includes two of the four nodes that are the most central to the network (see size of nodes in the network), having a

⁴ Blondel, V.D., Guillaume, J.L., Lambiotte, R., and Lefebvre E. (2008). Fast Unfolding of Communities in Large Networks. *Journal of Statistical Mechanics: theory and Experiments*, DOI: 10.1088/1742-5468/2008/10/P10008.

strong structural effect on it, namely Thales (Security & Defence) and Alcatel-Lucent (Technology Hardware & Equipment). Their high centrality scores, as well as the strong link between them, are in large part attributable to their joint venture, the III-V Lab for research in the field of III-V semiconductors.

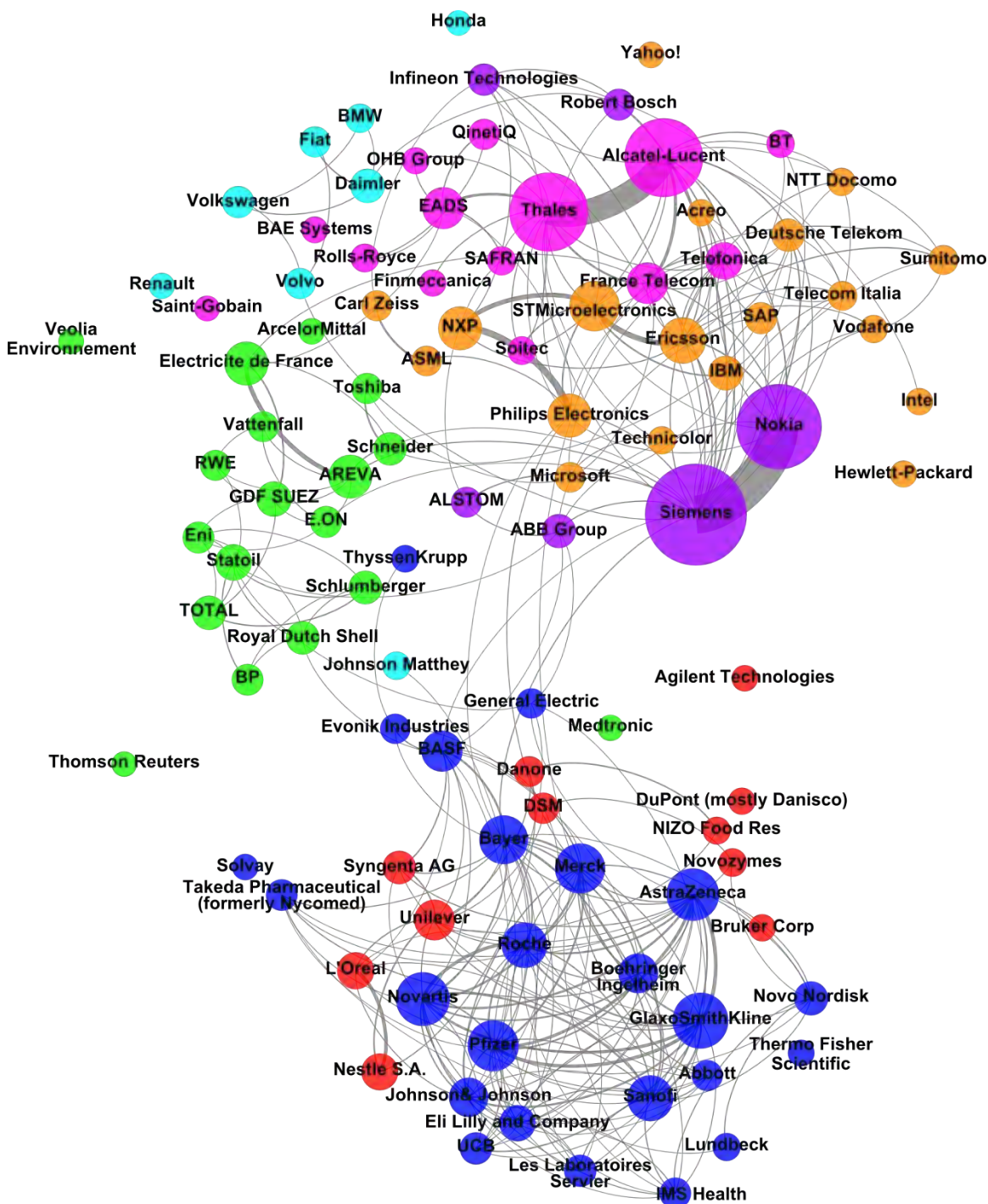


Figure 1 Collaboration network of the 100 most publishing companies within the ERA (2007–2011)

Note: Each company is represented by a node in the network and identified using a label. The size of nodes is proportional to the centrality (i.e., weighted PageRank) of companies. The width of links is proportional to the number of co-publications between any pair of companies. Only links representing at least 5 co-publications (i.e., one per year on average) are shown. Companies were clustered using an algorithm for the rapid unfolding of communities in large networks.

Source: Computed by Science-Metrix using Scopus

It turns out that Alcatel-Lucent is an important bridge connecting the Aerospace & Defence sector (light purple cluster) with the second largest cluster in the network (orange cluster, see below). This cluster is composed of companies mostly active in the Technology (ICB 953 and 957; e.g., STMicroelectronics, Ericsson, NXP, Intel, Microsoft) and Telecommunications (ICB 653 and 657; e.g., Deutsche Telekom, NTT Docomo, Telecom Italia) industries. The tight association of this cluster with the Aerospace & Defence cluster also reflects the relevance of ICT to the developments taking place in the Aerospace & Defence Sector, and vice versa; in fact, they partly share the same market for commercial applications in the telecommunication market and there are a number of telecommunication firms in the light purple cluster as well (e.g., BT, France Telecom, and Telefonica).

Close to this cluster is a very small group of companies (dark purple), two of which are among the four most central nodes in the network, namely Siemens and Nokia. Although Siemens is not classified in either the Technology or Telecommunications industries based on the ICB, it does have a large portion of its activities dealing specifically with telecommunications. Nokia is part of the Technology industry based on the ICB but also has important activities in telecommunications. For instance, the clustering of both companies near the orange cluster, their high centrality scores, and the strong link between them are to a large extent attributable to their joint venture: Nokia Siemens Network, which is the world's specialist in mobile broadband. Siemens is actually the most central player in the collaboration network of the most publishing firms in the ERA, with connections to all clusters. This is not surprising since it is a multinational company specialising in the Electronic & Electrical Equipment sector (ICB 273) for the ICT, Energy, Health Care, and Industry markets. It is actually an important bridge between companies performing research in the life sciences (bottom part of network, Figure 1) and those performing research in the physical, material, and engineering sciences (upper part of network, Figure 1).

Note that in the dark blue, orange, and both purple clusters, firms generally have a higher degree of connections than firms in the green, light blue, and red clusters; i.e., they are connected to a larger number of firms with which they co-authored at least five papers in the past four years. (Look at the density of links within each community.) Thus, there appear to be more firm-to-firm collaborations above the five co-publication threshold (in absolute terms) in the following clusters: Pharmaceutical & Biotechnology (dark blue), Technology & Telecommunications (orange), Aerospace & Defence (light purple), and Siemens' cluster (dark purple). The propensity of companies to collaborate with academia, RPOs, and the private sector across industrial sectors is studied in detail later in this section.

