

**Abdullah Gok<sup>1</sup> and Jordi Molas-Gallart<sup>2</sup>.**

<sup>1</sup>Manchester Institute of Innovation Research (MIOIR), Manchester University, UK.

<sup>2</sup>INGENIO (CSIC-UPV), Spain.

***Science, Technology and Innovation Policy Evaluation: An Isolated Academic and Practice Field***

**Abstract:** Introduction

Evaluation has long been an important component of the policy process and is the concern of a broad area of academic enquiry. Academic researchers on evaluation are very close to practitioners and, together, have established a number of national and international associations, conferences and academic journals. As any other research community it revolves around a core set of generalist evaluation theory and methods, around which more specialised schools have developed around specific methodologies and approaches, or particular policy fields (education, health, social and development policy, etc.). Evaluation is also a topic typically covered in introductory textbooks on public policy, policy analysis, public administrations and the policy sciences. The result is an active community with different internal “schools” and a lively internal debate. An expression of this diversity is the annual conference organised by the American Evaluation Association (AEA) which draws thousands of practitioners and researchers from a wide variety of policy areas. This broad community encompasses a variety of schools, evaluation methods and different ways of understanding the role of evaluation in the policy process. From instrumental views of evaluation revolving around measurement-led, experimental research techniques, to outlooks like “Fourth Generation Evaluation” that sees evaluation as part of a process of building shared understandings of challenges and solutions among all policy stakeholders.

The evaluation of Science, Technology and Innovation Policies (STIP evaluation) appear to be one constituent component of this broader community. STIP evaluation started to become commonplace long after the core evaluation literature had developed. References to this literature can be found in some STIP evaluation reports and academics working in the field were active in the broader evaluation community. Currently “Research, Technology, and Development Evaluation” is one of the many “Topical Interest Groups” in the annual AEA evaluation conferences. STIP evaluation has its “own” academic journal (Research Evaluation), like many other policy fields (mention some examples). STIP evaluation seems to draw from and contribute to the overall policy evaluation community.

Yet, a more detailed analysis of the profile of STIP evaluation suggests a very different and surprising situation. Evidence indicates that STIP has developed on the margins of the other evaluation communities and has seldom drawn from the insights, approaches and practices they have developed. Instead STIP evaluation is characterised by its own dominant research approaches, technical developments on indicators, and a conviction that STIP evaluation is somehow essentially different from evaluation in other policy fields. STIP evaluation is not alone in having developed a somewhat distinct approach from the mainstream evaluation community and literature (e.g. development policy evaluation), but the extent to which STIP

appears isolated from mainstream evaluation is, as we will argue below, unparalleled. This paper presents evidence supporting this view through a bibliometric study. We also explore the reasons for the divide between STIP evaluation and generalist and other sectoral evaluation communities and literature. Finally, we present a roadmap to widen the scope for STIP evaluation to learn from and contribute to the core evaluation literature and practice.

### Data and Method

In this paper, we employ a mixed methods research design. On the quantitative side, we conduct a bibliometric analysis of the evaluation literature and of social media references to evaluation. We constructed a database consisting of over 20,000 publications related to policy evaluation. The database includes all articles published in 13 generalist and sector-specialist evaluation journals as well as evaluation related articles published in non-evaluation journals identified through a bespoke search strategy.

We conduct cluster analysis of the STIP evaluation literature, generalist evaluation literature and other sectorial evaluation literatures on the basis of their abstracts, their references, and other articles that cite them. We also use science overlay maps.

Social media references to evaluation are analysed using Twitter. Social media has increasingly been used to explore communities around professions and practices. Twitter data will help in studying the relationship between STIP evaluation practice and other sectorial evaluation practices. We have starting accumulating tweets related to policy evaluation in different sectors and we aim to analyse a total of 30,000 tweets. Twitter data will also be utilised in a topic modelling exercise to be able to locate the STIP evaluation practice within the evaluation practice community.

On the qualitative side, we conduct a systematic review of the concepts, frameworks and tools in STI evaluation and compare it with the trends in other policy areas and core evaluation literature.

### The Isolation of STIP Evaluation: Preliminary Results from the Bibliometric Exercise

The preliminary results from the analysis of the bibliometric data show that STI policy evaluation is very distinct from the core evaluation literature. Analyses on the basis of the abstract topic modelling, cited references networks and articles citing the dataset all corroborate this preliminary finding. Health, Social and Educational policy evaluation literatures are very close to the generalist core around, while development policy evaluation is relatively more distinct. STI evaluation, however, is very distinct from this network. The paper will present the results in detail; yet, some of the anecdotal evidence is already quite telling. For instance, there are only 21 citations from Research Evaluation to the American Journal of Evaluation, and most of these cited papers are STIP related anyway.

Twitter analysis shows a very similar picture. Health, Social, Educational Policy practices are very closely linked, while development policy is less related to this core and STIP is almost completely distinct from it.

### An interpretation of the evidence: roots and consequences

The preliminary bibliometric evidence presented above suggest some possible reasons for the separation of STIP evaluation. A substantial share of STIP evaluation revolves around indicators that are unique to this sector and, which have developed very actively over the past three decades. The importance of bibliometric indicators and associated research techniques has generated a specialised field with several associated journals (Scientometrics, JASIST, Technometrics, and others), conferences and associations (ENID, RICYT). STIP evaluation has become closely associated with these communities as shown by the data to be presented in our analysis of bibliometric evidence.

Additionally, it is common in STIP evaluation to refer to the uniqueness of the problems faced by this task. It is often emphasized that the effects of scientific research over the economy and society are often long term and unexpected; the result of complex interactions among different actors along protracted periods of time. In this context, the problem of attributing observed socio-economic “effects” to their original causes in scientific research is particularly daunting. In fact some of the initial seminal evaluations of the effect of basic and applied research developed specific techniques to identify such effects and measure their influence (projects TRACES and HINDSIGHT). Whether STIP evaluation faces exceptional problems is debatable. There are other areas of public policy where social-economic impacts are long-term and difficult to attribute. The belief in exceptionality may reinforce the isolationist tendencies prompted by the existence of a vast dataset and associated powerful research techniques.

We can identify two main problems generated by the isolation of STIP evaluation. First, there is a tendency in STIP evaluation to reinvent the wheel: issues emerge in the STIP evaluation literature as if they were novel when they have already been addressed in other fields of evaluation. For instance, “systemic evaluation” has long been implicit in general evaluation theory, but its introduction in STIP evaluation is relatively new and has generated substantial interests: a 2005 article in *Research Evaluation* has become the most cited paper in this journal.

Second, and perhaps more important STIP evaluation has gravitated towards a particular approach in policy evaluation: the use of measurements to provide instrumental answers within the context of summative evaluations. Other traditions in policy evaluation (Fourth Generation, participative evaluation, usefulness focused evaluation, etc.) have not found much audience in the STIP evaluation and practice, even when their close relatives in the broader policy environment (participative policy processes, etc) have been widely used in many STIP areas (like for instance energy and sustainability research). There is therefore a need to link STIP evaluation theory and practice both to the mainstream evaluation theories and practice.