Astrophysics and Space Science Proceedings

Alberto Accomazzi *Editor*

Future Professional Communication in Astronomy II

Springer

Astrophysics and Space Science Proceedings

For further volumes:
http://www.springer.com/series/7395

Future Professional Communication in Astronomy II

Editor

Alberto Accomazzi

Harvard-Smithsonian Center for Astrophysics, Cambridge, USA

Springer

Editor
Alberto Accomazzi
Harvard-Smithsonian Center for Astrophysics
60 Garden St.
Cambridge, MA 02138
USA
aaccomazzi@cfa.harvard.edu

ISSN 1570-6591 e-ISSN 1570-6605
ISBN 978-1-4419-8368-8 e-ISBN 978-1-4419-8369-5
DOI 10.1007/978-1-4419-8369-5
Springer New York Dordrecht Heidelberg London

Library of Congress Control Number: 2011926010

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

In memory of our friend and colleague John P. Huchra

Editorial

This volume contains the papers presented at the Second Colloquium on *Future Professional Communication in Astronomy*, which was held at the Harvard-Smithsonian Center for Astrophysics, in Cambridge, Massachusetts (USA) on April 13–14, 2010. This meeting provided an open forum to discuss the state and evolution of professional communicating and in particular publishing in astronomy.

The meeting was attended by 44 people representing nine different countries, including representatives from the major astronomical publishers and learned societies, editors, librarians, scientists, as well as archive and information system managers.

The program included 19 lively talks covering a wide range of topics, including: the state of the art in publishing and the role of learned societies, the future of scholarly communication in the era of the web, communicating astronomy to the public, the role of journals and archives in data curation and preservation, the future of librarianship, bibliometrics and other evaluation criteria, and scholarly recommendation systems.

The keynote speaker of the Colloquium was John Huchra, the Robert O. & Holly Thomis Doyle Professor of Cosmology and the Senior Advisor to the Provost for Research Policy at Harvard University, who spoke about "Astronomical Publishing: Yesterday, Today and Tomorrow." John revisited the role of publishing in astronomy and asked the question "Why do we publish," setting the stage for a lively debate during the remainder of the conference.

Publishers and officers of learned societies described their current efforts and future plans to enrich the electronic scholarly papers, respond to open access mandates, control costs, and collaborate with external archives to provide valuable services to the astronomical community.

Scientists, librarians and archivists voiced their thoughts and desires on wide-ranging topics including open access models, the evolving role of libraries and services in astronomy, and the experience of the 2009 International Year of Astronomy.

Both this colloquium and its precursor, FPCA-I, were the brainchildren of Prof. André Heck, who, as the prolific author of over 1,500 papers, knows a thing or two about scholarly communication. Over the past two decades, André has authored and edited several seminal books about the sociology of organizations and collaborations in astronomy and related fields. This colloquium and book would not have been possible without his intellectual and material contribution, for which we are all very grateful.

We are grateful to the following sponsors who provided much needed material support for the meeting: the Smithsonian Astrophysical Observatory, the American Astronomical Society, EDP Sciences, Wiley-Blackwell, IOP Publishing, Springer, and Elsevier.

The Local Organizing committee worked tirelessly to make sure the Colloquium would be a success. LOC members were Alberto Accomazzi, Giovanni Dimilia, Paul Grant, Michele Hall, Edwin Henneken, Wendy Roberts, Donna Thompson and Su Tuttle.

Cambridge *A. Accomazzi*
April 2011

Contents

Additional Colloquium Talks

Four of the talks given at the colloquium did not produce contributions included in this book. A brief overview of the talks is given below.

Astronomy Journal Publishing: Serving the Community

Ray Boucher, ***Wiley-Blackwell***

Wiley-Blackwell publishes journals of behalf of more than 750 publishing partners which include the journals of the Royal Astronomical Society. The talk outlined how we interact with our respective communities: authors, libraries, readers and societies. Current trends and future opportunities were also discussed.

Publishing in Greater Depth: Article of the Future and Beyond

David Clark, ***Elsevier Publishers***

With Elsevier's Article of the Future, we have begun the step of changing how the research article is presented. Elsevier are focusing both on how articles are prepared and how the material can be developed, including greater support for data formats and mathematics, to deeper interlinking with the rest of the literature. We are also looking at how we can better work with others to make scientific articles, and their derivatives, more discoverable and usable.

The Present and Future of the Astrophysical Journal

Ethan Vishniac, ***McMaster University***

In the last decade the form and organization of The Astrophysical Journal have changed significantly in response to changes in the expectations and needs of the astronomical community. These include the gradual loss of interest in the printed version of the journal, the need for successful publication of more complicated data products, the increased demand for transparency in data processing, and the steadily increasing size of the journal. The talk discussed some of the more significant changes and their success, the major challenges, and the prospects for continued change over the next decade.

The SCOAP3 Open Access Publishing Model

Salvatore Mele, ***CERN***

The High-Energy Physics community has spearheaded Open Access with over half a century of dissemination of pre-prints culminating in the arXiv system and is now pushing forward with an Open Access model which goes beyond the present, often controversial, proposals with a novel practical approach: SCOAP3. In the SCOAP3 model, libraries federate to explicitly cover the costs of peer-review rather than implicitly supporting it via journal subscriptions. The talk described the history of innovations in Open Access in High-Energy Physics, the advantages they have generated as a context for the SCOAP3 model, as well as the status of this initiative, in the wider context of a user-driven evolution of scholarly communication in the field.

List of Participants

Abt, Helmut
NOAO
Box 26732
Tucson, AZ 85726-6732
USA

Accomazzi, Alberto
SAO MS-83
60 Garden Street
Cambridge, MA 02138
USA

Bailer-Jones, Coryn
Max Planck Institute for Astronomy
Königstuhl 17
D-69117 Heidelberg
Germany

Balli, Enrico Maria
Sissa Medialab
Via Stock 2/2
I-34135 Trieste
Italy

Barnes, Jonathan
Astronomical Society of the Pacific
Conference Series
1201 E. 8725 S.
Sandy, UT 84094
USA

Biemesderfer, Chris
American Astronomical Society
2000 Florida Avenue NW
Washington, DC 20016
USA

Blake, Michael
Harvard-Smithsonian Center
for Astrophysics
MS-56, P-107
60 Garden Street
Cambridge, MA 02138
USA

Blaylock, Joseph
SLAC National Accelerator
Laboratory
Mail Stop 82
2575 Sand Hill Road
Menlo Park, CA 94025
USA

Bohlen, Elizabeth
SAO MS-83
60 Garden Street
Cambridge, MA 02138
USA

Boucher, Raymond
Wiley-Blackwell
11 Marlborough Close
Chichester, PO19 7XW
England

Brooks, Travis
SLAC National Accelerator Laboratory
2575 Sand Hill Road
Mail Stop 82
Menlo Park, CA 94402
USA

Chaix, Marie-Louise
EDP Sciences
17, avenue du Hoggar
F-91940 Les Ulis
France

Clark, David
Elsevier
Radarweg 29
NL-1043 NX Amsterdam
Netherlands

Di Milia, Giovanni
SAO MS-83
60 Garden Street
Cambridge, MA 02138
USA

Eichhorn, Guenther
Springer
233 Spring Street
New York, NY 10013
USA

Fienberg, Richard Tresch
American Astronomical Society
2000 Florida Avenue NW
Suite 400
Washington, DC 20009
USA

Grant, Carolyn
SAO MS-83
60 Garden Street
Cambridge, MA 02138
USA

Graves, William
Harvard-Smithsonian Center for Astrophysics
John G. Wolbach Library
60 Garden Street
Cambridge, MA 02138
USA

Grothkopf, Uta
European Southern Observatory
Karl-Schwarzschild-Straße 2
D-85748 Garching
Germany

Heck, André
Strasbourg Astronomical Observatory
11, rue de l'Université
F-67000 Strasbourg
France

Henneken, Edwin
SAO MS-83
60 Garden Street
Cambridge, MA 02138
USA

Holtkamp, Annette
CERN
Route de Meyrin
CH-1211 Genève 23
Switzerland

Huchra, John
Center for Astrophysics
60 Garden Street
Cambridge, MA 02138
USA

Jensen, Joseph
Astronomical Society of the Pacific Conference Series
MS 179 Utah Valley University
800 W. University Parkway
Orem, UT 84058-5999
USA

Jorda, Jean-Paul
EDP Sciences
17, avenue du Hoggar
F-91944 Les Ulis
France

Kroffe, Kerry
IOP Publishing
1325 G Street NW
Suite 1010
Washington, DC 20005
USA

Kurtz, Michael
SAO MS-19
60 Garden Street
Cambridge, MA 02138
USA

Lagerstrom, Jill
Space Telescope Science Institute
3700 San Martin Drive
Baltimore, MD 21218
USA

Longden-Chapman, Elaine
IOP Publishing
Dirac House
Temple Back
Bristol, BS1 6BE
England

Luker, James
SAO MS-83
60 Garden Street
Cambridge, MA 02138
USA

Mahoney, Terence
Instituto de Astrofísica de Canarias
C/ Va Láctea s/n
E-38200 La Laguna
Tenerife
Spain

Mele, Salvatore
CERN
385 Route de Meyrin
CH-1211 Genève 23
Switzerland

Muench, August
SAO MS-42
60 Garden Street
Cambridge, MA 02138
USA

Muñoz Franco, Lucia
Elsevier
Radarweg 29
NL-1043 NX Amsterdam
Netherlands

Rots, Arnold
SAO MS-67
60 Garden Street
Cambridge, MA 02138
USA

Sachs, Kirsten
DESY
Böttcherkamp 1
D-22549 Hamburg
Germany

Schmelz, Lynne
Harvard University
Cabot Science Library
1 Oxford Street
Cambridge, MA 02138
USA

Simko, Tibor
CERN
Route de Meyrin
CH-1211 Genève
Switzerland

Smith, Tim
CERN
Route de Meyrin
CH-1211 Geneva
Switzerland

Thiell, Benoit
SAO MS-00
60 Garden Street
Cambridge, MA 02138
USA

Thompson, Donna
SAO MS-83
60 Garden Street
Cambridge, MA 02138
USA

Vigen, Jens
CERN
385 Route de Meyrin
CH-1211 Genèva
Switzerland

Vishniac, Ethan
McMaster University/The Astrophysical Journal
Department of Physics and Astronomy
Hamilton, ON L9G 4M1
Canada

Winkelman, Sherry
SAO MS-67
60 Garden Street
Cambridge, MA 02138
USA


PHILLIPS
AUDITORIUM

List of Contributors

Helmut A. Abt
Kitt Peak National Observatory
National Optical Astronomy
Observatory, P.O. Box 26732, Tucson
AZ 85726-6732, USA
abt@noao.edu

Alberto Accomazzi
Harvard-Smithsonian Center
for Astrophysics, 60 Garden Street
Cambridge, MA 02138, USA
aaccomazzi@cfa.harvard.edu

Jonathan Barnes
Salt Lake Community College
4600 South Redwood Road
Salt Lake City, UT, 84123, USA
Jonathan.Barnes@slcc.edu

Chris Biemesderfer
American Astronomical Society
2000 Florida Avenue NW, Suite 400
Washington DC 20009, USA
chris.biemesderfer@aas.org

Elizabeth Bohlen
Harvard-Smithsonian Center
for Astrophysics, 60 Garden Street
Cambridge, MA 02138, USA
ebohlen@cfa.harvard.edu

Marie-Louise Chaix
EDP Sciences, 17, avenue du Hoggar
B.P. 112, P.A. de Courtaboeuf
F-91944 Les Ulis, France
marie-louise.chaix@edpsciences.org

Giovanni Di Milia
Harvard-Smithsonian Center
for Astrophysics, 60 Garden Street
Cambridge, MA 02138, USA
gdimilia@cfa.harvard.edu

Guenther Eichhorn
Springer, 233 Spring Street
New York, NY 10013, USA
yguenther.eichhorn@springer.com

Richard Tresch Fienberg
American Astronomical Society
2000 Florida Avenue NW, Suite 400
Washington DC 20009, USA
rick.fienberg@aas.org

Carolyn Grant
Harvard-Smithsonian Center
for Astrophysics, 60 Garden Street
Cambridge MA 02138, USA
cgrant@cfa.harvard.edu

Uta Grothkopf
European Southern Observatory
Karl-Schwarzschild-Straße 2
D-85748 Garching, Germany
esolib@eso.org

André Heck
Strasbourg Astronomical Observatory, 11, rue de l'Université
F-67000 Strasbourg, France
andre.heck@astro.unistra.fr

Edwin A. Henneken
Harvard-Smithsonian Center for Astrophysics, 60 Garden Street
Cambridge, MA 02138, USA
ehenneken@cfa.harvard.edu

John Huchra
Harvard-Smithsonian Center for Astrophysics, 60 Garden Street
Cambridge, MA 02138, USA
huchra@cfa.harvard.edu

Joseph B. Jensen
Utah Valley University
800 W. University Parkway
Orem, UT, 84058-5999, USA
jjensen@uvu.edu

Michael J. Kurtz
Harvard-Smithsonian Center for Astrophysics, 60 Garden Street
Cambridge, MA 02138, USA
kurtz@cfa.harvard.edu

Jill Lagerstrom
Space Telescope Science Institute
3700 San Martin Drive, Baltimore
MD 21218, USA
lagerstrom@stsci.edu

Jay Luker
Harvard-Smithsonian Center for Astrophysics, 60 Garden Street
Cambridge, MA 02138, USA
jluker@cfa.harvard.edu

Terence J. Mahoney
Instituto de Astrofísica de Canarias
C/ Vía Láctea s/n
E-38200 La Laguna
Tenerife, Spain
tjm@iac.es

Kevin B. Marvel
American Astronomical Society
2000 Florida Avenue NW, Suite 400
Washington DC 20009, USA
kevin.marvel@aas.org

J. Ward Moody
Brigham Young University, N484
ESC, Provo, UT, 84602, USA
jmoody@byu.edu

Stephen S. Murray
Harvard-Smithsonian Center for Astrophysics, 60 Garden Street
Cambridge MA 02138, USA
smurray@cfa.harvard.edu

Paula Szkody
University of Washington
Box 351580, Seattle, WA
98195-1580, USA
szkody@astro.washington.edu

Donna Thompson
Harvard-Smithsonian Center for Astrophysics, 60 Garden Street
Cambridge, MA 02138, USA
dthompson@cfa.harvard.edu

1

FPCA-II Opening Comments

André Heck

Strasbourg Astronomical Observatory, 11, rue de l'Université,
F-67000 Strasbourg, France
andre.heck@astro.unistra.fr

Summary. This introductory talk reflects on history, sets up pieces of context, and puts forward a few issues to be hopefully addressed during the meeting.

1 Introduction

The first FPCA meeting took place 3 years ago at the Royal Academy in Brussels where it was organized in collaboration with Léo Houziaux. It had been motivated by a convergence of facts and trends in the world of publishing and more generally in the realm of astronomy professional communication. There were also comments and interrogations heard more and more frequently from within the astronomy community that could have been summarized by one question: *Where are we heading to in terms of publishing our results and communicating what we are doing?*

The Brussels FPCA-I colloquium has been historical in the sense that, for the first time in astronomy, virtually all major players in the field were gathered together and talked to each other: publishers, editors, archive managers, officers of learned societies, as well as scientists and librarians involved and/or concerned by the evolution of the professional communication processes.

2 History

Of course, this was not the first time such matters were tackled. For instance, the series of volumes entitled *Organizations and Strategies in Astronomy (OSA)* (Heck 2000–2006) included a number of chapters dealing with astronomy communication in the broad sense. There were also specific books on information handling and communication in astronomy (Heck 2000, 2003; Heck & Madsen 2003). Quite a few FPCA-I attendees had contributed to those masterpieces. There were also dedicated meetings, such as the 1996 colloquium on *Strategies and Techniques of Information in Astronomy* (Heck & Murtagh 1996).

A. Accomazzi (ed.), *Future Professional Communication in Astronomy II*,
Astrophysics and Space Science Proceedings 1, DOI 10.1007/978-1-4419-8369-5_1,

As to electronic publishing itself, the story had started earlier with the first international colloquium on the theme[1] held about 20 years ago in Strasbourg (Heck 1992). Other reviews on electronic publishing were produced subsequently (see e.g. Heck 1997 and chapters in the OSA series).

At FPCA-I in Brussels, many interesting issues and technicalities were discussed – then on a background of renewal of contracts between learned societies and publishers. The proceedings (Heck & Houziaux 2007[2]) include also the summaries of an Editor's Forum moderated by Helmut Abt and of a Publisher's Forum moderated by Terry Mahoney, as well as *Notes from the Meeting* by Mike A'Hearn who was then Chairman of the AAS Publications Board.

From these, I feel appropriate to echo here a point often forgotten: *There was widespread agreement, after much discussion, that the biggest "cost" of publishing is in the time of the scientists who write the papers and the time of the scientists who referee the papers. For a variety of reasons, these "costs" are never accounted for in the "cost of publishing".* I shall come back to this later on.

3 Why Another Meeting?

While FPCA-I could be considered a success, I was still in the dark regarding a number of basic questions I had beforehand – and that have not been cleared up since. This is partially motivating, 3 years later, this second meeting, most kindly organized by Alberto Accomazzi and his team.

Let me first point out that professional communication is much broader than just publishing. Sharing of knowledge encompasses also lectures, courses, demonstrations, press conferences, for instance. Astronomy professional communication can be schematized as in Figs. 1 and 2.

The main pending issues coming to my mind are the following ones:

- Again, where are we heading to in terms of publishing? Were the business models presented at FPCA-I applied? Were they successful? It seems that the main purpose of our commercial partners is quite legitimately to make more money – while giving us still more work though. I'll tell hereafter a little story about this.
- Open Access models failed to convince me so far, perhaps because OA is a label used to cover quite different things. For instance, a couple of weeks ago, I was demonstrated a system boasted as OA while it was in fact an institutional bibliographical system with formatting capabilities. Openness was left to individual authors in charge of securing authorization for accessing each of their papers if they wanted to make them visible.

[1] The buzzword shifted from *desktop publishing* to *electronic publishing* between the time the meeting was launched and when it took place.

[2] Available online at http://astro.u-strasbg.fr/~heck/fpca_toc.htm

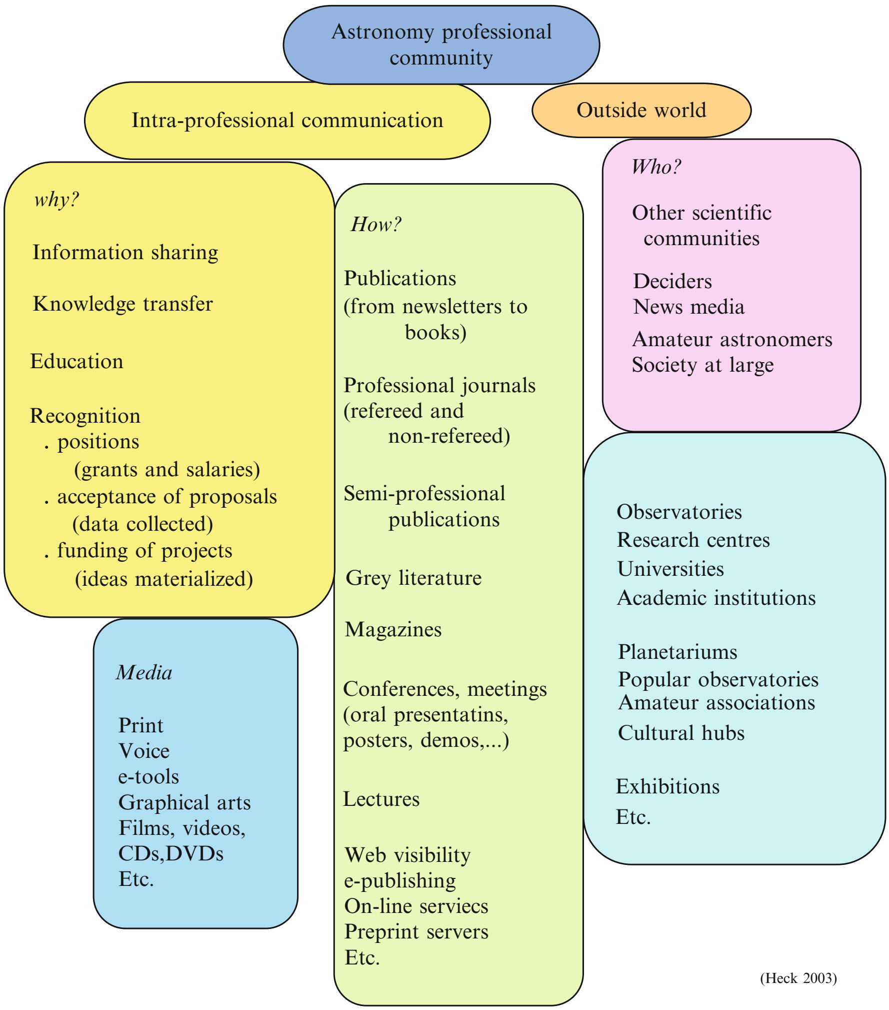


Fig. 1. The astronomy-related communication process (Adapted from Heck 2003).

