# Developing research career indicators using open data: the RISIS infrastructure[1]

Carolina Cañibano* Richard Woolley* Eric Iversen** Sybille Hinze*** Stefan Hornbostel*** Jakob Tesch***

*cacaisan@upvnet.upv.es; ricwoo@ingenio.upv.es
INGENIO (CSIC-UPV), Universitat Politècnica de València, Valencia, Spain.

**eric.iversen@nifu.no
NIFU, Oslo, Norway

***hinze@dzhw.eu; hornbostel@forschungsinfo.de; tesch@forschungsinfo.de
DZHW (German Centre for Higher Education Research and Science Studies), Berlin, Germany.

## ABSTRACT
This paper introduces the research infrastructure for rsearch and innovation policy studies (RISIS) and its ongoing work on the development of indicators for research careers. The paper first describes the rationale for developing an information system on research careers. It then uses and example to demonstratate the possibilities arising from aggregating open data from different datasets within the RISIS platform to create new information and monitoring possibilies with regard to research careers.

## INTRODUCTION
The need for developing research careers indicators is driven by three main perceptions: that the science and research systems are changing in ways that have profound impacts on research careers; that systematised knowledge about research careers is lacking; and that transformation in data access opens new avenues for research into scientific research careers.

A very large literature exists that attempts to describe and analyse the changing „contract" between scientific research and society. Much of this literature has emerged in the past two or three decades. A set of common themes run through this literature:

- there has been a transformation in the organisation of scientific knowledge production (Gibbons et al. 1994, for a survey see Hessels & van Lente 2008);
- public sector organizations involved in the production and utilization of scientific knowledge have experienced transformations in their missions and their expected response to societal expectations (Etzkowitz & Leydesdorff 2000);
- the matrix of authority relations involving governments, funding agencies, research performing organizations and researchers has been in a state of semi-permanent reform or revolution (Whitley, Gläser & Engwall 2010);
- the organization of research has become more bureaucratized and industrialized, with resource allocation and governance moving increasingly to the project level (Walsh & Lee 2015);

[1] This paper was developed as part of the RISIS project funded through the 7th Framework Program (DOW RISIS (313082) 2013-11-20-1). Funding from RISIS made possible the travel and workshops that allowed the paper"s collaborative development. RISIS also directly funded some of the research time devoted to the development of this work.

This work is licensed under a Creative Commons License: Attribution-NonCommercial-NoDerivatives 4.0 International.

- the framework conditions of science in a „dynamic steady state" (Ziman 1994) with increasing competition for finite resources has led to substantive problems with the operation of scientific labour markets (Stephan 2013).

The combination of these changes has apparently had two main effects on research careers. First, an enlarged cohort of researchers clustered at the post-doctoral level has developed in some scientific fields. In many cases these researchers will employed on consecutive post-doctoral contracts funded by project „soft money" (Stephan 2014). Second, global emphasis on the circulation of knowledge as a driver of innovation and socio-economic development has contributed to the promotion of scientific mobility through collaboration policy agreements and support mechanisms (Jonkers 2010; OECD 2010). In the case of Europe, promotion of an integrationist policy framework (including the European Research Area (ERA) that modifies national labour market and other rules facilitates researchers moving between Member States (MS) to work in research. These effects are likely to lead to a reduction in the dominance of the single-organization research career form and to an enlarged „tracking task" for research into research careers (Franzoni, Scellato & Stephan 2012, 2014; Geuna et al. 2015).

Second, the state-of-the-art research on research careers is significantly fragmented and exhibits substantial gaps, including:

- the lack of a comprehensive approach to researching research careers. The complex nature of research careers means that research tends to focus either on institutional determinants or individual choice, but attempts are rarely made to link the two approaches;
- there is a conspicuous lack of attention to research careers outside of academic research careers – and this research generally neglects that university researchers also have significant teaching, administration and other roles;
- comparability between cohort studies of research careers is lacking, not least because there have been very few attempts to construct variables based on universal categories (Kaulisch & Salerno 2005);
- a limited capacity to account for temporality and the evolution of institutions due to the emphasis on cross-sectional data;
- a lack of clarity about the range of research jobs that exist (Miller & Feldman 2014). Some research roles appear to be „invisible" or appear only as „stepping stones" to the established career pathway when there is evidence that these roles may be persisting and form significant element of contemporary careers (Miller & Feldman 2014; Stephan 2013);
- a general lack of understanding of mixed and hybrid careers (Lin & Bozeman 2006), although some recent attempts have been made to understand the impacts on researchers' career aspirations resulting from working in hybrid public/private organisational forms (Garrett-Jones, Turpin & Diment 2013; Gray et al. 2011; Lam 2011).

