Studying Scientific Impact of Research Infrastructures

A Methodology for Data Integration and Sharing

Renaud Fabre & Daniel Egret

Joachim Schöpfel

Daten. Politik. Zukunft. 9| KSWD

Konferenz für Sozial- und Wirtschaftsdaten

27.–28. März 2023 in Berlin

Authors and Contributions

Daniel EGRET and Renaud FABRE:

National Survey « **PUBLIMETRIQUES**: Apprécier l'impact scientifique des TGIR/OI »

http://perso.obspm.fr/daniel.egret/Publimetriques.pdf

Joachim SCHÖPFEL and Renaud FABRE

Data practices in French labs at large scale: https://www.emerald.com/insight/content/doi/10.1108/DTA-01-2017-0005/full/html ,

Digital technologies for science : one generation with internet https://journals.openedition.org/hrc/6448

D.EGRET, R.FABRE, J.SCHÖPFEL, O.AZEROUAL:

Scientific Impact of Research Infrastructures: https://direct.mit.edu/qss/article/2/1/42/97575/Evaluating-the-scientific-impact-of-research

Studying scientific impact of research infrastructures: Content

Research questions for a survey

Sample and approach

Results

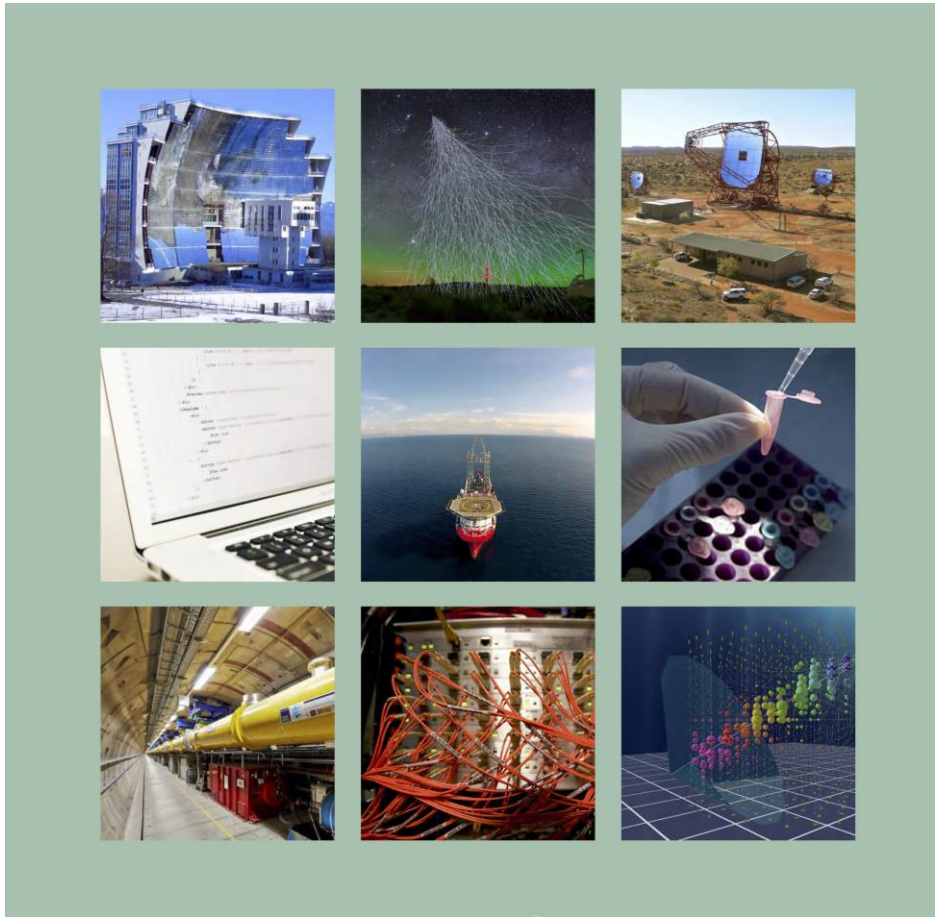
Discussion

Further works

Research questions for a survey

- How scientific results from main research *infrastructures* could be *designed* and *shared* for better *uses* and *benefits*?
- Which *categories* of recommandations could be produced in this way?
- Which *typology of recommendations* could be proposed in those categories?
- Which *methodology* could be followed to *implement* those recommendations?

The French *TGIR*



The French road map (99 infrastructures) includes, among many others the “very large” RIs (*TGIR*) as, the GENCI Company for high-performance computing, the CERN Large Hadron Collider, the CTA Cherenkov Telescope Array, the SOLEIL Synchrotron, the OpenEdition scientific publishing platform for the social sciences and humanities, and the Huma-Num digital humanities platform.

