The ESO Telescope Bibliography at your Fingertips

Uta Grothkopf[a] and Silvia Meakins[a]
[a]European Southern Observatory (ESO), Karl-Schwarzschild-Str. 2, 85748 Garching near Munich, Germany, library@eso.org

ABSTRACT

Bibliometric studies have become increasingly important in evaluating individual scientists, specific facilities, and entire observatories. The ESO Library has developed and maintains the Telescope Bibliography (telbib), a database of refereed papers that use observational data generated by ESO's facilities.

Recently, a new public interface has been released. In addition to classical queries for bibliographic and facility-related information, it provides advanced features like faceted searches and filtering, autosuggest support for author, bibcode and program ID searches, hit highlighting as well as recommendations for other papers of possible interest.

An additional tool offers the possibility to create graphical statistics on the fly based on user-defined criteria. The ESO Telescope Bibliography is available at http://telbib.eso.org/.

Keywords: Bibliometrics, Observatory bibliographies, Publication statistics, Science metrics, Libraries

1. INTRODUCTION

1.1 Background

Like many other observatories, ESO uses a variety of metrics to evaluate performance and impact of the organization. The library contributes to this task by providing bibliometric studies, i.e., it creates reports that help evaluate the productivity and impact of ESO's observing facilities by investigating the number of publications and citations, respectively. For many years, lists of papers published by the ESO users' community have been published in the ESO Annual Report in order to keep track of the scientific output from ESO telescopes and instruments. Starting in the mid-1990ies, a database was developed that allowed to not only print the respective lists, but also query the content based on bibliographic as well as observation-related parameters [1]. This database has been considerably extended and improved over time and has by now evolved into the ESO Telescope Bibliography.

Nowadays, many observatories maintain similar databases. A list of some ground and space-based observatory bibliographies is available at http://www.eso.org/libraries/publicationlists.html. Results from a recent survey on methodologies and practices applied by various observatories including their respective criteria for inclusion of papers, FTE employed, and possible interoperability, can be found in Lagerstrom et al. in these proceedings [2].

1.2 The ESO Telescope Bibliography (telbib)

In order to be able to provide consistent and reproducible statistics about ESO's scientific output, the Telescope Bibliography (telbib) was developed. telbib is a database of refereed papers published by the ESO users community. All papers use partly or exclusively data from ESO facilities. Statistics derived from the ESO Telescope Bibliography include only papers based on data from telescopes and instruments for which observing time was evaluated by the ESO OPO (Observing Programmes Office).

The database is used for a number of purposes, including

- to interconnect resources (from proposals to papers and from papers to data in the archive)
- to measure ESO's scientific output (productivity and impact)
- to evaluate the performance of telescopes and instruments
- to put ESO in context with other observatories
- to define guidelines for future facilities

It is important to note that bibliometric studies are only one aspect of these aims. Other metrics need to be used in addition to obtain a more complete picture. More information about the purpose, coverage, and completeness of the ESO Telescope Bibliography is provided on the "telbib Methodology" page at http://www.eso.org/libraries/telbib_methodology.html.

1.3 telbib workflow

The library has developed and maintains two essential tools in order to keep statistics and reports up-to-date, correct, and as complete as possible: the semi-automated full-text search tool FUSE and the telbib content management system. Both tools rely on the NASA ADS Abstract Service (http://adsabs.harvard.edu/) for bibliographic metadata. FUSE and telbib form part of a workflow that links published literature with data located in the ESO archive.

The telbib workflow starts with access to the scientific literature. Electronic versions of a defined set of refereed journals are monitored. FUSE provides retrieval methods for these journals that allow us to obtain the PDFs and convert them to searchable text. Full-texts are only used for text inspection and are not stored on our server. The system scans the texts for ESO-defined keywords and highlights hits in context on the results page. Blatantly wrong findings are removed immediately from this list.

All other papers that contain any of the ESO-defined keywords are inspected visually to examine whether or not they use observational data from ESO's facilities. Instrumentation and other engineering papers that do not actually use data are excluded, as well as papers that use images as overlays without further analysis. For those records that are to be added to the telbib database, bibliographic information as well as other metadata are imported from the ADS Abstract Service.