Third, a growing momentum exists within science and policy communities with regard to organising the practice and communication of science in a more „open" fashion.

- There is an increasing focus on making research data more openly accessible, both for reuse and to facilitate reproduction studies and validation trials (Crouzier 2015; Moedas 2015).

This work is licensed under a Creative Commons License: Attribution-NonCommercial-NoDerivatives 4.0 International.

- The development of significant online architectures for the distributed production, diffusion and use of research data is accelerating in many fields, including the social sciences.

The Research Infrastructure for Research and Innovation Policy Studies (RISIS) is a project that takes as its primary objective the building of a distributed infrastructure of data relevant for research and innovation dynamics and policies (RISIS nd). It aims to achieve this objective through three complementary activities: opening and developing datasets on issues of continuing relevance to research and innovation research and policy; developing open platforms for building, hosting, curating and developing ad-hoc datasets on research and innovation issues; and developing free and open standards facilitating the interconnection and aggregation of heterogeneous existing datasets (RISIS nd). Current datasets within RISIS cover five dimensions: ERA dynamics; firm innovation dynamics; public sector research; research careers; and a repository on research and innovation policy evaluations (RISIS nd).

RISIS thus provides a single access point to different types of data and existing information relevant to studying research careers. The available data and information includes:

- ETER – database of characteristics of higher education institutions in Europe;
- Leiden Rankings of universities:
- MORE1 and MORE2 – surveys of researcher mobility inside and outside Europe; and
- ProFile – panel data on training conditions and preferred occupational activities of doctoral candidates in German universities (Hauss, Kaulisch and Tesch 2015).

The aim of this paper is to provide an initial demonstration of using the RISIS platform to facilitate the aggregation of these datasets to conduct new enquiries into research careers. The open access to multiple data collections has the potential to lead toward the construction of new indicators that can monitor the evolution of systemic change in research careers.

## THE ROLE OF MOBILITY IN RESEARCH CAREERS: AN ILLUSTRATIVE EXAMPLE

Policymakers are interested in the phenomenon of mobility in research careers. Researcher mobility is important in the interest of capturing investment in highly skilled human capital, benefiting from knowledge and technology transfer and for research system capacity building (Cañibano & Woolley 2015). From the point of view of individual careers, international job mobility is generally the marker of a critical juncture that reconfigures the contexts and networks in which researchers work and therefore strongly conditions their careers.

A potential descriptive research question is: in what circumstances will researchers move to another country to work? Addressing this question will require information on the personal and institutional push-pull factors that are most important and most commonly aligned in facilitating mobility.

From the individual perspective, empirical research questions can be formulated for mobile researchers:

- What motivated you to move to another country to work?
- What obstacles did you encounter in making this move?

