Trends and developments in VLT data papers as seen through telbib

Dominic Bordelon[*a], Uta Grothkopf[a], Silvia Meakins[a], Michael Sterzik[a]
[a]European Southern Observatory (ESO), Karl-Schwarzschild-Str. 2, 85748 Garching near Munich, Germany

ABSTRACT

The ESO Telescope Bibliography (telbib; http://telbib.eso.org) is a database of refereed papers published by the ESO users community. It links data in the ESO Science Archive with the published literature, and vice versa. Developed and maintained by the ESO library, telbib also provides insights into the organization's research output and impact as measured through bibliometric studies. Numerous reports, statistics, and visualizatons derived from telbib help to understand the way in which the user community uses ESO/VLT data in publications. Based on selected use cases, we will showcase recent trends and developments.

Keywords: Bibliographic databases, telescope bibliographies, bibliometrics, publications, impact, observatories

1. INTRODUCTION

In many scientific areas, bibliographies and bibliometric studies are among the main tools used to measure, evaluate, and understand research performance. In astronomy, many observatories have established "telescope bibliographies", i.e., lists or databases that contain the observatories' scientific publications or, more specifically, those papers that present, analyze, or use data obtained with the respective facilities. An (incomplete) list of telescope bibliographies can be found at [1].

At the European Southern Observatory, the ESO Telescope Bibliography, or telbib for short, is maintained and further developed by the ESO Library. telbib's main purpose is to interconnect resources by establishing links between observing proposals, published papers, and data in the Archive. Once these links are implemented, telbib users can access the ESO Archive directly from the paper's telbib record; similarly, Archive users find direct links into telbib, leading to the articles that have already deployed the data.

In addition, telbib provides the basis for a variety of reports, statistics, and visualizations that give important insights into ESO's scientific output in general, and more specifically into the publishing behavior and trends among the ESO user community.

During the past years, various new features have been built into the system. Some of them are available through telbib's public interface [2], while others can only be accessed by the librarians through the system's backend. After giving a brief overview of the telbib workflow and curation, this paper will showcase some use cases to highlight telbib's capacities.

2. TELBIB WORKFLOW

The telbib workflow starts with the published literature. A set of refereed astronomy journals is screened during weekly searches. A text-mining software developed at ESO (the fulltext search program "FUSE") highlights those papers that mention any of the ESO-defined keywords (e.g., instrument and telescope names, surveys, etc.) or, through the use of regular expressions in FUSE queries, ESO observing program identifiers (program IDs). Each paper identified through FUSE is visually inspected by the librarians to ensure that the telbib selection criteria are met. First and foremost, papers must use partly or exclusively ESO data. These can be proprietary data (i.e., obtained by the authors) or observations retrieved from the ESO Science Archive.

* library@eso.org; http://www.eso.org/libraries/

ESO's data publication policy [3] states that authors must include the following statement in a footnote or the acknowledgements:

> *Based on observations collected at the European Organisation for Astronomical Research in the Southern Hemisphere under ESO program(s) PPP.C-NNNN(R).*

The place-holder PPP.C-NNNN(R) should be replaced with the actual program(s) that were used in the study. Papers that merely quote results from the literature, mention ongoing projects, suggest future observations or use data in models or simulations merely as examples are not included. Following this definition, telbib also excludes papers that describe existing or future instruments, along with those that use images only as a visual reference without actually analyzing the data. These policies are defined by the ESO Management.

Once it has been decided that a paper qualifies for inclusion in telbib, the curators import bibliographic information as well as author affiliations, the number of citations, and other metadata from the NASA Astrophysics Data System (ADS) [4]. It is then the responsibility of the telbib curators to identify all programs used in the research, and add the respective program IDs including information about the observing mode (service or visitor mode), program type (e.g., Normal, Target of Opportunity, or Large programs), and instruments that were used to obtain the data. Additional tags, for instance describing observing surveys or special observations (like use of commissioning or science verification data) are added. At this stage the telbib curators also indicate for each program whether the observations were obtained by the author team (i.e., whether one or several of the authors are also PI or CoIs of the observing program) or whether the data were downloaded from the ESO Science Archive.

