

Use Cases of the ESO Telescope Bibliography

Uta Grothkopf^{1,*}, Silvia Meakins¹ and Dominic Bordelon¹

¹European Southern Observatory, Karl-Schwarzschild-Str. 2, Garching near Munich, Germany, library@eso.org

Abstract. ESO's mission is to enable front-line research in astrophysics by operating and maintaining a wide range of world-class telescopes and state-of-the-art instruments. Various key performance indicators are used to understand whether ESO is achieving these ambitious goals. In order to assist the ESO Management, the Library builds and maintains the Telescope Bibliography (telbib). telbib provides insights into the performance of ESO's facilities and helps to understand publishing trends among the user community. This paper will highlight some use cases and recently added enhancements.

1 Introduction

In this paper, we will show various use cases of the ESO telescope bibliography (telbib)¹ and will highlight some recent additions. The ESO librarians spend a lot of time, energy, and ideas on the development of telbib, our database of refereed papers that use ESO data. By now, telbib has developed into a sophisticated tool that does not only list all the papers, but links them to the observations stored in the ESO Science Archive and provides the opportunity to add additional tags and identifiers. A multitude of reports and statistics can be derived from telbib, allowing stakeholders to answer a large variety of questions about how observational data are used in papers. The ESO Library continuously enhances telbib in order to enable users to obtain further metrics.

The overview of this paper is as follows: In section 2 we will explain why it is important that telescope bibliographies are curated. In section 3 selected use cases will be presented. We conclude with a description of some questions that cannot be answered at this time.

2 Curated bibliographies

Why do we need curated telescope bibliographies? In order to understand the importance, we take a look at the ESO telbib curation process. Our workflow starts with running a text mining software which finds scientific articles that contain at least one ESO-related keyword. Very importantly, each and every paper is visually inspected by the librarians. This inspection is the first evidence of human curation. It is essential in order to make sure that all telbib papers fulfill our main selection criterion,

*e-mail: uta.grothkopf@eso.org ORCID: 0000-0001-6830-0702

¹<http://telbib.eso.org>

i.e., they must use (study, analyze) ESO data. If papers only quote results from the literature, or describe future observations, or even if they discuss instrumentation, they are not part of telbib. Further information about telbib can be found on the web² and in various publications.³

Once the decision is taken that a paper should be included, a record is created in telbib. Bibliographic details obtained from the SAO/NASA Astrophysics Data System (ADS) are the basis of all records. In a second step, the Librarians assign a multitude of tags, labels, and annotations to characterize each entry, for instance regarding author affiliations, observational surveys from which data originated, or ESO press releases that featured specific papers. However, in particular, we add metadata regarding the observing programs, including the program ID(s), observing mode and type, and the instrument(s) used for the observations. Very importantly, we also track whether the data were originally observed by the authors or whether they were obtained from the ESO Science Archive.

As a result, reports and statistics can be obtained for this multitude of parameters. telbib is the only place where scientific papers and ESO's observational data are linked.

Observational details in telbib are identified through the program IDs. However, only approximately 70% of the authors mention these identifiers in their papers. For the remaining 30%, the IDs need to be retrieved through other ways, for instance by querying the Archive for observed objects, instruments, and observing dates, or through communication with the authors. Without this human curation, the ESO data used in 3 out of 10 papers could not be identified, and the papers would not be linked to the observations.

3 telbib use cases

Which questions can we answer using telbib? What are the main use cases? The statistics we provide can be divided in four categories: (1) Productivity and impact, (2) Author demographics, (3) Publishing trends, and (4) Instrumentation management. In the following sections, we will show examples for each category.

3.1 Productivity and impact

Productivity and impact are the most common performance measures that are derived from telescope bibliographies, and very likely everybody who maintains bibliographies provides such statistics.

Productivity is often measured by looking at the number of papers that were published, and impact by the number of citations. This can be the total number of papers or citations, or it can be a group, for instance the fraction per instrument, telescope, observing program, observing type, etc. Since productivity and impact are the most obvious and frequently applied metrics, we will only mention them briefly here.

One example is to study the number of papers per journal (Fig. 1). In the case of the ESO telescope bibliography, more than 60% of the papers were published in the journal *A&A* in the early years of VLT (Very Large Telescope) data papers. Over time, this fraction has gone down to (still remarkable) 40%, while more papers were published in particular in *MNRAS*. One explanation for this shift might be that the UK joined ESO in 2002, and therefore publications in a UK-based journal have increased. It is always important to put information retrieved from bibliographies in context with what is happening in the organization at large.