At ESO, the ultimate decision regarding telbib policies, specifically about inclusion or exclusion of papers, rests with the Director for Science who defines policies in cooperation with the librarians. The ESO librarians are responsible for applying them consistently and reproducibly. Each telbib record receives tags and annotations such as telescope and instrument names, standardized descriptions of surveys, tags to indicate use of archival data, and other information. Most importantly, identifiers of all programs (program IDs) that provided data for the research are added to the record.

Following ESO's policy for acknowledging use of data, many authors mention the respective program IDs in the footnote of papers. However, a large number of manuscripts only mentions some or no identifiers. In these cases, the ESO librarians try to determine the missing ones by querying the ESO observing schedule and the ESO archive based on observing dates, instruments, and other information provided by the authors. If this does not lead to a complete result, authors are contacted, asking them to clarify which data were used for their research. All author-provided program IDs are verified through the ESO archive in order to eliminate typos and other mistakes.

The program IDs link telbib records with the corresponding data in the ESO archive. Through the public telbib interface, users can access the archive server from where the data can be requested. Likewise, search results in the archive indicate the number of papers that have so far been published using the data.

More information about telbib's front-end is provided below.

2. THE PUBLIC ESO TELBIB INTERFACE

Since the first public interface of the Telescope Bibliography was made available, the ESO archive group was in charge of its layout and functionality. This changed in the final quarter of 2011. Since then, in addition to the back-end also the telbib front-end has been in the hands of the ESO librarians. The previous interface has been completely revised, and an updated, user-friendly version has been implemented, using state-of-the-art technology and layout features.

2.1 Search interface

Various search options are available to query the database, including searches by bibliographic details (for instance authors, title words, author-assigned keywords, publication year, etc.), by ESO-related observing facilities (instruments, telescopes, observing sites), and by program IDs. The main search screen provides immediate access to the top 5 journals and instruments. The number of records contained in the database for the respective facets are indicated. In addition, the most recent five publication years are displayed, together with the number of papers per year.

A spellchecker is available for certain search fields (for instance author names, title words). In case search terms entered in these fields do not lead to any hit, the system will provide hints towards search terms that will turn up results ("Did you mean...?"). In addition, queries for authors, bibcode, and program IDs are supported by an autosuggest feature that offers search terms which exist in the index after at least two characters have been entered.

2.2 Results list

Papers that fulfill the search parameters are presented in tabular format. For each paper, the publication year, first author, title, instruments, program IDs, and bibcode of each paper are shown. Titles are linked to the detailed view of records. Program IDs lead to the ESO observing schedule and from there to the archive from where the data can be requested. Bibcodes are connected to the full-texts at ADS which are available either from the publishers or from the astro-ph/arXiv eprint server.

Faceted filters can be used to refine search results. Facets exist for publication years, journals, and instruments. The available list of five items per filter can be expanded to show the entire list. Fig. 1 shows an example of a results page.