Although program ID citation offers significant automation advantages, human curation is still needed. One major issue is the fact that many authors do not cite program IDs despite ESO's data publication policy described above. Every program ID indicated by the authors is verified to make sure that (a) it is correct and resolves to existing observations in the ESO Archive, and (b) matches the instrument usage mentioned in the paper. Only approx. 70% of the telbib papers published since 2011 list correct and complete program IDs of the respective observations (including a few cases in which Archive request numbers are mentioned). For papers with incomplete or wrong IDs, the librarians search the Archive for observing dates, instruments, or observed objects to find the observing programs. They also inspect cited papers (e.g., "observations as in Paper I") and communicate with ESO scientists who advise on the appropriate application of ESO's data publication policy. Finally, the librarians frequently communicate directly with authors to get further information in order to classify papers properly. This intensive collaboration both ESO-internally as well as with the wider user community is important to assure consistent tagging of telbib records, and to explain to authors the added value and interconnectivity provided by telbib. Due to this extensive effort, we are confident that telbib is essentially complete (at a level of 95%) with respect to major astronomy journals [5].

Once a telbib record has received all *program ID-to-Archive* links, along with further tags and annotations, it is released by the curators and will be retrievable through the public interface. Further background information about the telbib workflow and methodology can be found on the web [6].

3. TELBIB USE CASES

Typical productivity and impact metrics derived from telbib such as the number of publications and citations are reported in the "Basic ESO Publication Statistics" document which is updated regularly and is available from the ESO Library website [7]. These statistics are not repeated here. Instead, we focus on the following selected use cases: (1) Number of authors and observing programs per papers; (2) Fraction of archival papers; (3) Overlap with papers included in the bibliographies of other major observatories; and (4) VLT paper impact relative to *Astronomical Journal* (AJ) papers. All statistics presented here refer to VLT/VLTI data papers (shortened to "VLT" below) and only include refereed publications.

3.1 Number of authors and programs per papers

The earliest papers using VLT data appeared in 1999, less than a year after First Light for VLT's first Unit Telescope (UT1). However, during this year only a small number of refereed publications appeared (29), therefore some of the trends that become visible in later years were not yet very pronounced in this initial set of papers.

Starting with publication year 2000, we see a clear trend towards an increasing number of authors per paper, along with more programs that are being deployed per paper. The average number of authors doubled, starting with 6.5 authors per paper for publication year 2000 and climbing to almost 13 authors on average for papers published in 2014. A decreasing fraction of single-author papers and a corresponding increase in the number of authors per paper has been noticed in astronomy for quite a while (see for instance [8]), hinting towards the fact that larger collaborations gain importance.

A similar development can be noticed for the average number of ESO programs that provide data for a given paper. From just above one (1.17) program per paper in the first year of VLT publications, this number has more than tripled and reaches on average 3.5 programs that provide data for papers published in 2014 (Fig. 1).

