

RoRI Working Paper No. 8

Career pathways in research: the current data landscape

Zeynep Anli, Josephine Bergmans, Inge van der Weijden
April 2022

Produced in partnership with:



**Universiteit
Leiden**
The Netherlands

About the authors

[Zeynep Anli](#) is a Researcher at the Centre for the Arts in Society (LUCAS), Leiden University.
z.anli@cwts.leidenuniv.nl; <https://www.cwts.nl/people/zeynep-anli>

[Josephine Bergmans](#) worked until 2021 as a Project Coordinator of several European projects at the Centre for Science and Technology Studies (CWTS), Leiden University.
j.e.bergmans@cwts.leidenuniv.nl; <https://www.cwts.nl/people/josephinebergmans>

[Inge van der Weijden](#) is a Senior Researcher and Coordinator of the Thematic Hub Academic Careers at the Centre for Science and Technology Studies (CWTS), Leiden University.
i.c.m.van.der.weijden@cwts.leidenuniv.nl; <https://orcid.org/0000-0001-5255-3430>;
<https://www.cwts.nl/people/ingevanderweijden>

For citation:

Anli, Z., Bergmans, J., and van der Weijden, I. (2022) *Career pathways in research: The current data landscape*. RoRI Working Paper No. 8., April 22. DOI: [10.6084/m9.figshare.19609512](https://doi.org/10.6084/m9.figshare.19609512)

Appendix 2: Supplementary data on 111 career pathway studies is published as a separate file on Figshare DOI: [10.6084/m9.figshare.19609542](https://doi.org/10.6084/m9.figshare.19609542)

Acknowledgements

We would like to express their gratitude to the members of the RoRI Pathways Steering Group, who provided valuable advice and guidance during this study, and to Anne-Marie Coriat for her overall leadership of this strand of RoRI's work.



<http://researchonresearch.org>

Contents

Introduction	2
2. Findings of the literature study	4
2.1 Methodology	4
2.2 Key findings	5
2.3 Representative examples	6
3. Findings of the interviews	8
3.1 Methodology	8
3.2 Key findings	8
4. ORCID: a promising data source	12
Conclusions	14
 Appendix 1: Interview topic guide	 15
Appendix 2: Supplementary data on 111 career pathway studies ¹	

¹ *Appendix 2: Supplementary data on 111 career pathway studies* is published as a separate file on Figshare
DOI: 10.6084/m9.figshare.19609542

1. Introduction

A diverse research community—with a broad range of perspectives, backgrounds, skills, knowledge and experience—is a stronger research community. We need a diverse research workforce to ensure that the research done benefits everyone, but there is much we still don't understand about career pathways in and through research.

Through its *Career Pathways in Research* project², the Research on Research Institute (RoRI), together with its partners, is investigating how data is collected about research careers, to work out how best to collect and analyse this data in the future. The project focuses on research career pathways in six countries: Austria, Canada, Denmark, Germany, the UK and the USA.

The project has been designed collaboratively with a number of RoRI partners: Austrian Science Fund (FWF), Canadian Institutes of Health Research (CIHR), Chan Zuckerberg Initiative (CZI), Howard Hughes Medical Institute (HHMI), Leiden University, Health Research BC, Novo Nordisk Foundation, UK Research and Innovation (UKRI), Volkswagen Foundation and Wellcome Trust. Its goal is to provide insights to help funders, researchers, institutions, and the wider sector better support careers in research and make them more diverse, reliable and equitable.

This report, prepared by a research team at the Centre for Science and Technology Studies (CWTS) at Leiden University, presents the results of the first phase of the *Career Pathways in Research* project. Based on a literature study and interviews with RoRI strategic partners, the report explores what data is available on careers in research. The following three research questions were addressed:

- Who is collecting career data and for what purpose?
- What framework is used to collect data, and how interoperable and open are the data?
- What data do we not have?

