



HAL
open science

POP: Publications by the Observatoire de Paris

C. Brocquet, F. Henry, A. Laurenceau

► **To cite this version:**

C. Brocquet, F. Henry, A. Laurenceau. POP: Publications by the Observatoire de Paris. Library and Information Services in Astronomy VII: Open Science: At the Frontiers of Librarianship, 2015. halshs-03473168

HAL Id: halshs-03473168

<https://shs.hal.science/halshs-03473168>

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.

POP: Publications by the Observatoire de Paris

Christophe Brocquet,¹ Florence Henry,² and Amélia Laurenceau¹

¹*Observatoire de Paris, UMS 2201, CNRS/INSU, PSL Research University ; 61 avenue de l'Observatoire, 75014 Paris, France*

²*LESIA, Observatoire de Paris, CNRS, UPMC, Université Paris Diderot, PSL Research University ; 5 Place Jules Janssen, 92190 Meudon, France*

Abstract. The paper presents “POP: Publications of the Observatoire de Paris,” a project of the Observatoire de Paris, which was launched to keep record of its activity by indexing its publications. It also describes the aims and the challenges of the project and the features of Publesia, the software that was adopted for the project.

1. Introduction

Astronomers and their respective institutions disseminate their work through scientific publication. It is therefore necessary to keep a record of this output. Moreover, as academics and institutions are subjected to multiple assessment processes, an index of these publications is expected. At the same time, the open access movement calls attention to the idea of freely available scientific and technical information and suggests that uploading academic papers onto open-access repositories is necessary. What kind of bibliographic tool would match all these assessment requirements and allow, at the same time, an increased visibility of publications?

In order to answer this question, the Observatoire de Paris launched “Project POP: Publications by the Observatoire de Paris.” Together with ESPCI ParisTech,¹ a Parisian chemistry and physics engineering school, the Observatoire decided to use Publesia, which is open source software initially developed at LESIA,² one of its laboratories. The software can interface with Astrophysics Data System, Web of Science or HAL³ databases, in order to automate data input as much as possible.

The main challenge was determining what single tool would answer the needs of the institution and its researchers and funding bodies, and, at the same time, initiate an organisational policy of open-access archive deposit? This paper will start with an introduction of the project in its national and local contexts. It will next present the

¹ESPCI ParisTech: Ecole supérieure de physique et de chimie industrielles de la Ville de Paris
<http://www.espci.fr>

²LESIA: Laboratoire d'études spatiales et d'instrumentation en astrophysique - Laboratory of space studies and instrumentation in astrophysics <http://lesia.obspm.fr>

³HAL: Hyper Articles en Ligne <http://hal.archives-ouvertes.fr/>

developments that were made of open-source software Publesia to create a multifunctional, easy-to-use database.

2. Context, origins and aims

Indexing publications: a necessity

The Observatoire de Paris and its laboratories and researchers all need up-to-date lists of their publications. Indeed, these lists have multiple uses including academic assessment, career development reports, results assessment, negotiations with funding bodies, and tender calls and funding applications. We can consider creating an index of all publications by the Observatoire de Paris as an aim in itself, as it is also something that trusteeships and other research funding bodies may require. Finally, the online availability of these lists could make the Observatoire de Paris's scientific production more visible, which is another of Project POP's aims.

Thus, for multiple reasons, academics are required to provide a list of their publications. POP is designed to provide them with a tool so that they only need to enter their publications into the database once. Once that has been done, the same tool will allow them to create lists of their publications and export them upon demand.

Indexing publications: the past

Between 1983 and 2001, two database indexes had successfully been created at the Observatoire de Paris to answer the above stated needs; the indexes referenced approximately 13.000 publications. The last database index was abandoned in 2002, when scientific departments were restructured and the current laboratories were created. Since that last database was abandoned, each laboratory re-created its own (more or less operational) indexing tools. These were created to answer the demands from assessment committees, which were constituted as a consequence of the creation of four-year contracts. Hence 12.750 publications were indexed as a result of these tools between the years 2002 and 2012. In parallel, at the Observatoire de Paris, former chairman Daniel Egret started to create publication lists using searches made in the ADS database. From 2007 onwards, the Observatoire de Paris has acquired a list of its own publications using the information gathered in WoS by the Observatoire des Sciences et Techniques. As WoS has a bigger thematic coverage than ADS, the data from WoS completes the list made using ADS data.