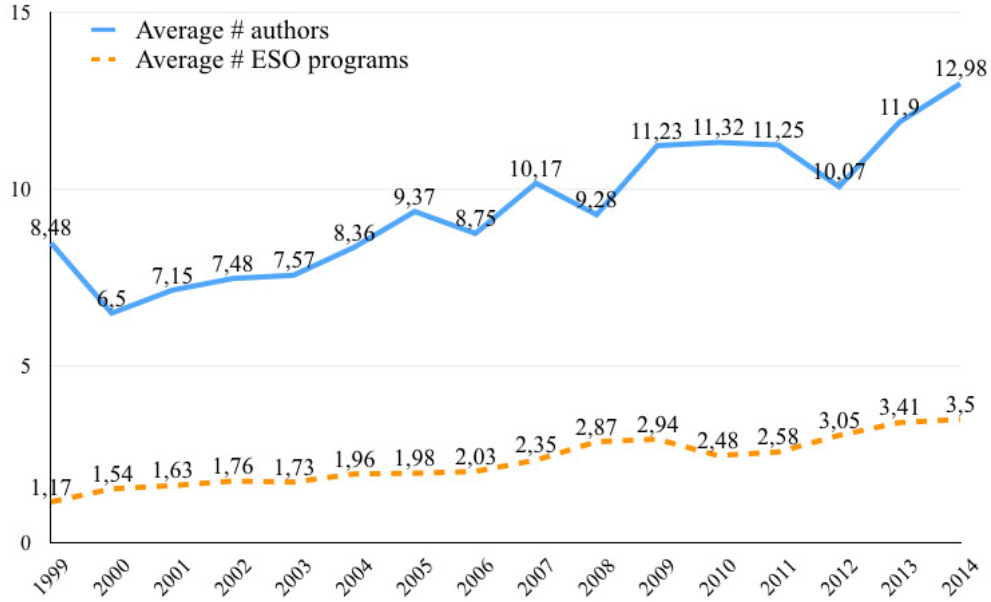

Figure 1. Average number of authors and ESO programs per paper. Publication years 1999-2014, total = 5,941 papers

3.2 Archival papers

Over time, the Archive of ESO's La Silla Paranal (LSP) Observatory has grown considerably. In addition to raw data generated by the LSP facilities, it contains selected processed data and data products from ESO's public surveys. Today, the Archive provides a powerful science resource for an ever growing user community [9]. This statement is fully supported by statistics derived from telbib. Papers that use partly or exclusively archival data, i.e., data from programs for which none of the authors was a member of the team of observers, has increased steadily. Since 2012, the Science Archive has provided data for approx. 25% of the VLT papers, more than any individual instrument.

A closer inspection of archival papers published between 2012 and 2014 reveals that the majority of these publications (14% out of 25%) use exclusively archival VLT data, while a slightly lower number (11% out of 25%) deploy data obtained from the Science Archive ("archival") along with observations obtained by the authors ("new") (Fig. 2).

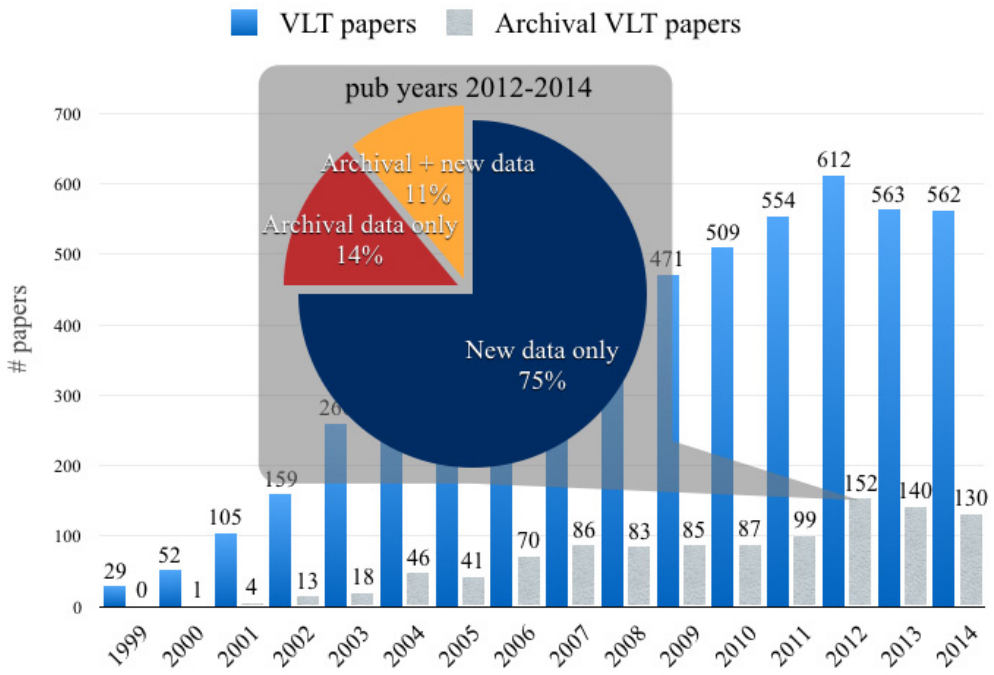

Figure 2. Number of papers using partly or exclusively archival data.

