



HAL
open science

Using ADS for Creating Bibliographies of Research Institutions

Daniel Egret, Amélia Laurenceau, Alberto Accomazzi

► **To cite this version:**

Daniel Egret, Amélia Laurenceau, Alberto Accomazzi. Using ADS for Creating Bibliographies of Research Institutions. *Library and Information Services in Astronomy VII: Open Science: At the Frontiers of Librarianship*, 492, pp.85-89, 2015. halshs-03473188

HAL Id: halshs-03473188

<https://shs.hal.science/halshs-03473188>

Submitted on 3 Jan 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Using ADS for Creating Bibliographies of Research Institutions

Daniel Egret,^{1,2} Amélia Laurenceau,³ and Alberto Accomazzi⁴

¹*LUTH, Observatoire de Paris, PSL Research Univ., 92190 Meudon, France*

²*UMR 8102 CNRS, Univ. Paris Diderot, 92190 Meudon, France*

³*Observatoire de Paris, UMS 2201, CNRS/INSU, PSL Research University; 61 avenue de l'Observatoire, 75014 Paris, France*

⁴*Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge MA 02138, USA*

Abstract. Institutions often have a need for maintaining “institutional bibliographies,” a list of all papers published by individuals affiliated with the institution. Using the Observatoire de Paris (France) as an example, we show how this can be done, using the services of the ADS and other bibliographic databases. In this specific case, we are able to quantify the success rate in creating the extensive list of the scientific output of the institution by each source database. Our analysis helps identify the current limitations and ambiguities in the contents and format of the affiliation/address field, as produced by the authors and modified by journal editors and database managers. We advocate for the development of an authoritative list of astronomical institutions that can be used by the ADS and publishers as a controlled field in order to significantly improve the efficiency of affiliation queries.

1. Introduction

Institutions often have a need for maintaining “institutional bibliographies.” For large institutions, such a list cannot be easily created by simply collecting individual lists provided by the authors themselves: full coverage of the scientific production of the institution requires programmatic, targeted searches using bibliographic databases such as the ADS (Kurtz et al. 2000). The biggest challenge of this approach is that the affiliation field (as it appears in the original article) is often insufficiently controlled and there are multiple variations of the same institution. The increased importance of university rankings and the recent acceptance of bibliometrics have increased pressure on authors and editors to address this issue.

First, it is important to establish the following two definitions.

Institution: A research group, university department, institute, or other official association identified in the signature of the article in the affiliation or address field.

Institutional bibliography: A list of all of the papers published by author affiliated to the institution during a given period, typically for a calendar year or a range of years.

It is important here to differentiate the selection criteria of institutional bibliographies with “observatory” and “telescope” bibliographies (Rots and Winkelman 2015,

this conference): the latter are based on usage of telescope data, rather than author affiliation. It is also important to note that, for a paper to qualify for inclusion in an institutional bibliography, the author must mention his/her affiliation in the signature of the paper. In the case of the Observatoire de Paris, an article is considered for inclusion in the institutional bibliography if and only if one of its authors provides an explicit affiliation with either the Observatoire, or one of the scientific departments or institutes of the Observatoire (for example, *Institut de Mécanique Céleste*, or *Station de Radioastronomie de Nançay*).

There are multiple objectives for building institutional repositories. These can include demonstrating the scientific output of an institution, as a part of an evaluation procedure; understanding the strengths and weaknesses of the scientific output of the institution; analyzing the collaboration network of the authors affiliated with an institution; as data to support preparation of an open access strategy; or publicizing and facilitating access to the scientific output of the institution by colleagues and/or the general public.

2. Using bibliographic databases to build institutional bibliographies

In order to create an institutional bibliography for a large institution, it is not enough to simply collect individual lists of publications provided by authors affiliated with the institution. It is therefore necessary to utilize specific bibliographic databases such as the ADS, or more general sources such as Web of Science, Scopus, or Google Scholar. It is important to note that institutional bibliographies do not aim to capture one hundred percent of all articles written by affiliated authors; a list that captures 95 percent of all papers may be considered acceptable.

