AGENCE NATIONALE DE LA RECHERCHE

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; yves.fort@anr.fr

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

INTRODUCTION: As part of ANR's "open science" initiative, the Agency ensures the circulation of knowledge on public grants by providing data for each funded project in standardized formats to facilitate their broad and easy use. These data provide information on the research funded by ANR since 2005 across all scientific fields. For illustrative purposes we propose three proof-of-principle analyses illustrating how these data could be exploited to evaluate the impacts of ANR-funded projects on the national scientific production.

THE DATABASES:

Extraction of the projects data from the internal Information System

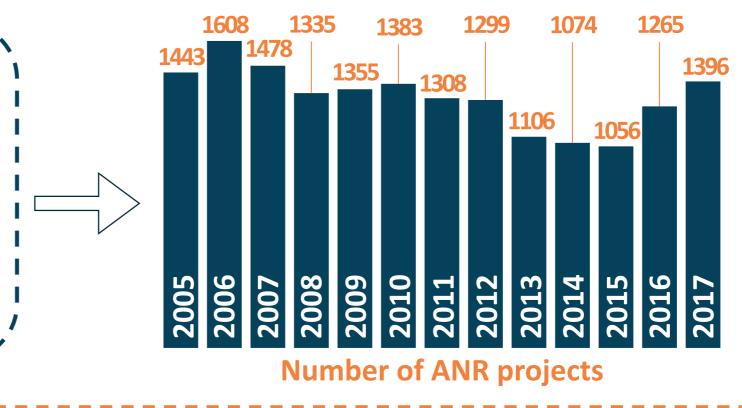
Between 2005 and 2017

> **EXTRACTION ON MAY 2018**

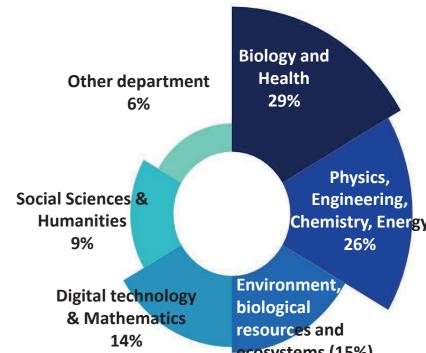
Clean Double-check Normalize Enrich

ANR-funded projects

Entries like grant number, project title, abstract, duration, i budget, coordinators, partners, geographic location etc. Number of funded projects: 17,106



ANR Scientific Departments



WEB OF SCIENCE

Web of Science is a scientific citation indexing service. It gives access to several interdisciplinary bibliographic databases

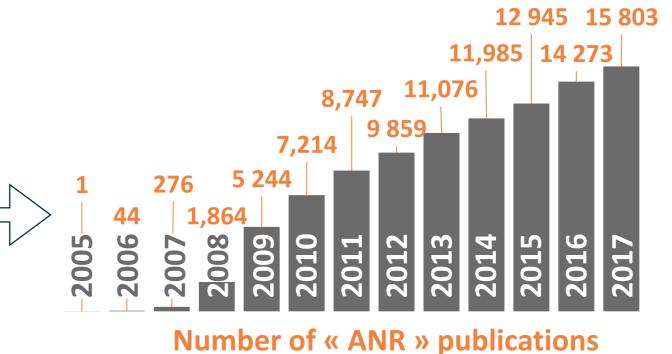
Between 2005 and 2017

EXTRACTION ON JUNE 2018

ANR

This corpus contains the publications that refer to the ANR in the acknowledgments. Number of documents:

99,330

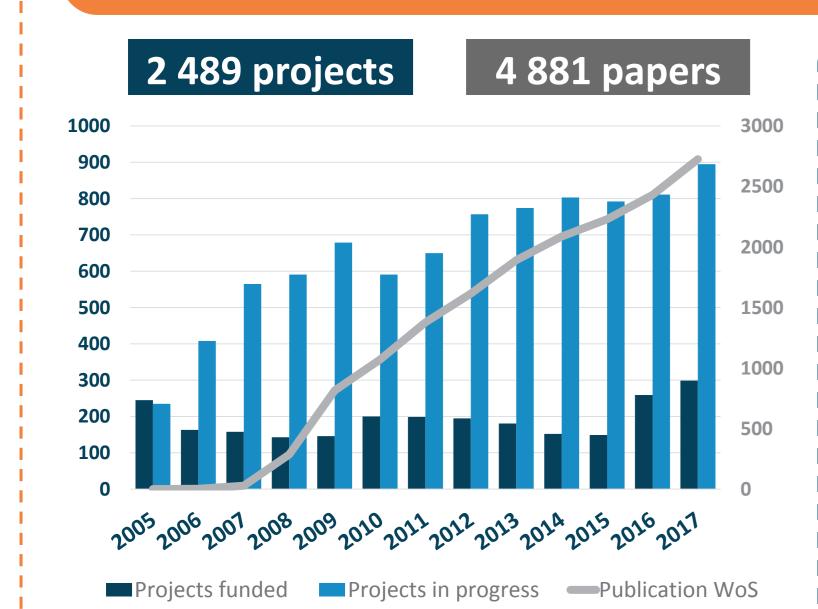


Disciplinary Areas Exact and Social Sciences &

METHODOLOGIES & TOOLS: Numerous treatments and harmonisation procedures were performed to ensure that all the ANR-funded projects (i.e. across all disciplines) since 2005 were described in a standardized and comparable way. Statistics obtained from the ANR dataset were confronted to similar metrics obtained from the Web of Science source (WoS). For this purpose, we identified a main pool of publications by querying the WoS database for all publications in which ANR was acknowledged. Three sub-corpus of publications (below) were subsequently extracted and focused respectively on: 1) the PI's funded through the JCJC instrument, 2) an institution recurrently funded by ANR (e.g. INRA), 3) the field of neurosciences. The figures were realized through "Excel", "Incites", and network analyses were conducted using "Cortext Manager" and "Gephi" open source tools. Despite some methodological limitations, we consider the WoS as an interesting source of data to conduct comparative analyses of the funded projects description and that of their output publications.

THREE EXAMPLES OF ANALYSIS:

1) FUNDING INSTRUMENT: JCJC



Question:

What is the dynamic of the scientific production of the young researchers funded I I through the JCJC instrument?

Primary results:

- Increasing number of publications over the years
- Apparent lag between project funding and publication

Perspective:

Consider a possible snowball effect of the I ANR funding leading to an increased productivity over time

For example, the case of Mr. Eric Cascales funded by ANR in 2010 through the JCJC instrument:



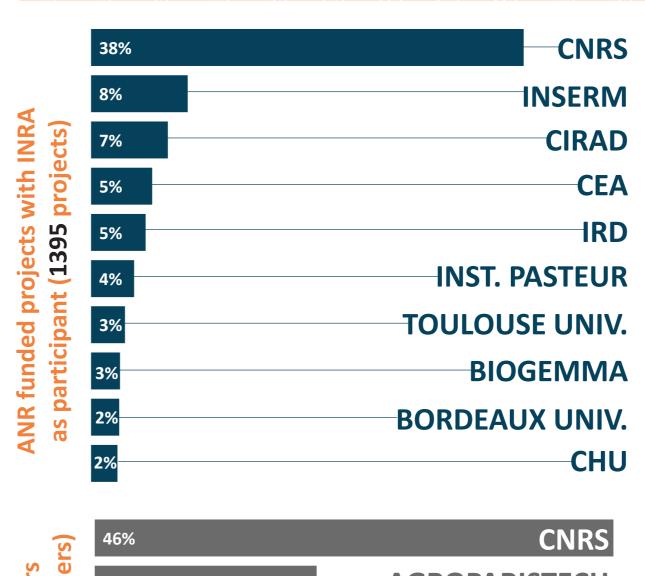
project (2014)

Participation in Coordinator of the B-WAR the TOOTHPAST **project (2015)**

Coordinator of the T6-PLATFORM **project (2017)**

(...)

2) CROSS-INSTITUTE PARTNERSHIPS: INRA



AGROPARISTECH TOULOUSE UNIV. INSERM CIRAD **TOULOUSE III UNIV.** -MONTPELLIER UNIV. INP TOULOUSE **SUPAGRO**

Question:

Are the cross-institute partnerships involved in ANR-funded projects also present in multiinstitutional publications?

Primary results:

 CNRS is the primary partner of INRA in the in ANR projects as well as in the ANR pool of publications including authors from INRA

Discrepancies are observed for lower-ranked collaborators

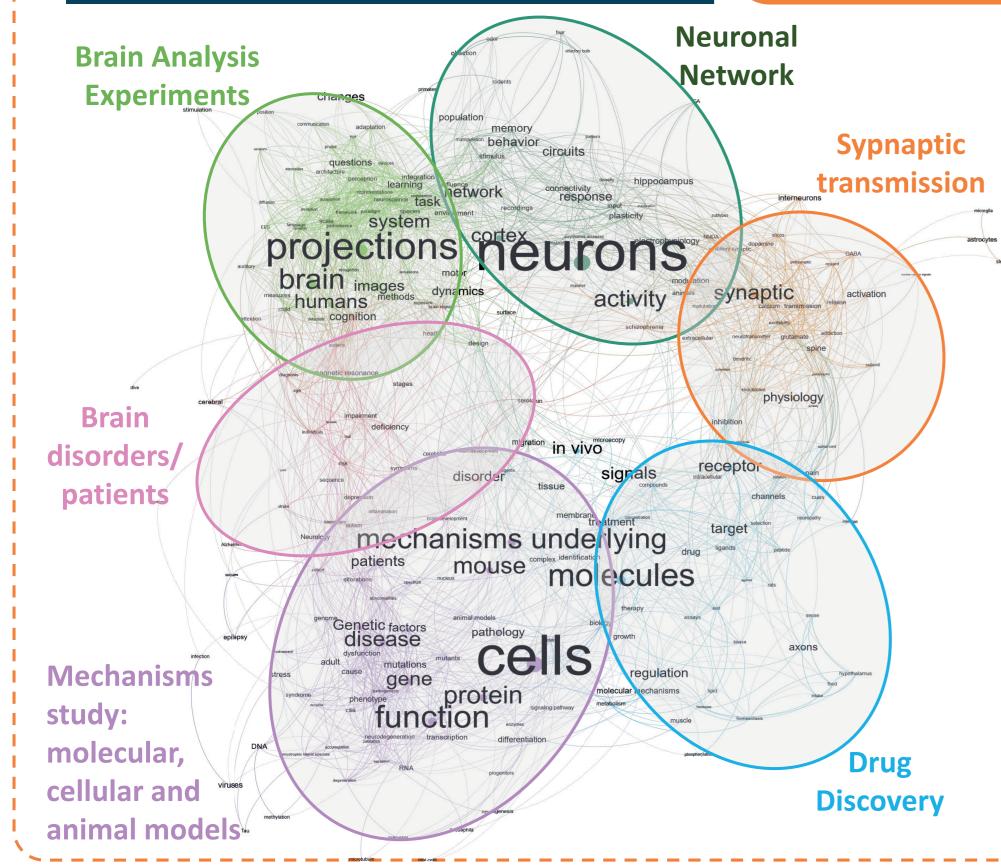
Perspectives:

Test if the initial cross-institute partnerships involved in the projects contribute to boost new national and international collaborations Improve the quality and comparability of the institutions' names in both sources of data

ANR abstract projects (nb: 1 101)

3) SPECIFIC FIELD: NEUROSCIENCES

WoS abstract papers (nb: 5 933)



Question:

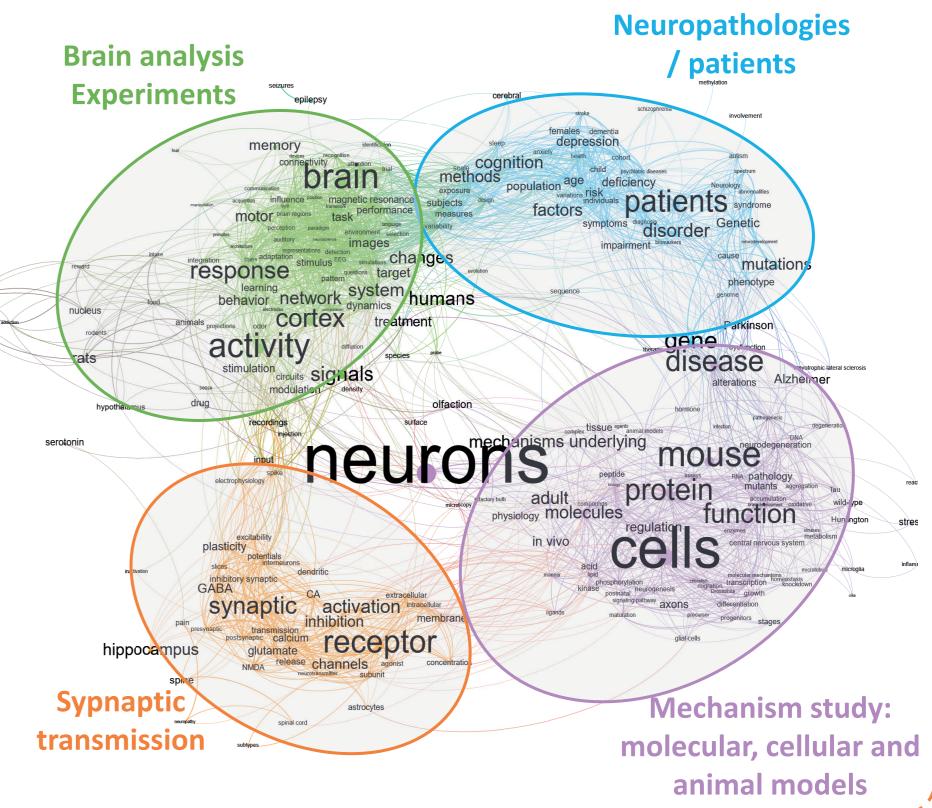
Are the output publications in line with the proposals aims and description?

Primary results:

- A consistent clustering of the "neurosciences" keywords in the two networks
- Much more relationships between the clusters in the network inferred from proposals information (left)
- Each paper separately covers a part of the scientific aims of the project (right)

Perspectives:

Follow the progression of scientific ideas and concepts from the beginning of the projects to the final production of new knowledge (towards longitudinal follow-up)



CONCLUSIONS AND PERSPECTIVES: The ANR dataset provides rich information on the research funded by ANR since 2005 and the three exploratory analyses show how such information could be completed by additional resources (bibliographic databases, open archives, European fundings, patents data, altmetrics data, ORCID...) to yield interesting results describing research funding and production. There are still methodological efforts to pursue, extension and refinements of these analyses to perform, and we hope that the wide dissemination of these data towards scientists, data analysts and educated-users will help identifying trends and evolution of research, dynamics in France, in Europe and worldwide.