



Universiteit
Leiden
The Netherlands

Open Data on ANR-funded research: a valuable resource to explore impact pathways in research funding

Virlon, B.; Bith, T.; Leon Escribano, R.M. de; Dupuis, J.; Fort, Y.

Citation

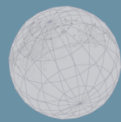
Virlon, B., Bith, T., Leon Escribano, R. M. de, Dupuis, J., & Fort, Y. (2018). Open Data on ANR-funded research: a valuable resource to explore impact pathways in research funding. *Sti 2018 Conference Proceedings*, 1595-1597. Retrieved from <https://hdl.handle.net/1887/65365>

Version: Not Applicable (or Unknown)

License: [Leiden University Non-exclusive license](#)

Downloaded from: <https://hdl.handle.net/1887/65365>

Note: To cite this publication please use the final published version (if applicable).



STI 2018 Leiden

*23rd International Conference on Science and Technology Indicators
"Science, Technology and Innovation Indicators in Transition"*

STI 2018 Conference Proceedings

Proceedings of the 23rd International Conference on Science and Technology Indicators

All papers published in this conference proceedings have been peer reviewed through a peer review process administered by the proceedings Editors. Reviews were conducted by expert referees to the professional and scientific standards expected of a conference proceedings.

Chair of the Conference

Paul Wouters

Scientific Editors

Rodrigo Costas
Thomas Franssen
Alfredo Yegros-Yegros

Layout

Andrea Reyes Elizondo
Suze van der Luijt-Jansen

The articles of this collection can be accessed at <https://hdl.handle.net/1887/64521>

ISBN: 978-90-9031204-0

© of the text: the authors

© 2018 Centre for Science and Technology Studies (CWTS), Leiden University, The Netherlands



This ARTICLE is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

Open Data on ANR-funded research: a valuable resource to explore impact pathways in research funding

Bérangère Virlon*, Tiphaine Bith**, Rosa Mariana de Leon Escribano**, Julie Dupuis**, Yves Fort***

* berangere.virlon@anr.fr

** tiphaine.bith@anr.fr; ** rosa.deleonescribano@anr.fr; ** julie.dupuis@anr.fr

Direction des Opérations Scientifiques, Unité "Etudes, Données et Analyses d'impact", ANR, 50 avenue Daumesnil, 75012 Paris (France)

*** yves.fort@anr.fr

Direction des Opérations Scientifiques, ANR, 50 avenue Daumesnil, 75012 Paris (France)

Introduction

ANR is the French national research agency founded in 2005 and responsible for the project-based research funding. It operates under the authority of the Ministry in charge of Research and Innovation and has to implement the national research strategy.

In compliance with the growing need for « open science », in order (i) to give a better visibility to ANR-funded research, and (ii) to create a favourable environment for studying and generating research-based impacts, the agency recently embraced and fostered « open access » and « open data » practices. ANR notably encourages the **principal investigators** to deposit their publications in an open archive, requires **grant holders and applicants** to prepare a Data Management Plan, and ensures the circulation of data and knowledge on **public grants** by providing data for each funded project in standardized formats to facilitate their broad and easy use.

In the present poster, we will first describe the type of data made available to the public which provides detailed information on the projects grants funded by ANR from 2005 to 2017. Using three proof-of-principle examples, we also illustrate how this data could be exploited and investigated, at different levels (institutions, PIs and research themes) to study the impacts of ANR on the national scientific production.

Materials and methods

Extraction of the ANR datasets:

Given the history of ANR and of its Information System, numerous treatments were performed to cover and describe all the projects funded since 2005 in a common and standardized format. This procedure includes several steps aiming at cleaning, consolidating,

normalising, and enriching the data, especially the fields describing each project. The successive treatments have been documented to ensure a better traceability and transparency.

Exploratory analyses from ANR datasets and additional data sources:

In a second part, we propose, as a proof-of-principle, three different exploratory analyses describing in time-resolved manner, from 2005 onwards, ANR-funded projects at three different levels;

- Our first example focuses on the description of cross-institute partnerships supported by ANR
- Our second analysis was conducted on a specific funding instrument supporting young researchers comparing the number of funded and active projects by calendar
- Finally, focusing on the field of neurosciences, we conducted a text mining exercise using the keywords describing the projects. Specifically, we used Cortext Manager, and Gephi to infer and represent co-occurrence networks.

In all three examples, results obtained from the ANR database were confronted to similar metrics obtained from external sources, in particular the Web of Science (WoS). Specifically, we identified a main pool of publications by querying the WoS database for all publications in which ANR was acknowledged (over 100,000 publications between 2006 and 2018), from which we excluded some publications related to specific calls (e.g. Investissements d'Avenir), which are not included in the present ANR database.

Results

The dataset contains information on 17,058 projects funded by ANR from 2005 to 2017, and includes the following information for each project: a unique grant number, a short description of the project (title, abstract), references of the call for proposals and funding instrument, duration and budget, identification of the coordinators and partners (at organisation and individual levels), geographic location where research is primarily conducted.

Our first descriptive analysis focuses on projects led by INRA, and explores the number of collaborations with other Universities and research institutes. These mainly include other large French research institutes (e.g. CNRS, INSERM, and CEA). In order to evaluate if the collaborative patterns extracted from grant proposals were also detected in resulting publications, they were compared to those inferred from co-authored publications. To this end we restricted the main ANR pool of publications from the WoS only to publications including authors from INRA. While this analysis consistently identified CNRS as a primary co-authoring institute of INRA, discrepancies were observed for lower-ranked collaborators. These may, at least partially, be attributed to labelling and referencing heterogeneity.

In our second example we compare, for each year between 2005 and 2018, the number of funded projects using the 'Jeune Chercheur' instrument to the number of publications including the grant awardee as a co-author. We clearly identify a time gap between the funding date and the publication, which reflects the natural 'time-to-publication'. Considering the (increasing) number of still active projects in each year, that gap becomes less apparent. Nevertheless, we observe an increasing number of publications over the years that vastly

exceeds the number of funded projects in that funding scheme, which may reflect longer-term production of ANR-JCJC funding scheme, and/or increased productivity in the recent years.

Our third example compares topological features of the co-occurrence network of the keywords from the grant proposals in the neurosciences field to that inferred from the corresponding publications from the WoS. While these two networks show markedly consistent clustering of the keywords, we observed much more relationships between clusters in the network inferred from the projects information. This likely reflects that each paper separately only covers a part of the scientific aims of the project.

Conclusion & Perspectives

The « open science » movement aims to make scientific data accessible and easy to use, understood and shared for everyone, from professionals to society. The dataset presented here provides rich information on the research funded by ANR since 2005, and the analyses we provide illustrate their utility. Our examples also illustrate how such information could be completed by additional resources (e.g. European funding, Web of Science database, patents data, open archives, ORCID, altmetrics data) to yield interesting results describing research funding and production.

As such, dissemination of these data, combined with other research data, and extension, refinements of these analyses should contribute to decipher and enlighten the trends of research, the dynamics of scientific networks in France, in Europe and worldwide.

References

(1) ANR's commitment for open access:

<http://www.agence-nationale-recherche.fr/missions-et-organisation/open-access/>

(2) Text of the French law for a « République numérique » (7 october 2016)

<https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000033202746&dateTexte=&categorieLien=id>