The biggest challenge to using the affiliation field as it appears in the original article is that this field is often uncontrolled and the same institution can be represented in various forms. Further, institutions that have a complex structure — such as joint research units — can be even more challenging. Authors may move between institutions, or may be temporarily affiliated with a different institution, again making it more difficult to rely on the affiliation in the signature field.

There are proposed solutions developed to disambiguate author names such as ORCID (Holmquist et al. 2015, this conference). It appears to be necessary for bibliographic databases to implement similar solutions for identifying institutions (and hierarchies of institutions). In Web of Science (developed by Thomson Reuters), there is the Organization-Enhanced (OG) field; in Scopus (developed by Elsevier), there is the Affiliation identifier (AF-ID). Both of these, however, are recent features in their respective databases and are still under regular development. This paper will show that ADS is an efficient and appropriate source for institutional lists in the context of astronomy and astrophysics; it also has developments related to uniquely identifying institutions, as described by Grant et al. (2015, this conference).

3. The scientific output of *Observatoire de Paris*

Observatoire de Paris is the largest astronomical institution in France, representing roughly 30% of the national research contribution in astronomy and astrophysics. It is also one of the largest astronomical institutions in the world. Its research staff con-

sists of about 250 researchers and permanent academic staff, approximately 130 PhD students and postdoctoral researchers, and several dozens of research associates and visiting researchers.

Because of its size and thematic diversity, the Observatoire is a good sample case for testing queries by affiliation in various bibliographic databases. The institute's total output amounts to approximately 750 refereed papers per year, mainly in the fields of astrophysics, planetary and space sciences, and physics. ADS also lists some 350 additional non-refereed publications that are not considered further in the discussion below.

3.1. Query results from the ADS

We constructed our query for the ADS as follows. An article is selected for inclusion in the institutional bibliography if and only if one of the paper's authors has officially listed the Observatoire de Paris (or one of its related scientific departments) as his/her affiliation. In ADS 2.0, we have used the `aff:` search syntax (see Grant et al. 2015). We have additionally used the filter for "refereed articles" available from the ADS. In order to assess the completeness of the results generated by this query strategy, we have compared these results with the results of searches for individual authors, as derived from a list of the academic and research staff of the science units at the Observatoire. Both of these result lists were eventually merged.

As illustrated in Table 1, the *Affiliation search* mode of ADS proves to be efficient in the context of Observatoire de Paris. The main restriction is that the address line is sometimes missing from the metadata for certain journals (or certain issues of a journal). It should be noted that, in most of the cases, the institution name is completely and correctly mentioned by the author, and is therefore present in the affiliation/address field. This may not be the case for other institutions where the institution's name itself is not always used as part of the address field.

Table 1. Observatoire de Paris: Query results from the ADS for year 2013

Query	Number of articles	Percentage of final list
<code>aff:"Observatoire de Paris"</code>	636 articles	83%
Other variations : +46	682 articles	89%
Completing with query by authors: +73 (mainly : affiliation field missing in the metadata)	755 articles	98%
Other refereed articles outside ADS	15 articles	2%
Grand total	770 refereed papers	100%

By querying other bibliographic services, such as Web of Science and Scopus, we were able to find some additional refereed articles (in 2013: 14 articles, i.e. 2%) where are outside of the ADS domains: these are, for example, papers published in mathematics or engineering journals. The number of additional papers found outside the ADS is small enough that the benefits of using the functionality provided by the ADS platform — query syntax and filters — largely outweigh the relative limitations of its coverage.