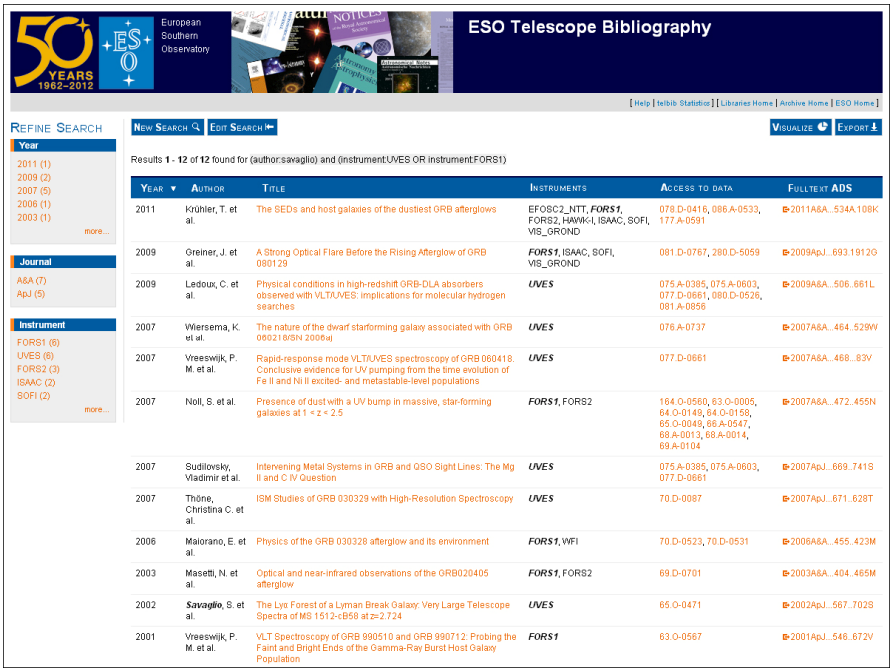

Figure 1. The ESO telbib results page shows the most relevant information of each paper. Access to the full-text is provided via the ADS Abstract Service where manuscripts can be retrieved either from the publisher or from the arXiv eprint server. In the upper right corner, the blue buttons can be seen that allow users to export or visualize results sets.

2.3 Detailed record display

The detailed record view shows more information about the ESO telescopes and instruments that provided data for the research. In addition, further descriptive tags like survey names or the use of archival data are indicated. Search terms are highlighted for easy identification. Instruments, telescopes, and observing sites are hyperlinked and will retrieve other papers that use the same facilities.

Each program ID offers two links with different functionality. The program ID itself is linked to all other papers in the telbib database that use data from the same program. Selecting the "access to data" option takes users to the ESO observing schedule and from there to the archive where the respective data can be requested, provided that the proprietary period has ended.

At the bottom of the page, recommendations for other papers that may be of interest are shown, based on certain factors that characterize the current paper. The detailed record page also provides information about the current number of citations this paper has gathered. Citations are loaded on the fly from the ADS; hence they are as complete and current as the citation numbers obtained directly from there.

The public telbib user interface has also been described in [3] and [4].

3. STATISTICS AND VISUALIZATIONS

Metrics derived from the telbib database are made available in various ways, for instance through the "Basic ESO Publication Statistics" document located on the libraries homepage (http://www.eso.org/libraries/) and through reports created by the librarians on request. Search results retrieved through the public telbib interface can also be exported into comma-separated (.csv) or tab-separated (.txt) format for further use. This functionality has been available since the very first version of the telbib website.

In order to enable users to obtain even more information independently, it is now possible to view graphs and visualizations of search results. These are either predefined "telbib Statistics" charts, covering all publication years of telbib papers (1996 to present), or "Visualizations" which are created on the fly, based on the criteria entered by users in the telbib search interface. Both tools allow users to get a graphical impression of previously only text-based results. The two display options are described in Sections 3.1 and 3.2.

All charts have two characteristics in common: they are animated, and they are interactive. Users can enable or disable items shown in the graph by clicking the corresponding name in the legend. In this way, the graph can be tailored as needed. In all charts, a mouse-over on a given data point shows the facility name (instrument, journal, etc.) as well as the value. Charts can be printed or downloaded in various formats.

telbib Statistics and visualizations are realized using software provided by Highcharts.com.

3.1 Visualizing results

Search results pages are equipped with a "Visualize" button that allows users to show the results of this specific search. The charts are created on the fly. Users can choose between four types:

1. number of papers per year
2. number of papers per journal
3. archive data usage
4. number of papers per instrument per year

Choosing option 1 results in a bar chart which shows the number of papers per year in the results set. Charts 2 and 3 are pie charts, indicating the number of papers per journal and the fraction of papers that use archival, new, and archival plus new data, respectively. Option 4 leads to a line diagram which shows all instruments used in the results set over time.

3.1.1 Examples of visualizations

In the example shown in Figure 2, telbib was searched for papers that use data from either ISAAC (Infrared Spectrometer And Array Camera) or SOFI (Son of ISAAC) published until 2011. The query leads to 1,205 papers. Using the "Visualize" button, users can now for instance initiate a graph that shows the number of papers per instrument over time. All papers in the results set contain ISAAC or SOFI data, but in addition several other facilities contributed data. By default, the 10 instruments that contributed to most papers are activated. Users have the option to select only those instruments they really want to see by clicking on all other facilities in the legend. This will grey-out these facilities until only the desired instruments (in this example ISAAC and SOFI) remain. Likewise, instruments that are currently not displayed can be switched on by users.