² <https://researchonresearch.org/projects#!/tab/273951116-4>

2. Findings of the literature study

2.1 Methodology

Eligibility criteria

To be included in our analysis, studies needed to be:

- (i) conducted on the theme of career pathways of researchers
- (ii) conducted in the USA, UK, Canada, Germany, Denmark or Austria

Information sources and search strategy

Searches for relevant articles were conducted in the Web of Science database in February-March 2021. The following search string were used:

- (i) TI=(career* AND data)
Refined by: COUNTRIES/REGIONS: (USA OR UK OR CANADA OR GERMANY OR DENMARK OR AUSTRIA)
Timespan: All years. Indexes: SCI-EXPANDED, SSCI, A&HCI, ESCI.
- (ii) AB=(career AND phd)
Refined by: COUNTRIES/REGIONS: (USA OR UK OR CANADA OR GERMANY OR DENMARK OR AUSTRIA)
Timespan: All years. Indexes: SCI-EXPANDED, SSCI, A&HCI, ESCI.

The identified articles were screened and the titles and abstracts were checked for eligibility.

Searches for relevant policy documents were conducted in the Overton database in March 2021. Overton is a database of policy documents, guidelines, think-tank publications and working papers. It allowed us to identify documents that are within the perimeter of policy making efforts. The simple string “%career%” was used to search the database.

Manual searches were also conducted to check projects as starting points. The emphasis was on keywords around the context of “career tracking researchers” and “career pathways”, which resulted in finding databases such as [Researchfish](#) and [DataCite](#).

Country focus

Of the 111 research articles and policy reports that were identified, 51 are from the USA, 9 from Canada, 1 from the USA and Canada, 4 from Germany, 2 from Denmark, 11 from the EU in general,

25 from the UK, 2 from the EU and the UK, 3 from multiple countries and another 3 are on a global scale. As shown in Figure 1, no career data studies from Austria were found.

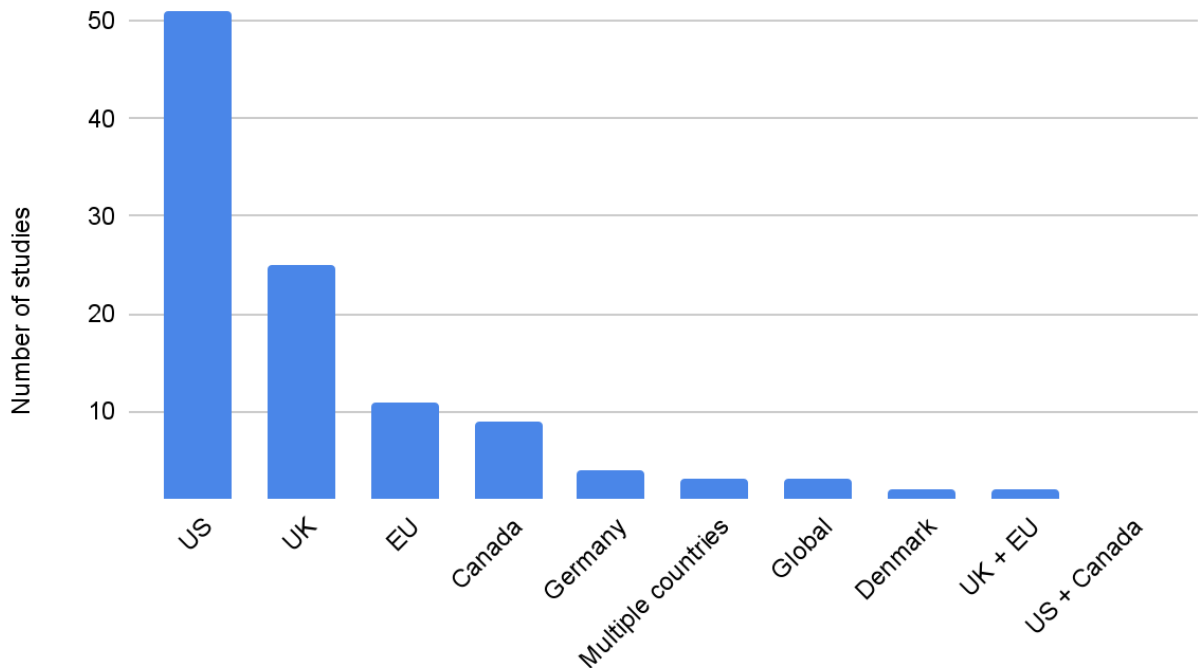


Figure 1: Career data studies from different countries and regions

2.2 Key findings

After screening approximately 1100 research articles and policy reports returned by our search strategy, 111 studies were identified that satisfied our inclusion criteria. These 83 research articles and 28 policy reports were included in our analysis. The most common reason for excluding a research article or policy report was the absence of a geographical match.