HAL or not HAL?

The Observatoire de Paris could have used HAL for publication management. HAL is a national archiving platform created in 2000 that all French research and higher education institutions adopted in 2013. In fact, a subset of articles on HAL is automatically transferred to the arXiv.org e-Print archive. One could assume that HAL would be enough to serve the Observatoire's needs. Yet few academics actually use HAL for their archiving purposes: most researchers claim that it is too complex to use. Academics also need to address open-access archiving issues, which they often have little information about. Finally, the version of HAL currently being used does not allow for the descriptive metadata management that the Observatoire de Paris requires.

Adopting a common tool

As requested by the Observatoire de Paris scientific board, a bibliometric indicators working group was formed. Its first recommendation was to look into indexing the Observatoire's publications. In 2012, after conducting a survey of researchers, the decision was made to adopt a new common tool, Publesia, to create POP: Publications by the Observatoire de Paris. POP was conceived with the following purposes:

- homogenising the collected data and allowing its management by all laboratories on the level of the institution;
- facilitating the identification and indexing of the institution's scientific editorial output;
- monitoring, publicising, and analysing the data on different levels — institution, laboratory, team, and author — and from different points of view — topic, project, and funding; and
- enhancing and promoting this scientific output by publicising it on different websites and via web services.

Publesia

Project POP: Publications by the Observatoire de Paris is founded on the use of Publesia. Publesia was chosen because the software was developed at LESIA by Florence Henry under CeCILL-B licence.⁴ It has been in production since 2006 and about a third of the Observatoire de Paris researchers have used it since then. Its initial aim was to take into account, as much as possible, the workflow of academics into its data input process. It has been built to interface with the ADS from the start, so as to automatically maximise automatic data gathering and to minimize manual data input by researchers.

Publesia is indexing software, not an open-access archive repository. For that reason, the Observatoire de Paris made the choice to utilize national infrastructure — specifically HAL — with which data in POP will be linked to in the future. Publesia is not a bibliometric analysis tool either, but bibliometric analyses can be conducted using the data collected using it.

Deployment phases

Project POP: Publications by the Observatoire de Paris is currently under deployment. In 2012, the software was adopted by ESPCI ParisTech, and since 2013, there has been shared development of the software. The Observatoire de Paris dealt more specifically with developments supposed to facilitate author and affiliations management. ESPCI Paris Tech worked on data import modules from ADS, Web of Science, and BibTeX file imports.

The first data imports started at the Observatoire de Paris in March 2014. Three subsequent data import phases have been planned. The 5.900 publications indexed by Daniel Egret between 2004 and 2012 were automatically imported into POP using scripts. After that, the 5.300 publications collected from Web of Science since 2007 will be imported into POP as well. The final data import will be of materials developed since

⁴CeCILL-B: http://www.cecill.info/licences/Licence_CeCILL-B_V1-en.html

2002 by the different Observatoire de Paris laboratories. At each step, data cleanup work will be done. Further development will enable academics to input their upcoming publications into POP.

3. Features of Publesia

Authors: the heart of the application

Each author in POP is in charge of uploading and updating their own publication list. Publesia includes automatic data import features from source such as ADS, Web of Science, HAL, or BibTeX files, which makes this task easier for academics. POP will eventually be able to automatically export publication lists into HAL (or serve as the data entry interface for HAL). Academics can manage all their publications in POP, including outputs not affiliated with the Observatoire de Paris. They can then export the lists in BibTeX or EndNote formats, or display them using a web service on their own website. The RTF file format was favoured for institutional reports, BibTeX and End-Note were generally used by authors, and the CSV format was favoured for analytical needs.

How are publications described in POP?