Using the same results set, Figure 3 illustrates visualization option 3. The pie chart shows the fraction of papers that are based exclusively on data obtained by the authors ("without Archive data", 76.76%), that use new as well as archival data ("Archive + new data", 10.29%), as well as those that are based exclusively on archival data ("Archive data only", 12.95%).

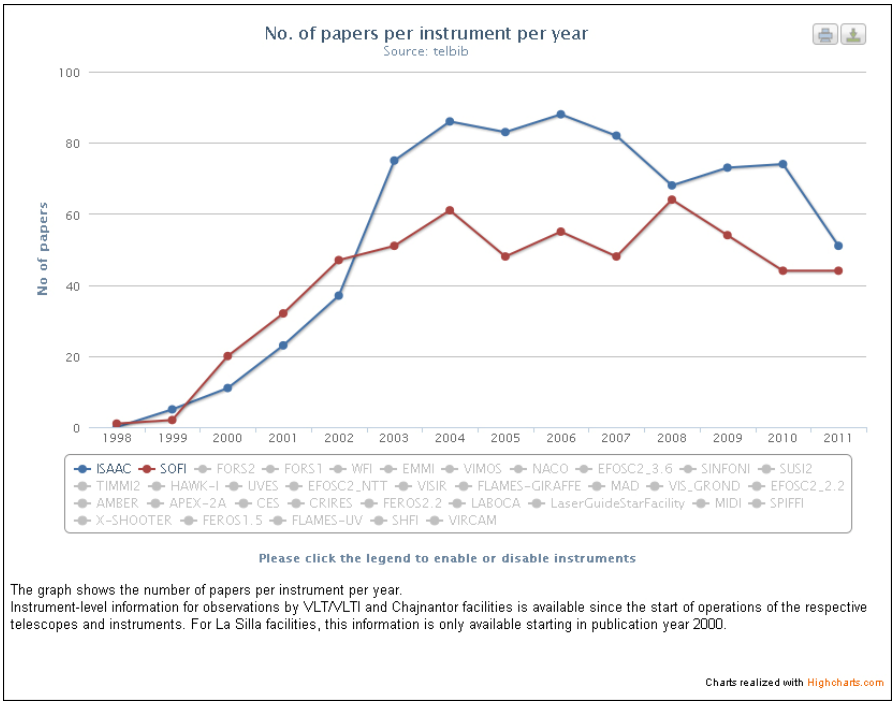

Figure 2. Example of visualizations of results sets. A line diagram shows the number of papers per instrument per publication year. Facilities that are not needed in the graph can be greyed-out by users.

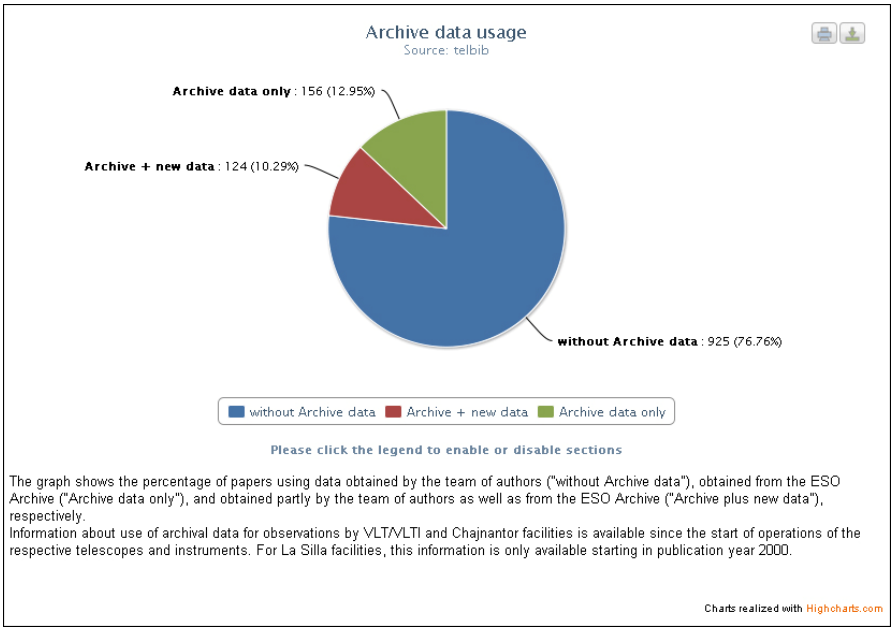

Figure 3. For the same results set, the fraction of papers that use only observations obtained by the authors ("without Archive data"), archival as well as new data, and exclusively archival data is shown.