In conclusion, inter-company collaboration, measured through the analysis of scientific co-publications, is strongest among organisations in the same or closely related industrial sectors, indicating that they perform research in similar fields. Nevertheless, the proximity between various sectors in the network allowed this study to highlight the cross-applicability of research findings from one sector to the next. In brief, these research-intensive industrial sectors appear to cover 9 of the 17 thematic priorities of the European Commission under the FP7. Indeed, there appear to be intuitive links between FP7 thematic priorities and the main clusters of ICB classes as revealed by the collaboration network of the 100 firms publishing the most within the ERA. For example, FP7-health and FP7-biotechnology link with the Pharmaceuticals & Biotechnology sector (ICB 457; dark blue cluster); FP7-energy links with the Oil & Gas industry, the Electricity sector, and the Gas, Water & Multi-Utilities sector (ICB 0001, 753, and 757; green cluster); FP7-ICT links with the Telecommunications industry (ICB 653 and 657; orange cluster) and the Technology

industry (ICB 953 and 957; orange cluster); FP7-space, FP7-aeronautics, and FP7-security link with the Aerospace & Defence sector (ICB 271; light purple cluster); FP7-Food, Agriculture & Fisheries link with the Food Producers sector and the Chemicals sector (ICB 357 & 135 [e.g., pesticides for agricultural purposes]; red cluster); and FP7-automobiles links with the Automobiles & Parts sector (ICB 335; light blue cluster).

Collaboration Patterns of Companies with Academia, RPOs, and the Private Sector

Table III presents the co-publication patterns of the 100 companies publishing the most within the ERA over the 2007–2011 period. Note that because the study focuses on the portion of these companies' scientific output that was produced within the ERA zone, the external partners of a company in the statistics presented in Table III include its affiliates located outside the ERA. Thus, intra-company publications are limited to those produced by affiliates located within the ERA zone. All percentages presented in Table III are based on FULL counting.

Thomson Reuters is by far the company with the highest proportion of exclusively intra-company publications (limited to the ERA); nearly all (91%) of its ERA papers were authored without any contribution from an external partner. This explains why the company is disconnected from the other companies in the network of the 100 most publishing firms within the ERA (Figure 1 above). It is followed by Vodafone (44%), then ThyssenKrupp, Technicolor, BAE Systems, ABB Group, RWE, and Deutsche Telekom (all between 30% and 37%). The above companies were also those with the highest percentage of exclusively intra-company publications with a *single affiliate*, along with Toshiba Corporation and SAP, rounding out the top 10. ThyssenKrupp and ABB Group are also the companies with the highest percentage of exclusively intra-company publications with *multiple affiliates* with rates of, respectively, 7.6% and 6.5%. These are followed by RWE (5.8%), Carl Zeiss (4.6%), and Les Laboratoires Servier (3.9%).

Turning to publications co-authored with external partners (External partners of a company include its affiliates located outside the ERA.), it logically follows that many of the companies that were near the top of the list when ranked by the previous indicators are now placed at the bottom. For instance, the affiliates of Thomson Reuters in the ERA zone are the least open to co-authorship with external partners, with only 9% of its papers being co-authored with at least one external partner.

On average, about 80% of the output produced by the 100 companies is co-published with at least one external partner (i.e., including affiliates located outside the ERA). In fact, 18 companies published 90% or more of their papers with at least one external partner, those at the top of the list being a diversified group including Medtronic, Danone, Renault, Eli Lilly and Company, Bruker Corp, ASML, and NIZO Food Research, all with 93% or more. Notice that some companies with high collaboration rates with external partners (e.g., Medtronic and Renault) show no linkages above the threshold set in Figure 1 above (i.e., minimum of five co-publications between any pair of companies over five years). In fact, when the threshold is removed, all companies have at least one connection in the network (figure not shown). This indicates that in spite of a high collaboration rate with external partners, these companies have very few connections with the 100 most publishing companies within the ERA, possibly because most of their partners are either outside the private sector or companies outside of the top 100. The former hypothesis is confirmed by the high proportion of their papers co-published with at least one external partner in the *academic or RPO sector* and the rather small proportion of their papers co-published with at least one external partner in the *private sector* (Table III).

Table III Co-publication patterns of the 100 companies publishing the most within the ERA (2007–2011)

Company	Industry (3-digit ICB)	Pubs (FULL)	Intra-Company Pubs & Co-Pubs				Co-Pubs with External Partners			
			EICP (%)	EICPSA (%)	EICPMA (%)	UEICP (%)	CPEP (%)	CPEP (Acad or RPO; %)	CPEP (Firm; %)	UCPEP (%)
Siemens	273	5,006	18.6%	16.5%	1.8%	0.3%	81%	71%	22%	11%
GlaxoSmithKline	457	3,965	14.3%	12.4%	1.8%	0.1%	86%	74%	29%	25%
Novartis International AG	457	3,888	14.1%	12.9%	1.0%	0.2%	86%	74%	31%	23%
AstraZeneca	457	3,561	16.6%	14.0%	1.9%	0.6%	83%	74%	22%	20%
Philips Electronics	374	3,483	19.3%	17.9%	1.1%	0.3%	81%	74%	18%	11%
Thales	271	3,362	16.5%	14.5%	1.5%	0.4%	84%	74%	15%	23%
STMicroelectronics	957	3,067	10.5%	9.7%	0.7%	0.1%	90%	77%	16%	21%
EADS	271	2,742	21.6%	18.9%	1.9%	0.8%	78%	67%	17%	23%
Merck	457	2,461	12.6%	11.6%	0.8%	0.2%	87%	74%	29%	23%
Nokia	957	2,454	22.7%	19.2%	3.1%	0.4%	77%	70%	19%	8%
Pfizer Incorporated	457	2,286	17.6%	16.8%	0.6%	0.3%	82%	64%	36%	27%
Bayer	135	2,196	13.8%	12.5%	1.2%	0.1%	86%	76%	26%	23%
International Business Machines	953	2,127	24.9%	24.2%	0.6%	0.2%	75%	56%	29%	9%
France Telecom	653	2,046	25.0%	24.2%	0.5%	0.3%	75%	60%	13%	22%
ABB Group	275	1,957	34.3%	26.9%	6.5%	0.9%	66%	51%	15%	12%
Alcatel-Lucent	957	1,929	22.3%	20.2%	1.6%	0.6%	78%	63%	27%	15%
Electricite de France	753	1,833	17.9%	15.1%	2.3%	0.5%	82%	68%	15%	28%
Roche Holding AG	457	1,772	13.9%	13.4%	0.4%	0.1%	86%	72%	33%	29%
Infineon Technologies	957	1,686	19.8%	16.2%	3.3%	0.4%	80%	68%	23%	13%
Novo Nordisk	457	1,623	9.2%	7.4%	1.6%	0.2%	91%	83%	21%	22%
Sanofi	457	1,566	11.7%	9.6%	2.0%	0.1%	88%	77%	26%	35%
BASF	135	1,497	15.4%	14.2%	0.9%	0.4%	85%	72%	27%	13%
Unilever	357	1,490	16.6%	15.4%	1.1%	0.1%	83%	78%	16%	20%
Johnson & Johnson	457	1,368	9.4%	7.3%	1.7%	0.4%	91%	72%	38%	30%
Microsoft Corporation	953	1,319	16.8%	16.3%	0.2%	0.4%	83%	77%	16%	7%
Ericsson	957	1,299	23.8%	21.4%	1.7%	0.7%	76%	64%	20%	13%
NXP	957	1,251	20.1%	17.3%	2.6%	0.2%	80%	67%	28%	19%
Deutsche Telekom	653	1,191	30.4%	28.3%	1.4%	0.7%	70%	62%	15%	10%
General Electric Co	272	1,167	17.3%	14.4%	2.7%	0.3%	83%	66%	24%	17%
Statoil ASA	053	1,140	13.1%	11.3%	1.3%	0.4%	87%	64%	25%	27%
Nestlé S.A.	357	1,054	22.3%	20.4%	1.7%	0.2%	78%	73%	13%	18%
Boehringer Ingelheim	457	997	12.8%	10.7%	1.5%	0.6%	87%	73%	37%	26%
SAP	953	987	29.2%	26.6%	2.0%	0.5%	71%	62%	11%	13%
Royal Dutch Shell	053	966	16.8%	14.7%	1.6%	0.5%	83%	59%	32%	24%
TOTAL	053	929	18.9%	16.6%	0.9%	1.5%	81%	59%	20%	29%
Robert Bosch	335	862	21.8%	19.1%	1.4%	1.3%	78%	70%	16%	10%
DSM	135	832	11.8%	10.9%	0.8%	0.0%	88%	81%	20%	14%
Eli Lilly and Company	457	823	5.2%	5.0%	0.2%	0.0%	95%	79%	45%	36%
Schneider	273	778	10.2%	8.9%	1.0%	0.3%	90%	76%	17%	19%
Cinetiq	271	773	26.5%	25.6%	0.8%	0.1%	73%	62%	19%	20%
AREVA	753	743	19.5%	15.7%	3.0%	0.8%	80%	62%	31%	27%
Rolls-Royce	271	743	13.2%	11.2%	0.9%	1.1%	87%	79%	14%	13%
Daimler	335	715	23.2%	21.3%	1.7%	0.3%	77%	65%	17%	10%
Volkswagen	335	684	20.6%	19.3%	1.0%	0.3%	79%	67%	20%	14%
ArcelorMittal	175	659	11.1%	8.2%	2.1%	0.8%	89%	76%	6%	24%
ALSTOM	275	656	20.0%	18.0%	1.5%	0.5%	80%	69%	19%	20%
Schlumberger Limited	057	652	25.9%	23.5%	1.1%	1.4%	74%	47%	25%	21%
Telefonica	653	608	16.9%	16.1%	0.5%	0.3%	83%	75%	25%	20%
Eni	053	606	28.2%	25.1%	1.3%	1.8%	72%	51%	21%	25%
SAFRAN	271	605	15.2%	14.0%	0.5%	0.7%	85%	75%	11%	24%