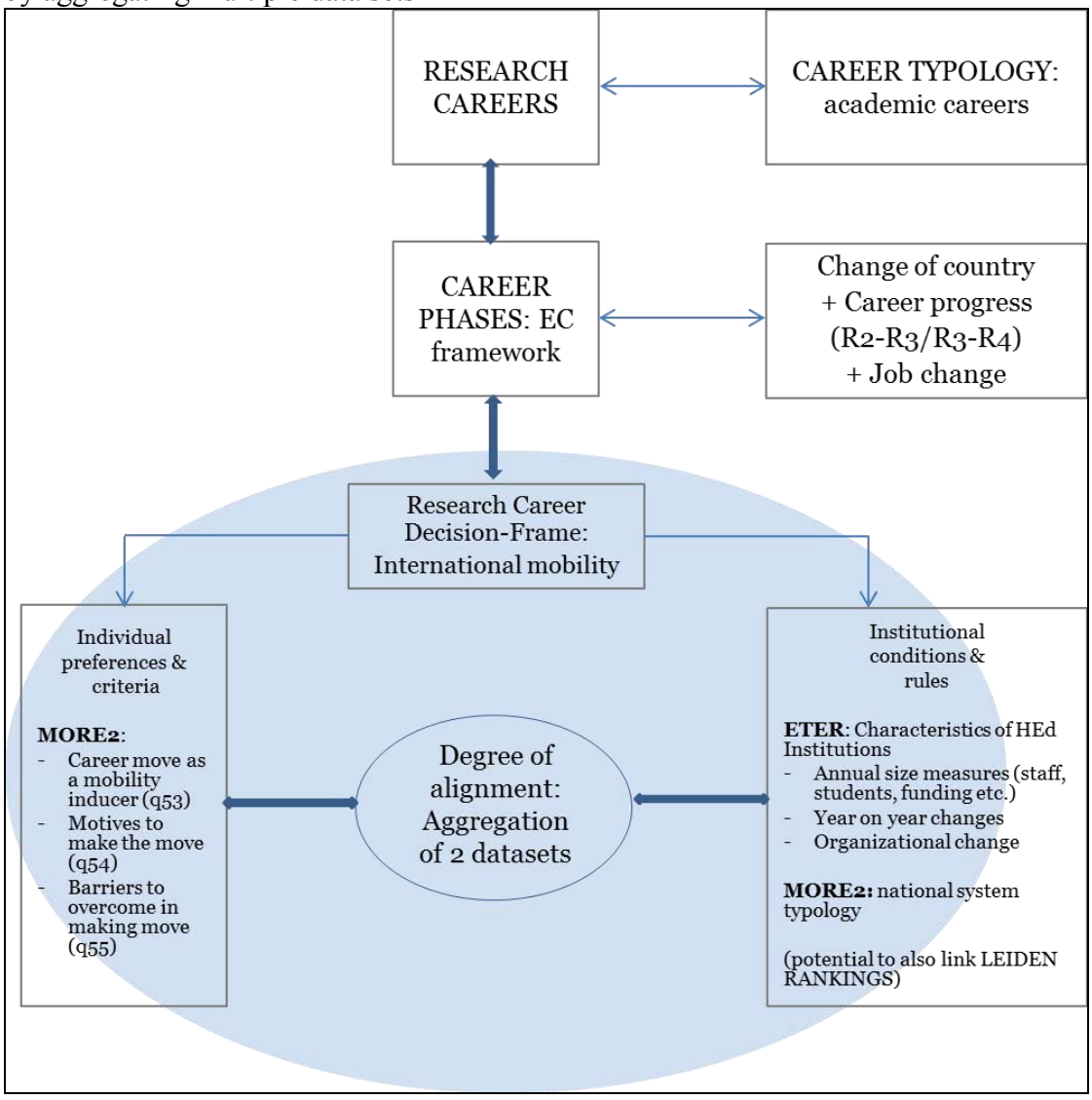
From the institutional perspective, empirical research questions can be formulated about the research performing organisations (RPOs) to which researchers move:

This work is licensed under a Creative Commons License: Attribution-NonCommercial-NoDerivatives 4.0 International.

- What are the characteristics of RPOs that are „magnets" for mobile researchers?
- Are the characteristics of „magnet RPOs" different to those of RPOs overall?
- Are there national system or scientific field effects?

The question „in what circumstances will researchers move to another country to work?" can be addressed using data available through the RISIS data infrastructure, in particular the MORE1, MORE2 and ETER datasets. These data address academic careers.

Figure 1 Using the RISIS data infrastructure to answer a question about researcher mobility by aggregating multiple data sets



