

Navigating from Publications to Astronomical Databases

François Ochsenbein

CDS, UMR 7550, Observatoire Astronomique de Strasbourg, France
francois@astro.u-strasbg.fr

Claude Bertout

Editor, Astronomy & Astrophysics
bertout@iap.fr

James Lequeux

Observatoire de Paris, France
james.lequeux@obspm.fr

Françoise Genova

CDS, UMR 7550, Observatoire Astronomique de Strasbourg, France
genova@astro.u-strasbg.fr

Abstract.

The implementation of journals on the Web has opened new possibilities for the scientific usage of published results because it is now possible to link published articles to other types of information. The availability of published information in electronic form also allows for new types of content validation, complementary to the referee's validation, and to the layout performed by the publisher.

For several years now, authors publishing in A&A are offered the possibility of quoting the astronomical objects they are studying directly in their latex manuscript (via the `\object` macro). Since April 2001, this macro is being translated into an actual link from the article to the SIMBAD database. This experiment is still a prototype, and its various aspects are presented here.

1. Introduction

Astronomical databases have now been operational for several decades — SIMBAD for instance has a history of over 30 years, and NED was launched in the late 1980s. These “generalist” databases provide data for millions of astronomical objects, including basic data such as position in the sky and brightness, as well as actual results extracted from important catalogues; however, the main originality of these databases is a complete *bibliography* of the astronomical objects, compiled from a list of the papers in which those objects are quoted.

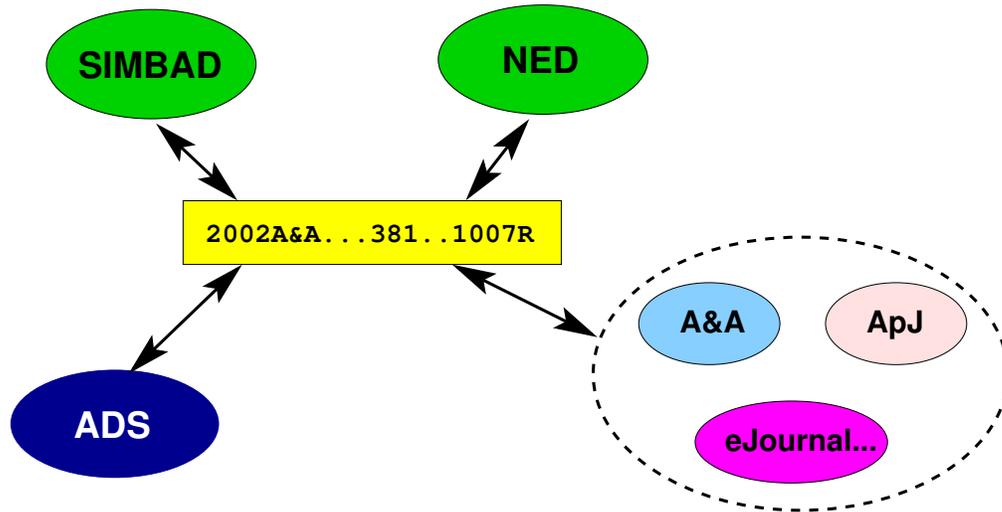


Figure 1. Diagram illustrating the role of the *bibcode*

The astronomical journals have been accessible on-line for five or six years; for *Astronomy & Astrophysics*, some parts of the journal were accessible electronically several years earlier. Abstracts and large tables have been stored on directories publicly accessible via Internet from 1993 (Ochsenbein & Lequeux, 1995). And around the same epoch – in the early 1990s – the ADS Abstract Service (<http://adswww.harvard.edu/>) was launched, enabling the retrieval of the abstracts of a paper from various words and keywords such as topics or authors' names.

2. The “bibcode” as a link

Bibliography therefore became, from the early stages, one of the basic commonalities shared by the databases of the discipline. The need for a common way of referencing an article led to the definition of the *refcode*, later also designated as the *bibcode*: it is a 19-digit number, like 2002A&A...381..1007R, understandable by the astronomer and easily processable. It was originally borne as an agreement between the SIMBAD and NED teams in the late 80's, and was later more formally described by Schmitz et al. (1995). The bibcode was adopted by the ADS group from its early stages enabling an immediate cross-referencing of the documents using or quoting the bibliography: the databases, ADS, and the astronomical journals as soon as electronic versions came into existence. Fig. 1 illustrates the role played by the bibcode, which is used both as an *input key* to access data (*e.g.* to get the full electronic paper from the Journal, retrieve the abstract in ADS, and find the objects quoted in NED or SIMBAD), or as an *output link* to get all related material (*e.g.* the papers referencing an article, all papers dealing with a peculiar source, etc...)