- At the time of FPCA-I, librarians were in an adaptive phase and we will certainly hear at this meeting what happened meanwhile.
- As to communication in the broad sense, was IYA2009 actually a success? Some voices claim discreetly that it came short of the expectations it raised, one of the explanations put forward being that, over the recent years, the public and the press grew weary of sensationalistic press releases about everything and nothing. Actually the concomitant Darwin Year 2009 did not fare better.
- I must confess that I have seen too many instances, including during IYA2009, where communicating astronomy with the public did not go

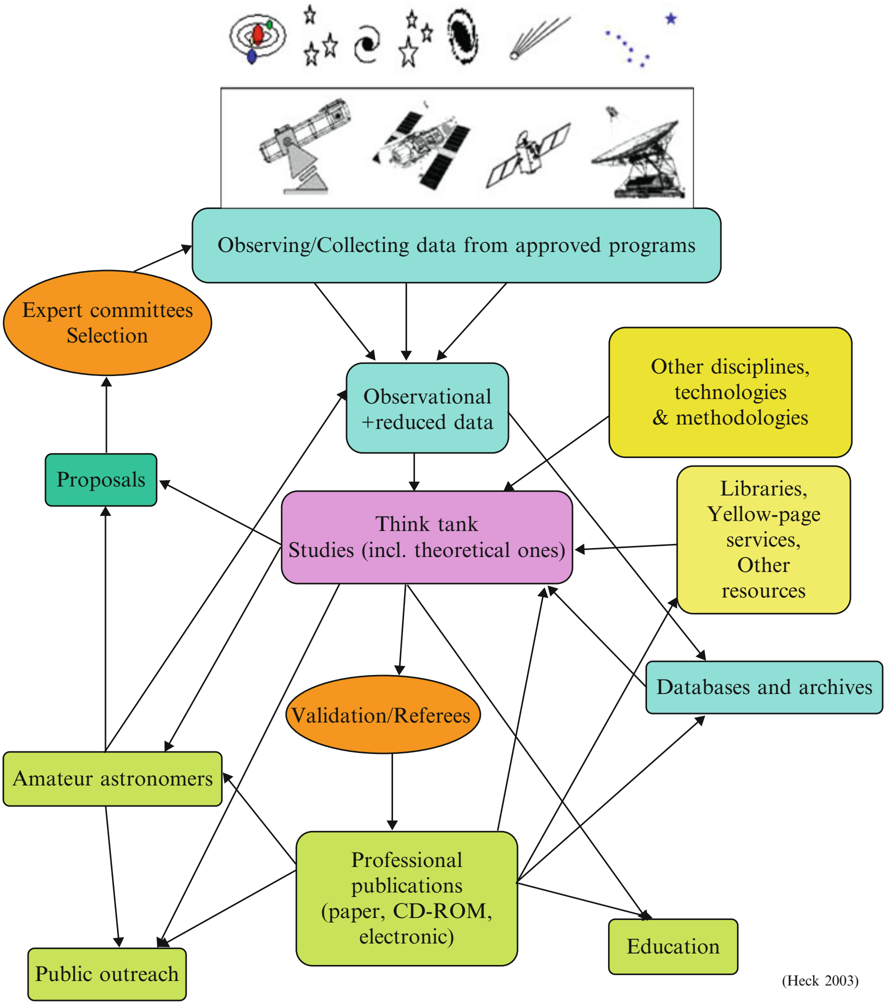


Fig. 2. A schematic view of the astronomy information flow (From Heck 2000).

beyond romantic gaping at the skies and extasis in front of beautiful images. This is a real issue as we have too many prospective students coming to us with a wrong idea of the day-to-day research activities. They are disappointed and leave.

Hence the need of educating students in the way research is actually carried out, and properly carried out. This is why I was very happy to see

AAS President John Huchra tackling those issues in his column in the *AAS Newsletters*.[3] Please do have also a look at the AAS Statement on Professional Ethics.[4]

Ethics can still be a real issue in our scientific publications. Let me tell a recent personal experience. Earlier this year, I had a paper submitted to a journal that looked quite normal, with an editor-in-chief and co-editors covering three continents. My paper went through two referees who did their job. They were right in asking me to clarify a couple of points and, as it is often the case, I had also to answer a couple of lousy comments. But my paper was easily accepted. This implied that its integrity should have been preserved from then on, except perhaps for a bit of language cosmetics to improve my English – something quite ok with me.

When getting the proofs however, I realized that some references had been added, references to papers not really relevant to my own article. Were they authored by friends of the editor-in-chief? I don't know. In any case, I withdrew my paper since, as a matter of principle, I am not publishing in journals where such practices are taking place. So ethics can still be an issue in this twenty-first century, and perhaps even more so because of the flexibility and potentially easy alteration of the electronic material.

This incident led me to investigate something else. What is the situation regarding ERA archives, i.e. archives of discussions between editors, referees and authors? I ran a quick survey. Not everybody answered, but I got a feedback from the main journals (Fig. 3). Obviously the situation could still be improved and streamlined.

4 Alsatian Maids in Paris

Since most of the talks at this meeting will be about papers in journals, let me say a few words about books – edited books or monographs.

One of my research projects dives into local history in Strasbourg. This implies decyphering old documents for which I had to follow courses of German paleography, a place where one can meet people from different backgrounds. One of these was a retired historian, Jean Haubenestel, who wrote a book telling the story of Alsatian maids in Paris (Fig. 4).

Who could be interested in the story of Alsatian maids in Paris? This is exactly what he was told by the publishers he approached and who turned him down. You might have experienced similar reactions with some of your projects. But this gentleman decided to go ahead on his own. He got a good printer-binder, advertized the book on his web site, left the usual 30–40% margin to book distributors, got things straight with the income tax office (where some 10% on earnings have to be left) and ... he is making money.

[3] See, e.g., *AAS Newsl.* **146** (May/June 2009) and **148** (September/October 2009).
[4] Cf. *AAS Newsl.* **151** (March/April 2010) p. 8–9.

ERA archives

(exchanges between Editors-Referees-Authors)

- A&A (Bertout/Quilbé)
  retained "some time" within MMS (Quilbé)
  archives in storage/Paris since 1995 (Bertout)
  willing to "do something" (Bertout)

- ApJ (Abt/Vishniac)
  correspondence under Morgan (1948-1953) & under Chandra (1953-1971) archived at U of Chicago
  correspondence under Abt (1970-1990) in storage lockers in Tucson (2M sheets of paper)
  later on: kept in database and accessible after 75y

- MNRAS (Carswell/Clube)
  "a lot" in S1M (no start date given – 2005?)
  (but some discussions outside S1M)

- PASP (Szkody)
  history file system since 2005

Fig. 3. ERA archives (between editors, referees and authors).

Just like for our specialized books in astronomy, his commercial publishers would have been quite happy to sell a few hundred copies on which he would have got peanuts in terms of royalties. Guess how much Haubenestel is selling? He is currently reaching 4,000 copies and he is going to print another thousand. For a book on Alsatian maids in Paris. And there is no blue literature inside, even if the book cover is a (possibly involuntary) masterpiece of subliminal messages.

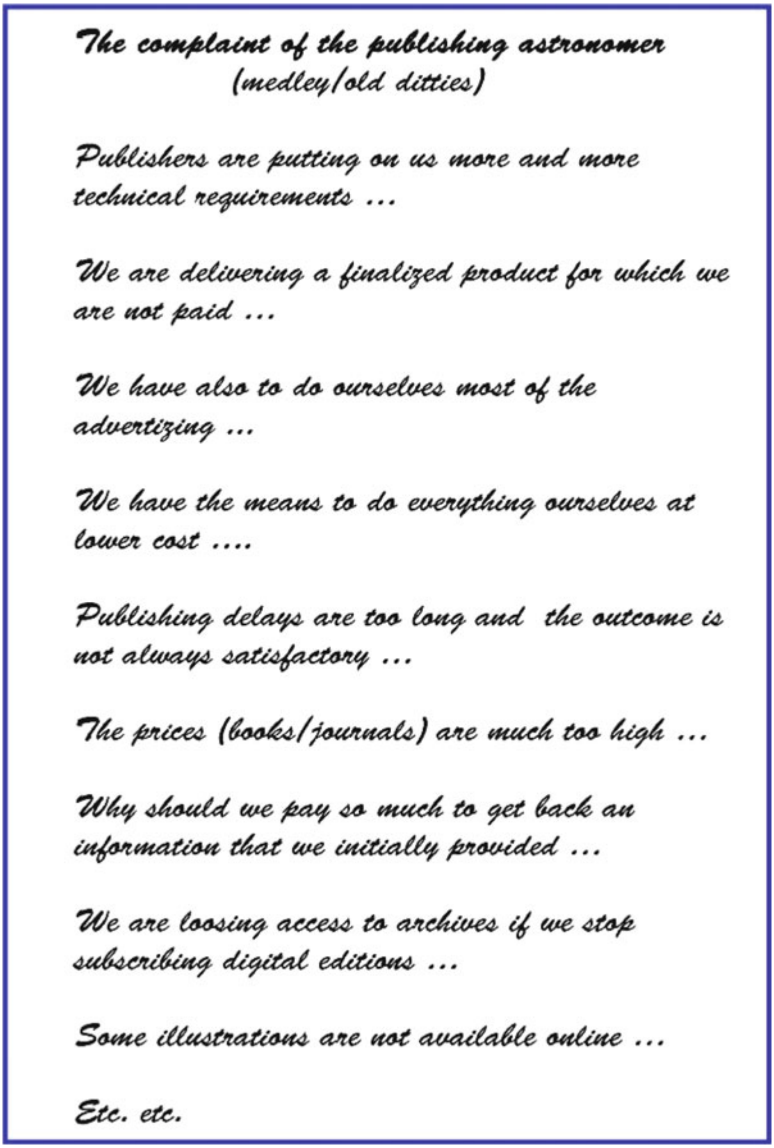
In a similar approach, I had asked one of my publishers what would be his conditions for producing a book I have currently in the making. In line with what many publishers are doing now (i.e. taking no risks and requesting a financial participation from the authors or from their institutions), he favored

Fig. 4. Alsatian maids in Paris: a very successful example of classical self-publishing (Reproduced with permission).

an electronic publication in exchange of few thousands Euros/Dollars. In the discussion, he claimed that another book of mine did not sell as much as they had hoped for, something understandable since they had run no significant advertizing campaign, considered as too expensive.

In the agreement proposed, I would have had no complimentary paper copy in my hand, nor any royalties, after handing over all my research, plus all the money requested. Remember what was said at FPCA-I about the cost of publishing being borne mainly by the scientists. So why should my work and money pay for the heavy machinery of a commercial publisher just to put my book on the web, something I can do myself?

Haubenestel's example with his Alsatian maids is something we should ponder.


The complaint of the publishing astronomer
(medley/old ditties)

Publishers are putting on us more and more technical requirements ...

We are delivering a finalized product for which we are not paid ...

We have also to do ourselves most of the advertizing ...

We have the means to do everything ourselves at lower cost ....

Publishing delays are too long and the outcome is not always satisfactory ...

The prices (books/journals) are much too high ...

Why should we pay so much to get back an information that we initially provided ...

We are loosing access to archives if we stop subscribing digital editions ...

Some illustrations are not available online ...

Etc. etc.

Fig. 5. The complaint of the publishing astronomer.

5 A Few Final Words

To conclude this, let me quote a recent comment from a contributor to the books I produced (and who has himself an extensive international experience as author, editor and translator): *Publishers are not noted for their common sense*, to which I would add that they seem frequently disconnected from the actual markets. And I could give more examples. All in all, my *Complaint of the Publishing Astronomer* (Fig. 5) remains valid today.

One might think that, perhaps because of my age, I am the one disconnected from the markets and the related phenomenology. A couple of weeks ago, I was in a train going from Aberdeen to Edinburgh in Scotland. A young lady sat in front of me, pulling out a baby Dell laptop and a book on neural networks. She was 21, German from Berlin, obviously bright, studying in Aberdeen, going for an interview in Edinburgh for a PhD position, being interested in condensed matter and other issues. To cut a long story short, at some stage, we started talking of this conference. And without me saying anything related, she said *No, electronic publications only? This would be a mistake. Paper is a support complementary to the other ones.* This is what I have been repeating myself for 20 years.

There are several other points from my introductory talk at FPCA-I (Heck 2007) that could be reminded here, such as

- the dramatic advances on brain research that will undoubtedly condition at medium term the way we communicate; our eyes-screens-hands trilogy might soon disappear;
- those magnetic bombs used now in any conflict to wipe out communications and memory storage of the enemy; paper can burn, but electronic information can vanish in a flash;
- the need for evaluation measures adapted to multimedia;
- the fight against hidden plagiarism facilitated by the flexibility of electronic material.

The talks at this FPCA-II conference, starting with the keynote address by John Huchra, are reviewing the current state of the art in astronomy professional communication, as well as providing sound insights into what is expecting us in the years to come.

Acknowledgement. I am grateful to Uta Grothkopf for a careful reading of this paper.

References

Heck, A. (Ed.) 1992, Desktop Publishing in Astronomy and Space Sciences, World Scientific, Singapore (ISBN 981-02-0915-0).

Heck, A. (Ed.) 1997, Electronic Publishing for Physics and Astronomy, *Astrophys. Sp. Sc. Library* **224**, Kluwer Acad. Publ., Dordrecht (ISBN 0-7923-4820-6) + *Astrophys. Sp. Sc.* **247** (1997) 1–254.

Heck, A. (Ed.) 2000, Information Handling in Astronomy, *Astrophys. Sp. Sc. Library* **250**, Kluwer Acad. Publ., Dordrecht (ISBN 0-7923-6494-5).

Heck, A. (Ed.) 2000–2006. Organizations and Strategies in Astronomy, Vol. 1-7, Kluwer Acad. Publ./Springer, Dordrecht, *Astrophys. Sp. Sc. Library* **256** (ISBN 0-7023-6671-9) + **266** (ISBN 0-7923-7172-0) + **280** (ISBN 1-4020-0812-0) + **296** (ISBN 1-4020-1526-7) + **310** (ISBN 1-4020-2570-X) + **335** (ISBN 1-4020-4055-5) + **343** (ISBN 1-4020-5300-2).

Heck, A. (Ed.) 2003, Information Handling in Astronomy – Historical Vistas, *Astrophys. Sp. Sc. Library* **285**, Kluwer Acad. Publ., Dordrecht (2003) xii + 294 pp. (ISBN 1-4020-1178-4).

Heck, A. 2007, Future Professional Communication in Astronomy: Questions and Challenges, in *Future Professional Communication in Astronomy*, Eds. A. Heck & L. Houziaux, Mém. Acad. Royale Belgique in 8°, **XXVII**, 19–36.

Heck, A. & Houziaux, L. (Eds.) 2007, Future Professional Communication in Astronomy, Mém. Acad. Royale Belgique in 8°, **XXVII** (ISSN 365-0936, ISBN 978-2-8031-0238-9).

Heck, A. & Madsen, Cl. (Eds.) 2003, Astronomy Communication, *Astrophys. Sp. Sc. Library* **290**, Kluwer Acad. Publ., Dordrecht (ISBN 1-4020-1345-0).

Heck, A. & Murtagh, F. (Eds.) 1996, Strategies and Techniques of Information in Astronomy, *Vistas in Astron.* **40** (1996) 361–440.

2

Astronomical Publishing: Yesterday, Today and Tomorrow

John Huchra

Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA 02138, USA
huchra@cfa.harvard.edu

Summary. Just in the last few years scientific publishing has moved rapidly away from the modes that served it well for over two centuries. As "digital natives" take over the field and rapid and open access comes to dominate the way we communicate, both scholarly journals and libraries need to adopt new business models to serve their communities. This is best done by identifying new "added value" such as databases, full text searching, full cross indexing while at the same time retaining the high quality of peer reviewed publication.

1 Introduction/History

My personal interest in astronomical publishing, beyond that of an author, goes back nearly 25 years to first a stint as chair of the American Astronomical Society Publications Board, followed by terms on the AIP Publishing Policy Committee, the PASP Editorial Board, a stint as a Scientific Editor for the *Astrophysical Journal*, a term on the Board of Directors of Harvard Business School Publishing, and also a stint as the Vice Provost for Research Policy at Harvard when open access to all scholarly publications came to the fore.

The history of publishing, especially scholarly periodicals, is a relatively short one in our usual astronomical terms of Gigayears and even short in terms of the history of writing. The earliest recorded writing or script dates back to the early Bronze Age, 6600–3500 BCE, with the Vinca script from southern Europe, cuneiform from Sumer ~3500 BCE, hieroglyphs from Egypt starting around 3200 BCE, and the Chinese Jia-gu wen from ~1500 BCE. The earliest "records" were primarily financial in nature (not scientific!) and date back to ~600 BCE in both cuneiform and hieroglyphics.

The first scientific periodicals date to fewer than 400 years ago; both the English *Philosophical Transactions of the Royal Society*, and the French *Journal des Sçavans* began publication in 1665. Astronomy did not get its own periodical for another century – probably the first astronomical periodical was

A. Accomazzi (ed.), *Future Professional Communication in Astronomy II*, Astrophysics and Space Science Proceedings 1, DOI 10.1007/978-1-4419-8369-5_2,

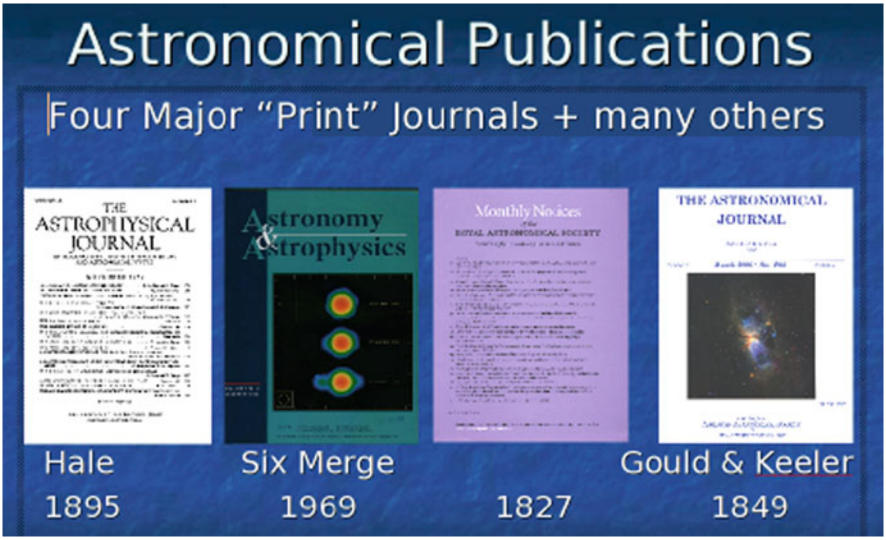


Fig. 1. The four major astronomical research journals.

His Majesty's Nautical Almanac, which began publication in 1767. The United States entered the game about a century after that with both the initiation of the US Naval Observatory *Nautical Almanac* in 1852 and the start of *the Astronomical Journal* by Gould and Keeler in 1849. There has been large growth in the number and diversity of scientific journals since the eighteenth century. Currently there are over 24,000 *refereed* journals and probably over 150,000 total periodicals including laboratory and observatory reports and popular magazines.

Today there are four major print journals in Astronomy (Fig. 1), *The Monthly Notices of the Royal Astronomical Society* which began publication in 1827, the aforementioned AJ, *the Astrophysical Journal* initiated in 1895 by George Ellery Hale and acquired by the American Astronomical Society in the 1970s, and *Astronomy & Astrophysics* which was formed from the merger of six European journals in 1969. There are over 300 additional astronomy and astrophysics related journals or periodicals, including those that cover astronomy education research. The Wolbach Library at CfA subscribes to 322 serials.

The growth in the field is also evidenced by the growth in the number of refereed papers. A nice study by John Hearnshaw (2009) documented the number of publications by decade through the twentieth century and then the growth in the number of papers by year and by journal for the last two decades (Figs. 2, 3 and 4). Astronomical publishing rates were fairly flat until the 1960s, when they significantly accelerated almost certainly due to the

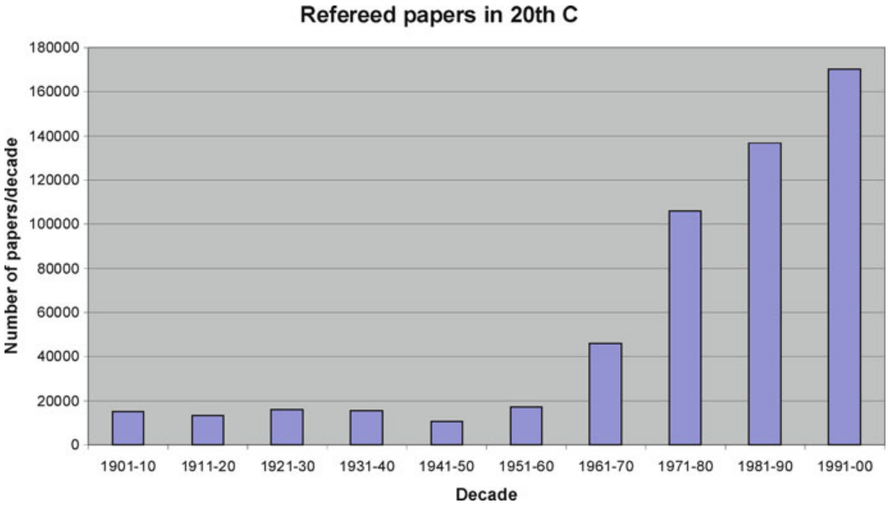


Fig. 2. Refereed papers per decade (From J. Hearnshaw (2009). With permission).