Information on the 111 career pathways studies was compiled in an Excel spreadsheet. Table 1 shows the attributes that we used to characterise the research articles and policy reports.

Attribute	Description
Dataset name	Title of the study
Dataset ID (PDF file name)	Unique identifier linking the study to the accompanying PDF file, if the file was accessible
Year	Publication date of the study
Owner	Authors and/or institutions
Partners	Partner institutions, if any
Country	Geographical location where the study took place

Target population	For whom the study was conducted
Data collection	Type of data collection method used
Sampling	What part of the target population was included, and in what way
Indicators	Employment, career satisfaction, mobility of respondents, etc.
Response rate	Percentage of respondents in the sample
Link	Link to the information page of the study or to the PDF file where available
Openly available	Whether the study is open access or not

Table 1: Attributes used to characterise the career data studies

Status of doctoral candidature, discipline of doctoral project, gender and age of respondent, end date of doctoral project, employment situation, career experience, academic positions, career transition, career satisfaction, performance markers, earnings, and career outcomes were among the most common indicators included in the 111 studies.

2.3 Representative examples

The studies below are presented as representative examples:

RISIS Doctoral Candidates and Doctorate Holders Study ProFile

“ProFile aims at identifying determinants of postdoctoral career development and providing information on conditions of doctoral education in a comparative perspective via a monitoring approach. Special attention is paid to the effects of structured doctoral programmes (Graduate Schools) on doctoral education, which have emerged increasingly during the past years.”³

The target population of RISIS consists of doctorate holders with dissertation data in two cohorts (2010 and 2014) in different countries (Germany, Spain, Israel, Norway, Netherlands), in total 5,468 doctorate holders, with a response rate between 5% and 75%, depending on the type of organisation providing the data. The data is openly available.

Postgraduate Career Progression: A Survey of Former SERC Funded Postgraduates, Prepared for the Engineering and Physical Sciences Research Council (EPSRC)

“This report provides a synthesis review of what is known about career choices and impact of UK PhD graduates. It was commissioned by the Economic and Social Research Council (ESRC) as part of its Science in Society strategy and the Research Councils UK (RCUK) Research Careers and Diversity Unit. The review responds to a growing interest in researcher career paths and the skills which doctoral graduates bring in cultural, social and economic terms.”⁴

The review targeted PhD graduates from the UK and made use of pre-existing data. The article is openly available.

³ <http://risis.eu/wp-content/uploads/2015/03/Report-Task-1-Profile.pdf>

⁴ <https://esrc.ukri.org/files/public-engagement/public-dialogues/full-report-phd-graduates-career-choices/>



Humanities PhD Graduates: Desperately Seeking Careers?

“This study addresses this gap by exploring the career experiences of Canadian PhD humanities graduates through descriptive statistics and narrative analysis. Specifically, it highlights the PhD experiences and post-graduation career trajectories of 212 Canadian humanities graduates from 24 universities who graduated between 2004 and 2014. The study offers insight into humanities career challenges, including during their PhD, the range of non-academic careers that humanities graduates find, as well as their work satisfaction and the perceived relevance of the PhD.”⁵

A Randomised Controlled Trial to Improve the Success of Women Assistant Professors

“Given the persistent disparity in the advancement of women compared with men faculty in academic medicine, it is critical to develop effective interventions to enhance women’s careers. We carried out a cluster-randomised, multifaceted intervention to improve the success of women assistant professors at a research-intensive medical school.”⁶ The study focused on US faculty in academic medicine, and more specifically on women assistant professors at a research-intensive medical school. The article is openly available.