In a similar way, citations can be investigated for specific entities, for instance for ESO program types. ESO observations are assigned to various program types, including Large Programs (LP),

²http://www.eso.org/sci/libraries/telbib_info.html

³http://www.eso.org/sci/libraries/useful_links/publications.html

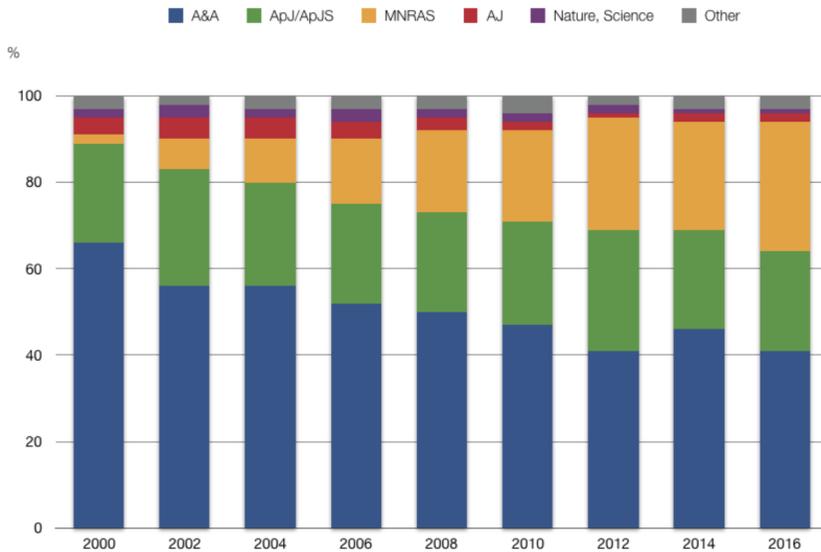


Figure 1. telbib papers by journal 2000-2016

Guaranteed Time Observations (GTO), Target of Opportunity observations (ToO), and others. Sterzik et al. (2015) [1] investigated the role of different observing types and modes in detail, using telbib data for an in-depth analysis of productivity and impact of VLT observations.

3.2 Author demographics

The second area for which statistics are compiled from telbib focuses on author demographics. Central questions are: How many distinct authors publish with ESO data, and what are their nationalities? In this context, we define nationalities as the country of the author’s affiliation.

We studied the number of distinct authors of telbib papers since 1999. A steady increase can be seen from approximately 1,300 authors in 1999 to almost 17,000 in 2016, so within approximately 15 years the number has increased by a factor of 13. These numbers are cumulative.

What are the nationalities of these approximately 17,000 authors? We have created a world map of 1st authors of telbib papers (Fig. 2) and found that they have affiliations from 64 countries (marked in dark blue). This constitutes approx. a third of all countries on this planet, so we can say that the ESO user community is truly global.

3.3 Publishing trends

The ESO Management is often very interested in learning about publishing trends. Lately, the Library investigated three topics: (1) Can we see any trends in the team size of authors, and has this changed over the years? (2) What is the fraction of papers that use archival data? and (3) Are there any synergies of ESO observations with data from other observatories?

Looking at the number of authors that publish with ESO data, we found that the first paper with over 100 authors was added to telbib in 2005. In 2011, where a slightly steeper increase can be found



Figure 2. World map of 1st author nationalities of VLT data papers

in comparison with previous years, four papers with more than 100 authors were added to telbib. Two years later, in 2016, three papers had even more than 600 authors, explaining the noticeable upward swing from 2015 to 2016. For author team sizes, measured by the average number of authors per telbib paper from 2000 until 2016, a duplication from 6.5 to almost 14 co-authors can be seen. The number for publishing year 2016 might be slightly inflated due to the three papers with more than 600 authors mentioned above. However, the general trend towards multi-author papers remains. In that context, collaborative writing along with supporting platforms like Authorea, Overleaf, and Writing Studio is gaining importance.

If the average number of authors increases, it is interesting to see whether the number of ESO programs used per paper has also changed. Indeed, for telbib papers, we found an increase by a factor of 2: in the early years of this millennium, authors used on average 1.5 programs in their papers. By 2016, this number has increased to over three programs that provide data for a given paper. A general trend can be seen towards papers with more authors that deploy observations from more programs.