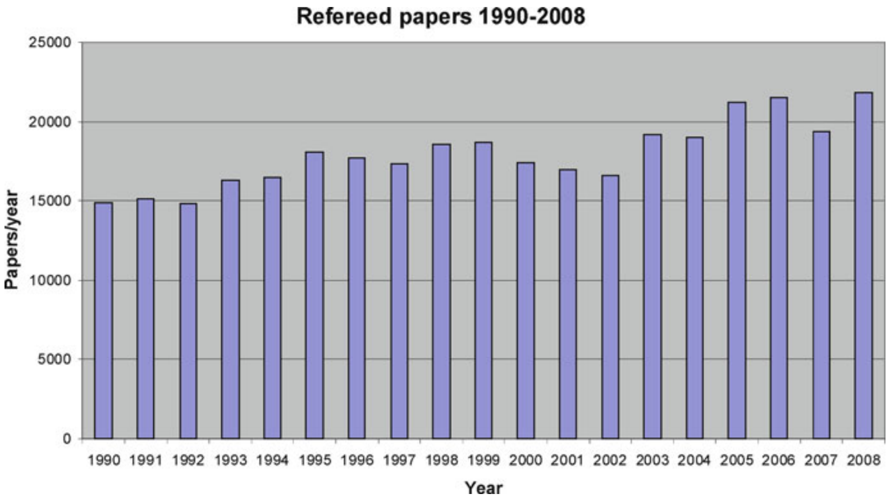


Fig. 3. Refereed papers published per year (From J. Hearnshaw (2009). With permission).

combination of the advent of space observatories, the growth in the numbers and types of ground based observatories and the availability of government support for the sciences that came with the founding of the NSF and NASA in the mid-late 1950s.

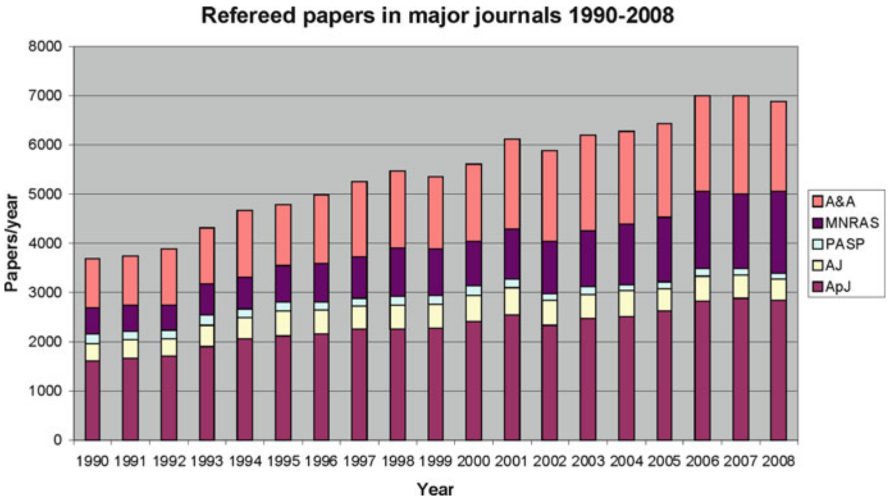


Fig. 4. Refereed papers in five major astronomy & astrophysics journals (From J. Hearnshaw (2009). With permission).

2 Why Do We Publish?

Any discussion of the future of astronomical publishing requires examination of why we publish. It is important to remember that the **Primary** reason is to transfer information. We are here at CfA which is part of the Smithsonian Institution, whose mission is the "Increase and Diffusion of Knowledge." The American Astronomical Society also recognizes this and now has that as part of its mission statement:

> The Society, through its publications, disseminates and archives the results of astronomical research. The Society also communicates and explains our understanding of the universe to the public.

Individuals have the usual additional reasons for publishing, which include career advancement (publish or perish), i.e. to obtain credit and precedence for research leading to employment and sometimes tenure. Publications are necessary to obtain support, e.g. grants, and, in some cases, cash (writing for hire). How much these drive the process of publication depends on both research field (instrumentalists do not publish much in the traditional sense, rather they publish in metal, plastic and semiconductors), and on location (astronomers working in industry or teaching colleges are subject to different criteria for advancement). Theorists publish, in numbers of papers per unit time, more than observers, who in turn publish more than instrumentalists. But, fundamentally, we publish to communicate – the title of this workshop.

How will this evolve? Do we expect change in the "why" of publishing? The primary reason won't, at least for the foreseeable future. The secondary reasons might, for example, due to demographic shifts or career shifts. The nature of tenure is changing, and most astronomers actually are not in traditional academic tenure or tenure track positions. The latest demographic studies for the Decadal Survey find that only ∼40% of astronomers work at research universities and an additional ∼15% are at small colleges. Even among those at research universities, a significant fraction are in research or soft money positions as opposed to line faculty jobs. The criteria for employment are changing as well, and, as we will discuss later, the way we do research is also fundamentally changing.

3 Publishing Models

Again, no discussion of the future of astronomical publishing would be complete without understanding the business models that are in use. It is also likely that these will have to evolve as additional stresses are placed on publishers, libraries and authors. There is a fundamental principle here that publishing is not free, especially if refereeing is required, copy editing is used to enforce stylistic criteria and check the grammar of authors and journals adhere to the principles of archiving and curation that are natural parts of the scientific process. There are five basic business models in use today:

1. Primarily subscription charges to libraries and individuals. This is a model often adopted by for profit publishers but also societies or organizations with limited other resources (e.g. the AIP).
2. User pays, primarily page or article charges.
3. Single Payer (often a benefactor or an organization that self publishes).
4. Mixed models (a combination of subscription income, page charges, abstract sales, e.g. the AAS).
5. Open Access (often also a benefactor, e.g. the DOE supporting ArXiv).

In examining these models, we need to understand the value added by the various services provided by the publisher (refereeing, editing, curation, electronic access, cross linking, the prestige of a refereed journal), and balance their costs against the available income sources. Editing costs money, good editing more. Publication speed also more, and archiving more too. Are authors and their institutions, and their governments willing to pay? My personal experience as an editor is that I've never seen a paper that could not be improved by the work of an impartial, knowledgeable referee. There will continue to be the challenges, though, of finding good referees, maintaining standards and keeping both time to print and the cost to libraries and users for the journals down.

It is also important for publishers to remember what is driving the open access movement, including the drivers for universities and other organizations to want to retain the copyright to material created by their faculty and staff, and the drivers for authors to use open access journals. The two largest for organizations are costs. Who can forget the "journals crisis" of the 1990s and early twenty-first century which brought long-term extreme increases in subscription costs (averaging between 10% and 20% per annum) for institutional subscriptions? Library budgets simply do not increase at those rates; at universities they are often driven by overhead on contracts and grants and the overall research budget for astronomy, at least in the US, has not grown at 15% per year.

Not to be forgotten are the per article charges to reproduce material, often papers you yourself have written, depending on the publisher's "author agreement," for courses. "Fair use" allows a small number (usually considered to be less than 10) of copies to be made for instructional purposes, but if you want to had out copies of an article to 200 freshmen, it could cost you or your university several thousand dollars. And even those author agreements that give you the right to reproduce generally will still require payment if you use not your own work but that of the woman down the hall.

Finally open access can offer benefits to individual authors. Prime among them is speed. Articles can be posted to ArXiV before being refereed so if there is a race for primacy, as in some fields of astronomy such as theoretical cosmology, this route may be preferred. But let us not forget altruism, in this case the desire to ensure that one's work is available to any interested reader, especially the academic poor, those not at rich and diverse institutions or those at third world institutions, as well as the public.

4 And Things Change

Even more important than business models and open access are other changes that will affect the future of our publications. One is just people. Figure 5 shows the membership of the AAS as a function of time. Note that the membership of the AAS has grown nearly 80% since 1984 while the US population as a whole has only grown 25% in that same period. As Helmut Abt will discuss later (Abt 2011), the number of papers is fairly well predicted by the number of astronomers.

Another change, also driving demographics, is resources for research. The balance around the world is changing and changing rapidly. The NSF's *Science and Engineering Indicators 2010* provide an indication of the changes to come. Figure 6 shows that while there has been measured growth in the number of researchers in the US and the EU, the numbers in the more advanced Asian nations are growing rapidly. In the 1990s the publication rate of the EU surpassed that of the US, but again, the publication rate of science and engineering articles from Asia is growing much faster than either (the US has essentially been flat for a decade) and is poised to surpass both regions in the next decade. And this is despite the lack of growth in Japan, one of Asia's previous powerhouses. Figure 7 shows why. The growth rate in R&D support in China, Malaysia and Singapore is more than triple that of the US and the EU. We get what we pay for.

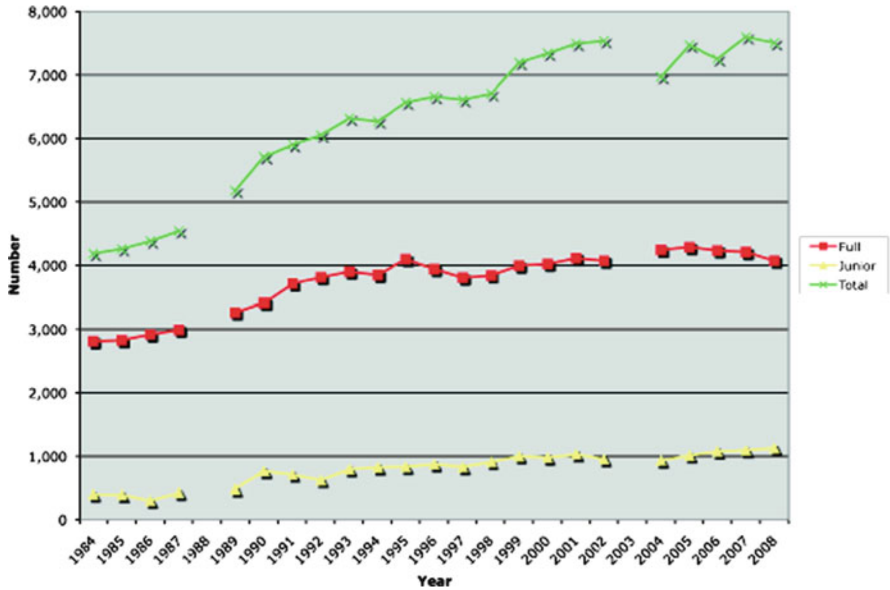


Fig. 5. Membership in the American Astronomical Society since 1984 (the gap at 2004 is due to an incomplete count that year). Total membership includes full, junior, associate, and emeritus, plus international and educational affiliates. Only total, full and junior are plotted here.

Astroph a.k.a. ArXiv has had a tremendous impact on astronomical communication, especially in rapidly changing fields like cosmology. The number of papers posted to astroph per year now exceeds the number of refereed papers published in the big four or five (Figs. 8 and 9) and has a doubling time of only ∼10 years. And the total number of papers contained in astroph as of April 12, 2020 was 108,557 with an additional 12,631 cross-referenced to astronomy, a total of over 120,000 papers freely accessible, many if not most of which are also in the refereed literature. But there are some, including some very heavily cited review papers, that *only* appear on astroph. And this trend is increasing in several subfields of astronomy, especially those where rapid publication is valued more than refereed publication.

The way we do science is also changing. One easily noticeable change is in the number of authors per paper. Since 1985 the average number of authors on an observational paper has jumped a factor of 2 from ∼4 to ∼8. Even theorists work in larger groups with the average number of authors per paper jumping from two to four over the same period. The major astronomical problems we now attack are more complex and require a broader range of skills and knowledge to solve. This drives us to work in larger groups. We are becoming more like physics every year. And despite the debates that may take place over the philosophy and sociology of astronomy (White 2007; Kolb 2007),

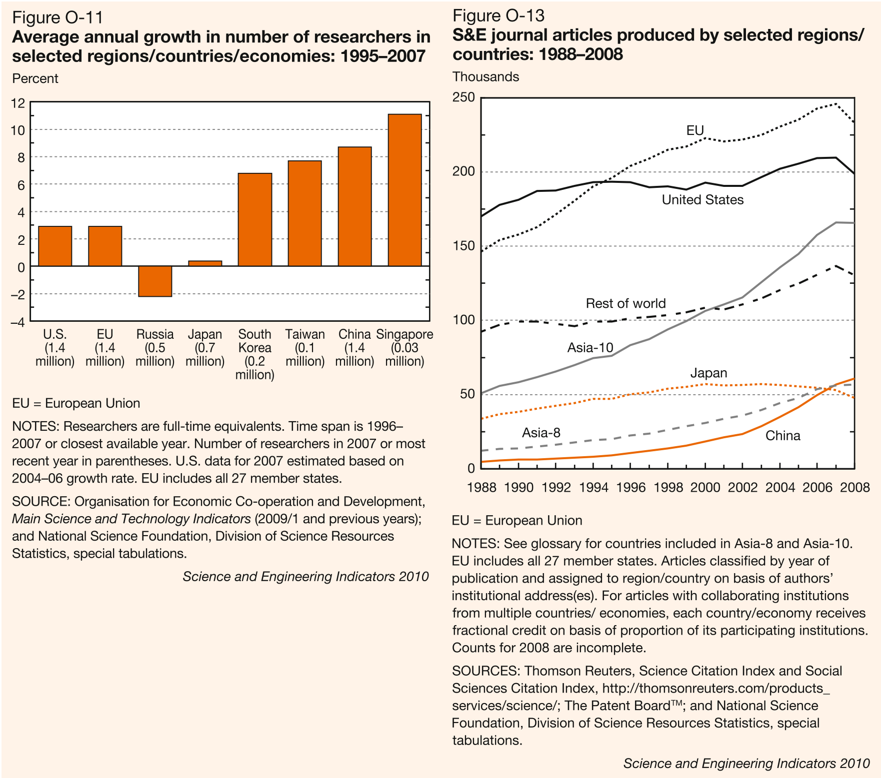


Fig. 6. Growth in the number of researchers (*left*) and in the number of articles produced (*right*) as a function of selected geographical regions (From the NSFs Science and Engineering Indicators 2010).

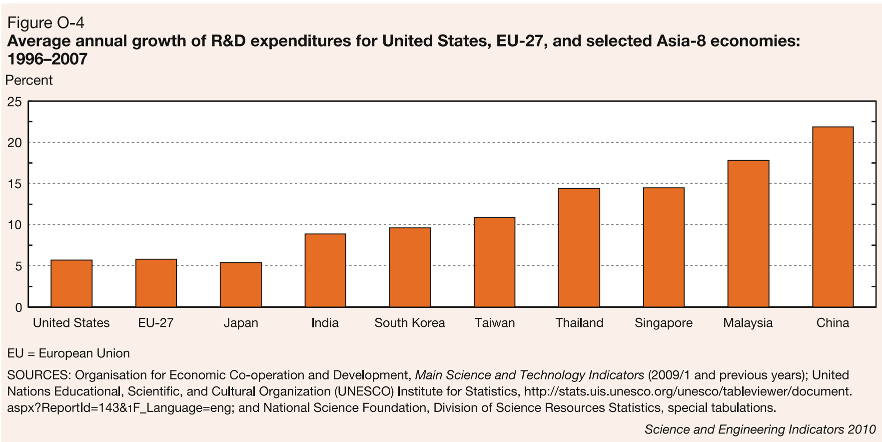


Fig. 7. Growth rate of R&D expenditures for selected countries (From the NSF's Science and Engineering Indicators 2010).

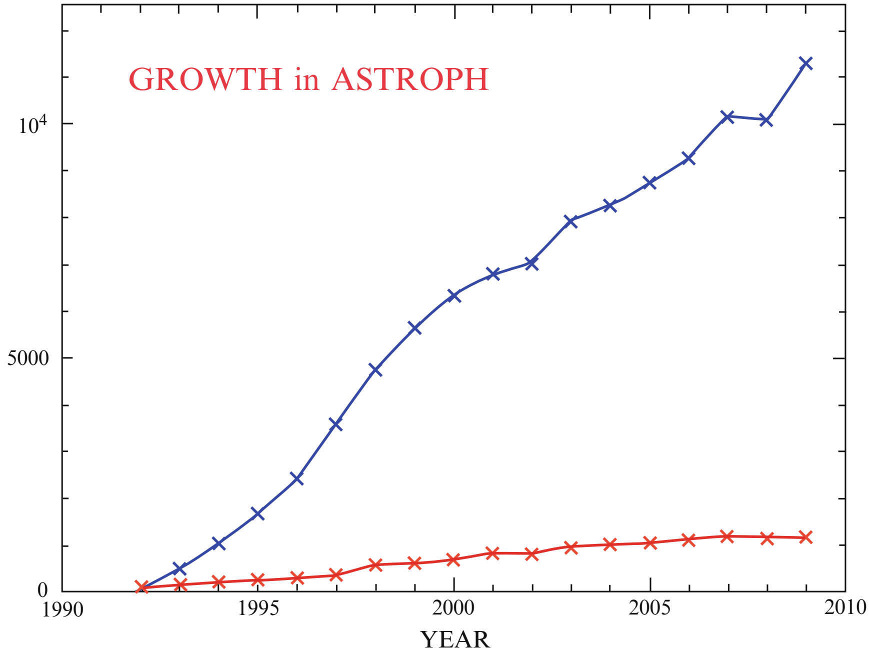


Fig. 8. The number of papers per year posted to astroph. *Top curve (blue)* are astronomy and astrophysics papers, *bottom curve (red)* are cross referenced papers, usually from physics.

Fig. 9. New methods of accessing the literature!

this change is happening and is successfully producing science. The days of the lone astronomer sitting underneath her or his telescope working at the forefront of the field are essentially over.

Lastly, and importantly, the way we access information is changing. Very few people read the journals in hardcopy, the number of individual subscribers to the AAS journals is less than 4% of the membership, and all reports are that library access in place has significantly declined. Most people read journals on line and many do not even know that the way they do this is through their library's subscription. Younger people are even further removed from the hardcopy. Look at the growth in ArXiv submissions and readership. We are turning into a nation of "Digital Natives" (Palfrey & Gasser 2008), connected 24/7. My teenage son is certainly there, my students are there and even my postdocs have crossed the line into total wireless connection.

Open access movements are growing, with more than 4,700 certified open access journals world-wide (see the Directory of Open Access Journals produced by the Lund University Libraries as reported in the Chronicle of Higher Education, Feb. 19, 2010). The economic downturn has driven universities to push hard for open access for the reasons noted earlier. And open access has great educational benefits as well as benefits for basic research. (A disclaimer: I work for Harvard University which now has a strong internal open access policy aimed at providing free use for non-profit purposes – education and research – to content produced by university staff and faculty). And the government has entered the debate by asking for comments on proposed regulations regarding access to content produced with Federal resources. Most people tend to forget that US Federal Employee and certain state employees are already "exempt" from copyright, i.e. their work cannot be copyrighted so by definition is freely available. This has not, as I can again tell you from personal experience as a Smithsonian employee and civil servant, hindered the ability of Feds to publish.

Congress, the Executive Branch and the Judiciary have been of two minds on this issue. On one hand they want to protect intellectual property rights (patents, copyrights) to provide proprietary use of material for gain and competitive advantage – its good for industry and individuals. On the other hand, they believe that there should be free access to information paid for by the taxpayer – which will also lead to economic gain and competitive advantage. It is quite possible in any compromise that we could get the worst of both worlds. Fortunately the problem is now under study and most (but not all) of the parties involved have gotten together to examine solutions to the open access question that will provide such access without destroying the economic basis for refereed scholarly publishing (see the report of the AAU Roundtable on Scholarly Publishing for a review of the problem and a list of recommendations, AAU 2010).

5 The Future (The Tomorrowland of Publishing)

So where do we go from here? And better, how can we prepare for the future? There are only a few drivers: New ways of Communicating, Changes in the Way We Work, Changes in the World, and lastly Changes in the Resources Available. Perhaps there will also be legal or legislative issues, but I cannot predict even in the short term what those might be. The others we understand and can plan to address them.

First we should remember (1) Why are we publishing? And (2) Who are the users? If we keep the answers to those questions in mind, we can have a much clearer view of what we need to do.

Next, we can predict, based on recent past performance, important trends.

1. There will be an increasing move towards open access and very rapid publication. "Tenure" based on only refereed publications is becoming less important, especially as more astronomers are employed outside major research universities. There are already major articles only on ArXiv.
2. Publishers will need to move towards mixed business models with page or article charges. This is not as crazy as you might think as we are moderately close to a single payer model now, either through page charges or through overhead supported library subscriptions.
3. Publishers should also identify "value added" components for their journals – e.g. full text searches, completely linked cross references, curation of data included in publications. Generally speaking, individual authors are not capable of archiving either their papers or their data, and universities lack the expertise (and will) to do this for all fields.
4. Publishers should continually work to identify new ways of providing services to their users. Activities like the value-added components noted above can provide publishers with the competitive advantage they need to produce a product of sufficient value that they can recoup the costs of publication and curation.

The Digital Natives are coming, it is only a matter of time! (and perhaps we should learn Chinese...)

References

Abt, Hemut 2011, in this volume.

Hearnshaw, J. 2009, Cosmic Diary "A Crisis in Astronomical Publishing," http://cosmicdiary.org/blogs/john_hearnshaw/?p=138

Kolb, R. 2007, *A Thousand Invisible Cords Binding Astronomy and High-Energy Physics*, Rept.Prog.Phys.70, 1583. ArXiV 0708:1199

National Science Board, 2010, *Science & Engineering Indicators 2010*, NSF Publication 10-01.

Palfrey, J & Gasser, U., 2008, *Born Digital*, (Basic Books).

Scholarly Publishing Roundtable Report, 2010, American Association of Universities. http://www.resourceshelf.com/2010/01/13/scholarly-publishing-roundtable-zreleases-report-and-recommendations/

White, S. D. M. 2007, *Fundamentalist Physics: Why Dark Energy is Bad for Astronomy*, Rept. Prog. Phys. 70, 883. Ar

3

The Emerging Scholarly Brain

Michael J. Kurtz

Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA 02138, USA
kurtz@cfa.harvard.edu

Summary. It is now a commonplace observation that human society is becoming a coherent super-organism, and that the information infrastructure forms its emerging brain. Perhaps, as the underlying technologies are likely to become billions of times more powerful than those we have today, we could say that we are now building the lizard brain for the future organism.