3.3 Overlap with other major observatories

A relatively new addition to the range of telbib features is the ability to track which papers are also included in the telescope bibliographies of other major observatories. For such a study, it is essential to ensure that all participating observatories apply comparable policies regarding the selection criteria for their respective bibliographies. Through collaboration and discussions with the librarians at the Space Telescope Science Institute (HST bibliography) as well as at Gemini, Keck, and Subaru, we are confident that the bibliographies of these observatories follow the same rigorous standards also applied by ESO. Of course this is also true for the ALMA bibliography which is jointly maintained by the librarians at ESO and NRAO (National Radio Astronomy Observatory), along with NAOJ (National Astronomical Observatory of Japan). Papers included in the HST, Gemini, Keck, and Subaru bibliographies were identified using the ADS Abstract Service (filter for specific bibliographic groups) [4]; the ALMA bibliography is available through telbib.

During publication years 2012 to 2014, 1,737 papers using VLT data were published. Of these, 9% also deploy Gemini data, 8% Keck and 4% Subaru data. In addition, we found that 1% of these papers are also included in the ALMA bibliography, despite the fact that the first ALMA papers just appeared in 2012. The largest overlap, and hence synergy, exists with the Hubble Space Telescope: 25% of the VLT papers also deploy data obtained by the HST. This result confirms findings presented in an earlier study which compared VLT and HST data papers and stated the growing interdependence of ground- and space-based observations [10]. Almost 8% of the VLT papers (134 out of 1,737) use VLT observations in combination with data from two, three, or four of the other observatories, including one paper deploying data from all observatories included in this study, except ALMA. In contrast, 63% of the VLT data papers do not overlap with Gemini, HST, Keck, Subaru, or ALMA (Fig. 3).

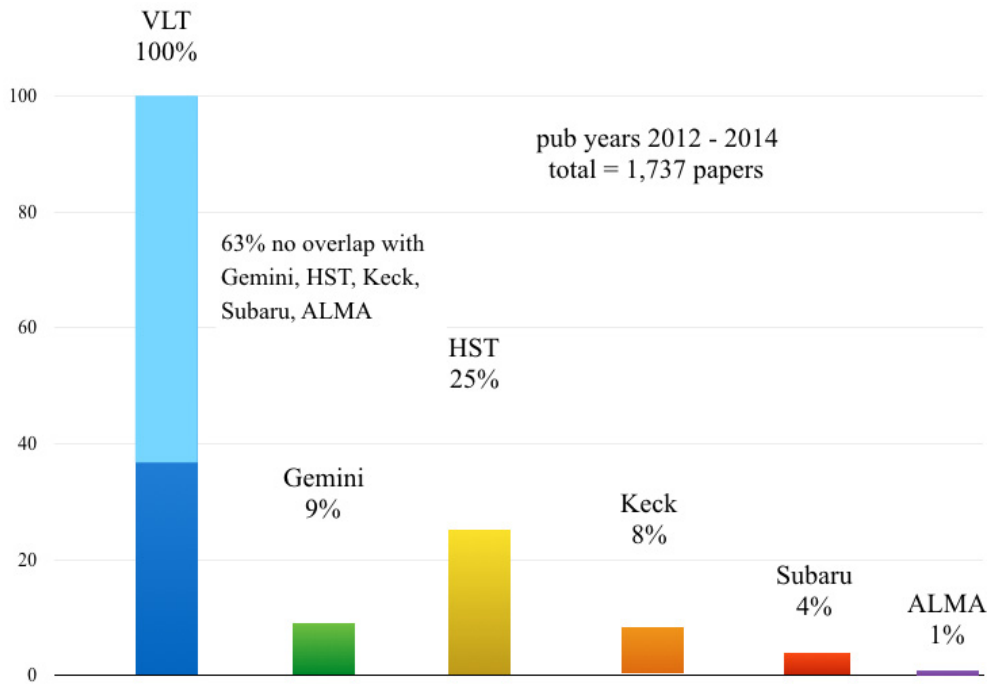

Figure 3. Overlap of VLT data papers with papers included in the Gemini, HST, Keck, Subaru, or ALMA bibliographies.