Note: Pubs (FULL) = number of publications based on full counting within the ERA; EICP (%) = % of exclusively intra-company publications (limited to the ERA); EICPSA (%) = % of exclusively intra-company publications with a single affiliate (limited to the ERA); EICPMA (%) = % of exclusively intra-company publications with multiple affiliates (limited to the ERA); UEICP (%) = % of unclassified exclusively intra-company publications; CPEP (%) = % of publications co-authored with at least one external partner (external partners of a company include its affiliates located outside the ERA); CPEP (Acad or RPO; %) = % of co-publications with at least one external partner in the academic or RPO sector; CPEP (Firm; %) = % of co-publications with at least one external partner from the private sector; UCPEP (%) = % of publications co-authored with an unclassified external partner.

ICB codes: 053 = Oil & gas producers, 057 = Oil equipment, services & distribution, 135 = Chemicals, 175 = Industrial metals & mining, 235 = Construction & materials, 271 = Aerospace & defence, 272 = General industrials, 273 = Electronic & electrical equipment, 275 = Industrial engineering, 279 = Support services, 335 = Automobiles & parts, 357 = Food producers, 374 = Leisure goods, 376 = Personal goods, 453 = Health care equipment & services, 457 = Pharmaceuticals & biotechnology, 555 = Media, 653 = Fixed line telecommunications, 657 = Mobile telecommunications, 753 = Electricity, 757 = Gas, water & multi-utilities, 953 = Software & computer services, 957 = Technology hardware & equipment.

Source: Computed by Science-Metrix using Scopus

Table III (Cont'd) Co-publication patterns of the 100 companies publishing the most within the ERA (2007–2011)

Company	Industry (3-digit ICB)	Pubs (FULL)	Intra-Company Pubs & Co-Pubs				Co-Pubs with External Partners			
			EICP (%)	EICPSA (%)	EICPMA (%)	UEICP (%)	CPEP (%)	CPEP (Acad or RPO; %)	CPEP (Firm; %)	UCPEP (%)
Hewlett-Packard Company	957	583	28.0%	26.4%	1.2%	0.3%	72%	59%	21%	10%
Abbott	457	580	10.0%	9.7%	0.3%	0.0%	90%	77%	28%	31%
Syngenta AG	135	570	11.4%	10.2%	1.2%	0.0%	89%	79%	23%	20%
Toshiba Corporation	272	550	29.3%	29.3%	0.0%	0.0%	71%	61%	20%	9%
Bruker Corp	453	522	6.5%	6.1%	0.0%	0.4%	93%	86%	21%	16%
Les Laboratoires Servier	457	519	14.8%	10.8%	3.9%	0.2%	85%	78%	15%	22%
Evonik Industries	272	502	16.1%	14.3%	1.4%	0.4%	84%	71%	25%	11%
Thomson Reuters	555	501	90.8%	90.4%	0.2%	0.2%	9%	4%	1%	5%
BT	653	499	25.1%	24.4%	0.0%	0.6%	75%	68%	13%	14%
Intel Corporation	957	497	18.1%	17.7%	0.2%	0.2%	82%	68%	31%	10%
Thermo Fisher Scientific	453	496	12.3%	11.3%	0.8%	0.2%	88%	81%	19%	23%
BP	053	485	18.6%	18.1%	0.4%	0.0%	81%	45%	35%	29%
BMW	335	443	22.1%	21.7%	0.5%	0.0%	78%	65%	19%	10%
UCB	457	440	16.4%	13.6%	2.5%	0.2%	84%	67%	37%	28%
Finmeccanica	271	431	23.2%	21.6%	0.9%	0.7%	77%	66%	13%	18%
Carl Zeiss	453	414	21.3%	15.9%	4.6%	0.7%	79%	53%	36%	18%
Takeda Pharmaceutical	457	390	12.3%	11.8%	0.5%	0.0%	88%	73%	30%	27%
Lundbeck	457	385	9.4%	8.8%	0.0%	0.5%	91%	83%	21%	19%
Solvay	135	382	13.1%	12.8%	0.3%	0.0%	87%	77%	15%	16%
NTT Docomo, Inc.	653	380	17.9%	17.6%	0.0%	0.3%	82%	73%	16%	7%
Sumitomo Corporation	272	376	25.8%	24.7%	0.5%	0.5%	74%	65%	22%	14%
DuPont (mostly Danisco)	135	363	11.3%	9.6%	1.7%	0.0%	89%	76%	27%	22%
Danone	357	360	3.1%	3.1%	0.0%	0.0%	97%	95%	18%	29%
Medtronic	453	350	2.6%	2.3%	0.3%	0.0%	97%	93%	15%	30%
Renault	335	349	5.2%	4.9%	0.0%	0.3%	95%	84%	9%	22%
ThyssenKrupp	175	344	36.6%	28.5%	7.6%	0.6%	63%	49%	17%	11%
Technicolor	555	342	35.7%	33.6%	1.5%	0.6%	64%	55%	15%	10%
GDF SUEZ	757	339	25.4%	25.1%	0.0%	0.3%	75%	55%	24%	35%
E.ON	757	330	27.0%	26.4%	0.3%	0.3%	73%	51%	26%	18%
Johnson Matthey	135	329	19.5%	18.2%	0.6%	0.6%	81%	70%	19%	5%
NIZO Food Research B.V.	357	326	7.4%	7.1%	0.0%	0.3%	93%	85%	14%	49%
Acreo AB	957	316	17.7%	17.7%	0.0%	0.0%	82%	69%	24%	18%
Novozymes	457	315	7.9%	7.6%	0.0%	0.3%	92%	88%	23%	13%
Volvo	275	315	9.8%	9.2%	0.3%	0.3%	90%	81%	29%	18%
BAE Systems	271	310	35.5%	34.5%	0.6%	0.3%	65%	52%	16%	12%
IMS Health	279	308	10.1%	9.4%	0.6%	0.0%	90%	74%	44%	39%
Vodafone	657	308	44.5%	44.2%	0.0%	0.3%	56%	43%	14%	12%
Saint-Gobain	235	307	17.6%	16.9%	0.3%	0.3%	82%	75%	11%	18%
Vattenfall	753	307	16.9%	13.4%	3.3%	0.3%	83%	57%	37%	27%
Fiat	335	303	10.2%	9.9%	0.0%	0.3%	90%	85%	17%	24%
ASML	957	302	7.0%	7.0%	0.0%	0.0%	93%	50%	51%	33%
Veolia Environnement	757	300	15.7%	12.3%	3.0%	0.3%	84%	65%	12%	38%
RWE	757	295	33.2%	27.5%	5.8%	0.0%	67%	49%	26%	21%
Yahoo! Inc.	953	292	15.1%	14.7%	0.3%	0.0%	85%	75%	23%	10%
L'Oreal	376	286	18.5%	17.1%	1.4%	0.0%	81%	63%	32%	27%
Honda	335	284	23.6%	23.6%	0.0%	0.0%	76%	72%	10%	5%
Agilent Technologies	273	283	11.3%	11.0%	0.4%	0.0%	89%	79%	24%	14%
Telecom Italia	653	282	19.5%	18.1%	1.1%	0.4%	80%	69%	24%	22%
Soitec	957	278	10.1%	10.1%	0.0%	0.0%	90%	82%	33%	27%
OHB Group	271	271	18.1%	15.9%	1.8%	0.4%	82%	69%	30%	30%

Note: *Ibid.* Takeda Pharmaceutical was formerly Nycomed.

