
SUPPORTING INFORMATION

A VALUE CREATION MODEL FROM SCIENCE-SOCIETY INTERCONNECTIONS

Irene Ramos-Vielba

Danish Centre for Studies in Research and Research Policy
Department of Political Science
Aarhus University
Aarhus, Denmark
iravi@ps.au.dk

Nicolas Robinson-Garcia

Department of Information and Communication
University of Granada
Granada, Spain
elrobin@ugr.es

Richard Woolley

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

March 29, 2022

ABSTRACT

Codebook and data description of datasets used in the study 'A value creation model from science-society interconnections'. It includes the description of the two datasets used in the study. The empirical part of this study tests the proposed value creation model by using a dataset of 9,190 cases which includes the responses of a sample of researchers affiliated to Spanish institutions, their associated research output, and the mentions such output has received in several non-academic outlets (news media, policy briefs and Twitter). The first dataset, *dataset_archetypes.csv*, includes researchers' responses as well as publication data. The second dataset *all_titles.csv* includes the list of titles of the researchers' outputs in order to study the thematic relation with the archetypes identified. These two datasets are unconnected in order to ensure respondents anonymity as they could easily be identified.

Description dataset 1 - *dataset_archetypes.csv*

This dataset includes 9,190 observations related to researchers surveyed with 11 variables attached to each observation. Variables are retrieved from four different sources: the EXTRA survey, Web of Science, Altmetric.com and Unpaywall. Table 1 includes the names of the variables and the data source from which they were retrieved.

Data sources

EXTRA survey. Survey derived from the EXTRA project [1], conceded by the Spanish Ministry of Science, and approved by the Spanish Research Council. The survey took place between June and July 2016, receiving a 21% response rate (11,992 valid responses). Respondents work in all fields of science including engineering and physical sciences (STEM), biology and medicine (BIOMED) and Social Sciences and Humanities (SSH). The data were analysed anonymously. Respondents were asked about research-related activities conducted in the 2013-2015 period.

Web of Science. Bibliometric database. Respondents from the EXTRA survey were matched with their publications using the CvE author name disambiguation algorithm [2]. Only publications associated to a respondent published within the 2013-2015 period and labeled as articles or reviews were included.

Table 1: Variables and data source

| Variable | Data source |
|---------------|----------------|
| Production | EXTRA survey |
| Engagement | EXTRA survey |
| Translation | EXTRA survey |
| Dissemination | EXTRA survey |
| Transmission | EXTRA survey |
| Discipline | EXTRA survey |
| OA | Unpaywall |
| PUBS | Web of Science |
| p_twitter | Altmetric.com |
| p_news | Altmetric.com |
| p_policy | Altmetric.com |

Unpaywall. Database including information regarding the open access status of an academic publication [3, 4]. This database is queried using the DOIs identified from the publications extracted previously in Web of Science.

Altmetric.com. Database including social media mentions and mentions from non-academic sources to scholarly literature. Similarly to Unpaywall it is queried using article identifiers. In our case, also DOIs were used to retrieve altmetric data.

Operational definition of variables

Bibliometric variables. These are those which are not derived from the survey. This data is extracted using the sql script *extra_archetypes.sql*. The variables produced refer to the number of publications in all cases, that is, total number of publications (PUBS), publications in Open Access (OA) or publications with at least a mention in a given altmetric source (p_twitter; p_news; p_policy).

Survey variables. We include 6 variables from the survey. These are the following and respond to the following questions:

- Joint research. Defined as the variety of stakeholders with whom the respondents have conducted joint projects. This variable ranges from 0 to 7 types of partners. These are: SME (small and medium-sized enterprises), large firms, government agencies, non-profit organizations, hospitals, associations, and international organizations.
- Engagement. Number of different types of stakeholders with whom formal interactions took place, up to 7 types.
- Commercialisation. Number of types of commercialization activities (patent licensing, spin-offs) in which researchers have participated
- Transmission. Frequency of promotion of research use (presentations in non-technical language, demonstrations or discussions with final users) in which researchers have participated.
- Dissemination. Frequency of use of analogue and digital communication tools to spread research findings among potential research users.
- Field. Selection of research field to which they belong. These are later aggregated to three large areas: STEM fields, SSH fields and Biomedical fields.

Description dataset 2 - *all_titles.txt*

This dataset includes 71,163 observations. Each observation includes the title of a publication (co)authored by respondents from the survey, indexed in Web of Science and published between 2013 and 2015. Linkages to each individual are not included in order to ensure anonymity. This dataset as well as subsets of it are used to create the co-occurrence maps in figures 4 to 7.

References

- [1] Irene Ramos-Vielba, Elena Castro-Martínez, and Pablo D’Este. Interactions with non-academic actors. Survey of researchers in the Spanish public research system, March 2018.

- [2] Emiel Caron and Nees Jan van Eck. Large scale author name disambiguation using rule-based scoring and clustering. In *Proceedings of the 19th international conference on science and technology indicators*, pages 79–86. CWTS-Leiden University, Leiden, 2014.
- [3] Dalmeet Singh Chawla. Unpaywall finds free versions of paywalled papers. *Nature News*, 2017.
- [4] Heather Piwowar, Jason Priem, Vincent Larivière, Juan Pablo Alperin, Lisa Matthias, Bree Norlander, Ashley Farley, Jevin West, and Stefanie Haustein. The state of oa: a large-scale analysis of the prevalence and impact of open access articles. *PeerJ*, 6:e4375, 2018.