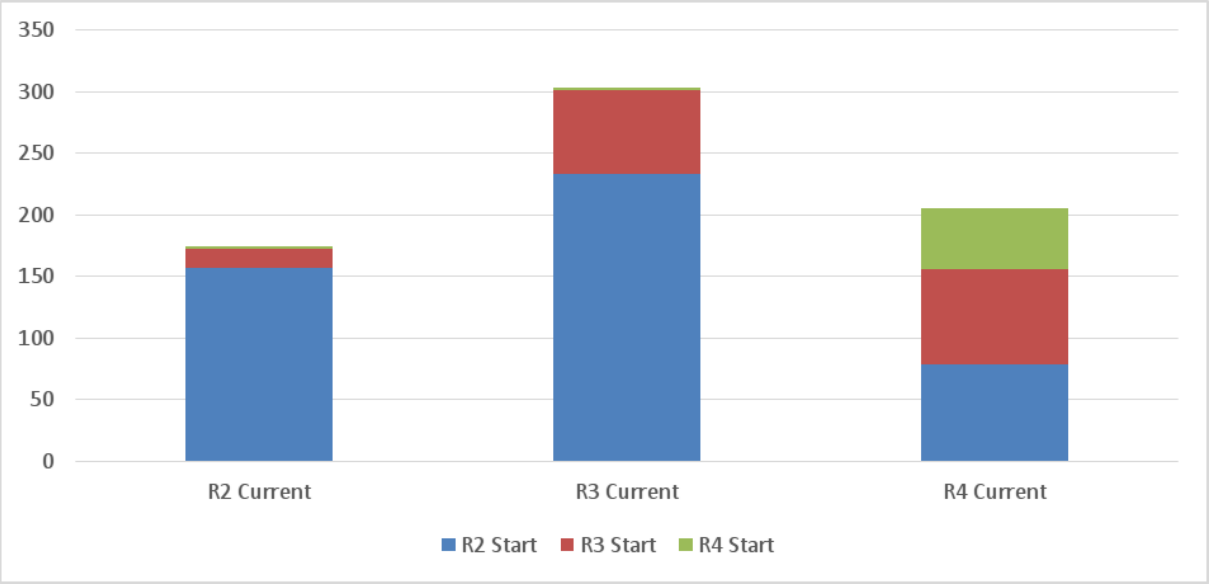

In relation to individual preferences and criteria, the MORE2 (IDEA Consult 2013) dataset can be used to shed light on complex questions associated with decision-frames, including decisions to move to another country. The example descriptive research question focuses on cases where researcher mobility is reported to coincide with critical junctures in career progression. The target population is then researchers who report a change in employer during the past 10 years (MORE2 questionnaire, q50) and/or where the latest episode of mobility coincides with the start-date of their current position (MORE2, q19). Mobile researchers are

This work is licensed under a Creative Commons License: Attribution-NonCommercial-NoDerivatives 4.0 International.

also included if they associate their latest mobility spell with some increase in functional status (MORE2, q53). These characteristics exemplify measures of career progression.

Defined in this (fairly restricted) way, the resulting target population provides a lens on decision frames as shaped by institutional factors. The following figure compares the target-population‟s current level of seniority (in terms of R2-R4) to that reported at the start of their latest mobility spell. The number of respondents in this sub-group (n=684) corresponds to a population of around 6.5 percent (or 82,000) researchers in Europe.

Figure 2 Current career stage of researchers versus their career stage at the beginning of their most recent instance of mobility (n=684)




Source: MORE2 dataset

Researchers who report having functioned as recognized researchers (R2) at the beginning of their most recent geographical move (blue) were generally at an early stage of their careers (and/or at a relatively low level of seniority).

These early stage researchers dominate (72 percent) our target population. We find that a third of these had remained at this level in their current positions, a half had moved to the stage of recognized researcher (R3), while a further 17 percent reported a current status as R4. Similarly, half of those who started their last mobility-stint at R3 had moved to R4 in their current position.

Meanwhile, of those currently in R2 positions (n=175), approximately 10 percent had commenced their latest geographical move in situations equivalent to more advanced career stages, either R4 (n=2) or R3 (n=16). This indicates a career regression which requires further investigation and explanation. In terms of motivations, 80 percent of the target-population reported having been driven by potential for career progression when embarking on a stint abroad as a researcher.

In MORE2 respondents report a great deal of information about their current position, their past positions abroad, and the motivations and barriers experienced along the way. Factors contributing to mobility events such as the types of contracts involved are also reported, in

This work is licensed under a Creative Commons License: Attribution-NonCommercial-NoDerivatives 4.0 International.

addition to information about the reseacher (age, gender, family situation) that may be correlated with particular career stages. Information is most complete for those respondents who report having international experience that is directly related to their career. Information about the geographical and disciplinary contexts can on this basis be analyzed. For example, one can compare the propensity for researcher mobility in different contexts (geographical, institutional, field) with the researcher‟s assessement of the home institution (current position).