Source: Computed by Science-Metrix using Scopus

External partners are most often from the academic or RPO sector but may also be private companies or may be unclassified. The companies with the highest proportion of co-publications with at least one external partner in the *academic or RPO sector* are Danone (95%), Medtronic (93%), Novozymes (88%), Bruker Corp (86%), NIZO Food Research (85%), Fiat (85%), and Renault (84%). Companies with the highest proportion of co-publications with at least one external partner in the *private sector* include ASML (51%), Eli Lilly and Company (45%), IMS Health (44%), Johnson & Johnson (38%), Vattenfall (37%), UCB (37%), and Boehringer Ingelheim (37%). Finally, those companies with the highest proportion of publications co-authored with an unclassified external partner are NIZO Food Research, IMS Health, and Veolia Environnement, all between 38% and 50%.

2.2.2 Comparative Analysis of the Propensity of the 100 Companies Publishing the Most within the ERA to Collaborate Internally or Externally across Industrial Sectors

To compare the propensity of the most publishing firms to collaborate internally (i.e., between affiliates of a company) as well as externally across industrial sectors, companies have been grouped based on their ICB classes. Furthermore, to ensure a proper group size by industry type, most sectors were grouped together based on their agglomeration in the collaboration networks of the 100 companies publishing the most in the ERA. (See discussion of Figure 1 above.) Thus, the groups of companies compared are as follows:

- **Aerospace & Defence:** Aerospace & Defence (ICB 271; light purple cluster in Figure 1);
- **Automobiles & Parts:** Automobiles & Parts (ICB 335; light blue cluster in Figure 1);
- **Basic Materials & Food Producers:** Chemicals (ICB 135; mostly dark blue and red clusters in Figure 1), Industrial Metals & Mining (ICB 175; dark blue and green cluster in Figure 1), Food Producers (ICB 357; red cluster in Figure 1), and Personal Goods (376; red cluster in Figure 1);
- **Health Care:** Health Care Equipment & Services (ICB 453; various clusters in Figure 1) and Pharmaceuticals and Biotechnology (ICB 457; almost exclusively dark blue cluster in Figure 1);
- **Oil, Gas & Utilities:** Oil & Gas (ICB 0001; green cluster in Figure 1), Electricity (ICB 753; green cluster in Figure 1) and Gas, Water & Multi-Utilities (ICB 757; green cluster in Figure 1);
- **Other Industrials:** Other sectors under the Industrials category (ICB 235, 272, 273, 275 and 279; various clusters in Figure 1);
- **Technology:** Software & Computer Services (ICB 953; orange cluster in Figure 1), Technology Hardware & Equipment (ICB 957; mainly orange cluster in Figure 1), and Leisure Goods (374; orange cluster in Figure 1); and
- **Telecommunications & Media:** Fixed Line and Mobile Telecommunications (ICB 653 and 657; mostly light purple and orange clusters in Figure 1) and Media (ICB 555; only two companies, one in the green cluster and the other in the orange cluster in Figure 1).

Table IV compares the scientific collaboration patterns of various industrial sectors within the population of the 100 firms publishing the most within the ERA zone. Generally, the most publishing ERA companies publish a great majority of their papers with external partners; when all industrial sectors are combined, companies co-authored, on average, about 80% of their publications with at least one external partner and this average score never drops below 66% regardless of the industrial sector. Furthermore, a far greater majority of papers co-published with at least one external partner involved the academic/RPO sector rather than the private sector (69% versus 22% when all industrial sectors are combined). Again, this tendency is the same irrespective of the industrial sector, but with some level of variation across industrial groupings.

In particular, it can be seen that the Telecommunications & Media industry is generally less open to collaboration with external partners compared to the other industrial groupings, its average percentage of exclusively intra-company publications as well as its average percentage of exclusively intra-company publications with a single affiliate being nearly twice as large as the corresponding averages for all sectors combined; its average percentage of co-publications with at least one external partner (External partners of a company include its affiliates located outside the ERA.) is also the smallest across all sectors, with a score about 20% smaller than the corresponding score for all sectors combined.

Similarly, its average percentage of co-publications with at least one academic or RPO partner is nearly 20% smaller than the average across sectors, being ranked last along with Oil, Gas & Utilities. This negative departure is even more pronounced in terms of its propensity to collaborate with external firms (i.e., about 30% smaller than for all sectors combined); in the Automobiles & Parts sector, the propensity to collaborate with external firms is the same.

Table IV Co-publication patterns of the 100 companies publishing the most within the ERA by grouping of industrial sectors (2007–2011)

Grouping of industrial sectors	N	AVG EICP (%)	AVG EICPSA (%)	AVG EICPMA (%)	AVG CPEP (%)	AVG CPEP (Acad or RPO; %)	AVG CPEP (Firm; %)
Aerospace & Defence	8	21.2%	19.5%	1.1%	78.8%	68.0%	16.9%
Automobiles & Parts	7	18.1%	17.1%	0.6%	81.9%	72.7%	15.1%
Basic Materials & Food Producers	14	15.1%	13.4%	1.5%	84.9%	75.0%	19.5%
Health Care	21	11.9%	10.4%	1.3%	88.1%	76.0%	28.2%
Oil, Gas & Utilities	13	21.3%	18.8%	1.9%	78.7%	56.3%	25.3%
Other Industrials	12	18.4%	16.6%	1.4%	81.6%	69.9%	22.8%
Technology	16	19.1%	17.7%	1.2%	80.9%	67.4%	24.3%
Telecommunications & Media	9	34.0%	33.0%	0.6%	66.0%	56.3%	15.1%
All Sectors	100	18.7%	17.1%	1.3%	81.3%	68.6%	22.4%

Note: N = number of companies among those selected; Avg EICP (%) = Average % of exclusively intra-company publications (limited to the ERA); Avg EICPSA (%) = Average % of exclusively intra-company publications with a single affiliate (limited to the ERA); Avg EICPMA (%) = Average % of exclusively intra-company publications with multiple affiliates (limited to the ERA); Avg CPEP (%) = Average % of publications co-authored with at least one external partner (external partners of a company include its affiliates located outside the ERA); Avg CPEP (Acad or RPO; %) = Average % of co-publications with at least one external partner in the academic or RPO sector; Avg CPEP (Firm; %) = Average % of co-publications with at least one external partner from the private sector. The colour gradient informs on the extent of negative (red) or positive (green) departure (in percentage points) from the average for all sectors combined (white); the darker the colour, the stronger the departure. (Note that the gradient is normalised relative to the strongest negative and positive departures in the table.)

Source: Computed by Science-Metrix using Scopus

Also interesting is the fact that the papers produced by companies in the Telecommunications & Media sectors do not involve multiple affiliates as often as in other sectors; however, this does not come as a surprise since the average number of affiliates (i.e., cities in which a company is located) in the Telecommunications & Media sectors was smaller than in any other sector (see Table II). Although the group size (N = 9) for this pool of industrial sectors is small, the conclusions remain mostly unchanged when the strongest outlier, Thomson Reuters, is removed; also, two thirds of this group are above the average for all sectors in terms of their propensity to produce exclusively intra-company publications. The only other sectors with a similar collaboration pattern include Aerospace & Defence and Oil, Gas & Utilities. However, departures from the averages for all sectors combined are not pronounced in these cases.

