

Digitalización de los primeros volúmenes del BAAA

E.C. Fernández¹, P. Méndez Moura¹, C.J. Nusch^{1,2}, M.R. De Giusti^{1,2}, L. Calamante¹,
L. Folegatto¹, Y. Aidelman^{3,4}, L.S. Cidale^{3,4} & R. Gamén^{3,4}

¹ *Proyecto de Enlace de Bibliotecas y Servicio de Difusión de la Creación Intelectual, UNLP, Argentina*

² *Centro de Servicios en Gestión de Información, Comisión de Investigaciones Científicas de la Provincia de Buenos Aires, Argentina*

³ *Facultad de Ciencias Astronómicas y Geofísicas, UNLP, Argentina*

⁴ *Instituto de Astrofísica de La Plata, CONICET-UNLP, Argentina*

Contacto / carlosnusch@prebi.unlp.edu.ar

Resumen / El Boletín de la Asociación Argentina de Astronomía (BAAA) es el medio oficial para registrar y difundir la actividad científica de los astrónomos argentinos. Las publicaciones de estos boletines, desde el volumen 40, están disponibles en forma digital, pero el acceso a los números anteriores, es posible sólo a través de algunas bibliotecas de instituciones astronómicas. Dado que el acceso a los mismos es limitado (o imposible), nos hemos propuesto digitalizar dichas publicaciones para su preservación y difusión en el repositorio institucional de la Universidad Nacional de La Plata, mediante el Servicio de Difusión de la Creación Intelectual. En este trabajo, describimos el procedimiento de digitalización y las características específicas de los productos generados.

Abstract / The Bulletin of the Argentine Astronomy Association (BAAA) is the official publication to record and disseminate the scientific activity of Argentinian astronomers. These bulletins, from volume 40, are available in digital form, but access to the previous numbers is possible only through some astronomy libraries. Since access to these documents is limited (or impossible), we have set out to digitalize these publications for their preservation and dissemination in the institutional repository of the National University of La Plata, through the Intellectual Creation Dissemination Service. In this work, we describe the digitization procedure and the specific characteristics of the products generated.

Keywords / publications, bibliography — astronomical databases: miscellaneous

1. Introducción

El Boletín de la Asociación Argentina de Astronomía (BAAA) es el medio oficial de la Asociación Argentina de Astronomía (AAA) abocado a registrar y difundir la actividad científica de los astrónomos argentinos. En sus volúmenes hay contribuciones de diversas áreas de trabajo, desde desarrollos instrumentales hasta trabajos de cosmología con estrellas RR Lyrae, pasando por los trabajos de búsqueda de sitio llevados a cabo en la década del 60 del siglo pasado (Muñoz & Cabrera, 1961, 1963) y la contribución sobre la distribución de brillo de una galaxia (alias “perfil de Sérsic”, Sérsic, 1963).

Lamentablemente, los artículos publicados en los boletines de la AAA están disponibles en forma digital sólo a partir del volumen 40 (año 1996). El acceso a los números anteriores era posible sólo a través de algunas bibliotecas de instituciones astronómicas. Algunos pocos artículos, publicados en un BAAA anterior al año 1996, se encuentran *on-line* gracias a la buena voluntad de quien se encargó de escanearlo y ponerlo disponible en Internet. Habida cuenta de que la consulta presencial presenta enormes limitaciones, se procedió a digitalizar dichas publicaciones para su preservación y difusión.

Este trabajo se realizó en el contexto del proyecto recientemente creado, en la Facultad de Ciencias Astronómicas y Geofísicas (FCAG) de la Universidad Nacional de La Plata (UNLP), para la REcuperación del

TRabajo Observacional Histórico (ReTrOH*). Este importante acervo patrimonial consta de registros meteorológicos, sísmicos, geomagnéticos y astronómicos que datan de principios del siglo XX. A través del proyecto ReTrOH se propone comenzar con el trabajo de recuperación y puesta en valor de todo dicho patrimonio.

A continuación brindamos detalles sobre el trabajo de digitalización y publicación de los boletines en el Repositorio Institucional de la UNLP, mediante el Servicio de Difusión de la Creación Intelectual (SeDiCI**). El cual es un servicio libre y gratuito creado para albergar, preservar y dar visibilidad a las producciones de las unidades académicas de la Universidad (De Giusti et al., 2018; Nusch et al., 2018a,b, 2019).

2. El proceso de digitalización

2.1. Estado de conservación

El proceso de digitalización depende de diversos factores entre los cuales se pueden destacar el estado de conservación del papel, el tipo de encuadernación y el formato de impresión. Por lo tanto es preciso, en primera instancia, hacer una inspección del material.