Sample – very large research infrastructures

- RIs are **“incredible engines of knowledge, attractors of talent**, catalysts for collaboration, bearers of scientific image and prestige […] not work tools like others,” because of their longevity, their ambitions, and their costs (largely **over €4.6 billion per year -** *Cour des Comptes*) .
- **24 very large Research Infrastructures** and international organizations. In 2018, the French Ministry of Higher Education, Research and Innovation commissioned a scientometric analysis for a shared assessment of their “scientific impact”.

TGIR: specificities in research management

- **Temporary accommodation:** The 24 *TGIR* host research teams from ***all institutional sources of scientific projects*** (universities and organizations with researchers from all countries, sometimes teams from ***private industrial research,*** etc.) and do not provide any permanent reception beyond a project, which ***generally spans a short period*** (less than or equal to 6 months).
- **"Self-service" experimentation of a project**: The experiment is carried out in ***"self-service"*** when the *TGIR* accepts the scientific project and validates the conditions for its realization, while defining the allocation of reception resources (technicians' time, adaptation of installations to experience, instrument time, computing time).
- **Technical assistance to experimenters**: The survey data, which cannot be developed here, show ***a very systematic adaptation of the TGIR to the needs*** for advice, expertise, and scientific support by the teams of permanent researchers of the *TGIR.*
- **Networking of means and results**: Technical assistance is frequently associated with ***networking***. In terms of resources, this is carried out by internationalization of similar resources. In addition, the current trend towards a ***more user-friendly evaluation system at various scales.***

Methodological approach

EXPECTED RESULTS, METHODOLOGY OF SURVEY:

- **Expected output is multiple**:
  - a better ***assessment of the scientific impact*** of each facility, at the level of disciplines and subdisciplines;
  - a better ***identification of research collaboration*** at the national, institutional, and individual levels;
  - the ***detection of emergent research topics***;
  - and a contribution to scientific foresight and advice, as ***part of the policy-making mechanism***.

Methodological approach

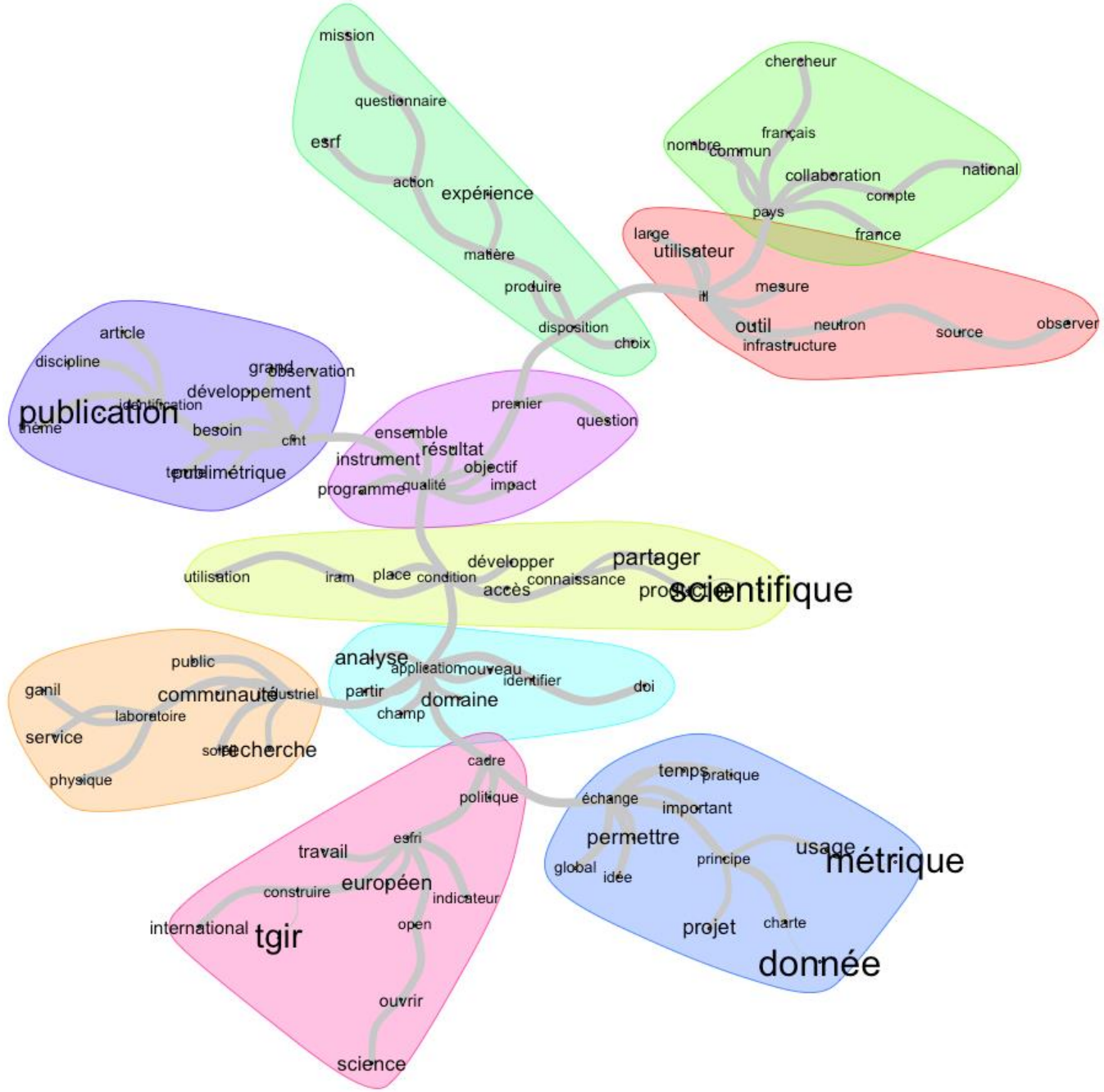
EXPECTED RESULTS, METHODOLOGY OF SURVEY:

- **Expected recommendations should be innovative:**
  - review of current ***scientometric practices***,
  - describes the *expectations and needs of infrastructure managers*;
  - proposals for the *development of shared impact measures,* and relevant indicators (“publimetrics”) to contribute to a conceptual and methodological framework for ***further harmonization and standardization of existing metrics***,
  - develop a common evaluation culture, in line with the ***requirements and standards of the European and international RI landscape***,
  - in particular the need for interoperability.

Methodological approach

EXPECTED RESULTS, METHODOLOGY OF SURVEY:

- **Survey methodology should be efficient and shared:**
  - simple and short ***questionnaire***,
  - interview with the ***head of the infrastructure***,
  - ***reviewing of answers by their authors*** to be cited in the report,
  - ***reviewing of the report by interviewed*** infrastructures.


mission
questionnaire
esrf
action
expérience
matière
produire
disposition
choix
chercheur
français
nombre
commun
collaboration
national
compte
pays
large
utilisateur
france
il
mesure
outil
neutron
source
observer
infrastructure
article
discipline
grand
observation
développement
publication
identification
thème
besoin
cnrt
publimétrique
premier
question
ensemble
résultat
instrument
objectif
programme
qualité
impact
utilisation
iram
place
condition
développer
partager
connaissance
accès
scientifique
public
ganil
communauté
industriel
laboratoire
service
recherche
physique
analyse
application
nouveau
identifier
doi
partir
domaine
champ
cadre
politique
travail
esrf
européen
construire
indicateur
international
tgir
open
ouvrir
science
temps
pratique
échange
important
permettre
usage
métrique
global
idée
principe
projet
charte
donnée

Results: three main recommendations

A. To organize the ***traceability*** of the RI results
B. To build a catalog of ***shared strategic indicators*** (publimetrics)
C. To create a ***network of these new metrics***

(1) Traceability

A: Organize the Traceability of Results

- Generalize the use of ***DOI*** and global traceability
- Harmonize the main ***performance indicators*** (domains, partnerships, equipment)
- Harmonize the ***terminology (classification)*** of research areas
- Develop new metrics for ***emerging research fields*** and monitor the genealogy of ideas
- Develop ***open science metrics*** for publications and data

NB 1: This first group of recommendations is particularly sensitive for TGIRs, which, unlike university or research institutions, lack visibility in the large bibliometric databases referencing scientific production.

NB2: The actors concerned by the recommendations are the persons in charge of the TGIR who must define, in an operational way, the contours of their scientific production: Indeed, this is most often not restricted to that of their teams but must also include that of their users, or even consider more broadly the production of knowledge that has directly benefited from the existence of the infrastructure.

NB3: The publishers of large databases are also concerned by these recommendations, who may seek to include specific metadata for instruments and infrastructures, and to develop the nomenclatures of research fields.