The use of VLT data in combination with data from other large observatories increases if we investigate papers using archival data. As a first step, we inspect papers published between 2012 and 2014 that use archival VLT data along with "new" (proprietary) observations obtained by the authors (422 papers). For this set of "partly archival" papers, we find a considerably higher overlap with Keck and HST. 13% of these VLT papers are also included in the Keck bibliography (vs. 8% if all VLT papers published 2012-2014 are taken into consideration), and one out of three archival VLT data papers also deploy observations from the Hubble Space Telescope (35% vs. 25% for all papers). Only 53% of the VLT papers do not appear in any of the other observatory bibliographies (vs. 63% for the whole set). Finally, considering only "purely archival" papers, i.e., those that use exclusively VLT data obtained from the Archive (228 papers published 2012-2014), 14% of the papers are shared with Keck, and an impressive fraction of 37% with the HST. Only just above half of the papers (51%) do not use Gemini, HST, Keck, Subaru, or ALMA data.

3.4 Impact (relative to AJ papers)

The impact of scientific publications is typically measured by the total or average number of citations that these papers have accumulated. Citations are seen as an indicator of the usefulness of a paper to other researchers. However, comparisons of citation counts (as well as comparisons of publication counts) have many caveats, such as differences in the publishing and citing behavior of various research communities, possible incorrectness or incompleteness of the citations database, or the often discussed issue of self-citations.

In this study, we apply a simplified version of the methodology described by Crabtree (2014). Instead of gauging impact by looking at the absolute numbers of citations, we apply the definition that "a paper's *impact* is determined by dividing the number of citations to the paper by the median number of citations to all Astronomical Journal (AJ) papers published in the *same* year" [11]. Accordingly, we define the impact of VLT papers as the median number of citations to VLT papers published in a given year divided by the median number of citations of AJ papers of the same year.

The citation statistics presented here were obtained in October 2015 from the ADS. For publication years 1999 to 2014, the median number of citations of VLT papers published in a given year was calculated and divided by the median number of citations of AJ papers of the same year. The result is shown in Fig. 4, treating the median AJ citation as a "measuring stick" (following Crabtree's methodology) and normalizing it to 1 for all years. Except for the first year of

VLT publications (during which only 29 papers were published, a low number for statistics to be really meaningful), the impact of VLT papers relative to AJ papers are on average higher by a factor of 1.7. In particular recent VLT papers (published 2013-2014) are cited approx. 2.5 times more often than an average AJ paper.