⁵ <https://files.eric.ed.gov/fulltext/EJ1189979.pdf>

⁶ <https://www.liebertpub.com/doi/pdf/10.1089/jwh.2016.6025>

3. Findings of the interviews

3.1 Methodology

In January 2021 we conducted interviews with nine research funding organisations. These organisations are all RoRI strategic partners. All interviews were semi-structured. We used a fixed list of questions that we asked to all respondents (see annex), but with the possibility for respondents to significantly discuss their replies and potentially raise relevant points not foreseen by the interviewer. In total nine interviews were conducted, which lasted between 30 and 51 minutes. Eight of the interviews were audio-recorded and automatically transcribed by Amber Script (www.amberscript.com). The interviews were analysed by coding them according to how pertinent they were to the main topics of the interview questions.

3.2 Key findings

Table 2 summarises the career data of grantees that is gathered by the nine funding organisations. The interviews show that career data is (to some extent) collected within the four different phases of awarded projects.

Funding organisation	Phases in project awarded by funding organisation			
	Before funding decision	Project lifespan (e.g. annually)	End of funded project	Beyond funding/grant/project
Volkswagen		Yes	Yes	Upon request 1 external study on impact of doctoral funding (2018) 1 external career tracking study (2009) survey, interviews, bibliometrics analyses
Novo Nordisk		Yes : Rfish	Yes	1 external career tracking study of PhDs and postdoc grant recipients (2020). Use of Web of Science and information on employment, income and travel from registries of Statistics Denmark

FWF	1 small study Data (career steps + mobility) from 1-2 years before funded project. Data collected by websites + research gate	Yes	Yes	2 small studies: 1 survey study 5 years after completion grant 1 survey study for special me every year
UKRI		Yes : Rfish	Yes	Vitae membership Currently establishing an early career research forum, could have some potential for career tracking
Wellcome Trust		Yes	Y: Rfish + 4 year PhD programme: Job following PhD training + career aspirations	Career tracker 2009-2014 1 external qualitative study of careers PhDs (2012)
IHR	CV with affiliation, awards and publications over the last 5 years. Data is not systematically analysed	Upon request	Yes Survey for two programmes	Internal survey 5 year after funding: 2 programmes on doctoral students and 1 on postdocs
CZI		Yes	Yes	Not yet Thinking about creating database for grantees as network function
HHMI/ Janelia		Annual survey to every employed lab head	Yes	Data on current position + institution from funded graduates and postdocs. Use of Google, LinkedIn, ORCID. Renewal review every 7 years: position right after completion training, filled in by lab head

				Searching on Google, LinkedIn and use ORCID
				Thinking about creating alumni network
Health Research BC (British Columbia)		Yes	Yes Survey by email: position after award finished. Since 2019 more detailed career information collected	5 years after grant finished. LinkedIn search: location/sector/job title/organisation

Table 2: Collection of career data

Funding organisations have the ambition to track the careers of its awarded researchers. Currently, tracking career data of researchers beyond the funding period is only done on a small scale, for instance for a specific funding programme. Most funding organisations would like to make further steps to come to a more systematic approach. The most reported aim is to get more information about the impact of a funding programme on science, society and careers. All funding agencies collect output data from funded projects, such as publications and patents, during the term of the projects. A few funding agencies use Researchfish (<https://researchfish.com/>) to collect publication and research impact data of its researchers. According to most interviewees, personal characteristics of grantees are important variables to include in career studies. In general, data on gender and nationality of grantees is known by the funding organisations. Data on race, ethnicity and disability is most of the time unknown.

Interviewees mentioned five categories of career information that are of interest for them to collect data on:


1. Career progress ('career steps, move on the academic career ladder')
2. Employment history (sectors, categories, contract)
3. International mobility
4. Non-academic career development, including reasons for leaving academia*
5. Satisfaction with position and prospects of developments (including income)*

*Survey or interview data is needed to gather information on motivation, attitudes and satisfaction regarding career choices of grantees.