1 Preamble

> ... the highest level of the ant colony is the totality of its membership rather than a particular set of superordinate individuals who direct the activity of members at lower levels.
>
> Höldobler and Wilson (1990)

2 The Super-Organism

For millennia the only means of intra-species cross-generational communication was chemical; information was transferred via genes in a Darwinian process. The structure for some very complex systems have been transferred by these means.

An example would be the reaction to high temperatures. Even single celled organisms avoid excessively hot regions; those which don't get cooked and don't survive. Fast forward a few eons and we have cooperative structures of 50 trillion cells (e.g. humans) which can very rapidly self mobilize to move a finger away from a frying pan.

A second example, indicating just how complex this chemical communication channel can be is that primates (including humans) are innately afraid of snakes (Hebb 1955). To quote Pinker (1997) "People dread snakes without ever having seen one. After a frightening or painful event, people are more prudent around the cause, but they do not fear it; there are no phobias

A. Accomazzi (ed.), *Future Professional Communication in Astronomy II*, Astrophysics and Space Science Proceedings 1, DOI 10.1007/978-1-4419-8369-5_3,

for electrical outlets, hammers, cars, or air-raid shelters." Ancient creatures which had an insufficient fear of snakes became lunch rather than becoming our ancestors. The agglomeration of this genetically transmitted information has been called the "collective unconscious" (Jung 1935).

Creatures can learn: bees, for example can learn which color flowers are the best sources of food, and they can communicate, as bees do with their "waggle dance," communicating the distance and relative position (to the sun) of a food source. The information needed to perform this dance appears to be entirely genetically transmitted, there is no evidence for a learned component (von Frisch 1967).

Humans, certainly, are capable of learning communication techniques. The development of language, aided by Darwinian processes (Pinker 1994) represents a phase transition in the information transfer process across time which is evolution. It became possible for parents to tell their children "eat this mushroom, but not that one!"

This phase transition enabled humans to form large coherent groups where information could be maintained and shared, thus problems could be solved by distributed efforts. Over time, sophisticated bodies of knowledge were built.

The next phase transition in the information transfer process was the development of writing. While oral transmission is powerful, and has provided us with, as an example, the core ideas behind religion, modern civilization is clearly too complex to be achieved without writing.

Writing, along with the two major enabling inventions, paper and printing, has formed the basis for the information transfer system for our globally connected human society. This society, viewed as an entity onto itself, a superorganism, has capabilities far beyond those of single individuals. An example of these capabilities might be the industrial revolution, which was designed and implemented without any central direction.

We are living in another period of phase transition, one that will happen much faster than the previous ones. The time span for phase transitions has been getting shorter by about a factor of 100 each time; implying that the current one will be consummated within a single human lifetime.

The technical underpinnings for the current phase transition are very rapid communication (beginning, perhaps, with the telegraph) and electronic storage. Together our electronic communication and memory network now form a coherent entity, a Super-Brain (Heylighten and Bollen 1996), for the superorganism which is human society.

Although pure communications networks have existed substantially longer (telegraph, telephone, radio, television), it is the addition of electronic memory which makes the computer network the fundamental technical advance. Within the past few decades computer networks have coalesced to form a single entity (in graph/network studies forming the giant component is usually viewed as a phase transition, Newman 2003).

With this the human society super-organism has developed the functional equivalent of a brain, but we are clearly just at the beginning of the process.

In the past 60 years the power of the information technologies behind the transition has increased by a factor of (about) a trillion; if this can continue for another 60 years, or indeed (Kurzweil 2001) increase more rapidly, the evolving super-organism might well be incomprehensible to us today.

The exponential growth of technical capabilities is not guaranteed, Moore's law, for example, appears to have reached its physical limits (in the number of transistors on a chip). The nano-bio-cyber convergence currently occurring suggests, however, that the growth will continue, or, following Kurzweil (2001), accelerate.

The nature of the changes are still in the realm of science fiction; Stapledon (1930) foresaw the fate of the "last men," after which (evolutionarily) humans became a new race:

> The system of radiation which embraces the whole planet, and includes the million million brains of the race, becomes the physical basis of a racial self...
>
> But chiefly the racial mind transcends the minds of groups and individuals in philosophical insight into the true nature of space and time, mind and its objects, cosmical striving and cosmical perfection....
>
> For all the daily business of life, then, each of us is mentally a distinct individual, though his ordinary means of communication with others is "telepathic." But frequently he wakes up to be a group-mind...
>
> Of this obviously, I can tell you nothing, save that it differs from the lowlier state more radically than the infant mind differs from the mind of the individual adult, and that it consists of insight into many unsuspected and previously inconceivable features of the familiar world of men and things.

3 Architecture

It appears that we humans, like the Krell (MGM 1956), will be (or indeed are) increasingly interconnected with an electro-mechanical system of almost unimaginable power. Already we communicate near telepathically with almost anyone on the planet, at almost any time, via cell phone; already our memories are tremendously augmented by internet search engines.

While we cannot know exactly what the future will bring, we can intuit some things concerning the elements of our new environment, based on past trends. The elements we construct will be extraordinarily complex; each with a multitude of design decisions. This is unlike the exact sciences, where there is a single correct description of a physical reality; it is more like architecture, there is no single exactly correct bridge or building design.

Certainly there are architectural elements which persist, such as a key stone in building arches. Perhaps the closest analog to the information entity we are now constructing, (the super-brain?) is a city. Cities are constructed by the long term and large scale combined efforts of people. Cities are not explicitly designed, but they grow as a result of numerous mostly independent efforts. Even without explicit design cities normally have a number of design elements in common, such as neighborhoods, or sidewalk cafes (Alexander et al. 1977).

Major elements in all cities are places to get things. These are often the end points of complex webs of activity. An example could be a bakery, which combines efforts by fertilizer providers, seed salesmen, farmers, millers, truckers, and others to provide bread. Major elements in the information structure we are building are places to get information. These also are the end points of complex webs of activity. An example could be the Smithsonian/NASA Astrophysics Data System (ADS; Kurtz et al. 1993, 2000, 2005). The ADS combines the efforts of lens designers, rocket scientists, astronomers, publishers and others to provide astrophysics research information.

There are substantial differences in how these design elements are instantiated. There can be bakeries, hardware stores, dress shops, toy stores, etc.; and there can be Wal-Mart or Target. Both systems have their advantages. In the world of information there can be specialized providers, such as ADS, PubMed, The Internet Movie Database, Flightstats, etc; and there can be Google or Bing. Again both systems have advantages.

4 Vectors of the Mind

Context is an important criterion; asking to get one's tank filled will get an altogether different result at Diver Jim's Belmont Scuba Co than it will if asked at the gas station next door. Likewise the query "plasma diagnostics" will get a different result on PubMed than on ADS.

Even when one can narrow down a request, for a product, or information, to a specific realm, context is still important. A typical hardware store, for example, will have hundreds, if not thousands of different fasteners (nails, screws, glue, tape, . . .); which product to purchase depends critically on exactly what the current problem is. Likewise the ADS has thousands of papers on a topic such as "weak lensing," which paper to read depends critically on the exact current problem.

Context is usually provided by people. A customer chooses to go to a hardware store (instead of, say, a cheese shop) to solve some problem, which is described to a clerk. The clerk has substantial expertise in the specific problem domain, and suggests one, or several possible solutions, taking into account the actual needs and abilities of the customer.

In the information world the customer still chooses between providers, ADS or PubMed, for example, but there is no equivalent to the clerk; this function is performed by the information service. Because the actual needs and abilities of each "customer" vary greatly for the exact same query (a beginning student will often be better served by different information than would be best for a world renowned expert) the information system will need to be able to adapt to, and perhaps sense, the larger context of queries.

4.1 Global Measures

Linear Algebra Techniques

Clearly information systems perform functions which were previously performed by humans, or by groups of humans. This requires techniques to model the thoughts and behavior of people and groups of people. Psychologists and Sociologists have been developing these techniques for decades.

Thurstone (1934) used eigenvector techniques to build an orthogonal vector space model for human thought. These techniques have become very widely used; psychometrics and marketing for example depend heavily on them. They have also been used to model the idea space of documents. Ossorio (1966) used a psychological testing approach to model individual opinions concerning the relevance of documents. Kurtz (1993) reverse engineered a set of classified documents to build a model of the thought process of a librarian classifier. Perhaps the most successful of these methods is latent semantic analysis (LSA; Deerwester et al. 1990) which builds the vector space based on the cooccurrence of words in documents.

These eigenvector techniques have also been used in many different classification problems (Murtagh and Heck 1987); such as with astronomical spectra. Kurtz (1982) first applied the method to stellar spectra, and Connolly et al. (1995) rediscovered it for galaxy spectra. While the methods have been used, they have not found great success, compared to partitioning line ratio diagrams (Baldwin et al. 1981) or color-magnitude diagrams (Golay et al. 1977). This is likely because, while the techniques are linear, the underlying physics is highly non-linear (Kurtz 1982).

Human thought is also highly non-linear; representing it by a linear vector space is bound to cause problems. Although the success of LSA as an indexing method for text demonstrated that very local measures of nearness can be effectively defined, more generally these are not metric spaces. As an example, human perceptions do not follow the Schwartz or triangle inequality. The perceptual distance between a dog and a vacuum cleaner is made shorter by the introduction of a third point, a mechanical dog; a mechanical dog is a kind of dog, and a mechanical dog is a kind of vacuum cleaner (Kurtz 1989).

Social Network Techniques

By analyzing the connections between entities (people, documents, molecules, traffic jams, ...) one can build powerful descriptive and predictive models (Barabasi 2003; Newman 2003). Many of the techniques were developed for the social sciences (Wassermann and Faust 1994) and are now widespread. Fifteen percent of the papers published in 2009 by *Physical Review E* address aspects of network problems, as do 26% of papers published in *PLOS Computational Biology*, for example.

The structure of information networks contains information about the intelligent processes which created them. Obtaining the underlying structure from the network is called community detection, and is similar to techniques for cluster analysis or classification.

Fortunato (2010) reviews these methods. Currently the most popular method is that of Girvan and Newman (2002) and the "best" (according to Lancichinetti and Fortunato 2009) is that of Rosvall and Bergstrom (RB 2008). RB have used their algorithm on citation data to show the interrelationships between the major fields of science; their map may very profitably be compared with the similar map of Bollen et al. (2009a) who show a similar structure based on usage data and *pre-existing* field classifications.

The RB algorithm has been used by Kurtz et al. (2007) and Henneken et al. (2009) to map the subfields of astronomy, based on both citation data and on shared keywords for journal articles. Attempts to build a similar map from usage data have not, thus far, been successful, perhaps because of the very broad readership patterns of many astronomers.

Another measure obtainable from a network graph is the "importance" of the individual nodes. Importance is normally called centrality in this context, and there are several different centrality measures. In a friendship network, where people are the nodes and they are linked to other people by friendship, the person with the most friends would be the person with the highest degree centrality. Note that friendship is directional: I may consider you my friend, but that does not mean that you consider me a friend; thus the concepts of in-degree and out-degree. In a citation network the paper with the highest in-degree is the most cited paper, while review articles would have very high out-degree.

Betweenness centrality is another "importance" measure. In a friendship network the most central people are those with friends in many different, otherwise autonomous cliques; in a journal to journal citation network the most central journals are the interdisciplinary journals (Leydesdorff 2007), like *Science* or *Nature*, which are between otherwise autonomous fields, such as astronomy and neuroscience. Betweenness centrality is a key measure in an information flow network; the high betweeness centrality nodes facilitate information transfer between fields.

The currently most used centrality measure is the expected occupation time for each node when visited by a random walk, where the agent randomly follows links from node to node. This is normally called eigenvector centrality, as the result is the same as the first eigenvector of the node-node connectivity matrix (Bonacich 1972). Google's famous Page-Rank algorithm (Brin and Page 1998) is essentially eigenvector centrality, with clever implementation details.

There are several other centrality measures. Kurtz and Bollen (2010) give a brief introduction; Koschützki et al. (2005) a detailed discussion.

4.2 Local Measures

While measures which solve for globally optimum measures are clearly desirable and useful, they are often not feasible. This is especially true in situations where the data is inherently highly multipartite. Journal articles, for example, may be connected by citations, but they may also be connected by having the same author, or having been read by the same reader, or concerning the same subject matter, or the same astronomical object, or ....

An interesting question is how much is lost if one uses local measures, instead of attempting a global solution. Some recent work discusses this.

Eigenfactor (West et al. 2010a) is a useful measure of scholarly journals; it is available in the Thompson-Reuters Web of Science, and finds frequent use by librarians in making purchasing decisions. The Eigenfactor is essentially eigenvector centrality, measured on the journal to journal citation graph.

Davis (2008) pointed out that the Eigenfactor journal rankings in a subfield (medicine is what Davis used) are very similar to rankings derived from simple citation counts. In their two dimensional comparison of 39 different network measures Bollen et al. (2009b) showed eigenvector centrality near to in-degree, but not identical. West et al. (2010b) showed convincingly that the Eigenfactor produces rankings which are significantly different from plain citation counts.

While eigenvector centrality is clearly different from simple in-degree, and in cases such as Eigenfactor likely better, the fact remains that the differences are small. This suggests that in many instances very similar results may be had using much simpler measures.

The ADS second order operators (SOO: Kurtz 1992; Kurtz et al. 2002) make use of local measures in the multipartite space to achieve specific results. The SOO are not a ranking or clustering methodology, per se, rather the SOO are relational operators which one uses to build custom rankings. Two examples of their use, both taken from the ADS Topic Search (`adsabs.harvard.edu/cgi-bin/TopicSearch`), will serve as an illustration. Here we will use the language of networks, Kurtz et al. (2005) use the language of lists of attributes. Both examples find information about a technical topic, while this can be almost anything, for concreteness we'll use the topic "weak lensing."

Example 1 starts by taking the article to article citation graph, where articles (nodes) are (directionally) connected by citations (links). First we take the subset of nodes which concern the topic, weak lensing; next we sort the nodes in the subset by their in-degree centrality, and retain only the top N (say 200). Next we form a super-node which is the sum of the individual top 200 nodes, and we sort the entire graph on the out-degree centrality to the super-node. The top of this list are articles which cite a large number of articles on weak lensing which are, themselves, highly cited. In other words we have found review articles on weak lensing.

Example 2 starts with the article-reader bipartite graph, where an article is connected to a reader if that person has read that article. Again we initially restrict the articles to ones concerning the topic weak lensing, then we sort that list on an attribute of the article (now that attribute is the publication date), and take the most recent 200. Again we form a super-node with the 200 articles, and sort the reader nodes by out-degree to this super node. We take the readers with high out-degree to the article super-node (persons who have read recent papers on weak lensing) and form a reader super-node. Finally we sort all the articles in the (non restricted) graph on in-degree from the reader super-node. The top of this list are the currently most popular papers among persons working in the field of weak lensing.

Notice how these constructions solve many of the problems associated with trust (Josang et al. 2007). Only directly relevant, and well regarded papers and individuals are used; ADS further restricts the readers by removing persons who come from external search engines, who usually do not share the goals of professional researchers (Kurtz and Bollen 2010).

The SOO were first implemented in the ADS in 1996; they are used throughout the system, and form much of the basis for the myADS notification service (`myads.harvard.edu`).

4.3 Recommender Systems

If direct queries to an information system, such as ADS, can be viewed as consciously recalling data from the collective memory, then recommender systems might be likened to having memories pop into your head. Recommender systems as a research and application field cover an enormous range of different areas (cf. `recsys.acm.org`) focusing mainly on commercial uses; here we will concentrate on the recommender systems for scholarly literature.

In a scholarly field, such as astrophysics, the information is denser, and the use sparser, than in many other fields. For example *The Videohound's Golden Movie Retriever* lists about 30,000 movies; the ADS contains more than four times this number of papers which contain the word cosmology in the abstract. The peak usage for a scholarly article is usually the first day it becomes available (so there can be no usage information) and decreases rapidly. The typical 10 year old article from a major journal is downloaded once per month.

The ADS is building a set of recommenders based on the article currently being read, in the future we will also create recommenders based on the article viewing history of the reader. While the exact details are yet to be determined, the systems use nearly all the techniques described in this section, as well as the CLUTO (Zhao and Karypis 2004) clustering software. Kurtz et al. (2009) and Henneken et al. (2010) describe the initial implementation; which is intended for use by active researchers.

Briefly, we use the text and references in an article to find a set of recent articles which are very similar to the target article in subject matter. These articles are then used with the SOO to find recommendations using the words/keywords, citations, usage, authors, and astronomical objects.

5 The Scholarly Brain

In addition to having immensely enhanced memories, we now also have immensely enhanced perception. The new data intensive science (Gray 2007) rests not simply in the mechanical extension of our perception, as begun by Galileo and van Leeuwenhoek, but on the automated perception and analysis of huge data sets. The ATLAS experiment at CERN has a raw data rate of 60 TB/s (Klous 2010), about 50 trillion times the information handling capacity of humans (Fitts 1954).

The underlying technologies are being developed to satisfy the needs of huge systems, like the LHC experiments, but ever larger systems of sensor networks are being created to take advantage of the new capabilities. For these systems to combine their impact, they need to be able to communicate with each other; they need to share a language (Kurtz 1989, 1992). The shared language need not be native to any particular system or experiment (Hanisch 2001), but needs to be universally understood, like medieval Latin, or English today. This standardization is taking place across all scientific and technical fields, examples are the International Virtual Observatory Alliance's Simple Application Messaging Protocol (SAMP; Taylor et al. 2009) or the Open Access Initiative's Object Re-use and Exchange (OAI-ORE) protocol (Van de Sompel et al. 2009).

Automated memory and automated perception are combining to form automated ideas. Standardized data descriptors along with semantic tagging of text (Accomazzi 2010) produce a new environment, in which inference engines, similar to the stochastic/syntactic procedures used to analyze bubble chamber tracts (Fu and Bhargava 1973) are able to discover new, important associations and patterns.

These systems will be able to model and predict the behavior of humans (Barabasi 2010). Indeed Ossorio (1967, 1977) spent a good fraction of his career using his methods to model the behavior of specific humans. The role of these systems will not be to model individual human behaviors, but will be to act as the functional core for our collective intelligence.

This collection of machines will not provide the totality of the brain for the emergent super-organism which is human society; the higher functions, the cognitive layer, will be provided by the sum of all people. The machines will provide the core functionality, the lizard brain. Analogous with our own evolution there will be innate capabilities, like our instant reaction to heat, and anomalies, like our fear of snakes.

Individual scholarly disciplines have long functioned as semi-autonomous super-organisms. The memory function (on paper) being journal articles in university libraries. With the advent of the internet scholars are rapidly transforming their work habits to take advantage of the possibilities (Borgman 2007). New technologies effecting the speed of information transfer (Ginsparg 1994), and the ability to directly and effectively access huge datasets (Szalay and Gray 2001) have already been created, and more are coming daily. The scholarly literature has been fully digital for more than a decade. The idea that a discipline, such as astronomy, already functions as a coherent super-organism, with electronic memory and perception systems is not at all far fetched.

Millions of years before he imagined it would occur, we are beginning to implement Stapledon's (1930) *racial mind*, beginning first not with the human race, but with the sub races of astronomers, bio-chemists, economists, particle physicists, etc. With this scholarly brain we are achieving "*insight into many unsuspected and previously inconceivable features of the familiar world of men and things.*"

Acknowledgement. This essay is dedicated to the memory of two extraordinary scientists who showed me great kindness. Peter Ossorio's work continues to inspire me in several ways. Jim Gray's (2007) fourth Paradigm is the clearest exposition of how the new science will actually function.

Also Andre Heck has, through the years, provided me, and many others, with venues to discuss the deeper meaning of some current trends. The references to this paper alone list several.

The ADS is funded by NASA Grant NNX09AB39G.

References

Accomazzi, A. (2010) Astronomy 3.0 Style. ArXiv e-prints arXiv:1006.0670.

Alexander, C., Ishikawa, S., and Silverstein, M. (1977) A Pattern Language, New York: Oxford University Press.

Baldwin, J. A., Phillips, M. M., Terlevich, R. (1981) Classification parameters for the emission-line spectra of extragalactic objects. Publications of the Astronomical Society of the Pacific 93, 5-19.

Barabasi, A-L (2003) Linked, New York: Plume.

Barabasi, A.-L. (2010) Bursts: the hidden pattern behind everything we do, New York: Dutton.

Bollen, J., Van de Sompel, H., Hagberg, A., Bettencourt, L., Chute, R., Rodriguez, M.A., Balakireva, L. (2009a) Clickstream Data Yields High-Resolution Maps of Science, PLoS ONE, 4, e4803

Bollen, J., Van de Sompel, H., Hagberg, A., Chute, R., Mailund, T. (2009b). A Principal Component Analysis of 39 Scientific Impact Measures. PLoS ONE 4, e6022.

Bonacich, P. (1972) Factoring and weighting approaches to status scores and clique identification, Journal of Mathematical Sociology 2, 113-120.

Borgman, C.L. (2007) Scholarship in the Digital Age: Information, Infrastructure, and the Internet, Cambridge: MIT Press.

Brin, S. & Page, L. (1998),The Anatomy of a Large-Scale Hypertextual Web Search Engine, Computer Networks and ISDN Systems, 30, 107.

Connolly, A. J., Szalay, A. S., Bershady, M. A., Kinney, A. L., Calzetti, D. (1995) Spectral Classification of Galaxies: an Orthogonal Approach. The Astronomical Journal 110, 1071.

Davis, P.M. (2008) Eigenfactor: Does the Principle of Repeated Improvement Result in Better Estimates than Raw Citation Counts?, Journal of the American Society for Information Science and Technology, 59, 2186.

Deerwester, S., Dumais, S., Furnas, G., Landauer, T., & Harshman, R. (1990). Indexing by latent semantic analysis. Journal of the American Society for Information Science, 41, 391.