3.2. Comments about other data sources

Web of Science (Thomson Reuters)

The Web of Science (WoS) field *Organization-enhanced* (OG) was, at the time of our query in early 2014, poorly populated for the entry OG=Observ Paris, and so a search for it only resulted in 276 articles, which was 36% of the final list. Scientific departments at the Observatoire de Paris are also generally joint research units with Centre national de la recherche scientifique (CNRS) and one of the Paris universities, making it quite difficult for bibliographers to identify which are the parent organization. This is why we had much better success by formulating a text query in the Address field. By querying the Address field with different variations¹ we find 720 entries, 686 of which are classified as refereed articles in the ADS. It should also be noted that some journals were not covered at the time of our query, including, *Journal of Geophysical Research*, *Lecture Notes in Physics*, *Experimental Astronomy* and the Web of Science also includes non-refereed conference proceedings and journals. In our query, this constituted 34 papers that were not considered refereed papers in the ADS.

Scopus (Elsevier)

The metadata in the Scopus database only includes one parent organization for any given affiliation. For the same reason as with Web of Science (science departments at the Observatoire are joint research units), a query by organization yielded only 388 articles, which was half of the final list. Here as before, by querying the address field with different variations,² we found 756 articles, 615 of which are classified as refereed papers in the ADS. Scopus's coverage includes 96 conference proceedings or book chapters that are classified as non-refereed in the ADS.

3.3. Managing the institutional bibliographies

We use ADS features in the following ways to manage the final list of publications. Personal libraries are useful for merging and comparing partial lists of articles. They are also useful for sharing data with colleagues. Personal libraries in ADS benefit from corrections and updates. ADS export tools, especially in a custom format, are helpful for producing different versions of the final list, for print or spreadsheets. It is also easy to perform comparisons with other databases using the Digital Object Identifier (DOI), which is common to all bibliographic databases.

4. Conclusion

Finally, here are some questions that institutional bibliographies help to answer. How does our institution compare with similar research centers? In which journals are the articles of our institution generally published? What proportion of articles are also submitted to arXiv? What proportion of articles have a first author from our institution?

¹AD= ("Observ Paris" OR "Observatoire Paris" OR "Obs Paris" OR "Paris Observ" OR "Observat Paris" OR "IMCCE") OR AD=(Observ SAME Meudon)

²AFFIL("Observatoire de Paris") OR AFFIL("Observatoire de Meudon") OR AFFIL(IMCCE) OR AFFIL("Obs. de Paris") OR AFFIL("Paris Observatory")

What are the trends in the number of authors per paper? For an analysis of French astronomical observatory publications, see Egret (2013).

We have shown that a simple affiliation query in the ADS provides a large but incomplete part of the institutional bibliography — about 83% in our sample case of Observatoire de Paris for the year 2013. The situation is similar, or even worse, with other major bibliographic services. This is a result of the current limitations and ambiguities metadata: the format of the affiliation/address field, as produced by the authors, and eventually modified by the journal Editors and the database managers, varies for the same institution. We advocate for the development of an authoritative list of astronomical institutions that will be used by the ADS and the publishers as a controlled field (see details in Grant et al. 2015, *Enabling meaningful affiliation searches in the ADS*, this conference).

Acknowledgments. This research has made use of SAO/NASA's *Astrophysics Data System*, funded by NASA grant NNX12AG54G.

References

- Egret, D. 2013, in SF2A-2013: Proceedings of the Annual meeting of the French Society of Astronomy and Astrophysics, ed. L., Cambresy, F., Martins, E., Nuss and A. Palacios, (Paris: SF2A), 9. <http://cdsads.u-strasbg.fr/abs/2013sf2a.conf...9E>
- Grant, C.S., Thompson, D.M., Chyla, R., et al. 2015, in ASP Conf. Ser., Vol. 492, LISA VII. San Francisco: ASP), 208 (these proceedings)
- Holmquist, J., Damon, J. and Erdmann, C. 2015, in ASP Conf. Ser., Vol. 492, LISA VII. (San Francisco: ASP), 323 (these proceedings)
- Kurtz, M.J., Eichhorn, G., Accomazzi, A., et al. 2000, A&AS, 143, 41 <http://dx.doi.org/10.1051/aas:2000170>
- Rots, A., Winkelman, S. 2015, in ASP Conf. Ser., Vol. 492, LISA VII. (San Francisco: ASP), 71 (these proceedings)