The most opposite collaboration pattern was observed for the Health Care sector, which is mostly made up of pharmaceutical firms (N = 21). It is in this sector that the average percentages of exclusively intra-company publications and exclusively intra-company publications with a single affiliate are the smallest, being 36% and 39% smaller, respectively, than the corresponding averages for all sectors combined. Alternatively, the average percentages of co-publications with at least one external partner, with at least one academic or RPO partner, or with at least one external firm are highest in this sector, although the positive departures, expressed in percentage from the corresponding averages for all sectors combined, are less pronounced, ranging from 8% to 26% (Table IV). The only other sector with a similar collaboration pattern is Basic Materials & Food Producers, which also involves an appreciable share of firms in the life sciences. However, departures from the averages for all sectors combined are not as pronounced as they were for

Health Care; in particular, the percentage of co-publications with at least one external firm is slightly below the average for all sectors combined (19.5% versus 22.4%).

2.2.3 Comparative Analysis of the Propensity of the 100 Companies Publishing the Most within the ERA to Collaborate within and Outside the ERA across Industrial Sectors

Table V compares the propensity of the most publishing firms, by grouping of industrial sectors, to collaborate externally within and outside the ERA with organisations in the academic or RPO sector, or alternatively with organisations in the private sector; the provided data is not intended to compare the propensity of companies to collaborate with academic/RPO organisations versus private firms. To allow the analysis of collaboration within and outside the ERA, the disaggregated data on the co-publications of the 100 selected companies by country in which the partner organisations are located were used. For example, the propensity of firms to collaborate with external firms within the ERA in a given grouping of industrial sectors (see 4th column in Table V) was computed as follows:

$$\frac{1}{N} \sum_{Firm=1}^N \left(\frac{\sum_{ERA=1}^I CPEP \text{ with Firms}_{ERA}}{\sum_{All Countries=1}^J CPEP \text{ with Firms}_{All Countries}} \right)_{Firm}$$

Where,

N = No. of firms in a given grouping of industrial sectors (see column 2 in Table V);

I = No. of ERA countries;

J = No. of countries in the world; and

CPEP with Firms = No. of co-publications with at least one external partner in the private sector.

Since unknown countries are not included in the computation, the propensity of firms to collaborate with external firms outside the ERA in a given grouping of industrial sectors can be obtained by taking one minus the corresponding score found in the 4th column of Table V, as in:

$$1 - \frac{1}{N} \sum_{Firm=1}^N \left(\frac{\sum_{ERA=1}^I CPEP \text{ with Firms}_{ERA}}{\sum_{All Countries=1}^J CPEP \text{ with Firms}_{All Countries}} \right)_{Firm}$$

Note that the co-publications of a given European company with external firms outside the ERA zone include co-publications with its affiliates located outside the ERA.

From Table V, it can be seen that the 100 most publishing companies within the ERA generally collaborate more extensively with external partners located within rather than outside the ERA regardless of whether these partners are classified in the academic/RPO or private sectors, or even whether they are unclassified; the percentage of co-publications involving countries within the ERA ranged from 70% (with other firms) to 83% (with academic/RPO partners) with only about 17% to 30% of co-publications involving countries outside the ERA. (Double counting of co-publications across countries occurs.) Note that the collaboration rate with external partners in the private sector outside the ERA might simply be higher due to the fact that affiliates of companies outside the ERA are considered as external partners.

From this table, it can also be seen that in two groupings of industrial sectors, namely Aerospace & Defence and Automobiles & Parts, ERA companies tend to collaborate more with external partners located within rather than outside the ERA to a greater extent than is generally the case

(i.e., when all sectors are combined); for instance, the propensity to collaborate with external partners in the academic or RPO sector within the ERA is 9% larger for Aerospace & Defence and Automobiles & Parts than the corresponding scores when all sectors are combined. They are, respectively, 24% and 14% larger in the case of collaboration with external partners in the private sector within the ERA. Although collaboration in Health Care remains more pronounced within than outside the ERA, the propensity to collaborate with external partners outside the ERA is more pronounced in this grouping of industrial sectors than is generally the case (i.e., when all sectors are combined); for example, the propensity to collaborate with external partners in the academic/RPO or private sector is, respectively, 4% and 15% smaller than the corresponding scores when all sectors are combined.

Table V Co-publication patterns within and outside the ERA of the 100 companies publishing the most within the ERA by grouping of industrial sectors (2007–2011)

Grouping of industrial sectors	N	$\text{AVG}_{\text{Firm}} (\frac{\sum_{\text{ERA}} \text{CPEP}_{\text{Acad/RPO}}}{\sum_{\text{All Countries}} \text{CPEP}_{\text{Acad/RPO}}})$	$\text{AVG}_{\text{Firm}} (\frac{\sum_{\text{ERA}} \text{CPEP}_{\text{Firm}}}{\sum_{\text{All Countries}} \text{CPEP}_{\text{Firm}}})$	$\text{AVG}_{\text{Firm}} (\frac{\sum_{\text{ERA}} \text{UCPEP}}{\sum_{\text{All Countries}} \text{UCPEP}})$
Aerospace & Defence	8	90.1%	86.1%	83.5%
Automobiles & Parts	7	89.9%	79.0%	90.7%
Basic Materials & Food Producers	14	80.3%	73.3%	75.9%
Health Care	21	79.0%	59.2%	70.6%
Oil, Gas & Utilities	13	84.6%	75.1%	73.2%
Other Industrials	12	79.9%	68.7%	71.4%
Technology	16	82.4%	58.3%	75.4%
Telecommunications & Media	9	83.4%	78.2%	74.8%
All Sectors	100	82.6%	69.5%	75.3%

Note: N = number of companies among those selected; $\text{AVG}_{\text{Firm}} (\frac{\sum_{\text{ERA}} \text{CPEP}_{\text{Acad/RPO}}}{\sum_{\text{All Countries}} \text{CPEP}_{\text{Acad/RPO}}})$ = Average across firms of the sum of co-publications with at least one external partner in the academic or RPO sector across ERA countries in which the partner is located divided by the sum of co-publications with at least one external partner in the academic or RPO sector across all countries in which the partner is located; $\text{AVG}_{\text{Firm}} (\frac{\sum_{\text{ERA}} \text{CPEP}_{\text{Firm}}}{\sum_{\text{All Countries}} \text{CPEP}_{\text{Firm}}})$ = Average across firms of the sum of co-publications with at least one external partner in the private sector across ERA countries in which the partner is located divided by the sum of co-publications with at least one external partner in the private sector across all countries in which the partner is located; $\text{AVG}_{\text{Firm}} (\frac{\sum_{\text{ERA}} \text{UCPEP}}{\sum_{\text{All Countries}} \text{UCPEP}})$ = Average across firms of the sum of co-publications with at least one unclassified external partner across ERA countries in which the partner is located divided by the sum of co-publications with at least one unclassified external partner across all countries in which the partner is located. Since unknown countries are not considered in the denominators for these indicators, the corresponding scores for co-publications with partners located outside the ERA are equal to one minus the scores in the table. The colour gradient informs on the extent of negative (red) or positive (green) departure (in percentage points) from the average for all sectors combined (white); the darker the colour, the stronger the departure. (Note that the gradient is normalised relative to the strongest departure in the table.)

Source: Computed by Science-Metrix using Scopus

3 CONCLUSION

Using bibliometric indicators for the 2007–2011 period, the present report first measures the scientific production of companies within the ERA and portrays the geographic distribution of their scientific activities (Section 3.1). It subsequently characterises their collaboration profiles (Section 3.2). The analysis is performed for all S&T fields combined (i.e., using the entirety of the peer-reviewed literature covered in the Scopus database). The report focuses on the 100 firms that have published the greatest number of publications during that period to report exclusively on the most research-intensive companies within the ERA zone; this includes multinational companies with activities outside the ERA, although only their production within this zone is considered.