In terms of the institutional conditions and rules that shape these mobilities observed in MORE2, the outcomes observed can be linked through RISIS to the **ETER** registry of European institutions to determine which types of institutions may be mobility magnets. The ETER data includes annual information (from 2011) about total (academic staff), number of full-professors, revenue, total expenditure, budget, enrolment, graduates, and other measures. Applying year-on-year data from ETER can be used to relative growth or decline according to a given proxy for growth. This information could be used to explore whether annual changes in the university (size, budget, number of students or staff, or performance) correlates with inflows of mobile researchers. The value-added of the RISIS data infrastructure is exploited in this example by linking different research mobility outcomes to certain organisation profiles, providing an overall picture of how individual and institutional factors align in observed mobility patterns. The distribution of poles of attraction for mobile researchers can not just be identified in terms of organisations/countries of destination, but also some additional qualities associated to the research and systemic environment of these mobility magents can start to be understood.

A further typological variable could also be potentially applied to these linked data. The MORE2 study categorises national academic research systems into a four-category typology depending on a range of institutional and labour market factors. A potential further link could be made through RISIS to a proxy for the overall relative performance, productivity (and hence prestige) of individual unversities, through the Leiden Ranking.

**DISCUSSION**

An important principle of an open science and innovation agenda is the capacity for new users to utilise existing datasets for their self-defined purposes. Opening access to existing datasets can expand the range of questions that these data are used to address. By linking to other datasets the quality of the results that can be generated may also improve. Accessing existing datasets also provides learning opportunities, regarding the types of data collection that might be replicated in a different national or regional context, for example. A case for repeated collection of specialized datasets on research can also more easily be made when the potential exists for these data to benefit users in diverse locations and national systems.

The empirical research questions highlighted in our example demonstrate how the RISIS infrastructure can facilitate such open science and innovation principles. The questions posed are initially addressed through RISIS by aggregating two datasets:

- MORE2 (and MORE1)
  What motivated you to move to another country to work?
  What obstacles did you encounter in making this move?
- and ETER
  What are the characteristics of RPOs that are „magnets“ for mobile researchers?
  Are the characteristics of „magnet RPOs“ different to those of RPOs overall?

This work is licensed under a Creative Commons License: Attribution-NonCommercial-NoDerivatives 4.0 International.

Are there national system or scientific field effects?
However, the potential to link to other datasets, such as ProFile, and information such as the Leiden Rankings, also exists within the RISIS data infrastructure.

The paper demonstrates how the RISIS infrastructure can be used to identify and exploit complementarities between datasets and thus improve our understanding of (changing) research careers. We pointed out several dimensions that can be exploited, including the geographical dimension. An opportunity exists for broad view cross-country surveys to communicate with the more specific national work; points of contact between MORE2 (EU level) and ProFile (Germany) will be explored in future work.

The paper also highlights how RISIS opens possibilities to exploit data complementarities between levels. Researcher-level observations (MORE2) are being enhanced with information about their home institution (ETER) and potentially its relative stature in the higher education/research landscape (Leiden Ranking). In addition, there is a temporal dimension that the infrastructure captures. Some datasets (such as MORE and ProFile) are periodic. Provided internal consistency is maintained, longitudinal information can be collated (MORE2 and the upcoming MORE3) and information accumulated. Indicators with a broad coverage that are fit for monitoring the evolution of research careers over time can thus be the eventual outcome of the RISIS development process. The open data platform approach ensures that use of the available data will be driven by (heterogeneous) research questions and hence open to a diversity of ongoing approaches.

There are of course numerous limitations to these various data and to the interpretations that can be made of the linked data construct at this stage in the development process. The purpose of this methodological paper is to demonstrate the linking of research questions and the RISIS data infrastructure in studying research careers. The paper has outlined the development of a new data sharing and aggregation approach to the development of information about research careers that has the potential to provide a basis for future indicator construction and monitoring development.

**REFERENCES AND CITATIONS**

Cañibano, C. and R. Woolley. 2015. “Towards a Socio-Economics of the Brain Drain and Distributed Human Capital.” *International Migration* 53(1): 115–30.

Crouzier, T. 2015. *Science Ecosystem 2.0: How will change occur?*, Directorate-General for Research and Innovation, European Commission, Brussels.

Etzkowitz, H. and L. Leydesdorff. 2000. “The Dynamics of Innovation: From National Systems and „Mode 2“ to a Triple Helix of University–industry–government Relations.” *Research Policy* 29(2): 109–23.

Franzoni, Chiara, Giuseppe Scellato, and Paula Stephan. 2012. “Foreign-Born Scientists: Mobility Patterns for 16 Countries.” *Nature biotechnology* 30(12):1250–53.