An important aspect for ESO is how archival data are used. Observational data obtained with ESO's facilities are stored in the Science Archive from where data are made available to astronomers worldwide on request. Therefore, the Archive is often seen as an observing facility of its own. It is essential for ESO to understand how many papers are published that use archival data. Statistics derived from telbib show that in particular during the past five years, the fraction of papers that use archival VLT data (exclusively or along with ESO data obtained by the authors) has been on average 23%, or almost a quarter of all VLT papers. Archival papers are divided almost equally between those that use exclusively archival data (12%) and those that deploy archival data in addition to proprietary data (11%).

Are ESO data mostly used alone in papers or in conjunction with data from other observatories? We investigated VLT data papers published between 2012 and 2016. Using information provided by other observatories and available through the ADS (see e.g., Accomazzi et al., in press [2],) we found that 24% of the VLT papers are also included in the HST bibliography. Also Gemini (8%), Keck (8%), or Subaru (3%) data are used in papers along with VLT data. Interestingly, on average 1% of

the papers use VLT together with ALMA data. This trend seems to be growing as in 2016, the overlap increased to 4%. In total, 36% of the VLT papers published in the past five years also deployed data from one or several of the other five observatories, while for 64% of the VLT data papers no overlap was found. These papers still may have used data from observatories other than HST, Keck, Gemini, Subaru, and ALMA (Fig. 3).

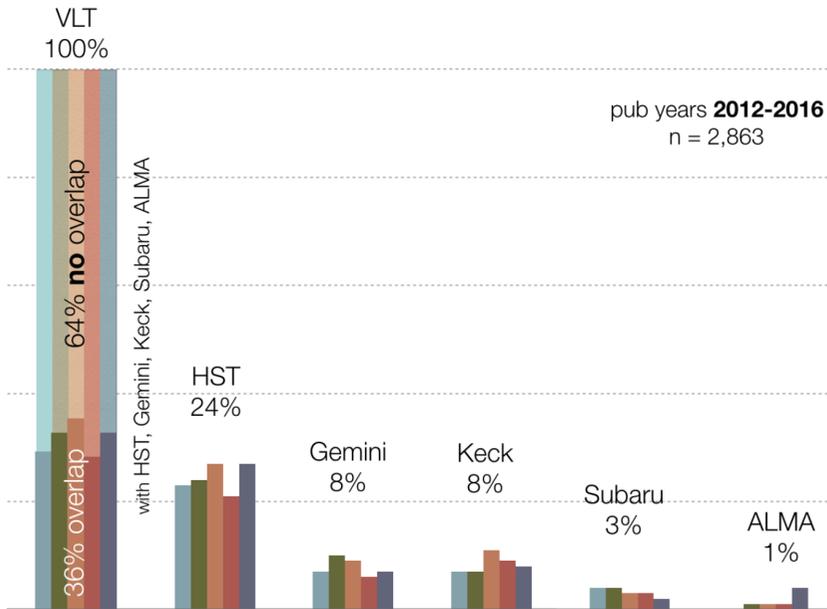


Figure 3. VLT data papers using also data from HST, Keck, Gemini, Subaru and/or ALMA

3.4 Instrumentation management

Instrumentation management is of key importance to ESO. The organization uses a variety of tools and metrics to make decisions about whether instruments need to be upgraded, refurbished, or even retired and replaced. Such tools include for instance the general knowledge of the community’s needs or the oversubscription rate, i.e., the demand for observing time. This information is coming from other sources, not telbib.

telbib information in this context focuses mostly on the (total or average) number of papers and citations. In addition, lately the Library adopted a new metric originally presented by S. Kulkarni [3]. Using this methodology, citations to data papers are grouped by the year of the citing papers. The resulting graphs are called citation flux curves. Kulkarni concludes that instruments with decreasing overall citation counts (citation flux curves) should be upgraded or replaced.