Fitts, P.M. (1954) The Information Capacity of the Human Motor System in Controlling the Amplitude of Movement, Journal of Experimental Psychology 47, 381-391.

Fortunato, S. (2010) Community detection in graphs. Physics Reports 486, 75-174.

Fu, K.S. and Bhargava, B.K. (1973) Tree Systems for Syntactic Pattern Recognition, IEEE Transactions on Computers 22, 1087-1099.

Ginsparg, P. (1994) First steps towards electronic research communication, Computers in Physics 8, 390-396.

Girvan, M., Newman, M. E. J. (2002) Community structure in social and biological networks. Proceedings of the National Academy of Science 99, 7821-7826.

Golay, M., Mandwewala, N., Bartholdi, P. (1977) Spectral classification of stars with the same colours in intermediate multiband photometry - The concept of photometric ‘star-box’. Astronomy and Astrophysics 60, 181-194.

Gray, J. (2007) Jim Gray on eScience: A Transformed Scientific Method, in The Fourth Paradigm, Data-Intensive Scientific Discovery, eds T. Hey, S. Tansley, and K. Tolle, Redmond, Washington: Microsoft Research (2009).

Hanisch, R. J. (2001) ISAIA: Interoperable Systems for Archival Information Access. Virtual Observatories of the Future 225, 130.

Heylighten, F. and Bollen, J. (1996) The World-Wide Web as a Super-Brain: from metaphor to model. in Cybernetics and Systems ’96 (ed. R. Trappl), p. 917, Vienna: Austrian Society for Cybernetics.

Hebb, D. O. (1955) Drives and the C.N.S. (Conceptual Nervous System), Psychological Review 62, 243-254.

Henneken, E. A., Accomazzi, A., Kurtz, M. J., Grant, C. S., Thompson, D., Bohlen, E., Murray, S. S., Rosvall, M., Bergstrom, C. (2009) Exploring the Astronomy Literature Landscape. Astronomical Society of the Pacific Conference Series 411, 384.

Henneken, E. A., Kurtz, M. J., Accomazzi, A., Grant, C., Thompson, D., Bohlen, E., Di Milia, G., Luker, J., Murray, S. S. (2010) Finding Your Literature Match – A Recommender System. ArXiv e-prints arXiv:1005.2308.

Höldobler, B. and Wilson, E.O. (1970), The Ants, Cambridge: Belknap Press.

Josang, A, Ismail, R. and Boyd, C (2007) A survey of trust and reputation systems for online service provision, Decision Support Systems 43, 618

Jung, C.G. (1935) Über die Archetypen des kollektiven Unbewussten, Zürich: Rhein-Verlag.

Klous, S, for the ATLAS collaboration (2010) Event Streaming in the Online System: Real-Time Organization of Atlas Data, report number ATL-DAQ-PROC-2010-017, Geneva: CERN.

Koschützki, D., et al (2005) Centrality Indices, in Network Analysis Methodological Foundations, eds U. Brandes and T. Erlebach, Lecture Notes in Computer Science 3418, Heidelberg: Springer.

Kurtz, M. J. (1982) Automatic spectral classification. Ph.D. Thesis, Hanover: Dartmouth College.

Kurtz, M. J. (1989) Classification and knowledge. in Knowledge Based Systems in Astronomy, Lecture Notes in Physics 329, 91-106.

Kurtz, M. J. (1992) Second Order Knowledge: Information Retrieval in the Terabyte Era. in Astronomy from Large Databases II, European Southern Observatory Conference and Workshop Proceedings 43, 85.

Kurtz, M. J. (1993) Advice from the Oracle: Really Intelligent Information Retrieval. in Intelligent Information Retrieval: The Case of Astronomy and Related Space Sciences, Astrophysics and Space Science Library (ASSL) 182, 21.

Kurtz, M. J., Karakashian, T., Grant, C. S., Eichhorn, G., Murray, S. S., Watson, J. M., Ossorio, P. G., Stoner, J. L. (1993) Intelligent Text Retrieval in the NASA Astrophysics Data System. in Astronomical Data Analysis Software and Systems II 52, 132.

Kurtz, M. J., Eichhorn, G., Accomazzi, A., Grant, C. S., Murray, S. S., Watson, J. M. (2000) The NASA Astrophysics Data System: Overview. Astronomy and Astrophysics Supplement Series 143, 41-59.

Kurtz, M. J., Eichhorn, G., Accomazzi, A., Grant, C. S., Murray, S. S. (2002) Second order bibliometric operators in the Astrophysics Data System. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series 4847, 238-245.

Kurtz, M. J., Eichhorn, G., Accomazzi, A., Grant, C. S., Demleitner, M., Murray, S. S. (2005) Worldwide Use and Impact of the NASA Astrophysics Data System Digital Library. Journal of the American Society for Information Science and Technology 56, 36.

Kurtz, M. J., Henneken, E. A., Accomazzi, A., Bergstrom, C., Rosvall, M., Grant, C. S., Thompson, D., Bohlen, E., Murray, S. S. (2007) Mapping The Astronomy Literature. Bulletin of the American Astronomical Society 38, 808.

Kurtz, M. J., Accomazzi, A., Henneken, E., Di Milia, G., Grant, C. S. (2009) Using Multipartite Graphs for Recommendation and Discovery. ArXiv e-prints arXiv:0912.5235; to appear in Astronomical Data Analysis Software and Systems XIX.

Kurtz, M. J., Bollen, J. (2010) Usage Bibliometrics. Annual Review of Information Science and Technology, 44, 3-64.

Kurzweil, R. (2001) The Law of Accelerating Returns, http://www.kurzweilai.net/the-law-of-accelerating-returns

Lancichinetti, A., Fortunato, S. (2009) Community detection algorithms: A comparative analysis. Physical Review E 80, 056117.

Leydesdorff, L. (2007) Betweenness centrality as an indicator of the interdisciplinarity of scientific journals. Journal of the American Society for Information Science and Technology, 58, 1303-1319.

Metro-Goldwyn-Mayer (1956) Forbidden Planet (film)

Murtagh, F. and Heck, A. (1987) Multivariate Data Analysis, Berlin: Springer Verlag.

Newman, M.E.J. (2003) The structure and function of complex networks, SIAM Review 45, 167.

Ossorio, P.G. (1965) Classification Space - A Multivariate Procedure for Automatic Document Indexing and Retrieval, Multivariate Behavioral Research 1, 479.

Ossorio, P. G. (1967). Attribute space development and evaluation (RADC-TR-67-640), Rome, NY: Rome Air Development Center.

Ossorio, P.G. (1977) “What Actually Happens”, The Representation of Real-World Phenomena, Colombia, South Carolina: University of South Carolina Press.

Pinker, S. (1994) The Language Instinct, New York: William Morrow.

Pinker, S. (1997) How the Mind Works, New York: W. W. Norton.

Rosvall, M., Bergstrom, C. T. (2008) Maps of random walks on complex networks reveal community structure. Proceedings of the National Academy of Science 105, 1118-1123.

Stapledon, O. (1930) The First and Last Men, London:Methuen.

Szalay, A. and Gray, J. (2001) The World-Wide Telescope, Science 293, 2037-2040.

Taylor, M., Boch, T., Fitzpatrick, M., Allan, A., Paioro, L., Taylor, J., Tody, D. (2009) SAMP - Simple Application Messaging Protocol, v 1.11, International Virtual Observatory Alliance.

Thurstone, L.L. (1934) Vectors of the Mind, Psychological Review 41, 1.

Van de Sompel, H., Lagoze, C., Nelson, M. L., Warner, S., Sanderson, R., Johnston, P. (2009) Adding eScience Assets to the Data Web. ArXiv e-prints arXiv:0906.2135.

von Frisch, K. (1967) The Dance Language and Orientation of Bees, Cambridge: Harvard University Press.

Wassermann, S., and Faust, K. (1994) Social Network Analysis, Cambridge (U.K.): Cambridge University Press.

West, J.D., Bergstrom, T.C., Bergstrom, C.T. (2010a) The Eigenfactor Metrics (TM): A Network Approach to Assessing Scholarly Journals, College and Research Libraries, 71, 236.

West, J., Bergstrom, T., Bergstrom, C.T. (2010b) Big Macs and Eigenfactor Scores: Don’t Let Correlation Coefficients Fool You, Journal of the American Society for Information Science and Technology, to appear: (DOI: 10.1002/asi.21374).

Zhao, Y. and Karypis, G. (2004) Empirical and Theoretical Comparisons of Selected Criterion Functions for Document Clustering, Machine Learning, 55, 311-331.

4

Communicating Astronomy Beyond IYA2009

Richard Tresch Fienberg

American Astronomical Society, 2000 Florida Avenue NW, Suite 400, Washington, DC 20009, USA
rick.fienberg@aas.org

Summary. The International Year of Astronomy 2009, celebrating the 400th anniversary of Galileo's first look at the night sky through a telescope, was arguably the most successful public-outreach event in the history of science. Thousands of professional and amateur astronomers, public-information officers, writers, publishers, teachers, and informal educators from 148 countries worked tirelessly to share the wonders of astronomy with millions of people, from rural schoolchildren to the President of the United States. How to build on IYA2009 was the theme of "Communicating Astronomy with the Public 2010," a conference held in Cape Town, South Africa, in March 2010. I report some of the highlights from CAP2010 and consider the extent to which the IYA2009 helped us address global challenges in astronomy communication.

1 The International Year of Astronomy 2009

At the triennial General Assembly of the International Astronomical Union (IAU) in July 2003, outgoing IAU president Franco Pacini of Arcetri Observatory proposed that 2009 be designated the International Year of Astronomy. His idea was to celebrate the 400th anniversary of Galileo's first look at the night sky through a telescope, which brought about a fundamental change in humankind's perception of the universe. Four hundred years of telescopic astronomy have led to a remarkably detailed understanding of the origin, structure, and evolution of the universe – and of Earth's place in it – something certainly worth celebrating!

Unlike other "years of science," such as the World Year of Physics 2005, the focus of astronomers' IYA2009 activities would not be global collaborations in research – after all, there were plenty of those in astronomy already – but a global effort to reach out to students and the public. Pacini envisioned IYA2009 as an opportunity to spread awareness of astronomy's role in enriching all cultures, to nourish a scientific outlook in society, and to increase

A. Accomazzi (ed.), *Future Professional Communication in Astronomy II*,
Astrophysics and Space Science Proceedings 1, DOI 10.1007/978-1-4419-8369-5_4,

public support for scientific research, improve science education at all levels, and attract young people to careers in science, technology, engineering, and mathematics.

Delegates at the 2003 IAU meeting supported Pacini's proposal, which ultimately led to a collaboration between the IAU and the United Nations Educational, Scientific and Cultural Organization (UNESCO), which endorsed the idea in 2005. UNESCO in turn presented the proposal to the UN General Assembly, which formally declared 2009 the International Year of Astronomy in December 2007.

Even before the official UN vote, astronomers worldwide had begun planning IYA2009 activities. The IAU Executive Committee, under the leadership of president Catherine Cesarsky of the European Southern Observatory, served as the overall coordinating body. The committee's role was not to tell others what to do, but to marshal the resources of the global astronomical community – both professional and amateur – and to serve as advisors and enablers of IYA2009 events and activities at the national, regional, and local levels. Early on, there was hope that the IAU would raise millions of dollars from governments, foundations, other institutions, and private donations, but those hopes were dashed when the global economy went into a tailspin in late 2008, just as the IYA2009 planning effort was reaching a crescendo. Still, the IAU was able to raise nearly EUR 663,000 (USD 860,000) to support a minimally staffed IYA2009 Secretariat and to partially underwrite a dozen major projects.

Astronomy is experiential – anyone with a telescope can see what Galileo saw. The IYA2009 slogan was "The Universe: Yours to Discover." So, naturally, a prime goal of IYA2009 was to give as many people as possible a chance to look through a telescope, especially those who'd never had such an opportunity before. Undoubtedly, the highest-profile "star party" of IYA2009 was the one held on October 7th at the White House in Washington, DC. President and Mrs. Obama welcomed hundreds of local schoolchildren to the presidential mansion to enjoy an evening of hands-on astronomy activities and stargazing.

But IYA2009 was so much more than a series of "Galileo moments." By the time the year-long celebration was over, 148 countries ("National Nodes") had become involved, with 111 of them building special IYA2009 websites in their local languages. This is remarkable, as the IAU has only 70 national members! Forty observatories, space agencies, and other organizations around the world also built dedicated IYA2009 websites. Some 33 commercial enterprises, 22 media organizations, and 11 task groups contributed to the effort too. One group, for example, labored to remind event organizers that 2009 was also the 400th anniversary of Johannes Kepler's monumental work *Astronomia Nova*, in which the brilliant contemporary of Galileo introduced his first two laws of planetary motion.

Amateur-astronomy clubs, planetariums, science museums, and even individual astronomers hosted countless small events, but much of the global effort was focused on 12 Cornerstone Projects and 16 smaller Special Projects. All are described online.[1] Full disclosure: I chaired the Galileoscope Cornerstone Project, in which we developed a high-quality, low-cost refractor capable of showing all the objects Galileo observed but much more clearly than he ever saw them. About 180,000 Galileoscopes were distributed worldwide during IYA2009, and through 2010 and into 2011 sales continued at a brisk pace.

In the aggregate, the countries, organizations, groups, and individuals involved in IYA2009 raised and spent nearly EUR 18 million (USD 23 million). How many people did we reach? During two global star parties – 100 Hours of Astronomy in April and Galilean Nights in October – some three million people looked through telescopes. An estimated ten million viewers saw the PBS-TV special *400 Years of the Telescope*. The main IYA2009 website attracted 1.3 million unique visitors; undoubtedly the 111 national and 40 organizational websites attracted millions more. At least seven million citizens participated in stargazing activities in Japan, and 45 million watched the July 2009 total solar eclipse in China on television or online. Most Galileoscopes went to schools, museums, planetariums, and astronomy clubs, where each was used by an average of about ten people during IYA2009, so around 1.8 million people glimpsed the Moon, Jupiter, Saturn, or another celestial spectacle with one of these small telescopes. Adding all the other projects pushes the number of people reached by IYA2009 activities to over 815 million worldwide, making IYA2009 arguably the largest and most successful public-outreach effort in the history of science (Russo and Christensen 2009).

2 Communicating Astronomy with the Public

If the thinking head of IYA2009 was the IAU Executive Committee and its IYA2009 Secretariat, the working arm of the effort was IAU Commission 55, Communicating Astronomy with the Public. Most of the planning at the national and international levels took place under the auspices of C55 through a series of meetings that trace their origin to a 2002 gathering in Tenerife, Canary Islands, organized by Terry Mahoney of the Instituto de Astrofísica de Canarias. That meeting, Communicating Astronomy, brought together a spectrum of astrocommunicators:

- *Producers* of astronomical information, i.e., mainly research scientists.
- *Public information officers* connected with observatories, space agencies, space-science missions, and universities.
- *Mediators*, e.g., science reporters and writers from print and broadcast media; staff members from museums, planetariums, and national parks; operators of commercial and nonprofit websites focused on astronomy; and educators at all grade levels.

[1] www.astronomy2009.org

A follow-on meeting in Washington, DC, in 2003 was called Communicating Astronomy to the Public and was focused more narrowly on astronomy education and public outreach (EPO). Editors and managers of astrophysics journals, abstract services, and other types of researcher-to-researcher communication who had attended the 2002 meeting sat this one out; they subsequently attended Future Professional Communication in Astronomy I and II, which dealt more directly with issues of concern to them. A principal outcome of the 2003 meeting in DC was the Washington Charter for Communicating Astronomy with the Public. I'll say more about this below; for now, I simply want to call your attention to the change of preposition: Communicating Astronomy *to* the Public had morphed into Communicating Astronomy *with* the Public, acknowledging that science communication (indeed, *any* communication) is effective only when it's a two-way street.

Commission 55 got its start around the time of the Communicating Astronomy with the Public (CAP) 2005 meeting in Garching, Germany, when the IAU approved the creation of a CAP working group. The following year, at the next IAU General Assembly, the working group was promoted to a commission under Division XII, Union-Wide Activities. C55's officers and organizing committee were given (and accepted!) responsibility for the now-approximately-biennial CAP meetings and, more significantly, for undertaking the worldwide planning effort for IYA2009. In addition, with support from the European Southern Observatory, C55 launched a new print and online quarterly magazine, *CAP Journal*, featuring best practices in astronomy education and public outreach.

Not surprisingly, the next CAP meeting, in Athens, Greece, in late 2007, was devoted entirely to preparing for IYA2009. By this point many countries had identified institutions to serve as National Nodes as well as astronomers at those institutions to serve as national leaders, or Single Points of Contact (SPoCs). Also, thanks to a crucial meeting at Garching in early 2007, many of the Cornerstone Projects had begun to take shape under the leadership of enthusiastic (some would say masochistic) volunteers. After CAP2007, nearly every large national gathering of professional and/or amateur astronomers, such as the meetings of the American Astronomical Society and the Astronomical Society of the Pacific (ASP), included sessions on getting ready for IYA2009. The proceedings of CAP2007 and the 2008 ASP meeting (Gibbs et al. 2008) served as IYA2009 handbooks for organizers, and the number of volunteers signed on to help with projects large, medium, and small soared into the thousands and, by the time IYA2009 was over, tens of thousands.

3 CAP2010: Beyond IYA2009

Even though IYA2009 would last only 1 year, its founders and organizers worked from the outset to forge a sustainable network of EPO professionals and volunteers whose efforts could continue indefinitely – though, for practical (and mental health!) reasons – not at the same level of intensity as during IYA2009 itself. This was one reason why C55 took on much of the work of co-ordination: there was already a sizable group of CAP "regulars" from around the world, many were key players in IYA2009 projects, and they'd be around afterward to carry on.

Thus, naturally, the subject of the CAP2010 meeting in Cape Town, South Africa, was how to build on IYA2009's unprecedented success and continue enhancing scientific literacy worldwide through astronomy education and public outreach. More than 150 participants – including many SPoCs and project chairs – attended, representing 45 countries and six continents (Fig. 2). A majority spent an extra few days and nights in the region to visit the South African Astronomical Observatory in Sutherland, where we were treated to clear, unseasonably temperate skies.

Over the course of more than 60 presentations at CAP2010, a small number of themes emerged as the main "lessons learned" during IYA2009:

– There is a worldwide hunger for astronomy information, especially in underdeveloped and still-developing countries.
– Stargazing has universal appeal, and people everywhere recognize the value of dark skies.
– The global "marching army" of amateur astronomers plays a crucially important role in communicating astronomy with the public.
– Money sure helps, but it isn't always necessary!
– Every medium can be effective, but real communication is two-way, so new media (e.g., Web 2.0) and in-person engagement between astronomy communicators and the public works best.

Among the greatest legacies of IYA2009 are the many alliances forged among the various astronomical communities involved in the effort, such as between professionals and amateurs. Event and project organizers enjoyed much success during IYA2009 and have no interest in losing momentum nor cutting back on their frequent and positive interactions with their collaborators across the globe. Thus another common theme at CAP2010 was a pledge to continue much of the work of IYA2009 indefinitely, branded with a revised logo in which the words "International Year of Astronomy 2009" are replaced by the words "Beyond International Year of Astronomy" (Fig. 1).

Most IYA2009 websites remain up and running as part of the year's legacy, and many of the projects introduced in 2009 are being continued in 2010–11 and, if all goes according to plan, beyond. To cite just one example, Astronomers Without Borders, the international organization that made 100 Hours of Astronomy happen April 2–5, 2009, stepped up to coordinate Global Astronomy Month in April 2010 and hopes to repeat the celebration annually.

The American Astronomical Society established a Working Group on Pro-Am Collaboration for scientific research in the late 1990s, and only a little more than a decade later, professionals routinely work with amateurs in many branches of astronomy, including asteroid studies, gamma-ray-burst follow-up,

Fig. 1. The logos for the International Year of Astronomy 2009 (*left*) and its legacy, Beyond International Year of Astronomy (*right*).

and exoplanet characterization. In the wake of IYA2009, when professionals and amateurs worked together on huge numbers of EPO activities, delegates at CAP2010 agreed that C55 should promote further growth in this area by establishing a Working Group on Pro-Am Collaboration for astronomy communication. A subcommittee of C55 is developing a proposal that will draw heavily from the experience of the highly successful Night Sky Network. This project of the ASP and NASA's Jet Propulsion Laboratory involves professional astronomy communicators giving members of astronomy clubs the training and tools they need to do effective informal education and public outreach.

But amateur astronomers aren't the only ones who need help learning how to be effective communicators. An increasing number of professional astronomers are being asked, or are volunteering, to explain astronomy to the public, the media, representatives of funding agencies, and/or government officials. Despite their good intentions, these scientists often find themselves poorly prepared to handle the questions they get asked. Training for a career in research and university-level teaching doesn't give one the skills needed to talk to ordinary citizens about science. At CAP2010 several speakers described programs whereby they provide communications training for scientists, for example, in summer workshops or weekend classes held in conjunction with scientific-society meetings. Many delegates left the conference intent on setting up similar programs in their own countries. Indeed, there was strong sentiment

Fig. 2. Attendees at the conference Communicating Astronomy with the Public 2010, hosted by the South African Astronomical Observatory in Cape Town, South Africa (Photo: Sze-Leung Cheung).

(not surprisingly) that communications training should be incorporated into regular graduate – and perhaps even undergraduate – education in astronomy.