Franzoni, Chiara, Giuseppe Scellato, and Paula Stephan. 2014. “The Mover‟s Advantage: The Superior Performance of Migrant Scientists.” *Economics Letters* 122(1):879–93.

This work is licensed under a Creative Commons License: Attribution-NonCommercial-NoDerivatives 4.0 International.

Garrett-Jones, S., Turpin, T. and K. Diment. 2013. "Careers and Organisational Objectives: Managing Competing Interests in Cooeprative Research Centres." Pp. 79–111 in *Cooerative Research Centres and Technical Innovation: Government Policies, Industry Strategies and Organizational Dynamics*, edited by C. Boardman, D. Gray and D. Rivers. New York: Springer.

Geuna, A., Kataishi, R., Toselli, M., Guzmán, M., Lawson, C., Fernandez-Zubieta, A. and B. Barros. 2015. "SiSOB Data Extraction and Codification: A Tool to Analyze Scientific Careers." *Research Policy* 44(9): 1645–1658.

Gibbons, M. et al. 1994. *The New Production of Knowledge: The Dynamics of Science and Research in Contemporary Societies.* London: Sage.

Gläser, Jochen and Grit Laudel. 2015. *The Three Careers of an Academic*.
Gray, D., Sundstrom, E., Tornatzky, L. and L. McGowen. 2011. "When Triple Helix Unravels: A Multi-Case Analysis of Failures in Industry–university Cooperative Research Centres." *Industry and Higher Education* 25(5):333–45.

Hauss, K., Kaulisch, M. and J. Tesch. 2015. "Against All Odds: Determinants of Doctoral Candidates‟ Intention to Enter Academia in Germany." *International Journal for Researcher Development* 6(2): 122-43.

Hessels, L. and H. van Lente. 2008. "Re-Thinking New Knowledge Production: A Literature Review and a Research Agenda." *Research Policy* 37(4):740–60.

IDEA Consult (2013) MORE2 Final Report. Support for Continued Data Collection and Analysis concerning Mobility Patterns and Career Paths of Researchers. Deliverable 8 prepared for the European Commission. Brussels, August 2013.

Jonkers, K. 2010. *Mobility, Migration and the Chinese Scientific Research System*. Routledge, Oxford.

Kaulisch, M. and C. Salerno. 2005. "Comparing Academic Careers Systems : The Cases of Germany, England and US." Pp. 1–21 in *30th ASHE conference*. Philadelphia.

Lam, A. 2011. "University-Industry Collaboration: Careers and Knowledge Governance in Hybrid Organisational Space." *International Journal of Strategic Business Alliances* 2(1/2):135.

Lin, M. and B. Bozeman. 2006. "Researchers‟ Industry Experience and Productivity in University-Industry Research Centers: A „Scientific and Technical Human Capital‟ Explanation." *Journal of Technology Transfer* 31(2): 269–90.

Miller, J. and M. Feldman. 2014. "The Sorcerer‟s Postdoc Apprentice: Uncertain Funding and Contingent Highly Skilled Labour." *Cambridge Journal of Regions, Economy and Society* 7: 289–305.

Moedas, C. 2015. Open Innovation, Open Science, Open to the World, speech to European Commission, 22 June, retrieved from https://ec.europa.eu/research/openscience/index.cfm

This work is licensed under a Creative Commons License: Attribution-NonCommercial-NoDerivatives 4.0 International.

OECD. 2010. “International Mobility.” in *Measuring Innovation: A New Perspective*. OECD Publishing. DOI:10.1787/9789264059474-25-en

RISIS (nd) RISIS: Background on the project, retrieved from http://risis.eu/about/

Stephan, P. 2014. “The Economics of the Postdoctoral Position.” Presentation to the National Science Foundation, April, retrieved from http://sites.gsu.edu/pstephan/economics-of-postdoctoral-position-stephan/

Walsh, J. and Y-N. Lee. 2015. “The Bureaucratization of Science.” *Research Policy* 44(8):1584–1600.

Whitley, R., Gläser, J. and L. Engwall. 2010. *Reconfiguring Knowledge Production: Changing Authority Relationships in the Sciences and Their Consequences for Intellectual Innovation*. Oxford University Press, Oxford.

Ziman, J. 1994. *Prometheus Bound: Science in a Dynamic “Steady State.* Cambridege University Press, Cambridge.

This work is licensed under a Creative Commons License: Attribution-NonCommercial-NoDerivatives 4.0 International.