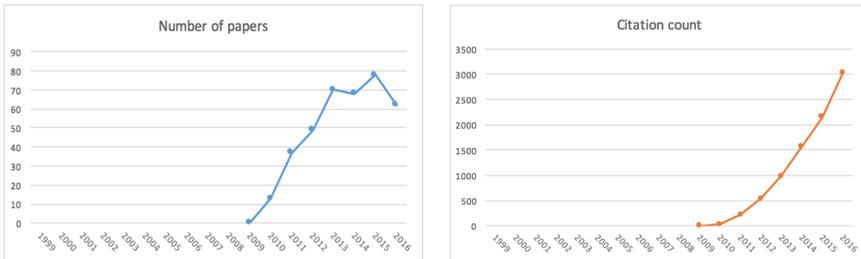
To assist with instrumentation management, the Library calculated citation flux curves for all VLT instruments. Here we present two examples (Fig. 4).

Example 1 studies the X-SHOOTER instrument, a fairly recent instrument that has seen a steep rise of papers as well as citations from the start. However, the paper count now seems to have “plateaued” at approx. 70 papers per year on average. As a consequence, some upgrades are now foreseen to prevent the instrument from aging.

The 2nd example is ISAAC, one of the oldest VLT instruments, which has reached a publication plateau several years ago. Even the citation flux curve has been flat during at least the past five years. Indeed, ESO has decommissioned ISAAC in 2013.

In this way, information from telbib can be used to confirm or even initiate instrumentation management decisions.

▶ Example 1: **X-SHOOTER**



▶ Example 2: **ISAAC**

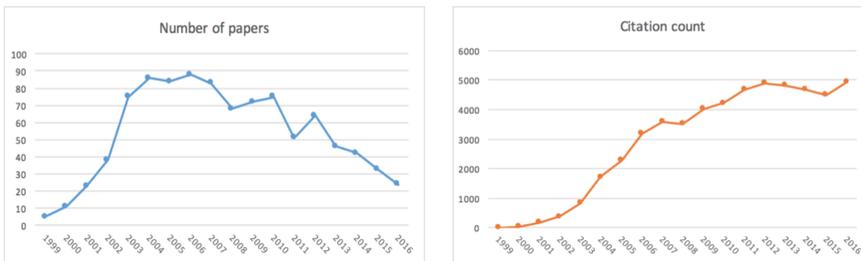


Figure 4. Paper counts and citation flux curves of two VLT instruments

4 Questions we cannot answer (yet)

While there are various strong use cases for the ESO telescope bibliography, there are also some questions that cannot currently be answered. For instance, telbib is not well suited to analyze the subject areas dealt with in papers, or general trends in topics. During previous telbib studies, we experimented with obtaining insights from the keywords authors assigned to their papers, combined with title and/or abstract words, and from the ESO observational categories⁴ to which the programs were assigned. However, the results were less meaningful than we had hoped for.

Gender studies are receiving a lot of attention these days, but since we do not record (or even know) the gender of telbib authors, we cannot run any statistics pertaining to this topic.

When papers are ingested into telbib, no “weights” are assigned to reflect the percentage of ESO data in comparison with data from other observatories in the same paper. Regardless of whether only

⁴<http://www.eso.org/sci/observing/phase1/p100/opc-categories.html>

one spectrum was used or a large data set, all papers are treated (and weighted) equally as far as statistics are concerned.

One considerable complication is that by now telbib contains so much and so varied information that results can become somewhat indistinct or “blurry” at times. This is true in particular for papers that deploy data from many programs with a large variety of different characteristics. At present, it is not possible to disentangle the various parameters.

5 Conclusion

The ESO Telescope Bibliography (telbib) is curated by the ESO Librarians. We interpret the word curation as *to take care of*. Our workflow ensures quality content, including, but not limited to, linking observations to papers, and provides the largest possible warranty for complete and correct data.

Thanks to the detailed annotations and tagging of telbib records, numerous statistics, reports, and visualizations on a large number of parameters can be provided. telbib started in 1996 so that data are available for more than 20 years. The database has therefore become an essential tool to understand publishing trends among the ESO user community.

Despite the many facets of telbib, stakeholders occasionally request statistics that at present cannot be obtained. In order to provide further insights, the ESO Librarians continuously enhance the system with new features.

References

- [1] M. Sterzik, C. Dumas, U. Grothkopf et al., *The ESO Messenger* **no. 162**, 2–8 (2015)
- [2] A. Accomazzi, M.J. Kurtz, E.A. Henneken et al., *Aggregation and linking of observational meta-data in the ADS*, in Proceedings of the ADASS XXV conference, arXiv:1601.07858, in press
- [3] S. Kulkarni, *Instruments on large optical telescopes – a case study*, arXiv:1606.06674 v.2, *subm.*