3.2 telbib Statistics

In contrast to the visualization options described above which are graphic interpretations of search results, the telbib Statistics option refers to all records contained in the database, i.e., from publication year 1996 to the present. telbib Statistics can be retrieved by following the link located in the navigation bar in the upper right corner of all telbib front-end pages, or by accessing them directly at http://telbib.eso.org/telbibstats/.

At present, the following graphs are available:

- VLT/VLTI instruments
- Other instruments of the La Silla Paranal (LSP) Observatory
- Chajnantor instruments
- VLT instruments by years of operation
- Archive data usage
- Average number of authors and programs per paper
- Average number of authors per paper by journal

Thumbnails on the main statistics page give a first impression of the individual graphs so that users know what kind of chart type to expect. All charts referring to instrument use from the various observing sites are line diagrams. Archive data usage as well as the average number of authors and programs per paper are illustrated through bar charts. The average number of authors per paper by journal is again a line diagram that shows the number of papers published in the major astronomy journals over time.

3.2.1 Examples of telbib Statistics

Those charts that illustrate the number of papers based on data from specific instruments follow the principles explained in Section 3.1.1 and Figure 2. Instead of reproducing them again, we have chosen to highlight here two other options.

Using the predefined telbib Statistics, it is possible to get an overview of the average number of authors and programs that contributed to papers over time. Instrument-level information for observations by VLT/VLTI and Chajnantor facilities is available since the start of operations of the respective telescopes and instruments. For La Silla facilities, this information is only available starting in publication year 2000. Using this tool provided by the ESO Telescope Bibliography, it can be shown how both the average number of authors and the average number of programs have increased over time (Figure 4).

The average number of authors per paper of articles published in the major journals can be visualized with the last telbib Statistics option. The output is limited to papers that appeared in *A&A*, *AJ*, *ApJ*, *Icarus*, *MNRAS*, *Nature*, *PASP*, and *Science*. By default, only five journals are displayed, but others can be enabled by users through the legend. Likewise, those journals shown by default can be disabled if users choose to do so (Figure 5).

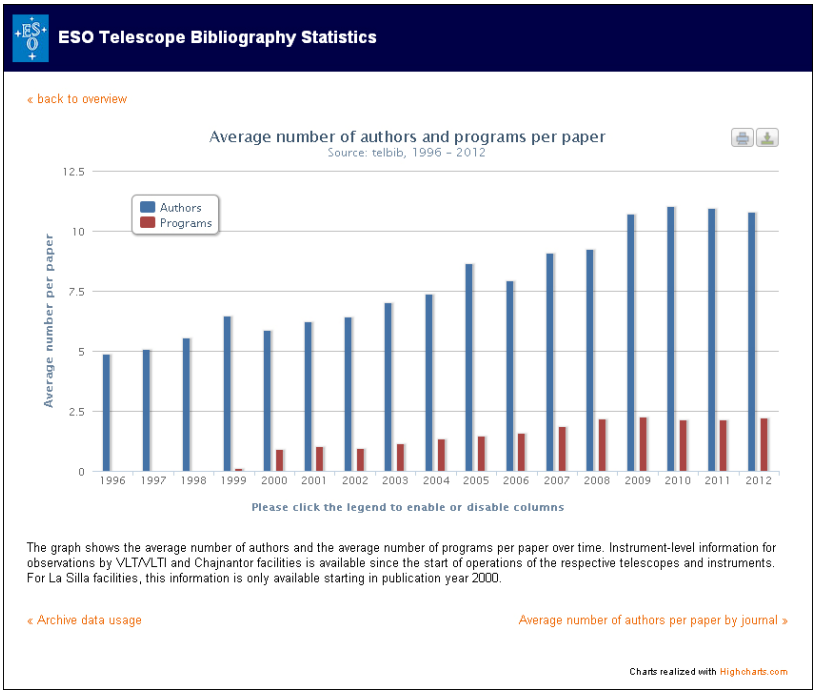

Figure 4. The average number of authors and average number of programs per paper are shown in a bar chart.