(2) Strategic indicators

B: Shared Strategic Indicators

- Drafting of a ***Guide with recommendations and best practices for the use of "publimetrics"*** (i.e., metrics of the scientific impact of the RI output [publications, data]). The framework of this Guide is drawn by the following list of recommendations:
  - ***List the rules for identifying publications*** and reaffirm the requirement of an explicit mention of the RI (affiliation)
  - ***Collect the shared scientific impact indicators*** based on the prior establishment of a Guide for publimetrics
  - Design an ***architecture of metrics practices*** by major purposes and build a typology of current metrics practices
  - Participate in a ***global modeling of the uses of publimetrics at a European scale***
  - Specify the ***organization and standards of publimetrics services***

NB: The second group of recommendations is based on the current practices and expectations regarding metrics of the RI activities (performance) and their scientific impact.

(3) Network of shared metrics

C: Create a Network of Shared Metrics

- ***Display the reference charters*** and support large RIs in their efforts to ***adhere to international declarations of good practice*** for the evaluation of scientific results
- Develop new metrics to support ***scientific foresight***
- Consolidate the ***scientific and professional deployment*** of publimetrics
- Initiate a **national/European metrics orientation approach**
- Set up a ***first experimentation process*** with a few large RIs

NB1: This third group of recommendations aims to promote a national network with the knowhow and skills for the implementation and monitoring of tracers and indicators of scientific impact. The actors directly concerned are therefore, in addition to those responsible for the TGIRs (and potentially for the RIs), the national research organizations (such as the CNRS, the CEA, and the IFREMER), the ministerial authorities, as well as the national evaluation and control bodies.

NB2: The publimetrics guide, mentioned in recommendation B, would be the means of bringing about the networking of metrics practices common to RIs, universities, research organizations, and other academic institutions.

Discussion: Data integration and metrics

Standard formats and metrics: *no appropriate tools and models: which ways for change with common goals and tools?*

- Way 1: CERIF data model in the EC
- Way 2: Research core data set (KDSF/RCD) plus CERIF/RCD
- ***Way 3: Compliance*** CERIF/RCD/Publimetrics

Further works

- Initiate a **National/European metrics orientation approach: a *first experimentation process*** with a few large RIs has been done (Strategy in http://perso.obspm.fr/daniel.egret/Publimetriques/)
- Develop a **Germany-France survey of Shared Indicators of RI**, their shared indicators and applications: first contacts began (J. Schöpfel) http://perso.obspm.fr/daniel.egret/Publimetriques/Publimetrics-FD.pdf
- Develop **Research and Foresight on Advanced Metrics and Tools** for efficiently digitalized research (ORKG, TIB Leibniz Information Centre) https://arxiv.org/abs/2206.01439 and https://www.mdpi.com/1999-5903/14/9/262

German partners (sample)

RI F	Name	Domain	Topic	Partner D	Other RI D
ECORD/IODP	Programme international de forage profond en mer	SSTE	Geology	Deutsche Forschungsgemeinschaft DFG	MARUM, Bremen ; AWI, Potsdam
IFREMER FOF	Flotte Océanographique française	SSTE	Oceanography	Helmholtz-Zentrum für Polar- und Meeresforschung, Alfred-Wegener-Institut (AWI), Bremerhaven	Deutsche Forschungsflotte
ICOS	Système Intégré d'Observation du Carbone	SSTE	Ecology	ICOS Germany	
SOLEIL	Source Optimisée de Lumière d'Energie Intermédiaire du-LURE	SMI	Synchrotron	DESY, Hamburg	
GANIL	Grand Accélérateur National d'Ions Lourds	PNHE	Particle accelerator	GSI Helmholtzzentrum für Schwerionenforschung, Darmstadt	
HUMANUM	Humanités numériques	SHS	Digital humanities	DARIAH-DE, Göttingen	

Conclusions:
Ahead with traceable web usage

- **A better understanding of scientific discovery and knowledge:** at the crossroads of information sciences and bibliometrics, research is advancing towards the construction of “global” traceable document paths that exists but are underexploited.
- **Information search behaviors have to be better traceable, safe, shared and modeled:** to optimize the semantics of the documentary choices made, stabilized by a solid “topic modeling,” based on a “topic analysis-based approach” built through innovative and exhaustive methods (Tsatsaronis, 2020).
- **A common requirement towards more traceability appears:** the paths allow discoverers and users of science to represent their path of hypotheses, discoveries, and ideas, in a more readable and traceable way.
- New logics are required for **discovery sharing approaches** (causal inference, transfer learning, multiverse analytics, meta-learning…) with both their outcomes to citizen science.

Thank you!

Contact: joachim.schopfel@univ-lille.fr