3.1 GEOGRAPHIC DISTRIBUTION OF THE SCIENTIFIC PRODUCTION OF THE MOST PUBLISHING COMPANIES WITHIN THE ERA

This section starts with an overview at the level of individual companies (Section 3.1.1) and ends with an overview at the level of industrial sectors aggregating companies' data only for the most publishing companies within the ERA (Section 3.1.2).

3.1.1 Company Level

Of the 100 companies publishing the most within the ERA, Siemens—a German multinational conglomerate company specialising in electronic and electrical equipment—leads, having published a quarter more scientific papers (i.e., about 5,000 publications in FULL) than the second most publishing firm from 2007 to 2011 in ERA countries. Coming next is a group of other Europe-based multinational firms whose number of publications (FULL) ranges from about 3,000 to 4,000. In the Pharmaceutical & Biotechnology Sector, they include GlaxoSmithKline (headquartered in the UK; nearly 4,000 papers), Novartis International (Switzerland; about 3,900), and AstraZeneca (UK; about 3,600). They also include Philips Electronics (Netherlands; about 3,500) in the Leisure Goods sector, Thales (France; about 3,400) in Aerospace & Defence, and STMicroelectronics (Switzerland; nearly 3,000) in Technology Hardware & Equipment.

Although 50% of the selected companies published their scientific contributions with affiliates distributed in at least six countries and 15 cities of the ERA, it should be noted that the research activities of private firms are generally highly concentrated in a few sites in Europe. Indeed, at least 65% of a company's scientific output was produced in only two countries and at least 40% of a company's scientific output was produced in only two cities for 80% of the selected companies. Also, the city in which the firm's affiliates published the most is responsible for 40% or more of the company's papers for 43% of selected companies, and it published at least twice as many papers as the second most important city for 52% of selected companies. The concentration of output is noticeable even when the scientific output is distributed over a wide range of sites; for instance, Pfizer, Philips Electronics, Novartis, BASF, Microsoft, Roche Holding, Hewlett-Packard, Fiat, Nestlé, Deutsche Telekom, Vodafone, Rolls-Royce, Takeda Pharmaceutical, and Boehringer Ingelheim each have affiliates in at least 10 cities while at least 50% of their output is produced in only one city.

Although there is a correlation between the number of a company's publications and the number of countries or cities in which they are distributed, the correlation coefficients are moderately high

with many outliers. As such, those companies with the most publications are not necessarily those with affiliates dispersed in the largest number of countries or cities.

When ranked by the number of cities in which a company's affiliates are located, the top ten companies are Siemens (106), ABB Group (56), and Thales (55); followed by General Electric, Merck, Pfizer, EADS, Ericsson, Robert Bosch, and Sanofi (all between 36 and 51 affiliates). Alternatively, those with the smallest number of affiliates include Thomson Reuters, BT, and Soitec (4 affiliates each); NTT Docomo (3 affiliates); Yahoo! (2 affiliates); and NIZO Food Research, ASML, and Honda (1 affiliate each).

The countries in which the affiliates of the 100 companies publishing the most (based on the leading two cities) within the ERA are most often located include, in descending order, Germany, the UK, France, Italy, Switzerland, the Netherlands, and Sweden.

3.1.2 Industry Level

Among the 100 companies having published the most within the ERA from 2007 to 2011, the most prevalent industry is Health Care (ICB 4), which accounted for about 24% of all publications (based on fractional counting) produced by the leading firms (i.e., the top 100). This industry is followed closely by Industrials (ICB 2; 22%).

The Technology (ICB 9) and the Consumer Goods (ICB 3) industries also appear to be highly research-intensive with, respectively, a 20% and 11% share of publications produced by leading firms. Other industries appearing among the 100 leading firms include Oil & Gas, Basic Materials, Consumer Services, Telecommunications, and Utilities.

At the sectorial level, Pharmaceuticals & Biotechnology (3-digit ICB: 457) clearly stands out with a 23% share of publications produced by leading firms, a 17% share of leading firms, and an average production of 610 papers per firm. Other sectors that stand out among the 100 firms publishing the most within the ERA include Technology Hardware & Equipment, Aerospace & Defence, and Electronic & Electrical Equipment.

Industries and their sectors do not differ profoundly in terms of the propensity of companies to disperse their research activities over multiple sites across countries and cities. Dispersion only appears slightly less than the average across industries in Consumer Goods, Telecommunications, and Utilities. However, in the latter two, the population of companies is less than 10.

3.2 COLLABORATION PATTERNS OF COMPANIES HAVING THE LARGEST SCIENTIFIC PRODUCTION WITHIN THE ERA

This section starts with an overview at the level of individual companies (Section 3.2.1) and ends with Section 3.2.2 and 3.2.3, which provide an overview at the level of industrial sectors aggregating companies' data only for the most publishing companies within the ERA.

3.2.1 Company Level

The scientific collaboration network of the 100 companies having published the most within the ERA allowed the identification of seven coherent sub-networks (or communities), each corresponding to organisations that are active in similar industrial sectors/markets. These clusters represent research-intensive industrial sectors in Europe and appear to cover nine of the 17 thematic priorities of the European Commission under the FP7. These groupings are as follows:

- **Aerospace & Defence:** Aerospace & Defence (ICB 271; light purple cluster);
- **Automobiles & Parts:** Automobiles & Parts (ICB 335; light blue cluster);
- **Basic Materials & Food Producers:** Chemicals (ICB 135; mostly dark blue and red clusters), Industrial Metals & Mining (ICB 175; dark blue and green cluster), Food Producers (ICB 357; red cluster), and Personal Goods (376; red cluster);
- **Health Care:** Health Care Equipment & Services (ICB 453; various clusters) and Pharmaceuticals and Biotechnology (ICB 457; almost exclusively dark blue cluster);
- **Oil, Gas & Utilities:** Oil & Gas (ICB 0001; green cluster), Electricity (ICB 753; green cluster), and Gas, Water & Multi-Utilities (ICB 757; green cluster);
- **Other Industrials:** Other sectors under the Industrials category (ICB 235, 272, 273, 275 and 279; various clusters);
- **Technology:** Software & Computer Services (ICB 953; orange cluster), Technology Hardware & Equipment (ICB 957; mainly orange cluster), and Leisure Goods (374; orange cluster); and
- **Telecommunications & Media:** Fixed Line and Mobile Telecommunications (ICB 653 and 657; mostly light purple and orange clusters) and Media (ICB 555; only two companies, one in the green cluster and the other in the orange cluster).

This clustering pattern in the network indicates that the research fields of highest relevance to companies in a given industrial sector/market are generally similar. Furthermore, the proximity between various clusters in the network allowed the study to highlight the cross-applicability of research findings from one sector to the next. For example, the Aerospace & Defence grouping is positioned near the Telecommunications & Media grouping. In fact, it even includes, at its periphery, a major player in the Telecommunications industry (France Telecom). This is not surprising since research progress in ICT is known to contribute importantly to the developments taking place in the Aerospace & Defence industry.

Siemens is the most central player in the collaboration network of the most publishing firms in the ERA, with connections to all clusters. This is not surprising since it is the most publishing firm within the ERA and because it is a multinational company specialising in the Electronic & Electrical Equipment sector for the ICT, energy, health care, and industry markets. It is actually an important bridge between the two poles in the network: companies performing research in the life sciences and those performing research in the physical, material, and engineering sciences.

Thomson Reuters is by far the company with the highest proportion of exclusively intra-company publications (limited to the ERA); note that because the study focuses on the portion of a company's scientific output that was produced within the ERA zone, the external partners of a company include its affiliates located outside the ERA. In fact, it has no link, above the minimum threshold of five co-publications over five years, to other companies in the network. It is followed by Vodafone (44%), then by ThyssenKrupp, Technicolor, BAE Systems, ABB Group, RWE, and Deutsche Telekom (all between 30% and 37%).