The following metadata fields are available in POP — title, author, affiliation(s), publication type, year, journal title, book title, conference data and place, publisher, scientific publisher, pagination, journal volume, journal number, external references, summary, university or specialty thesis, and thesis type. Publications are also assigned institution-specific metadata, which can include teams or laboratory sites; each lab is free to use the metadata fields that match their own priorities. Finally, each publication is given an AERES⁵-normalised category. Using these AERES-based categories for publication is helpful when answering funding bodies' requests, though these categories are not exhaustive and do not cover every category of research output. The Observatoire de Paris decided to index all publication activity on POP. The "Other Publications" category allows researchers to enter any type of material as long as it has been published or broadcast, including radio shows, software, databases, and other types of output. There is also the ability to assign keywords so that publications that are part of the same project or funded by the same entity can be grouped together.

Managing authors and affiliations

How are author names and affiliations managed when publications are entered into POP via the web interface? The first version of Publesia limited this type of management to LESIA members only. When considering exporting publications to HAL it was clear that all authors who are associated with a publication have to be referenced. One challenge is that there is currently no database that provides a unique identifier for each author or affiliation. Publesia cannot synthesise the different forms under which an affiliation or an author may appear without human intervention. A "de-duplication" mechanism was developed to address this issue. When something is published on POP

⁵AERES: Agence d'évaluation de la recherche et de l'enseignement supérieur - Evaluation agency for research and higher education <http://www.aeres-evaluation.fr/>

via the web interface, Publesia compares the character strings entered for authors and affiliations with those that are already in the database. If no matches can be found, a new author or a new affiliation is created. If there are potential matches, the submitter of the publication can choose among the various suggestions. In this case, an author or affiliation can be associated with an already-existing character string in POP. This can be complicated if the academic submitting a publication to POP lacks the information needed to disambiguate author strings, however.

Publesia's web service

Finally, Publesia uses a web service to publish to institutional or laboratory websites, or on an academic's personal webpage. All publication lists can be displayed as lists. The lists can be refined by author, year, sub-entity, publication type, or other facets. It is also possible to search by fields such as title, ADS bibcode, DOI, and more.

4. Conclusion: improvements and prospectives

Connection to HAL

Because French research and higher education institutions chose HAL as their national open-source archiving platform, POP: Publications by the Observatoire de Paris aims at serving as a data-input interface to the national open archiving platform, so that researchers don't have to input publications multiple times. There are automated data export tools that are compatible with the third version of HAL which was launched on Oct. 14, 2014,⁶ and which use HAL's new API,⁷ also under development.

Connection with external reference tools

The first import of data in March 2014, which was executed from a list of 5.900 ADS bibcodes, showed the need for single author and affiliation IDs. Indeed, since ADS does not assign a unique ID to authors or to affiliations, the automatic import scripts generated many variations of the same field value. For instance, more than 600 different forms have been stored in POP's affiliations table for LERMA,⁸ one of the Observatoire de Paris's laboratories, and these occurrences had to be cleaned up manually. If the ADS includes ORCID IDs,⁹ this may help alleviate these problems, though it will also require the cooperation of journal editorial boards. Affiliation information is still prob-

⁶See *See Mise en production de HAL v3.0 : la migration de vos données*:
<http://blog.ccsd.cnrs.fr/2014/10/mise-en-production-de-hal-v3-0-la-migration-de-vos-donnees/>
See also HAL 3 <https://hal.archives-ouvertes.fr/>

⁷See the documentation about the submitting APIs in HAL (SWORD or SOAP):
<http://api.archives-ouvertes.fr/docs>

⁸LERMA: Laboratoire d'études du rayonnement et de la matière en astrophysique - Laboratory of the study of radiation and matter in astrophysics <http://lerma.obspm.fr/>

⁹ORCID: Open Researcher and Contributor ID <http://orcid.org/>

lematic, particularly on an international level. French laboratories can still be identified in the HAL system, however.¹⁰

¹⁰See *See Référentiel Structures de recherche*, 2011, <http://api.archives-ouvertes.fr/docs/ref/resource/structure>