El estado general de los volúmenes, aún de los más

*<http://retroh.fcaglp.unlp.edu.ar>

**<http://sedici.unlp.edu.ar>

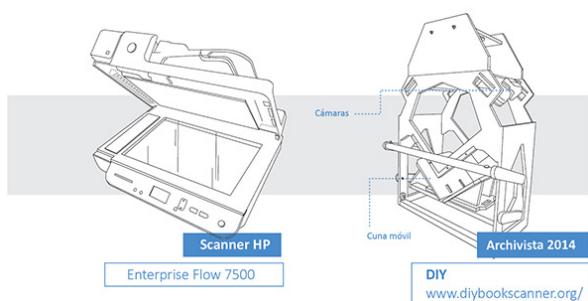


Figura 1: Vista gráfica de los escáneres HP Scanjet Enterprise Flow 7500 (izquierda) y de libros DIY Archivist 2014 (derecha). Dibujos realizados por Lucas Folegotto.

viejos era bueno. Algunas hojas se encontraban amarrotadas pero sin manchas ni desprendimientos. Algunos ejemplares habían sido mecanografiados y otros presentaban encuadernaciones ajustadas que no permitían la apertura correcta del boletín. Todos estos factores fueron determinantes para la elección de los escáneres a utilizar.

2.2. Captura

La captura de las imágenes se realizó con un escáner plano HP Scanjet Enterprise Flow 7500 y un escáner de libros DIY Archivist 2014. Este último se utilizó en los volúmenes que presentaban encuadernaciones ajustadas, con el objetivo de asegurar la preservación física de los ejemplares que por cuestiones de la encuadernación mencionada no podían tener un ángulo de apertura de 180° (ver Fig. 1).

3. Edición

3.1. Manipulación y ajustes posteriores

Una vez realizada la captura de las imágenes es necesario realizar un proceso de manipulación y edición de los archivos.

Esta tarea se llevó a cabo con el software SCANTAILOR ADVANCED el cual permite aplicar las mejoras necesarias para una óptima visualización y prepara los archivos para el proceso de reconocimiento de texto. Este software fue elegido por sus prestaciones y por ser libre. Entre las acciones que fueron realizadas para cada captura podemos mencionar la identificación de páginas (división de páginas, detección del contenido y sus márgenes), el enderezado de las mismas y ajuste de color y contrastes. Todo este proceso se ilustra esquemáticamente en la Fig. 2.

3.2. Reconocimiento óptico de caracteres

Una vez editadas las capturas el proceso de digitalización termina convirtiendo la imagen a un formato adecuado para su visualización en dispositivos electrónicos. Para ello se empleó el software de reconocimiento óptico de caracteres (OCR, por sus siglas en inglés) ABBY FINEREADER. El cual permite convertir documentos es-

caneados, imágenes, PDF y fotografías digitales en archivos electrónicos compatibles con la búsqueda en la *web*. Su elección, en este caso, se debe a que a pesar de ser un software pago presenta ventajas tales como:

- Detección de tablas, textos, imágenes y los valores contenidos en éstas.
- Reconocimiento de la estructura del texto, títulos, encabezados y pies de página.
- Reconocimiento de tipografías.
- Guardado de archivos en formato PDF/A^{***}, que garantiza la preservación y correcta visualización a largo plazo de documentos electrónicos.
- Permite que el usuario pueda enseñarle a reconocer caracteres considerados ambiguos y aplicarlo desde allí en adelante.

Para ilustrar la potencialidad que tiene este proceso de escaneo y edición de los boletines, para el reconocimiento de tablas impresas, utilizamos el trabajo de Sisteró (1965). Allí se presenta fotometría visual del sistema tipo Algol S Velorum (CPD -44 3864). Mediante la simple acción de “copiar-pegar” en el archivo PDF/A se pudieron obtener los datos observacionales publicados, i.e. día juliano, fase y magnitud, y graficarlos con la efemérides actual del sistema, como se muestra en la Fig. 3.

4. Preservación y acceso a los contenidos

4.1. Preservación

Se guardan tres archivos, un archivo TIFF sin comprimir, un archivo PDF/A sin compresión y otro archivo PDF/A con compresión. Los primeros dos tipos de archivos de alta calidad de imagen están destinados a la preservación digital y el tercero de ellos, de formato comprimido, se utiliza para la difusión en el repositorio. El último archivo puede implicar la pérdida en la calidad del dato, a fin de lograr archivos de menor tamaño, más amigables para el usuario final y para la indexación por parte de *crawlers*. Detalles sobre esto en Nusch et al. (2019).

4.2. Acceso a las publicaciones electrónicas

Como se dijo, los boletines digitalizados están almacenados y disponibles, en su totalidad, a través del Repositorio Institucional de la UNLP, mediante el SeDiCI^{****}, de acceso abierto y bajo Licencia Creative Commons Atribución-NoComercial-CompartirIgual 2.5 Argentina (CC BY-NC-SA 4.0).

Actualmente, estamos trabajando para la incorporación de todos estos trabajos en el SAO/NASA

^{***}El PDF/A es un formato de archivo para el almacenamiento a largo plazo de documentos electrónicos. Está basado en la Versión de Referencia 1.4 de PDF de Adobe Systems Inc. (implementada en Adobe Acrobat 5 y versiones posteriores) y está definido por la ISO 19005-1:2005. Este estándar intenta asegurar que los documentos digitales puedan ser reproducidos con exactitud en el futuro.