Turning to publications co-authored with external partners, the affiliates of Thomson Reuters in the ERA zone are the least opened to co-authorship with external partners, with only 9% of its papers being co-authored with at least one external partner. Eighteen companies published 90% or more of their papers with at least one external partner, those at the top of the list being a diversified group including Medtronic, Danone, Renault, Eli Lilly and Company, Bruker Corp, ASML, and NIZO Food Research, all with 93% or more. Companies with the highest proportions of co-publications with at least one external partner in the academic or RPO sector are Danone (95%), Medtronic (93%), Novozymes (88%), Bruker Corp (86%), NIZO Food Research (85%), Fiat (85%), and Renault (84%). Companies with the highest proportion of co-publications with at

least one external partner in the private sector include ASML (51%), Eli Lilly and Company (45%), IMS Health (44%), Johnson & Johnson (38%), Vattenfall (37%), UCB (37%), and Boehringer Ingelheim (37%).

3.2.2 Comparative Analysis of the Propensity of the 100 Companies Publishing the Most within the ERA to Collaborate Internally or Externally across Industrial Sectors

Using the above seven grouping of industrial sectors (see Section 3.2.1), the propensity of the most publishing firms to collaborate internally (i.e., between affiliates of a company excluding affiliates outside the ERA) as well as externally was compared across industrial sectors.

Generally, the most publishing ERA companies publish a great majority of their papers with external partners; when all industrial sectors are combined, companies co-authored, on average, about 80% of their publications with at least one external partner and this average score never drops below 66% regardless of the industrial sector. Also, a far greater majority of papers co-published with at least one external partner involved the academic/RPO sector rather than the private sector (69% versus 22% when all industrial sectors are combined). Again, this tendency is the same irrespective of the industrial sector, but with some level of variation across industrial groupings.

The Telecommunications & Media industry is generally less open to collaboration with external partners compared to the other industrial groupings. Also interesting is the fact that the papers produced by companies in the Telecommunications & Media sectors do not involve multiple affiliates as often as in other sectors; however, this does not come as a surprise since the average number of affiliates (i.e., cities in which a company is located) in the Telecommunications & Media sectors was smaller than in any other sectors. The only other sectors with a similar collaboration pattern include Aerospace & Defence and Oil, Gas & Utilities. However, departures from the averages for all sectors combined are not pronounced in these cases.

The most opposite collaboration pattern was observed for the Health Care sector, which is mostly made up of pharmaceutical firms. It is in this sector that the average percentages of exclusively intra-company publications and exclusively intra-company publications with a single affiliate are the smallest. Alternatively, the average percentages of co-publications with at least one external partner, with at least one academic or RPO partner, or with at least one external firm are highest in this sector. The only other sector with a similar collaboration pattern is Basic Materials & Food Producers, which also involves an appreciable share of firms in the life sciences. However, departures from the averages for all sectors combined are not as pronounced as they were for Health Care.

3.2.3 Comparative Analysis of the Propensity of the 100 Companies Publishing the Most within the ERA to Collaborate within and Outside the ERA across Industrial Sectors

Using the above seven grouping of industrial sectors (see Section 3.2.1), the propensity of the most publishing firms to collaborate externally within and outside the ERA with organisations in the academic or RPO sector, or alternatively with organisations in the private sector, was compared across industrial sectors.

The 100 most publishing companies within the ERA generally collaborate more extensively with external partners located within the ERA than those outside the ERA regardless of whether these partners are classified in the academic/RPO or private sectors; the percentage of co-publications involving countries within the ERA ranged from 70% (with other firms) to 83% (with academic/RPO partners), with only about 17% to 30% of co-publications involving countries outside the ERA. (Double counting of co-publications across countries occurs.)

In two groupings of industrial sectors, namely Aerospace & Defence and Automobiles & Parts, ERA companies tend to collaborate more with external partners (either academic/RPO or private partners) located within rather than outside the ERA to a greater extent than is generally the case (i.e., when all sectors are combined).

Although collaboration in Health Care remains more pronounced within than outside the ERA, the propensity to collaborate with external partners (either academic/RPO or private partners) outside the ERA is more pronounced in this grouping of industrial sectors than is generally the case (i.e., when all sectors are combined).

4 METHODS

The bibliometric indicators in this report were produced using the Scopus database (Elsevier). Science-Metrix hosts an in-house version of Scopus in the form of an SQL-relational database and is licensed to use this database to produce bibliometric data and reports. Science-Metrix has carefully conditioned the database for the purpose of producing large-scale comparative bibliometric analyses. This edition of the report covers the period from 2007 to 2011. A subsequent update will follow, covering the 2007–2012 period.

Section 4.1 details the procedure used in identifying the 100 companies publishing the most within the ERA zone. Subsequently, Section 4.2 presents the indicators produced at the level of companies, whereas Section 4.3 presents the approaches used to report the data.

4.1 STANDARDISATION OF AUTHOR ADDRESSES FOR FIRMS AND SUBSIDIARIES

The method used to identify and standardise the author addresses for the leading companies is based on two approaches. First, a matching approach was applied based on the list of firms found in the 2012 EU Industrial R&D Scoreboard.⁵ The second, and complementary, approach is based on a frequency ranking of the publication output of companies found in Scopus. This involves cleaning the preliminary lists of institutions identified as companies, beginning with those that publish the most. The merger of those two approaches results in a list of identified companies, from which subsidiaries are identified, extracted from the publication database, and associated with their parent companies. The steps involved in these two approaches are described below.

When the addresses of leading firms and their subsidiaries have been standardised and validated via quality control processes, the unique and standardised organisation names are reincorporated into the database, where they can be used to produce robust bibliometric data and statistics at the institutional level.

4.1.1 Delineating the Dataset

As per the scope of the study, the dataset used to identify leading firms included all papers published from 2007 to 2011 for all countries within the ERA. Note that, as in the rest of the study, only peer-reviewed articles were included in the dataset. Author addresses associated with schools, universities, academies, colleges, or other public and academic institutions were then removed from the dataset based on multi-lingual keywords. Care was taken during this step not to remove companies that would have these words in their names. The resulting dataset (i.e., the pool of resulting author addresses) was then considered the target list within which to identify and clean company names.

⁵ JRC & DG Research and Innovation (2013). EU R&D Scoreboard: The 2012 EU Industrial R&D Investment Scoreboard, 126 pages.

4.1.2 List Matching

The European Commission provided Science-Metrix with a list of leading companies based on their research and development (R&D) investments.⁶ This list was used as a starting point, since high investment in R&D could indicate that a company is likely to publish scientific papers. In brief, this approach sought to identify all papers for which the author addresses matched the names of the companies on the list provided by the European Commission. Note that, prior to being used in the matching process, this list was cleaned to prevent common words (e.g., “the”, “inc.”, “corp.”) interfering with the similarity scoring and ranking of matches.

A customised matching process was elaborated and an application was developed to aid this process. This application used three different strategies to enhance the efficiency and precision of the matching process. The first strategy was to match company names that consist of short words and acronyms. The matching using this strategy was based on exact matches only. Indeed, while typos frequently occur in company names (e.g., “Siamens” instead of “Siemens”), such typos are much more unforgiving in acronyms (e.g., “SNP” instead of “SAP”). Thus, for acronyms and company names of fewer than four characters, an exact match was used. The match was further validated by looking at the geographic location of the matched addresses.

The second strategy used a Jaro-Winkler fuzzy matching algorithm. This algorithm was developed by the US Census Bureau to link names found in different and heterogeneous databases. This algorithm was therefore a perfect tool to use in this case. Tests showed that a similarity threshold of 85% gave good quality and repeatable results. The resulting matching list was manually revised and cleaned. Note that the Jaro-Winkler algorithm has some limitations. In particular, it does not perform well when matching several words and it heavily penalises typos or missing words at the beginning of the matching or matched expression. Thus, a third strategy had to be developed to match the remaining occurrences of companies in the dataset.