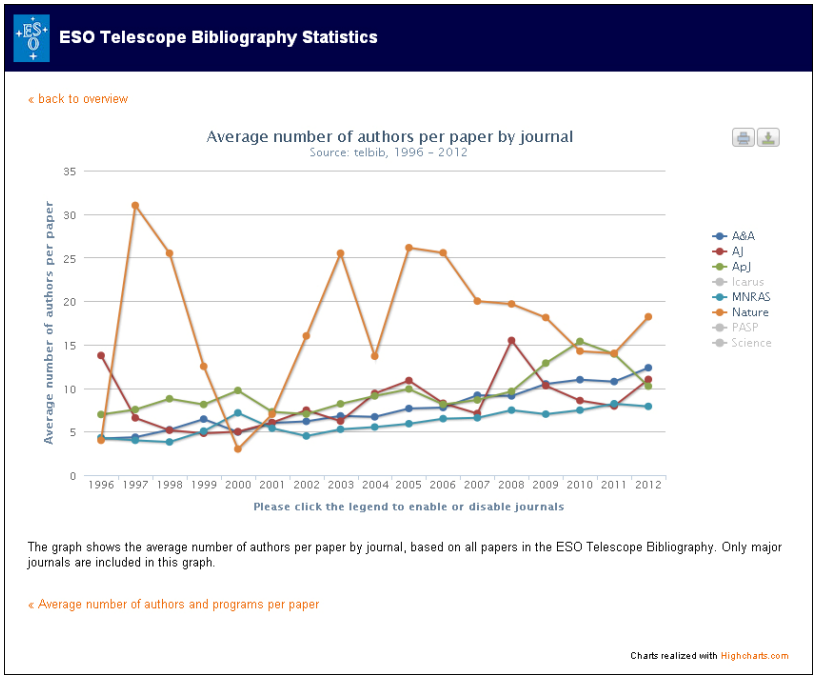

Figure 5. The average number of authors per paper by journal is displayed. Only major journals are included in the graph.

4. CONCLUSION

Statistics derived from telescope bibliographies have become an intrinsic part of evaluations of specific observing facilities as well as observatories at large. In this paper, we have given an overview of the bibliometric tools and workflow implemented at ESO in order to keep track of the organization's scientific output and impact. In particular, we have described the ESO Telescope Bibliography (telbib), a content management system used to track refereed papers that use partly or exclusively ESO data, store additional metadata, and generate statistics and reports.

The public ESO telbib interface has undergone a complete transformation and a state-of-the-art website has been implemented to provide new features and sophisticated search functionality. telbib is maintained by the ESO Library.

Lately, tools to visualize telbib search results have been added. Two options are available: (a) the tools can either be evoked through the "Visualize" button on results pages in order to illustrate the specific search result, or (b) refer to the entire content of the database, covering all publication years since 1996 ("telbib Statistics"). These new features provide graphical interpretations of information that was previously only available in text-based format. All charts available through the visualization and telbib Statistics areas are interactive and can be tailored to the users' specific needs, making search results more informative, visual and, ultimately, fun to use.

REFERENCES

[1] Erdmann, C. and Grothkopf, U., "Next generation bibliometrics and the evolution of the ESO Telescope Bibliography," ASP Conference Proceedings 433, 81-88 (2010).
[2] Lagerstrom, J.P., Winkelman, S.L., Grothkopf, U. and Bishop, M.J., "Observatory bibliographies: current practices", Proc. SPIE 8448, these proceedings (2012).
[3] Meakins, S. and Grothkopf, U., "Linking publications and observations – the ESO Telescope Bibliography," ASP Conference Proceedings (ADASS 2011), in press (2012).
[4] Grothkopf, U. and Meakins, S., "ESO Telescope Bibliography: new public interface," The ESO Messenger 147, 41-43 (2012).