4 The Washington Charter

As the conferees at CAP2010 discussed how to capitalize on the success of IYA2009 and further the cause of communicating astronomy with the public, they realized that they had already laid the groundwork for the next step back in 2003, when they created the Washington Charter. The document commits individuals and institutions involved in the global astronomical enterprise to support public communication, whether or not they are directly involved in it themselves. In 2003 and 2004 a modest effort was undertaken to attract signatories to the Charter but without much success as various parties objected to some of the wording. Taking their concerns into account, a small group at CAP2005 revised the text into its present form, reproduced in its entirety here:

The Washington Charter for Communicating Astronomy with the Public

As our world grows ever more complex and the pace of scientific discovery and technological change quickens, the global community of professional astronomers needs to communicate more effectively with the public. Astronomy enriches our culture, nourishes a scientific outlook in society, and addresses important questions about humanity's place in the universe. It contributes to areas of immediate practicality, including industry, medicine, and security, and it introduces young people to quantitative reasoning and attracts them to scientific and technical careers. Sharing what we learn about the universe is an investment in our fellow citizens, our institutions, and our future. Individuals and organizations that conduct astronomical research – especially those receiving public funding for this research – have a responsibility to communicate their results and efforts with the public for the benefit of all.

RECOMMENDATIONS

For Funding Agencies:

- Encourage and support public outreach and communication in projects and grant programs
- Develop infrastructure and linkages to assist with the organization and dissemination of outreach results
- Emphasize the importance of such efforts to project and research managers
- Recognize public outreach and communication plans and efforts through proposal selection criteria and decisions and annual performance awards
- Encourage international collaboration on public outreach and communication activities

For Professional Astronomical Societies:

- Endorse standards for public outreach and communication
- Assemble best practices, formats, and tools to aid effective public outreach and communication
- Promote professional respect and recognition of public outreach and communication
- Make public outreach and communication a visible and integral part of the activities and operations of the respective societies
- Encourage greater linkages with successful ongoing efforts of amateur astronomy groups and others

For Universities, Laboratories, Research Organizations, and Other Institutions:

- Acknowledge the importance of public outreach and communication
- Recognize public outreach and communication efforts when making decisions on hiring, tenure, compensation, and awards
- Provide institutional support to enable and assist with public outreach and communication efforts
- Collaborate with funding agencies and other organizations to help ensure that public outreach and communication efforts have the greatest possible impact;
- Make available formal public outreach and communication training for researchers
- Offer communication training in academic courses of study for the next generation of researchers

For Individual Researchers:

- Support efforts to communicate the results and benefits of astronomical research to the public
- Convey the importance of public outreach and communication to team members
- Instill this sense of responsibility in the next generation of researchers

A concerted effort to attract endorsements began after CAP2005, netting support from 20 professional societies and 11 research institutions, but unfortunately no funding agencies. Then things ground to a halt as everyone associated with C55 turned their full attention to IYA2009. There has been essentially no promotion of the Washington Charter for the last 5 years. That is about to change, and in the wake of the extraordinarily successful IYA2009, we expect a renewed emphasis on the Charter to generate many more endorsements in the coming months and years. As a result, communicating astronomy with the public will go from being a "necessary evil" to being a fundamental part of what astronomers do in the course of their careers – to the benefit of both the public and the profession.

5 Links to More Information

Almost all the presentations made at CAP2010 are available online as PDFs or PowerPoint files. They can be downloaded from the Programme page of the CAP2010 website:
http://www.communicatingastronomy.org/cap2010/programme.html

The same is true for presentations made at CAP2007, which focused on preparations for the International Year of Astronomy 2009:
http://www.communicatingastronomy.org/cap2007/programme.html

More general information about IAU Commission 55 and the Communicating Astronomy with the Public meetings is available on the C55 website:
http://www.communicatingastronomy.org

Communicating Astronomy with the Public Journal:
http://www.capjournal.org

Main International Year of Astronomy 2009 Website (with links to all IYA2009 projects): http://www.astronomy2009.org

ESCOnet Science Communications Courses:
http://www.esconet.org/Visitors/Welcome.html

References

"International Year of Astronomy 2009 Final Report," edited by Pedro Russo and Lars Lindberg Christensen, International Astronomical Union, 2009 http://www.astronomy2009.org/resources/documents/detail/iya2009_final_report/

"Preparing for the 2009 International Year of Astronomy: A Hands-on Symposium," edited by Michael G. Gibbs, Jonathan Barnes, James G. Manning, and Bruce Partridge, ASP Conference Series, Vol. 400, 2008; http://www.aspbooks.org/a/volumes/table_of_contents/?book_id=433

5

Trends in Scientific Publishing at Springer

Guenther Eichhorn

Springer, 233 Spring Street, New York, NY 10013, USA
guenther.eichhorn@springer.com

Summary. Scientific publishing has undergone tremendous changes in the last decade, and it is still evolving rapidly. This article describes some of the issues that are facing scientific publishers, and shows some examples of what one of the commercial publishers, Springer, has done and is working on, in order to stay abreast of these changes and to embrace the new technologies that become available. Springer has moved rapidly into the digital age and has by now digitized almost all its journal content and a significant part of its book content. We have developed new capabilities that make use of the new technologies available and are in the process of further utilizing these new possibilities. Web products like AuthorMapper, SpringerProtocols, and Social Networking sites explore some of these new capabilities. We will continue to explore enhancements of our scientific publishing efforts to provide new possibilities for communicating scientific research.

1 Introduction

This article describes some of the past, current, and projected efforts at Springer to utilize the new capabilities provided by the recently emerged Internet technology.

Springer is an old publishing company, founded in 1846 as a family business. It is by now the largest publisher of scientific/technical books and the second largest publisher of scientific/technical journals.

In recent years, the publishing system has changed dramatically due to the Internet. The whole publishing paradigm is changing, because of new publishing capabilities and access methods, that are available to almost everybody. Springer has adopted these new technologies very quickly by providing electronic articles for its journals and books, and by collaborating with Abstracting & Indexing (A&I) Services very early on. The following sections describe some of the efforts in moving into the Internet age and how it affects traditional publishing.

2 Going Digital

The most obvious change was to provide content in digital form that traditionally was published on paper. By 1996, Springer was publishing all new journal content

A. Accomazzi (ed.), *Future Professional Communication in Astronomy II*,
Astrophysics and Space Science Proceedings 1, DOI 10.1007/978-1-4419-8369-5_5,

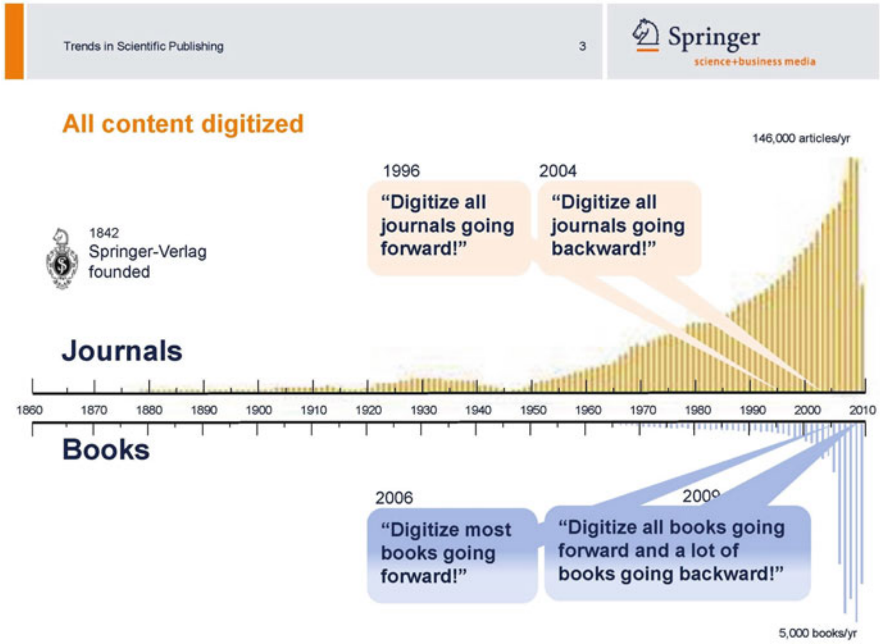


Fig. 1. Development of electronic content at Springer.

in electronic form. By 2004, we were digitizing all back issues of our journal in order to provide the complete Springer journal literature in electronic form. By 2009, all our books were available in electronic form and we were working on digitizing older books as much as feasible (see Fig. 1).

One aspect of the availability of content in electronic form was the possibility to make the content available much sooner than the printed version. We provided this early access through our Online-First system, where articles were put online before they were even bundled in issues, right after they were accepted for publication, and the typesetting was completed. This provided our readers with much earlier access to their literature.

Another important aspect of the digital era are the online A&I services. Springer collaborated very early on with these online A&I services and provided electronic metadata in a consistent format. This greatly facilitates the capability of the A&I services to index our data quickly and accurately.

Online search systems are very important for a publisher. They provide our users in different disciplines dedicated search capabilities that are tailored to the particular field. Overall however, one search system has become the most used: Google. On the order of 60% of all referrals to content on the Springer website come from Google.

3 The Changing Role of Publishers

In the print era, the publishers were the main (if not only) means of publishing scientific articles. Scientists would either subscribe to relevant journals themselves, or visit their library to read the journals. To a small degree mailing preprints to fellow scientists helped spread the news about an article, but that was small compared to the distribution of the literature through journals.

With the Internet all that changed. It is now very easy to publish articles online. In some disciplines (e.g. Physics and Astronomy) most articles are published first online in a preprint or institutional manuscript repository, not by commercial or society publishers. This has forced the publishers (both commercial and society) to review their position and to look at other services that they can provide to add value to their journals. There are two broad categories of services that publishers provide as added value:

1. Improvements on the article content.
2. Other services.

3.1 Improvements on the Article Content

a. Typesetting/Formatting/Content Presentation
   Publishers were always involved in providing the formatting for presenting articles in a well formatted style. This includes copyediting, formatting of tables, reference lists, figures, etc. This continues to be an important part of the publishing process. This part of the publishing process lends itself to improvements based on new technologies like animated content, downloadable data, active calculations, etc (see Sect. 4).
b. Content
   This part of the publishing process involves ensuring reliability and accuracy of the content of the articles. This activity includes the peer review process. The peer review process is an essential part of the scientific publishing endeavor, provided by the publisher. The scientific community monitors what is being published through this process, helps to improve published articles, and tries to prevent publication of unsubstantiated or incorrect information.
   Another part of monitoring the content of scientific journals is the detection of plagiarism. In recent years it seems as if plagiarism has been happening more frequently. This is probably due to the fact that the availability of electronic content makes it very easy to copy material that was published elsewhere. Springer, together with other publishers is working on a system that can detect plagiarism in submitted articles, before they are being published. This is done under the umbrella of CrossRef and called CrossCheck.
c. Metadata
   Providing consistent and accurate metadata for published articles is another service that publishers provide. By now the established format for metadata is XML. Providing XML metadata that are properly parsed and formatted is an important aspect of publishing, since the A&I services rely on these metadata to provide their search capabilities accurately and in a timely fashion.

3.2 Web Services

The second broad area that publishers are involved in are additional services. Following are some examples of services that Springer is providing.

a. Springer.com
   This is the platform that contains the general information for all our books and journals. It provides a search capability on the book and journal level. It has subscription information, information about editorial issues, information for librarians, and much more. It is the first stop for any user who is looking for content or article submission information.
b. SpringerLink.com
   This is the platform that holds the content of our journals and books. It provides a sophisticated search capability for all Springer content. The new version of SpringerLink has many new capabilities relating to content discovery, like finding related articles, citations, etc.
c. New Web Services

   1. AuthorMapper
      This service allows users to execute any query for content in SpringerLink, and then map where the authors of the found articles are located (see Fig. 2). This allows you to find geographic locations of research groups and individual authors. This system is freely available for everybody.

Fig. 2. Screen shot of AuthorMapper search for "Superheavy Elements", clearly (among others) showing the locations of the Joint Institute for Nuclear Research, Dubna, Russia, GSI Darmstadt, Germany, Fermilab, Illinois, USA, and SLAC, Stanford, California, USA, the four major sites for research on superheavy elements.

Fig. 3. Screen shot of SpringerProtocols.

2. SpringerProtocols
   SpringerProtocols is the largest subscription-based electronic database of reproducible laboratory protocols in the Life and Biomedical Sciences (see Fig. 3). Compiling protocols from Humana's successful book series *Methods in Molecular Biology*, *Methods in Molecular Medicine*, *Methods in Biotechnology*, *Methods in Pharmacology and Toxicology*, and *Neuromethods*, as well as from a vast number of Laboratory Handbooks, such as *The Biomethods Handbook*, *The Proteomics Handbook*, and the *Springer Laboratory Manuals*, SpringerProtocols offers researchers access to nearly thirty years worth of time tested, easily reproducible, step-by-step protocols for immediate use in their lab. Written by renowned experts in the field with an emphasis on both comprehensiveness and ease of use, each protocol is organized in an easily reproducible recipe style, and offering helpful tricks of the trade in an invaluable "notes" section, SpringerProtocols is an invaluable resource for the modern research laboratory. All SpringerProtocols are written by experts in the field identified and reviewed by distinguished editors in their specific subject areas. Prior to publication each protocol is reviewed for clarity, accuracy, and consistency by the subject editor, a member of our prestigious editorial board, and Springer's in-house editorial staff. There are currently over 20,000 protocols in the database.
3. SpringerMaterials
   SpringerMaterials is based on the Landolt-Börnstein New Series, the unique, fully evaluated data collection in all areas of physical sciences and engineering (see Fig. 4). SpringerMaterials also comprises the Dortmund Data Bank Software & Separation Technology, a Database on Thermophysical

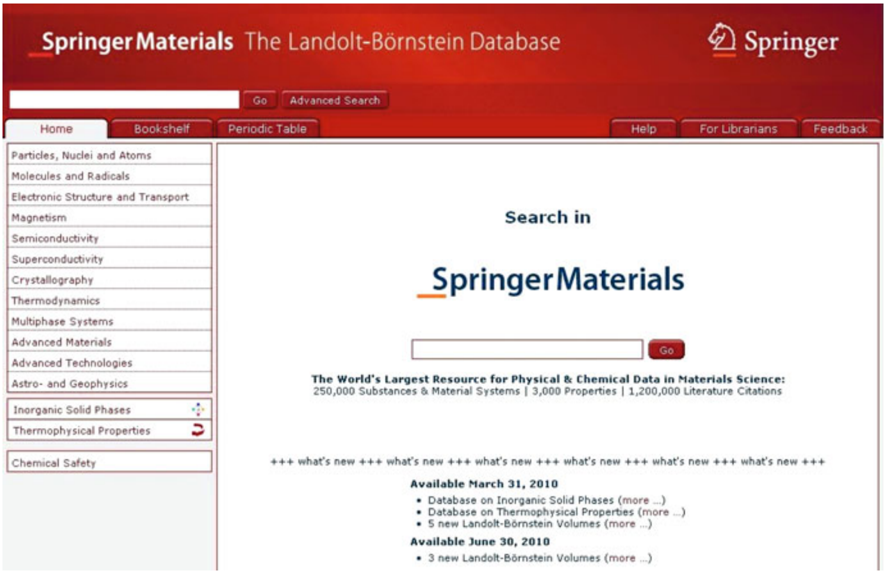


Fig. 4. Screen shot of SpringerMaterials.

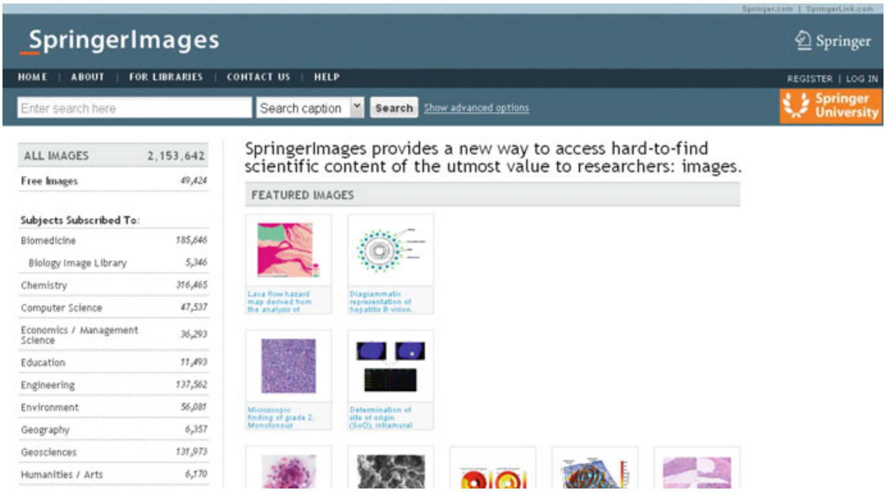


Fig. 5. Screen shot of SpringerImages.

Properties and the Linus Pauling Files, a Database on Inorganic Solid Phases and chemical safety data. SpringerMaterials is expanded and updated on a quarterly basis and offers advanced search & navigation functionalities.

4. SpringerImages
SpringerImages is a growing collection of scientific images that spans the scientific, technical and medical fields, including high-quality clinical images from images.MD (see Fig. 5). The continually updated collection –

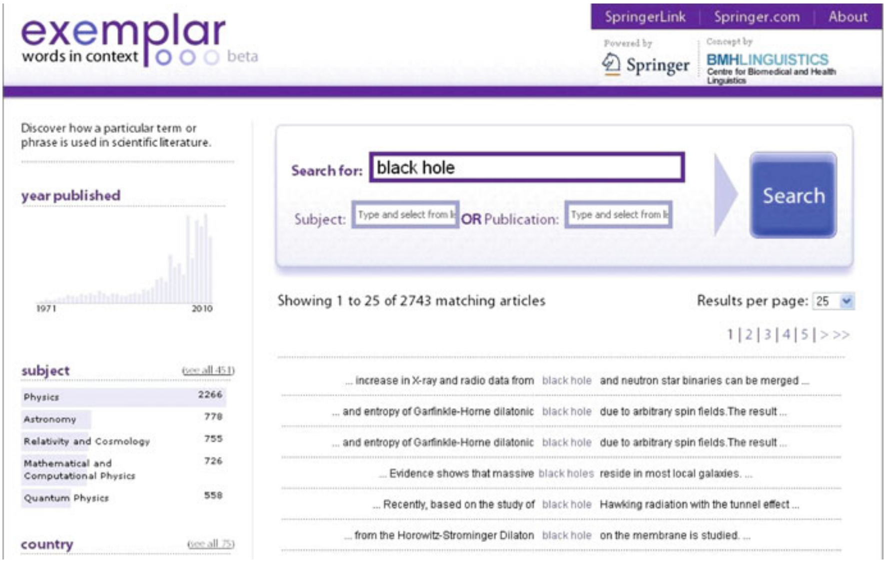


Fig. 6. Screen shot of the Exemplar output for the query “Black Hole”.

currently over 2 million images – gathers photos, graphs, histograms, figures, and tables, and is available to libraries and their patrons via a searchable online database. The SpringerImages interface enables users to search faster, more broadly and more accurately, through captions, keywords, context and more, even jumping from the image to the source article. Users can create personalized image “sets,” and can easily export images for use in their own presentations or lectures.

5. Exemplar
   Exemplar searches over 1,900 journals from Springer’s collection to find authentic examples of how a word or phrase is used in published literature (see Fig. 6). Comprehensive coverage includes both current and archival content in all major subject areas including the life sciences, medicine, engineering, mathematics, computer science, business, and law, contributed by some of the world’s leading academics in these fields. Exemplar is continuously updated with new content as it is published.
   Exemplar is for:
   Authors who want suggestions on how to express themselves or want to confirm how a particular word or phrase is used in published literature.
   Teachers who are looking for authentic phrases and sentences for teaching and testing.
   Editors and reviewers who want to confirm whether or not a particular turn of phrase is justified by usage.
6. SpringerNetworks
   Springer has developed a generic social network platform. Societies that are interested in using such a social network site for their users can utilize this

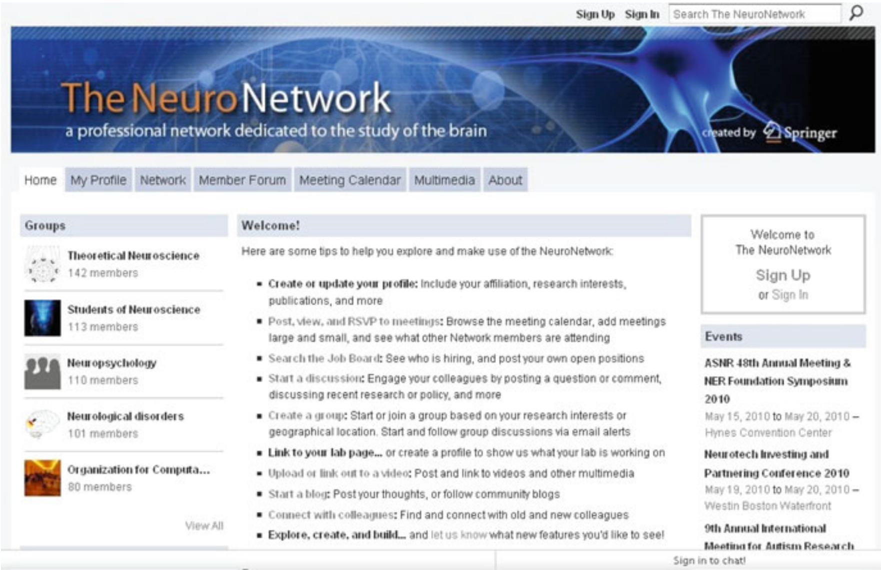


Fig. 7. Screen shot of the society website "The NeuroNetwork".

capability. A couple of examples are "The NeuroNetwork", a social network for studies of the brain (see Fig. 7), and the JMS Network a site for the Journal of Materials Science.