The third strategy was similar to the second one, in that a custom algorithm heavily influenced by the Jaro-Winkler algorithm was developed. Instead of matching and calculating a score for the entirety of the expression (as the Jaro-Winkler algorithm does), this custom algorithm split the expression into individual words. The words from the list were then matched with the words from the dataset. If a similarity score of 95% was obtained for every word of the expression, a similarity score was computed for the expression as a whole. This composite score was the product of all of the matches (i.e., for each of the words) in the expression. This strategy allowed for matches to be made on long expressions, including those with typos, while achieving good recall and precision. The results of these three strategies were merged and validated.

4.1.3 Frequency-Based Extraction

The list-based approach was useful to extract European companies that published in Europe. However, this approach overlooked several foreign companies that also publish within the ERA. To identify these companies, a frequency-based method was needed, in which affiliations are grouped and ranked by their number of publications.

⁶ Ibid.

Starting with the same dataset (described in Section 4.1.1) and excluding the companies that had already been cleaned using the first approach, a raw list of institutions was put together based on their number of papers. Manual cleaning of this raw list was conducted to remove author addresses for colleges, universities, and other inappropriate affiliations that remained. Redundant information was also manually removed until a list of company names was obtained. This list was compared to that provided by the European Commission to ensure that it only contained companies not previously identified. Furthermore, efforts were made using various online resources (see below) to ensure that this list contained only parent companies, not subsidiaries. Using this list, the same list-matching approach as described above was applied to match additional companies. The results were merged with those from the first approach to create a master parent company list.

4.1.4 Retrieval, Cleaning, and Matching of Subsidiaries

Based on the master parent company list, a subset of companies ranked by their publication numbers was constructed from the merger of two lists obtained using the preceding methods. A list of these companies' subsidiaries had to be retrieved, cleaned, and matched to their publications.

Various online resources (i.e., company repositories such as Kompass and Mergent Online) were used to identify subsidiaries linked to these companies. Unfortunately, these databases generally only identify subsidiaries of public companies and a limited number of private companies. For private companies, a reasonable level of effort was made to find their subsidiaries (e.g., by consulting annual reports, financial filings, and press releases). Ultimately, a list of about 17,000 subsidiaries was compiled.

As seen in the case of the parent companies, this list of subsidiaries contained several words and acronyms (e.g., "the", "inc.", "corp.") that could impede the matching precision and recall and were therefore removed. This list was then uploaded to the matching tool to be matched against the dataset. The resulting list of matched subsidiaries was then cleaned and subsidiaries were associated with their parent companies. As a result, the publications produced by companies (i.e., the parent and its subsidiaries) could be counted in order to produce the list of the 100 companies which published the most in the ERA from 2007 to 2011.

4.1.5 Identifying the Main Affiliates of Each Company within the ERA

To investigate the geographic distribution of each company's research output within the ERA, the cities appearing in the author addresses matched to a given company were cleaned and used as a proxy to identify a company's affiliates. Please note that even in the case of multinational firms, only those papers produced by affiliates located within the ERA are considered. Furthermore, only cities in which a company's affiliates have published at least one paper (based on fractional counting prior to cleaning) are considered. All remaining cities and the countries in which they are located are classified as "Unknown". These were left out since it does not really make sense to consider the scientific production of cities in which a company published less than one paper (based on fractional counting) in the past four years to appreciate its geographical spread. For instance, cities with less than one paper (based on fractional counting) could result from erroneous entries in the database or from noise in a company's distribution pattern (e.g., one or a few punctual publication(s) with no continuation of research activity in the corresponding city).

4.2 BIBLIOMETRIC INDICATORS

In producing bibliometric data, only documents published in refereed scientific journals are retained, as these documents have been reviewed by peers prior to being accepted for publication. The peer-review process ensures that the research is of good quality and constitutes an original contribution to scientific knowledge. In the context of bibliometrics, these documents are collectively referred to as “publications”.

Bibliometric indicators were calculated using either full counting (i.e., each paper was counted once for each entity listed in the address field) or fractional counting (i.e., each author/entity is attributed a fraction of the paper, so that the total across entities adds up to the total number of papers), as specified.

- **Number of publications:** Number of peer-reviewed scientific publications written by authors located in a given organisational entity (e.g., a company) or its affiliates within the ERA. Both full and fractional counting are used.
- **Number of exclusively intra-company publications:** Only those publications produced by the company without any contribution from authors in external organisations or from its affiliates located outside the ERA.
- **Number of exclusively intra-company publications with a single affiliate:** Only those publications produced exclusively by authors from a single affiliate (i.e., city) within the ERA of the considered company.
- **Number of exclusively intra-company publications with multiple affiliates:** Only those publications produced by authors from at least two affiliates (i.e., cities) within the ERA of the considered company without any external partner. (External partners of a company include affiliates of the company located outside the ERA.)
- **Number of co-publications:** Those publications of the considered company that were co-authored with at least one external partner. (External partners of a company include its affiliates located outside the ERA.)
- **Number of co-publications (Acad or RPO):** Those publications of the considered company that were co-authored with at least one external partner in the academic or RPO sector.
- **Number of co-publications (Firm):** Those publications of the considered company that were co-authored with at least one external partner in the private sector. (External partners of a company include its affiliates located outside the ERA.)
- **Number of co-publications (Unclassified):** Those publications of the considered company that were co-authored with at least one external partner for which the sector is unknown.

4.3 GRAPHIC REPRESENTATION OF DATA

To ease the visualisation and interpretation of the collaboration patterns of companies, Science-Metrix has produced a collaboration network of the 100 firms which published the most within the ERA from 2007 to 2011.

Collaboration networks are used to help illustrate the scientific collaboration between entities (e.g., companies). Based on a matrix (either a symmetric or asymmetric matrix in full or sparse format) cross-linking the number of co-publications by the relevant entities, the GEPHI software program was used to produce a visual representation of the strength of the relationships between the selected 100 companies. More specifically, a force atlas layout algorithm was used to establish the relative locations of the entities in the graphic representation.

In the network, each company is represented by a node and identified using a label. The size of nodes is proportional to the centrality (i.e., weighted PageRank) of companies. The width of links is proportional to the number of co-publications between any pair of companies. Only links representing at least five co-publications (i.e., one per year on average) are shown. Companies were clustered using an algorithm for the rapid unfolding of communities in large networks; the resulting clustering has been displayed by colouring nodes in the network.

The advantage of the collaboration network is that it allows the rapid visualisation of major “hubs” in the network, i.e., entities for which the centrality is highest. It also allows for the rapid visualisation of ties between companies.

European Commission

Scientific Output and Collaboration of Companies Publishing the Most in the ERA

Luxembourg: Publications Office of the European Union

2013 — I-VIII, 28 pp — 21 x 29,7 cm

ISBN 978-92-79-33014-8

doi:10.2777/37043

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/eurodirect/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>).

Priced subscriptions:

- via one of the sales agents of the Publications Office of the European Union (http://publications.europa.eu/others/agents/index_en.htm).

This report is part of a broader study aimed at characterising the scientific performance - including impact and collaboration patterns - of countries, regions, and research performers (such as universities, public research institutes, and companies) with an emphasis on the European Research Area (ERA). The analyses presented in the report rely on a series of indicators designed to take into account national and sector specificities. These indicators also allow for a comprehensive analysis of the evolution, interconnectivity, performance, and impact of national research and innovation systems in Europe. In measuring progress towards past and current objectives, this information aims to support the coherent development of research policies for the ERA.

The current report measures the scientific production of research-intensive companies within the ERA zone, portrays the geographic distribution of their scientific activities, and characterises their collaboration profiles. Data are presented for the top 100 firms which published the most within the ERA zone from 2007 to 2011. While this includes multinational companies with activities outside the ERA, only their production within this zone is considered.

Studies and reports



Publications Office



doi:10.2777/37043