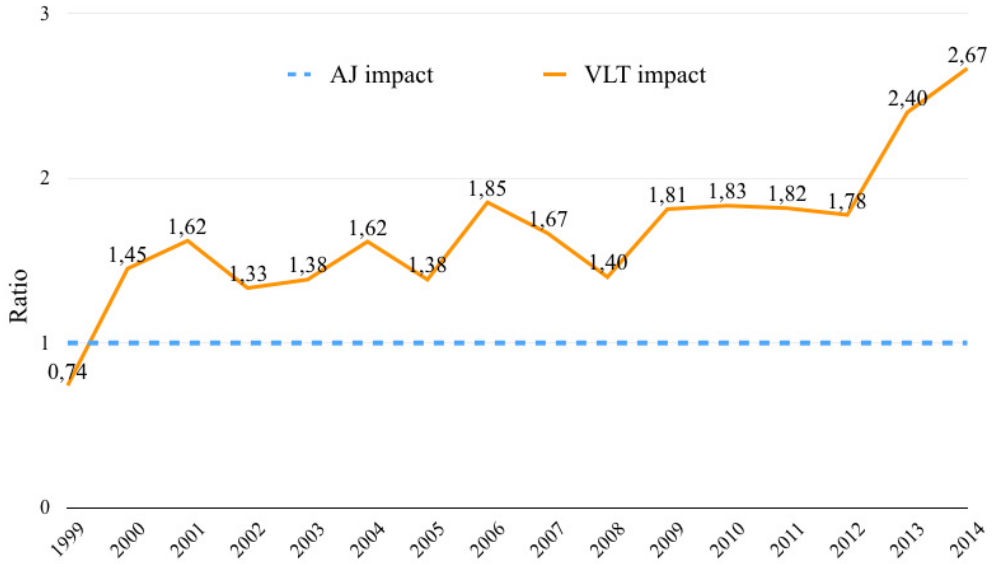

Figure 4. Impact of VLT papers relative to AJ papers.

4. CONCLUSION

In this paper, we show that the ESO Telescope Bibliography (telbib) is an important tool that helps ESO Management as well as instrument scientists and other stakeholders to evaluate the scientific performance of ESO's facilities, to discover and analyze publishing trends among the ESO user community, and possibly to define guidelines for future instrumentation. We conclude that

- telbib is curated
  The ESO librarians are the *curators* of telbib in the original sense of the word, i.e., they *take care of* the complex and sophisticated workflow to ensure that the system provides quality content. This includes, but is not limited to, associating program IDs with the corresponding data in the ESO Archive.
- telbib provides insights
  The database includes a large range of parameters that provide the basis for numerous statistics, reports, and visualizations that support ESO stakeholders in understanding publishing trends.
- telbib has many facets
  The telbib curators maintain and further enhance the system by adding new features, be it in response to specific requests or by applying new technologies that enable exploration of interesting new aspects. This constant development ensures that telbib remains an essential tool for ESO and its user community.

ACKNOWLEDGEMENTS

This research has made use of NASA's Astrophysics Data System Bibliographic Services; the authors wish to thank the ADS team for their excellent service to the astronomy community.

REFERENCES

[1] List of ground-based and space-based observatories and their telescope bibliographies: http://www.eso.org/sci/libraries/useful_links/publicationlists.html

[2] telbib public interface: http://telbib.eso.org

[3] ESO data publication policy: http://www.eso.org/sci/observing/policies/publications.html

[4] NASA ADS Abstract Service: http://www.adsabs.harvard.edu/

[5] Journals regularly screened by telbib/FUSE are A&A, AJ, ApJ, ApJS, AN, EM&P, Icar, MNRAS, Nature, NewA, NewAR, PASP, P&SS, Science

[6] telbib methodology: http://www.eso.org/sci/libraries/telbib_methodology.html

[7] Basic ESO Publication Statistics: http://www.eso.org/sci/libraries/edocs/ESO/ESOstats.pdf

[8] Schulman, E. et al., "Trends in astronomical publication between 1975 and 1996," *PASP*, 109, 1278-1284, (1997); doi: http:10.1086/134008

[9] Romaniello, M. et al., "The growth of the user community of the La Silla Paranal Observatory Science Archive," *The Messenger*, no. 163, 5-9 (2016), http://www.eso.org/sci/publications/messenger/archive/no.163-mar16/messenger-no163-5-9.pdf

[10] Grothkopf, U. et al., "Comparison of science metrics among observatories," *The Messenger*, no. 119, 45-49 (2005) http://adsabs.harvard.edu/cgi-bin/nph-data_query?bibcode=2005Msngr.119...45G&link_type=ARTICLE&db_key=AST&high=

[11] Crabtree, D., "A bibliometric analysis of observatory publications for the period 2008-2012," *Proc. SPIE* 9149, Observatory Operations: Strategies, Processes, and Systems V, 91490A (2014); doi:10.1117/12.2054058