4 Future Directions

4.1 Business Model

As explained above, the whole publishing area has been changing radically in the last decade. This means that we as a publisher have to re-evaluate our business model. One of the most important aspects of this re-evaluation is how publishers get paid for the value added services that they provide. A large part of this evaluation centers on the possibility of Open Access models for journals. In this model, the authors (or societies or libraries or funding agencies affiliated with the authors) pay for the cost of publishing. They retrain the copyright for the articles published with Open Access, and the articles are made available for everybody worldwide without access restrictions.

Some web services (like SpringerProtocols) are made available on a subscription basis. Others (like AuthorMapper and Exemplar) are available for free as a general service by Springer.

The changes in the business model of publishing are slow and will take much more time before we will reach a stable new publishing environment. It will require a lot of cooperation between authors, societies, libraries, funding agencies, and publishers.

4.2 New Technologies

There are many new technologies available already now, and many more are being developed. All major publishers are working continuously to utilize these new capabilities. There is too much being developed to describe it here, and everything is too much in flux to provide any kind of comprehensive or detailed descriptions. Here are just a few highlights of things that are being developed or thought about:

1. All data tables downloadable in database formats.
2. Spreadsheet-like active data tables.
3. Active mathematical formulas. You can plug in your own parameters and get the results of the calculation with your parameters.
4. Animations of simulations with user specified parameters.
5. 3D images.
6. Semantic networking and clustering for finding related articles.

5 Conclusions

The world of publishing is changing rapidly. All the players in this field need to cooperate to make sure that the direction in which we are going leads us to a better publishing model. The new technologies have the potential to provide a much better vehicle for disseminating the results of scientific research, we just have to find the right model for utilizing these technologies. They also have the potential to enable scientists from developing countries to be much better involved in cutting edge research. Hopefully we will be able to use these new capabilities in order to further research and enable everybody who is interested in participating.

6

Publishing Astronomy and Astrophysics: Article Numbering, Electronic First

Marie-Louise Chaix

EDP Sciences, 17, avenue du Hoggar, B.P. 112, P.A. de Courtaboeuf,
F-91944 Les Ulis, France
marie-louise.chaix@edpsciences.org

Summary. Quick dissemination of validated scientific results is an important aspect of academic publishing. Over the years, the electronic editions of journals have become the primary vehicle for achieving that objective. To streamline the production of the journal *Astronomy and Astrophysics*, it has appeared useful to change from the classical "page-numbering system" to a more flexible "article-numbering system". In this paper, we recount the setting up of the "page-numbering system" in the publishing process of A&A. We also provide a quick review of publishing trends for the journal.

1 Introduction

EDP Sciences is a medium-sized French publishing company. It is a subsidiary of learned societies that works closely with the scientific world. It is involved in the communication and dissemination of science to specialist audiences (researchers, engineers, students etc.) and non-specialist audiences alike (general public, decision makers, teachers etc).

EDP Sciences produces and publishes international journals, books and Internet sites with a predominantly scientific or technical content. The editorial activities of the company cover astrophysics, applied and fundamental physics, mathematics, electronics, materials sciences, life sciences, and medical fields. More information about EDP Sciences may be found in the previous proceedings of the colloquium on the Future Professional Communication in Astronomy, Brussels, 10–13 June 2007, published by the Belgian Royal Academy of Sciences, Letters and Fine Arts in its in 8° Memoirs series (Vol. XXVII, n° 2047, 2007).

Since 2001, the journal *Astronomy and Astrophysics* has been published by EDP Sciences.

A. Accomazzi (ed.), *Future Professional Communication in Astronomy II*,
Astrophysics and Space Science Proceedings 1, DOI 10.1007/978-1-4419-8369-5_6,

2 About A&A

A&A was founded forty one years ago by the joining of some European journals. Now the A&A consortium includes 23 countries from Europe and South America, and each of them nominates a Director. The journal is owned by the Board of Directors with the copyright ESO. Scientists from sponsoring countries do not pay publication costs when publishing in A&A.

A&A publishes all new results of astronomical research, regardless of the technique used to obtain them, and accepts papers from any origin, i.e., it is not restricted to serving the sponsoring countries of the Journal.

The whole review process is overseen by two editors-in-chief and an editorial team of 10 scientific editors, assisted by language editors because most papers in A&A were written by non-native English speakers.

Astronomy and Astrophysics is a broad journal:

– About 2,000 articles from 60 countries
– 17,000 printed pages (about 19,000 published pages but some sections are electronic only)
– the *Astronomy & Astrophysics* ISI impact factor for 2008 is 4.153
– A&A is read in over 160 countries around the world.

A&A also has a dedicated website[1] with around 56,998 visits per month (15% Direct Traffic, 42% Referring sites and 43% Search Engines) and 80,000 PDF files/120,000 abstracts downloaded per month.

3 A Living Journal

It is interesting to note that the website is now a vector of life for the journal.

For the last few years, A&A has had a press release service that both produces its own press releases and collaborates on other institutes' press releases. The texts of the press releases, as well as the related articles, are freely available on the Press Releases web page (Fig. 1). The PRs are available on the A&A web site, via the A&A emailing list dedicated to journalists, via the web sites.[2]

With the agreement of the A&A Board of Directors, A&A publishes A&A Highlights. The highlights are intended to complement the press releases. They are directed at the astrophysical community rather than the general public, as press releases are. These are papers selected by the Editors of A&A, who wish to attract the readers' attention to some work in the issue that they find particularly exciting and/or intriguing for those outside the speciality. The highlights are available at the same time as the article is published electronically,

[1] http://www.aanda.org
[2] lphgalileo.org and urkalert.org

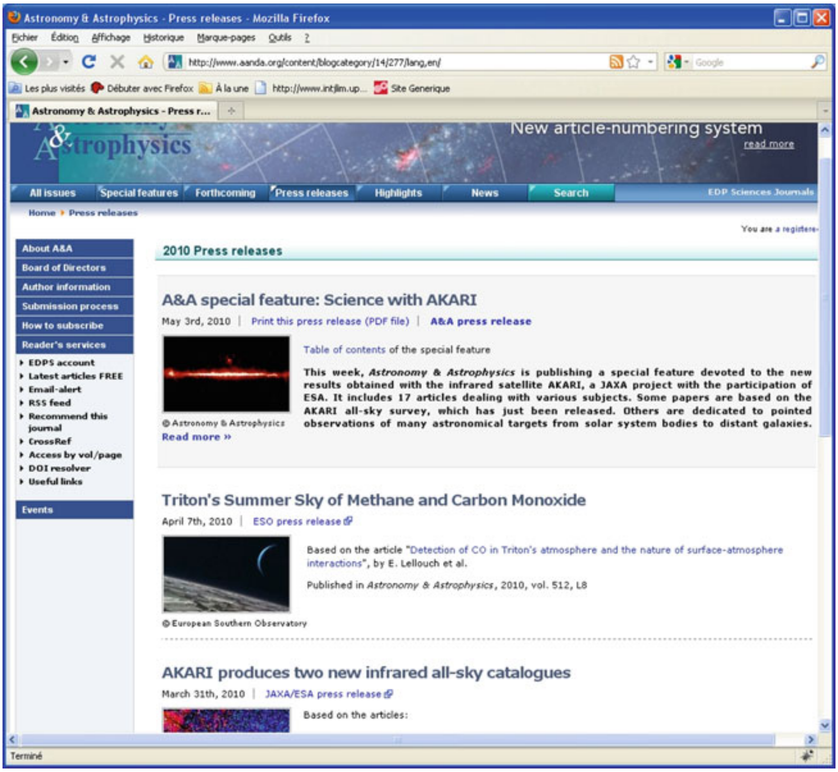


Fig. 1. Screen capture of the press releases service.

in both HTML and PDF formats. They are available from the homepage of the A&A web site, as well as directly through a link in the table of contents or in the TOC email (Fig. 2).

The editors regularly choose some figures that may be printed on the cover of the print volume. The editorial office has proposed to improve the visibility of the figures on the Web table of contents and to keep on selecting these figures. There are now several enhanced figures on the table of contents, which is formatted in two columns.

We can also cite the experience of the SWYA school.[3] SWYA was an *"Astronomy and Astrophysics School"* whose direct purpose was to teach students how to express their scientific results and how to write scientific papers for different purposes (journals, proceedings, thesis manuscripts, etc.).

All of these services contribute to increasing the visibility of the Journal.

[3] http://www.swya.org/

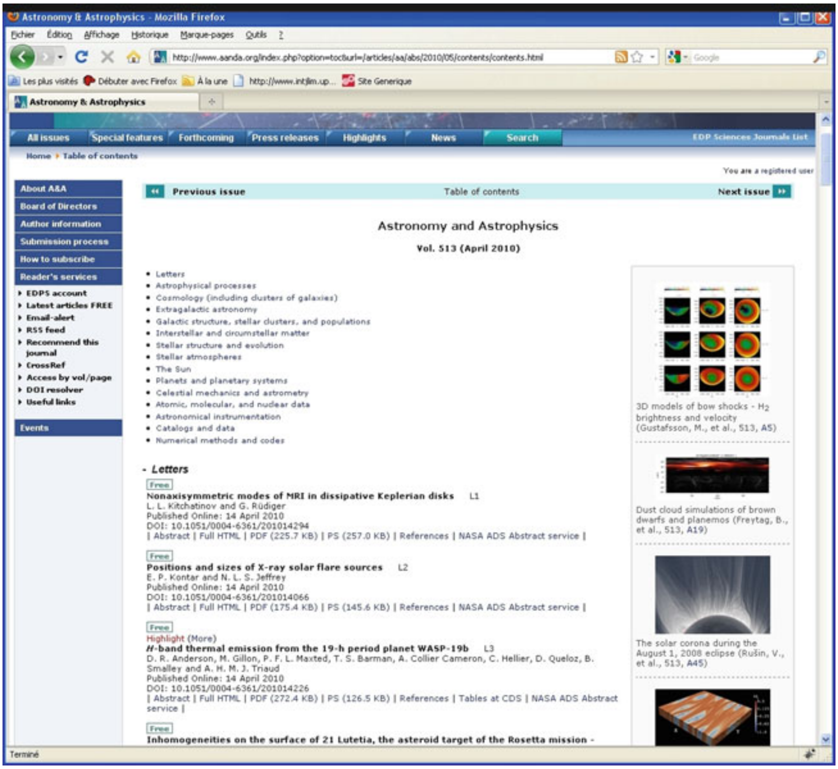


Fig. 2. Screen capture of the table of contents.

4 Astronomy and Astrophysics: Electronic First

Early on, A&A articles received a DOI identifier and were indexed in CrossRef. But there was also a permanent collaboration with the main ADS and CDS databases, for example with links to stellar objects or electronic tables provided by the CDS or links to the ADS data input.

Over the years, the electronic version of the journal has become the main source of fast access to validated results. Since the Scientific community considers the Electronic version of the journal as the main version, we have started to transfer certain sections of the journal to an exclusively electronic version. We also have subscriptions for the online version only, and the number of such subscriptions is rising (Fig. 3).

The A&A site has more and more services for readers. We can point out: "Articles from the same author" or "Related articles" which provides a hyperlink to articles having a common author or common keywords in the EDP Sciences database; "Alert me when my article is cited" and so on.

Another example is Dexter. The ADS has developed the Java applet Dexter, which helps users in the process of generating data points from

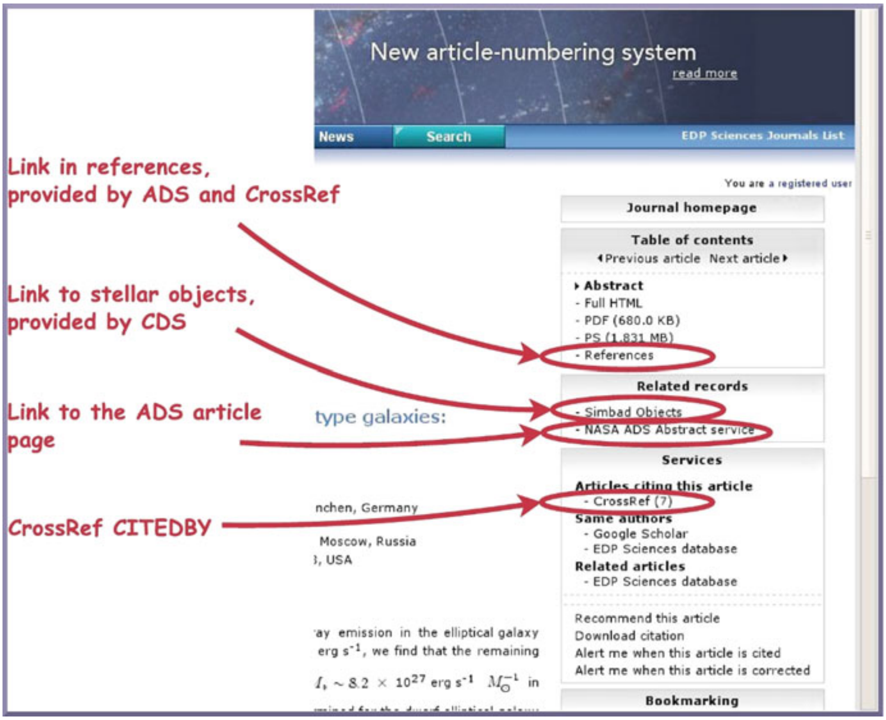


Fig. 3. Some of the specific services are highlighted on this screen capture.

published figures. Dexter has been adapted by EDP Sciences to work with A&A articles.

All the exchanges between the main databases and the EDP Sciences platform are summarized in Fig. 4. The exchange format is XML. All data are in XML with different DTD, depending on the different applications. Different tools have been developed to take into account these different DTDs.

5 A&A Has Made the Transition to Article Numbering

Quick dissemination of scientific results is an important aspect of academic publishing. To streamline the production of A&A, it has been useful to change from the classical "page-numbering system" to a more flexible "article-numbering system".

In the new article numbering system, the page number is simply replaced by an article number. Manuscripts accepted for publication appear in the currently open A&A volume as soon as they are ready to be made public. A&A users can see on the Journal website how the open volume is being filled up every day with newly published manuscripts. The A&A website displays

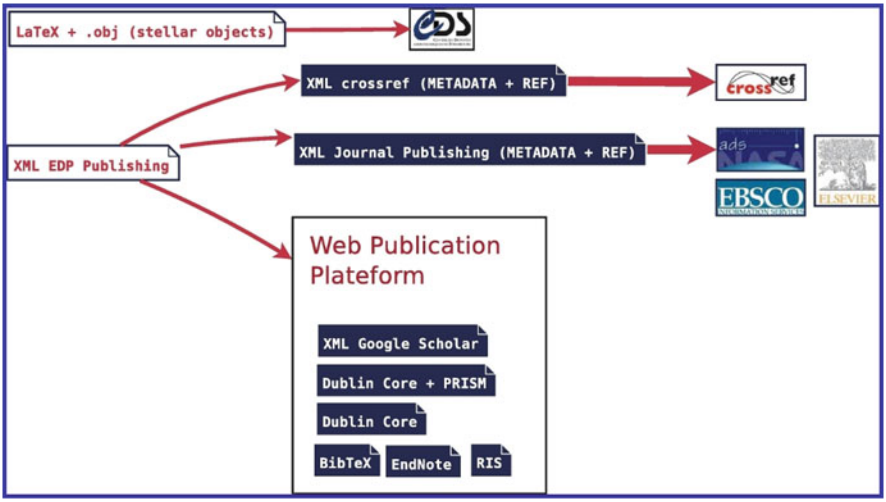


Fig. 4. Metadata spreading.

the published papers in the usual manner as before, with a table of contents divided into sections. At the closure date of the volume, its content is sent to the printer.

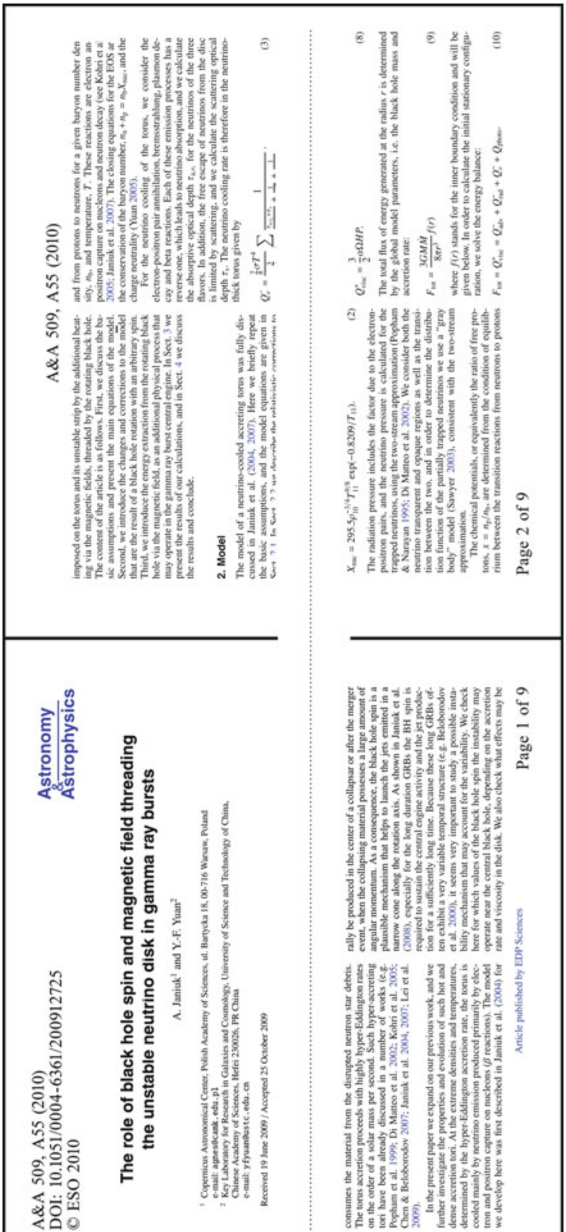
The Board of A&A have chosen to mark articles according to their class of publication using a roman letter: L for Letter, A for Article, E for Editorial, C for Corrigendum, etc. This letter is associated with a number based on the sequence in which the article is ready for publication in that volume of A&A.

Example:
L2 for the second letter in volume 509
A55 for the 55th article in volume 509
Citation
Jutzi, M., Michel, P., & Benz, W. 2010, A&A, 509, L2
Janiuk, A., & Yuan, Y-F., 2010, A&A, 509, A55

Citation is nearly the same, but the page number is replaced by the article number: L2, A55. This is a number preceded by a roman letter signifying the publication class.

The page numbering is internal to each manuscript. The first page is always 1. The page number is located at the bottom of each page and takes the form "Page *n* of *N*" where *N* is the total number of pages in the manuscript. The paper's reference is repeated as the running title of even-numbered pages.

Consequences:
on the tools:
– ADS, CDS take into account this new notation in the notation of the bib-code: 2010A&A...509A...55Z


A&A 509, A55 (2010)
DOI: 10.1051/0004-6361/200912725
© ESO 2010

Astronomy & Astrophysics

The role of black hole spin and magnetic field threading the unstable neutrino disk in gamma ray bursts

A. Janiuk[1] and Y.-F. Yuan[2]

Received 19 June 2009 / Accepted 25 October 2009

Article published by EDP Sciences

Page 1 of 9

A&A 509, A55 (2010)

2. Model

Page 2 of 9

Fig. 5. The new page layout for articles.

– Bibtex:

```
...
volume = 510,
pages = {A2-+},
...
-/+ in LaTeX means that the pagination is not a simple pagination
like 1-10.
```

The electronic version of an issue was already available around 3 or 4 weeks before the subscribers receive the printed issue. With the article number, articles are online around 10–15 days earlier. A&A is now printed by volume, approximately once a month. There is no longer a weekly issue.

6 About Open Access

Open Access (OA) means free, unlimited access to published peer-reviewed results of scientific research. A&A has already made a first move (Fig. 5):

– articles are in OA for 7 days after their initial publication.
– letters and electronic-only sections are in OA
– three years after publication, the contents of A&A are in full OA
– the discussion about Toward full open access is ongoing, and the Board is examining this question.

7

The Future of the ASP Conference Series

Joseph B. Jensen,[1] Jonathan Barnes,[2] J. Ward Moody,[3] and Paula Szkody[4]

[1]Utah Valley University, 800 W. University Parkway, Orem, UT, 84058-5999, USA
jjensen@uvu.edu

[2]Salt Lake Community College, 4600 South Redwood Road, Salt Lake City, UT, 84123, USA
Jonathan.Barnes@slcc.edu

[3]Brigham Young University, N484 ESC, Provo, UT, 84602, USA
jmoody@byu.edu

[4]University of Washington, Box 351580, Seattle, WA, 98195-1580, USA
szkody@astro.washington.edu

Summary. The Astronomical Society of the Pacific (ASP) has been publishing the proceedings of conferences in astronomy and astrophysics for more than 20 years. The ASP Conference Series (ASPCS) is widely known for its affordable and high quality printed volumes. The ASPCS is adapting to the changing market by making electronically published volumes available to subscribers around the world, including papers in the Astrophysics Data System (ADS) database, and allowing authors to post papers on e-print archives. We discuss the role of the printed book in our future plans, and how electronic publishing affects the types of products and services we offer. Recently there has been increasing pressure in the academic world for open access (electronic copies of scholarly publications made freely-available immediately after publication), and we discuss how the ASPCS is responding to the needs of the professional astronomical community, the ASP, and humanity at large. While we cannot yet provide full open access and stay in business, we are actively pursuing several initiatives to improve the quality of our product and the impact of the papers we publish.

1 Background

The Astronomical Society of the Pacific Conference Series (ASPCS) publishes the proceedings of professional conferences and symposia in astronomy and astrophysics. Scholarly publishing is changing rapidly, and the ASPCS is adapting to the changing market in a number of ways described in this contribution. We also reported on the ASPCS's future initiatives and policies regarding open access at the recent Library and Information Services in Astronomy meeting (LISA VI) held in Pune, India on 12–13 February 2010 (Jensen, Moody, & Barnes 2010).