Barriers to collecting career data

Interviewees mentioned six barriers to collecting career data:

1. No alumni network or platform; no contact with grantees after the end of their grant

- 
2. Email addresses need to be searched on the internet (e.g. ResearchGate, LinkedIn), which is time consuming
 3. Lack of information on the project members working/employed on a grant, next to the principal investigator
 4. Limited resources available at the funding organisation
 5. Outsourcing: work cannot always be done by in-house staff members because of lack of data analytics experience and knowledge
 6. Data security issues

Missing career data

Nine types or dimensions of career data were mentioned by interviewees that are currently not collected but are of interest to collect in future:

1. Career data of grantees before they applied to the funding organisation (see the column 'before funding decision' in Table 2)
2. Data on people that are employed by the principal investigator of a grant
3. Career data beyond publication metrics (e.g. journals, publications)
4. Data extending more than five years after a grant has ended
5. International mobility data
6. Non-academic pathways: career data on people who left academia
7. Data on reasons why people stopped applying for funding
8. Longitudinal data on grantees
9. Community building data seen from funder's perspective

Career data sources

Interviewees mentioned four data sources which could be used to track careers of researchers:

1. Tracking people by publications, citations and scientometric data (e.g. Dimensions);
2. LinkedIn (time consuming and GDPR issues);
3. ORCID (Open Researcher and Contributor ID) (see Chapter 4);
4. Vitae (UK; membership is needed);

In the next section we present the results of a more in-depth study of potential uses of ORCID.

4. ORCID: a promising data source

To analyse employment and career paths of researchers, public ORCID data could be a promising data source. ORCID is an open, non-profit, community-driven organisation that provides a unique and persistent identifier to researchers.⁷ ORCID connects researchers to their work in a reliable,

⁷ <https://info.orcid.org/what-is-orcid/>

unambiguous way. The datafile provided by ORCID in 2020 contains a total of 9,584,122 ORCID profiles of researchers from all over the world. This makes ORCID the largest open data source currently available for the analysis of academic career and employment transitions in the world.

ORCID ID

- Unique identifier for individual contributors
- Each ORCID ID is linked to an ORCID profile

ORCID profile

- Profile in which information about a researcher is collected
- Includes information on employment, education, publication activities, funds received
- Can be public, visible to pre-defined organisations, or private

ORCID profiles may include education and employment data, but this data is not always available. In the 2020 file there are approximately 2.4 million profiles which provide at least one employment role. About 2.2 million profiles record an educational role. From ERA countries alone, a total of 728,156 profiles can be found recording at least one employment role and 618,778 recording at least one educational role. Figure 2 shows an example of a public ORCID profile with education and employment information.

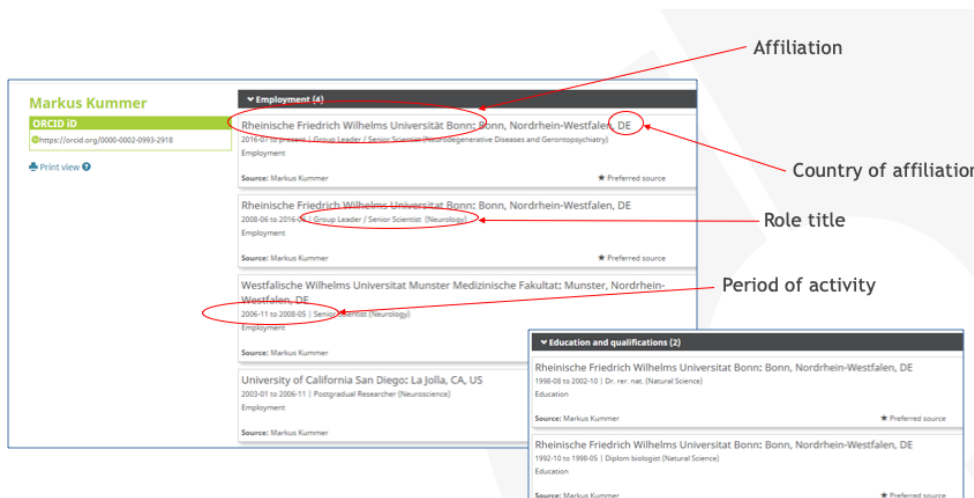


Figure 2: An example of an ORCID record. Source: Costas, Dudek & Francisco-Borrueal, CWTS, 15-04-2021

There is considerable diversity in employment and education role titles. The vocabulary of roles and education events has recently been harmonised by colleagues at CWTS, Leiden University. There are eight levels, which are connected to career stages (R1-R4, used by the European Commission) and seniority. The whole spectrum of academic roles – graduate, PhD, postdoc, researcher, lecturer and professor – as well as non-academic roles is covered. Examples of non-academic roles are professional and administrative roles. Examples of (new) professional roles are project engineer, lab manager, senior physician and biostatistician. Examples of administrative roles are dean, director and (vice)rector.