3. Links to Astronomical Objects

The existence of the bibcode enabled an immediate transformation of any cited paper into a dynamical *link* bringing up the complete information with just a mouse click.

A similar problem arises in publications dealing with individual astronomical sources. The presence of a *link* to databases containing the known details about those objects would avoid multiple repetitions of the basic object properties like positions, brightnesses and morphologies; such links moreover bring up the *actual* details of the sources, rather than a frozen set of parameters. Since the databases are providing links to other data sets like digital images from different epochs, different bandpaths, and different observatory archives, a complete knowledge becomes just a few clicks away from the reader.

Such links are convenient not only for the reader, they also bring up other interesting features:

- from the *author's* perspective: it becomes unnecessary to add the minimal details which are required for a presentation of the individual objects used in a demonstration; the link can also be used to check the correctness of the object designation and to compare his results to the knowledge already stored in the databases;
- from the *Editor's* perspective: the existence of such links are powerful checking tools and can minimize the length of papers by avoiding some repetitive descriptions;
- from the *database maintainer's* perspective: the existence of these links brings up facilities for a more automated scanning of the bibliography. These links can also be an interesting element in the measurement of a paper's impact relative to the objects being studied (Lesteven, 2002).

4. The `\object` macro

A special `\object` macro has been existing in the *A&A* latex macro package for about 5 years. For instance, an author may quote his objects with the following original latex source:

```
\title{Spectral analysis of the exciting star of \object{Ps 1},
a planetary nebula in \object{M15}}
\begin{abstract}
We present the analysis of the sdO star \object{K648}, the central
star of the planetary nebula \object{Ps 1} discovered in the
globular cluster \object{M15}.
\end{abstract}
```

where the author decides to quote the objects in the title and the abstract. The author is of course free to insert the `\object` macro anywhere in his article, including *e.g.* the tabular material. Another `\listofobjects` macro also exists in the *A&A* latex macro package whose role is to list a summary of the `\object` usage in the article.

The authors have therefore complete control as to which objects have to be marked and on their exact location in the text of their paper.

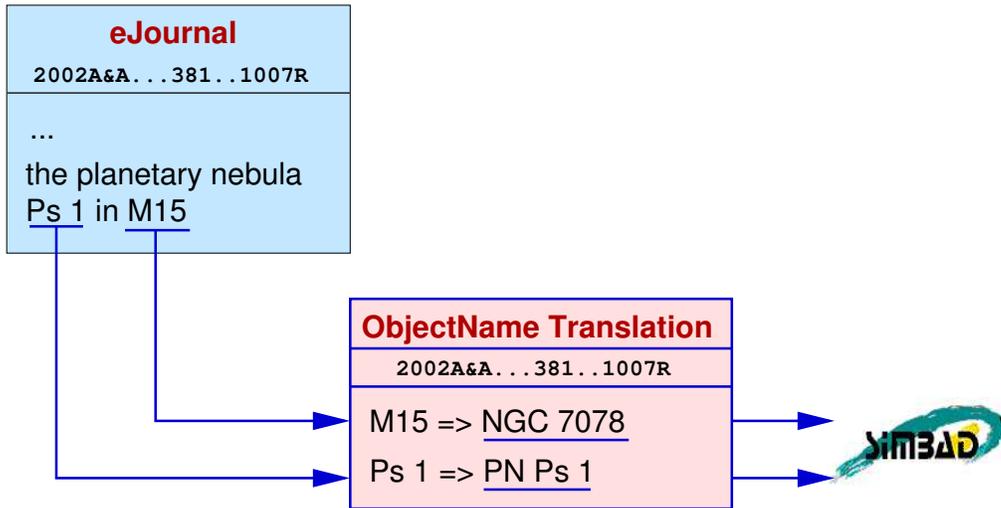


Figure 2. Diagram showing the path from an object name cited in a paper to SIMBAD via a *translator* of names adapted to each *bibcode*