^{****}<http://sedici.unlp.edu.ar/handle/10915/72927>



Figura 2: El software SCANTAILOR ADVANCED permite la manipulación de las imágenes obtenidas. Dibujos realizados por Lucas Folegatto.

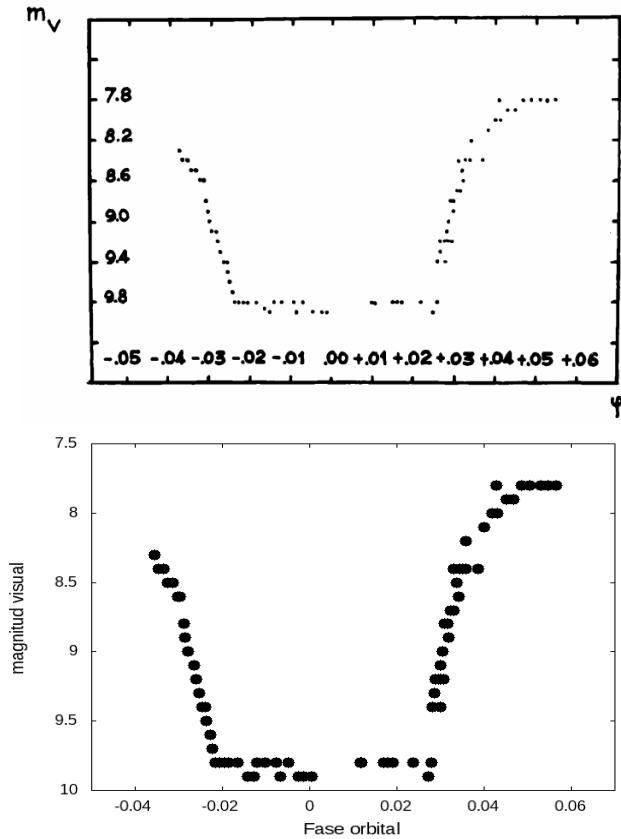


Figura 3: Arriba: Fotometría visual de S Vel publicada en Sisteró (1965). Abajo: Mismos datos recuperados del archivo PDF/A y graficados según efemérides actuales, i.e. $T_0=2452504.0250+5.9336475E$ d.

Astrophysics Data System (ADS). Si bien algunos boletines ya están incorporados (especialmente desde el volumen 40), muchos otros tienen sus respectivas entradas incompletas (falta de resumen y/o títulos, autores, etc.) y fundamentalmente no tienen el acceso al trabajo correspondiente. Es oportuno señalar que la entrada al trabajo de Sisteró mencionado, no existe en el ADS.

5. Conclusiones

La tarea realizada por el equipo de trabajo ha contribuido no solo a la conservación de la totalidad de los volúmenes publicados por la Asociación Argentina de Astronomía desde sus inicios, sino también su difusión y acceso libre por parte de todos los investigadores, docentes u otro tipo de usuarios interesados en el tema.

El estándar utilizado para su preservación digital, el formato PDF/A, permite el copiado y la transformación del texto por el usuario final y la extracción e indexación por parte de *crawlers* y repositorios lo que optimiza y multiplica las vías de acceso a los materiales sin ninguna barrera económica o digital.

Cabe mencionar que así como el trabajo de Sisteró (1965) hay muchas otras publicaciones que a partir de este trabajo de digitalización se encuentran disponibles para la comunidad astronómica internacional. La posibilidad de disponer de información histórica en formato digital permite recuperar datos de varias décadas atrás, para analizarlos con herramientas modernas y contrastarlos con los actuales. Por todo esto, creemos que la digitalización de los BAAA potencia y vuelve a poner en valor el trabajo de los primeros astrónomos argentinos.

Agradecimientos: Agradecemos a las autoridades de la FCAG por confiarnos este proyecto.

Referencias

De Giusti M.R., Pinto A.V., Nusch C., 2018, *Archivos Audiovisuales Universitarios: actualidad y desafíos*
 Muñoz F.F., Cabrera L.A., 1961, BAAA, 4, 61
 Muñoz F.F., Cabrera L.A., 1963, BAAA, 6, 44
 Nusch C., et al., 2018a, *XXVIII Asamblea General del ISTE-TEC 2018 - Congreso Internacional "La influencia de la tecnología en las comunidades del conocimiento"*
 Nusch C., et al., 2018b, *VIII Conferencia Internacional sobre Bibliotecas y Repositorios Digitales BIREDIAL-ISTEC*
 Nusch C., et al., 2019, *IX Conferencia Internacional sobre Bibliotecas y Repositorios Digitales de América Latina*
 Sércis J.L., 1963, BAAA, 6, 41
 Sisteró R.F., 1965, BAAA, 10, 81