Use of ORCID profiles by research funders

ORCID profiles can be used to map transitions between roles and career stages of researchers. Such information can be used to investigate transitions of researchers from one position to the next. ORCID is supported by a global community of member organisations, including research institutions, publishers, funders, professional associations, service providers and other stakeholders in the research ecosystem. As an example, the Swiss National Science Foundation (SNSF), one of RoRI's strategic partners, has been a member of ORCID since November 2017. Over 2500 ORCID users have added their ORCID ID in mySNF, the web platform of the SNSF. MySNF enables interaction between applicants, reviewers, research institutions and the administrative offices of the SNSF. It facilitates the submission, evaluation and administration of applications and projects. As a result, ORCID IDs are available also in the SNSF project database. For reporting purposes, researchers can import publications from their ORCID profile into mySNF. Projects funded by the SNSF can be added to a researcher's ORCID profile using a wizard.

Limitations

There are a few limitations to consider when working with ORCID data. Firstly, some researchers do not have an ORCID profile, they have an incomplete profile, or they have not made their profile public. Data is self-reported and has not been harmonised. Secondly, ORCID use is skewed toward early career researchers (PhDs and postdocs) and toward specific countries. ORCID adoption is more prevalent in Europe and less in the US and Asia, although this is currently changing. In some countries (e.g. Italy, Spain and Brazil) there are public mandates which seem to work to increase uptake.

5. Conclusions

In this short paper, we have explored what data is available on careers in research. Based on a literature study and an interview study, as well as further discussions with the members of the steering group of the *Career Pathways in Research* project, some initial conclusions can be drawn.

Our literature study shows that researcher career data is collected by different stakeholders. National bureaus of statistics collect researcher career data at the national level. Universities and research institutions collect data to track the careers of researchers (e.g. cohort of doctorate holders) graduated from their organisations. In addition, alumni associations or platforms (e.g. Marie Curie Alumni Association) track the career development of its members. Also, different types of researcher career data are collected, ranging from register-based labour market data to thematic topics such as degree of career satisfaction. Often the collected researcher career data is not linked to existing data sources.⁸

Surveys are the most common method used to track careers of researchers. It is important to be aware that surveys provide data snapshots. Full insight into the career trajectory of researchers is still lacking.

Our interviews with nine RoRI strategic partners and the discussion with the members of the Career Pathways steering group show that research funders are currently collecting career data of funded researchers on a small scale, for instance only for grantees from a specific funding programme. Most funding organisations have the ambition to take further steps to adopt a more systematic approach.

The Excel spreadsheet with the overview of the literature study includes a number of open access publications. However, the collected researcher career data is usually not made openly available. This means that the data cannot be freely used, reused and redistributed. As suggested in Section 4, public ORCID data could be a promising open data source for future studies of career pathways of researchers.

⁸ However, the US is currently in the process of linking NSF early career doctorates surveys to IRIS administrative data.

Appendix 1: Interview topic guide

Background information

- Working experience (organisations and positions) interviewee
- Goals funding organisation

Research career information available at funding organisation

- Does the funding organisation you work at actively gather research career pathway information? Which data, how, why for what purposes? If not, why not? Plans to do so in the near future?
- What framework is used to collect research career data? Which variables/items? Which research method? Which steps do you take to collect the data?
- When do you start collecting research career data? What is the time frame?
- How interoperable and open are the research career data? What are reasons for you to open up or not open up your datasets?
- What research career data do you not have but should be collected?
- Which challenges exist in collecting the research career data?
- Which advantages and disadvantages do you see from collecting research career data specifically for your funding organisation

General research career information

- Which (openly available) research career datasets are you aware of? In your country or organisation such as LERU, OECD, or the European Commission?
- Does your organisation use these research career datasets? If yes, for what purposes?
- Which general research career pathways can you deduct from this data?
- Which general advantages and disadvantages do you see from collecting research career data?
- Which other organisations could also gather research career data?



<http://researchonresearch.org>