4.1. The difficulties

Despite its long existence, the `\object` was activated into an actual link only in April 2001. There are several difficulties which have to be solved in this exercise:

- the designation of an object is frequently *relative* to a paper's context, where very short abbreviations are the rule: for instance, in the context of a study of the distance of a globular cluster based on its RR Lyr variables, the designation *V3* can be used to designate one of the variables catalogued in this cluster. This designation is unambiguous in the context of the paper, but requires other details like the name of the cluster and the reference of the catalogue to become an *absolute* designation processable by the database;
- even in the case of unambiguous names, the interpretation of astronomical designations may be relative to databases — SIMBAD and NED do not always agree on the designation of an object. Such difficulties are unavoidable, even though we are working to minimize their number.
- the designation of an object is subject to evolution — a name does not live for ever, the increasing resolution of instrumentation has for instance an impact on the astronomical nomenclature;
- and finally errors may occur — especially in the spelling of astronomical objects' names — but the original paper cannot be modified.

4.2. The current linking implementation of the `\object`

In view of these difficulties, the way the link from the journal to the SIMBAD database was installed is illustrated in Fig. 2: the name of the object as quoted by the author is associated to the article's *bibcode* — a straightforward way of

designating an article, as shown in section 2— and these 2 parameters are used to resolve the possibly ambiguously named object into an *absolute designation* of the object which can be understood by SIMBAD. The actual way of making this relative-to-absolute name designation is currently using an *ObjectName Translation* table.

It is worth noting that the current implementation assumes that the two parameters — the *bibcode* and the *object name* quoted by the author — are enough to uniquely identify one single object. This assumption would fail in articles where the author uses the same name to designate different objects: such cases should be discouraged and are hopefully extremely rare. If that happens, a click on the ambiguous name would bring up the list of the different objects studied in the article sharing this name.

The current implementation also proposed a solution for dealing with the detected errors: the association of the article's *bibcode* and the wrong designation in the *ObjectName Translation* table can contain a text explaining the origin of the error. This possibility was already used a couple of times.

The practical implementation consists of the following: the list of the `\object`'s macros — in fact the result of the `\listofobjects` — are sent by the publisher to the CDS, who then checks the names, and implements the *ObjectName Translation* of the quoted objects into *absolute* names. Some of the links are not immediately operational — the inclusion of the newly discovered objects into SIMBAD requires minimal delays — but become effective as soon as all the questions about the quoted objects could be solved, without any intervention of the journal publisher.

4.3. The A&A experience

Since the `\object` has been in use now for about one year, it was possible to make some statistics for the period April 2001-May 2002 (volumes 369 to 386) among the 2049 papers published during this period,

- 13.7% (276) were marked by the authors (containing at least one `\object` macro, the average being 9.4 *different* objects marked in one paper)
- 72.4% have quoted objects reported in SIMBAD.

This means that currently about 20% of the authors are well aware of the usefulness of this macro; its usage is progressively increasing.

From the database maintainer's point of view, this result also means that it is currently difficult to rely on the author's markup. Even in the 13.7% papers containing objects marked by the `\object` macro both the list of objects and the names quoted are not yet reliable. The most frequently encountered difficulty is the *variability* of the astronomical nomenclature mentioned in section 4.1: the names of the objects are frequently truncated — the names of the popular objects being not surprisingly the most severely truncated.

5. Concluding remarks

The experiment of the `\object` linking, done in close cooperation with the journal scientific Editor, finally works with good reliability once the *Object-*

Name Translation table is created and maintained. It is however difficult today to envisage a possibility of relying on the author's markup to update the databases.

References

- Lesteven, S. 2003, "Identifying the important papers for one astronomical object" (*Library and Information Services in Astronomy IV, this volume*)
- Ochsenbein, F., Lequeux, J. 1995, *Vistas in Astron.* **39**, 227
- Schmitz, M., Hérou, G., Dubois, P., et al. 1995, "NED and SIMBAD Conventions for Bibliographic Reference Coding", in: D. Egret & M.A. Albrecht, ed., *Information & On-line Data in Astronomy*, Kluwer Acad. Publ., p. 259; alternatively see <http://cdsweb.u-strasbg.fr/simbad/refcode/refcode-paper.html>