A. Accomazzi (ed.), *Future Professional Communication in Astronomy II*,
Astrophysics and Space Science Proceedings 1, DOI 10.1007/978-1-4419-8369-5_7,

The Astronomical Society of the Pacific, an international, nonprofit, scientific and educational organization, has served the professional, amateur, and educational astronomical communities since 1885. The ASP was born following a meeting between astronomers from the Lick Observatory and members of the Pacific Coast Amateur Photographic Association shortly after the New Year's Day total solar eclipse in 1889. Edward Holden, Lick's first director, complimented the amateurs on their service to science and proposed to continue the good fellowship through the founding of a Society "to advance the Science of Astronomy, and to diffuse information concerning it." The ASP continues this mission today, partly through the Publications of the Astronomical Society of the Pacific (PASP), the ASP Conference Series, and Mercury magazine.

In 1988, Harold McNamara, the PASP editor at the time, founded the ASP Conference Series at Brigham Young University. Over more than 20 years, the ASPCS has published 420 volumes. We currently publish about two Conference Proceedings volumes per month. These volumes now contain more than 30,000 papers by 35,000 authors from countries around the world. ASPCS contributions describe a wide array of research covering all aspects of astronomy and astrophysics, and include volumes on public education and outreach in astronomy, astronomical publishing and archiving, computer software and data processing, summaries of space missions, and so forth. The ASPCS represents a unique publication venue for a range of topics, not all of which are usually published in the regular journals. For many of our authors, the ASPCS article is the only published version of their work, making a professional, high-quality publication very desirable. Recently the ASPCS has initiated a Monograph series, where appropriate reference materials can be published. The ASPCS has developed a reputation among professional astronomers for their affordable, high-quality printed volumes.

The way the professional astronomical community uses published materials today is vastly different than when the ASPCS printed its first volume more than 20 years ago. Most astronomers access materials through on-line subscriptions to journals, find bibliographic information and links through the Smithsonian's Astronomical Data System (ADS) funded by NASA, and read pre-prints on the `arXiv.org` server. Astronomers today rarely venture to the library to retrieve a specific article that can be found more quickly on-line. In spite of the popularity of electronic articles, the printed ASPCS volumes remain popular. It is still easier to come up to speed in a field by browsing one of the topical ASPCS books than it is to get the same overview by searching for and retrieving individual articles on-line. Astronomical conferences provide a timely snapshot of progress in a field, and the ASPCS volumes capture that snapshot both for meeting participants and those who did not attend. Although on-line use of the ASPCS volumes has increased dramatically, the demand for printed books has remained relatively stable. While many libraries have cut their subscriptions for printed books to save money or space, almost all have maintained subscriptions to our electronic books, and conference attendees still like to have the printed books on their shelves.

2 The Advantages of Open Access

The goal and desire of most astronomers is to have their work reach as wide an audience as possible. Open access, as we use the term in this paper, refers to rapid and public distribution of an article, so that anyone in the world with a computer and an Internet connection could easily find and read the article as soon as it is published. Such open access benefits the author, the astronomical community, and the rest of society. The typical astronomer works in academia or in the public sector; those organizations usually reward their employees with higher ranks and increased salaries (e.g., by awarding tenure) based, at least in part, on the number and impact of their scholarly publications. It is therefore in the author's best interest to have her work seen, read, and cited as often as possible. Open access enhances the impact of the author's work, particularly in places around the world that may not have subscriptions to the journal in which the article was published.

The astronomical community also benefits from open access. Organizations could save money on subscriptions if articles were available for free. More importantly, astronomers without subscription access would no longer have to wait 2 or more years for articles to enter the public domain. In the words of Heather Joseph in the LISA V conference, "open access to the results of scholarly research has the potential to help individual researchers vastly increase the visibility, usefulness and impact of their work. Perhaps even more importantly, its potential benefits extend far beyond individual researchers to institutions, the scholarly community, and to society as a whole" (Joseph 2007).

The astronomy community is relatively unique in the academic world in that it has well-developed open-access services that are already in wide use. Astronomers can freely post and access electronic copies of their papers ("e-prints") via `arXiv.org`, and the ADS provides an extensive searchable database of the astronomical literature. Access to these services has not reduced demand for the scholarly journals, which still provide an authoritative version of the final work blessed by the peer-review process and the good names of the journals and the organizations that publish them. The journals, which generally hold the copyright on the articles, have not prevented authors from posting versions of their work on-line with `arXiv.org`.

In the book "The Access Principle: the Case for Open Access to Research and Scholarship," Willinsky (2006) makes a strong case for open access. He starts by defining the "open access principle" as follows: "a commitment to the value and quality of research carries with it a responsibility to extend the circulation of this work as far as possible, and ideally to all who are interested in it and all who might profit by it." In his book, Willinsky outlines the advantages of open access to scientists and to humanity in general, and points out that open access increases the impact of research. Willinsky even argues that open access is a fundamental right, and that the benefits of science to humanity should never be restricted to those with the most resources.

It is important to remember, however, that open access alone is not sufficient. For the full impact of scholarly work to be realized, we must have the technology to interface to the material. Not only do we need computers and Internet connections, but we also need archive servers, data protection systems, sophisticated indexing and database software, and cooperation of people in the computer and publishing worlds.

The ASPCS falls somewhere between the less-formal e-prints and the final journal articles. ASPCS articles are formally published and cited, so authors are expected to produce a polished and well-written proceedings article. Articles are not generally refereed; rather ASPCS articles are selected by editors and conference organizers, who choose the contents of their meetings and are responsible for the quality of the proceedings submissions. The ASPCS maintains high standards for the quality of the publications, and our volumes have a strong reputation as being the definitive conference series for the astronomical community. ASPCS articles are widely circulated and cited among the astronomical community.

Because conference presentations often represent work in its early stages, it is a unique window into the status of a field at the cutting edge. Conference Series articles are often the first publication of a new result or data set. We expect that many authors will publish the final results in refereed journals after the results have been refined and confirmed, but many other types of conference contributions only appear as Conference Series articles. It is therefore crucial that ASPCS articles be published quickly and be widely accessible, because their value and relevance will often decline with time, particularly if the final versions are printed in refereed journals. Other types of contributions, such as review articles or papers on topics like software development or public outreach, are not usually published in refereed journals; for these contributions, the ASPCS represents the only publication of the work, and it is therefore important for these to be made publicly available to maximize their impact. The ASPCS allows authors to post their papers with e-print servers, and frequently grants permission for authors to reprint copies of their work elsewhere, while at the same time maintaining proprietary subscription or purchase access for a period of 2 years. After 2 years, all ASPCS articles are available for free via the ADS server.

The ASPCS recognizes the value of open access both to the astronomical community and to humanity. Immediate open access of our volumes would increase the impact of a meeting, effectively sharing the results with a wider audience. It would allow participation in the field by people who would not normally be invited to a meeting or could not attend the meeting for financial or other reasons, such as students, interested amateurs, or the public. Increased readership would potentially increase support for the ASP as well, by drawing increased attention to its other missions (e.g., education and public outreach). It is not economically feasible, however, to provide full open access to Conference Series volumes at the present time. We have reduced our proprietary time from 3 to 2 years, but further reductions in the proprietary time will require revisions to our business model to maintain the revenues required to continue publishing a high-quality volume. In Sect. 6 we discuss the detailed implementation of open-access plans for the ASPCS, and identify some of the difficulties with a full open-access business model.

Open access is necessarily a digital endeavor, since no other technology for distributing information can approach electronic publishing if the goal is rapid, free, and universal access. This places open access issues squarely at the interface between traditional printing and electronic publishing. It is therefore helpful to review how other organizations and governments are dealing with questions of open access and electronic publishing.

3 U.S. Government Policy on Open Access

In the United States, there is increasing pressure to make work sponsored by the government using public funding freely available to the public within a year of the original publication. This seems reasonable enough, provided that public funding is also provided to cover the costs associated with publishing the work. Many journals are published and marketed by non-profit scholarly organizations or for-profit companies, and publishers in general have resisted efforts by the U.S. government to take control of the products they currently produce and market. Companies and non-profit organizations add value to the final published work in several important ways. They provide reviews of the work by peer referees, publish and market printed versions, make electronic versions accessible via web services, and provide long-term archiving and reprinting services. Some of these services can be provided by the public sector, but at present, publishers cannot stay in business without charging authors and readers for the products and services they provide.

The U.S. federal government has instituted an open access policy with research funded by the National Institute of Health (NIH). Initially, the NIH asked authors to post electronic copies of their papers on an NIH server on a voluntary basis. After a trial run of several years, they found that compliance was quite low. Authors were frequently confused about the legality of posting papers for which the journal publishers held the copyright. Last year, NIH began requiring authors to post e-prints to the PubMed Central open access server as soon as they are accepted for publication; papers are made freely available no more than 1 year after publication. Scientists in many fields are watching this experiment with great interest. If it turns out to be successful, it is possible similar policies could be applied to other U.S. government-funded research, such as that funded through the National Science Foundation (NSF) or National Aeronautics and Space Administration (NASA).

In January 2010, the U.S. government Office of Science and Technology Policy released their "Report and Recommendations from the Scholarly Publishing Roundtable" (2010). The report identified the following shared principles: peer review, adaptable business models, increased access, sustainable archives, and creative reuse of results. The report stopped short of recommending full open access, but rather advised that the stakeholders (public and private) work together to develop policies. The report recommends free access to the "version of record" after an embargo period of 1 year or less. The report further encourages innovation in establishing new business models and ensuring the digital preservation of materials. Many of these suggestions are already very much a part of publishing in astronomy, with the notable exception that the embargo period for the U.S. astronomical journals published by the American Astronomical Society (AAS) is 2 years. The ASPCS and PASP embargo period was recently changed to be 2 years as well.

4 Open Access and Publishing in the Private Sector

Publishing scholarly work involves much more than just printing an author's paper or posting it on a web site. "Publishers are there … to support the authority, quality, accessibility, longevity, and recognition of scholarship" (Willinsky 2006). These are important tasks, and they do not come free. It takes the efforts of dedicated professionals to ensure that scholarly works are published, marketed, and preserved. In the astronomical community, the cost of these services is usually covered by a combination of charges to the authors and to the readers.

The AAS publishes two of the preeminent journals in astronomy: the Astrophysical Journal (including the Letters and Supplement series) and the Astronomical Journal. The AAS is a non-profit organization that publishes these journals on behalf of the international astronomical community. To fund the editorial process including referee reviews and printing, authors pay page charges, and the AAS sells subscriptions to pay for distribution and archiving. The AAS does not use journal revenues to pay for other Society expenses. To maintain a viable subscription service, AAS journals have a 2 year proprietary period. As of the present time, the AAS has not prevented authors from posting electronic versions of their papers on eprint servers such as `arXiv.org` (Huchra 2010).

The pressure on the AAS to comply with a potential U.S. government mandate along the lines of the NIH policy prompted John Huchra, President of the AAS, to address publication issues on the front page of the AAS Newsletter (Huchra 2010). In the same newsletter, Kevin Marvel wrote that "a government mandate will disrupt our publishing model and require significant changes to how we manage our journals" (Marvel 2010). In the newsletter article, Huchra argues that the AAS cost split between author and subscribers is fair and close to optimal, and acknowledges the need for quality in publication (including peer referees) and for providing long-term, reliable access. He notes that the contributors and readers of the AAS journals are distributed around the globe.

The ASP faces many of the same challenges as the AAS. Like the AAS journals, the PASP pays for publication with a combination of author page charges and subscription revenues. The ASP Conference Series depends on income from conference participants, subscribers, and additional sales to pay for publication. The ASP is an international organization, and it is unclear what effect future U.S. government policy changes will have on ASP publications.

The issue of electronic publishing and open access (or pricing for electronic access) is much larger than scientific or scholarly publishing. Recent newspaper headlines recount the battles between book publishers, who need to make money, and book sellers, who discount book prices as low as possible to attract buyers. The crux of the issue is that consumers seem to think that electronic publications should be free – or as close to free as possible – while authors and publishers must recover their costs and make a profit to survive. Many readers, even in the astronomical community, fail to recognize that the majority of the cost of publishing is not in printing costs, which are typically only a minor fraction of the cost of bringing a publication to market (in whatever form). As reported in the New York Times just before the LISA VI conference, "publishers have managed to take control – at least temporarily – of how much consumers pay for their content" (Rich 2010). It seems at present that book sellers like Walmart or Amazon will not always be able to reduce prices for electronic books at publishers' and authors' expense.

5 The Future of the ASP Conference Series

The ASP Conference Series is used extensively by professional astronomers all around the world. ASPCS volumes are the recognized standard in astronomy conference proceedings. Many volumes present a comprehensive overview of the current status of research in a particular field and contain cutting-edge results not found elsewhere. The ASPCS Conference Series is often used to collect results from specific satellite missions, astronomical instruments, or experiments. Organizers of recurring meetings often choose the ASPCS for consistency in quality and support. The PASP publishes peer-reviewed articles covering a wide array of topics in astronomical research. It is particularly popular as a location to publish review papers, overviews of astronomical instrumentation, new observing or data processing techniques, and results of scientific missions or surveys.

A survey of 639 refereed articles published by Harvard Center for Astrophysics (CfA) astronomers conducted in 2008 by Michael Kurtz of the CfA showed that the PASP was the seventh most cited source among astronomical journals and the ASPCS was the eighth most cited, and the only conference series in the top ten. Approximately 60% of the articles surveyed made reference to one or more ASPCS articles. A recent survey of ASP members and professional astronomers showed that as many people read the printed Conference Series books as access electronic versions. The distinctive blue and white hard-back volumes remain popular. We also track visits to the electronic article repository on the web (http://www.aspbooks.org and http://www.aspmonographs.org). People from 195 countries and territories visited the ASPCS web site last year, which provides access to electronic versions of all our articles. Most visits were from the United States, but significant numbers also came from the United Kingdom, Germany, India, Canada, Italy, Japan, France, Spain, Australia, and many other countries, for a total of nearly 150,000 visits in a 12 month period. The ASP publications are reaching a wide international audience.

The ASPCS publishes on average 24 volumes per year. While the variation from year to year is significant, we appear to have a stable demand on the part of conference organizers. During 2009 and 2010, because of the global economic recession, fewer astronomical conferences were organized. We planned for a reduction in the number of published volumes, but are finding robust demand in spite of the economy. It appears that the conferences that are going forward are still publishing proceedings volumes, and that the conference organizers recognize the printed ASPCS volume as a way to guarantee the impact of a conference. We expect to publish approximately 30 volumes in 2010.

The ASP Conference Series is completing a period of transition. During the past couple of years, the ASPCS has added and enhanced access to electronic volumes and provided subscriptions to libraries interested in both electronic and print editions. Last year, Joseph Jensen took over as the new Managing Editor of the Series, relieving J. Ward Moody, who served as Managing Editor from 2004 to 2009. The transition of Managing Editor was accompanied by a relocation from the office at Brigham Young University where the series was born to a new office at Utah Valley University, a few miles down the road from BYU. Dr. Jensen brings to the ASPCS a renewed commitment to the quality and relevance of the Conference Series. His primary goals, which are shared by the entire ASPCS staff, are to publish volumes more quickly, to reduce the amount of work required of authors and editors, and to maintain the high quality for which the Conference Series is well-known.

To achieve these goals, the ASPCS has released new versions of the LaTeX style files and templates, including a new updated set of detailed instructions to go with them. These new files will allow authors and editors to spend more time assembling their volume and less time trouble-shooting formatting errors. ASPCS will continue to use LaTeX for publication of its volumes because it produces a high-quality and consistent printed product. To further improve the quality of the publication, ASPCS is encouraging more use of color in printed volumes, and is taking advantage of reduced printing costs for color books. To help get books published more quickly, the ASPCS has made various changes to the standard contract form to encourage editors to complete their work in a timely fashion. The new expectations are implemented in a set of financial incentives (e.g., price discounts) for authors and editors who complete their work and publish within 6 months of the conference or symposium. This will benefit authors and meeting organizers by increasing the impact of their meeting, reducing its cost, and preventing the editing process from dragging on for an extended period of time, impacting other aspects of their work.

The ASP has recently conducted a survey of professional astronomers, educators, and other astronomy enthusiasts. Of those that read, reference, and publish Conference Series articles, most refer to ASPCS articles a few times per year and publish once every few years. Interestingly, just as many ASPCS readers pick up printed volumes – either their personal copies or from the library shelves – as read electronic versions on-line. We suspect that the persistent popularity of the printed book in a world with nearly instantaneous access to electronic searching and reading is the fact that flipping through a printed volume is still the most enjoyable and efficient way to get an overview of a particular field in astronomy. This is particularly important for graduate students looking for thesis ideas, post-doctoral fellows entering a new field of research, or senior astronomers catching up on a field neglected due to administrative or teaching duties. The ASPCS will continue to print and market the high-quality volumes that have been so instrumental in establishing the reputation and brand recognition of the Conference Series. At the present time, we do not plan to provide "print on demand" services, although printing small numbers of the books is becoming increasingly affordable given advances in digital printing technology. The on-line versions of ASCPS articles are particularly popular when accessed through the ADS archive and search tools. The ASPCS will continue to develop its products for use through this important research tool, and look for ways to improve access to those who find the article titles and abstracts freely available through ADS but do not have access to the full volumes during the first 2 years through an institutional subscription.

The survey also provided important feedback to the ASPCS staff regarding which qualities and services the community finds most important and useful. Among the most important characteristics, our users widely agree that the value of the ASPCS is in providing an archival record of the conference or symposium. Equally important to our readers is electronic open access. These characteristics were far more important to our users than such things as the type of binding or format of the printed book.

For future development, survey respondents suggested that open access is more important than providing meeting proceedings in other formats (transcripts, audio, or video recordings). At the present time, the ASPCS does not plan to provide services such as web-hosting of presentation materials for practical reasons. Maintaining publication quality and avoiding copyright infringement would make publishing presentation materials or slides legally and technically challenging, while providing little of real value to the readers of our series; presentations are easy for authors to provide, but difficult for readers to interpret and practically impossible to reference properly. We are now actively promoting full-color printing of volumes, which has recently become more affordable due to advances in printing technology. We also offer publication of ancillary materials, such as poster contributions, in the electronic version of the volume. The ASPCS is simplifying the process of producing a volume to help authors and editors get their work into circulation faster, thereby enhancing its impact and value to humanity. By constantly upgrading and enhancing our printed and electronic products, we expect to remain a popular choice among meeting organizers for many years to come.

6 The ASP Conference Series and Open Access

How is the ASP Conference Series responding to the increasing pressure to provide free access to our articles? In some ways, publication of our proceedings volumes is no different than other professional journals. The majority of the cost of publication is not printing, but providing all the other services related to assembling the volume in a format that can be printed, archived, searched, and accessed electronically. Like other professional journals in astronomy, the ASPCS shares publication costs fairly equally between authors and readers. Making publications available for free would require redistributing production costs and/or significantly reducing revenues. As a non-profit organization, the ASP depends on Conference Series revenues to help fulfill its institutional objectives, including publishing scholarly articles, supporting astronomy education, and building bridges between the professional astronomical community and the public. For the ASPCS to become a fully open-access publication, a new source of revenue would have to be identified, and a new business model adopted. Without sales income, the ASPCS, if not the ASP, would go out of business.

At present, the ASPCS revenue comes from three roughly-equal sources: sales to conference participants or conference organizers, institutional subscriptions, and sales of books (electronic or print) after the conference. This distribution of cost-sharing between authors and readers is achieved by having a 2-year period during which articles (including abstracts) are listed in the ASP and ADS databases, but are only fully accessible to purchasers of the book or institutions with subscriptions to the series (the subscription access period has now been changed to 2 years). If we were to make articles freely available from the date of publication, we would lose half to two-thirds of our revenue stream. Subscriptions would have no value, and most post-conference readers would choose not to buy a book they could read for free on-line. The situation for the PASP is similar. About half the PASP revenue comes from subscriptions, so full open-access would result in a 50% loss of income.

Could we pass the full cost of publication on to conference organizers and participants? If the price per book for conference participants were doubled or tripled, we can be fairly certain that a significant fraction of conference organizers would choose not to publish. Our price would no longer be competitive with commercial publishers. The survey of ASPCS authors and readers showed that the top two reasons people choose the ASP Conference Series for their meeting proceedings publication were the low cost of publishing with us and the reputation of the publisher. We publish about 24 volumes per year, and reducing this number significantly to the handful of conferences that could get funding to cover the full cost of publication would likely drive the Conference Series out of business. It is just not equitable to expect the authors and conference participants to shoulder the full burden of paying for the proceedings volume.

Could we reduce publication costs enough to make open access feasible? Only about one-third of our publication cost is for printing, so going to an electronic-only format cannot save enough money to make up for the 50% or 60% shortfall in revenue that would result from loss of subscriptions and sales. Such a large cut would require a significant reduction in other services, many of which are still required to support an open-access electronic publication. Open access is not just about posting articles on a web server; it requires searchable, durable archives, inclusion in databases like ADS, and promotion to make sure the best conferences are represented in the series. Cutting these kinds of services would do more damage to the community than good. In fact, development of the hardware and software needed to support a full open-access publication could easily cost as much on a continuing basis as we would save by discontinuing the printed books.

Can the ASP find other sources of income to pay for publication expenses? First of all, the institutional libraries that support the ASP via subscriptions have very limited budgets, and a factor of two or three increase in subscription prices for a product that would be available for free is an obvious dead-end. It is possible that other public-minded non-profit organizations could make contributions to support the ASP, but at present we know of no such organization willing to pick up the tab for the readers of our books and journals. We would also need such support to be committed for the long-term if we want to remain in business through the inevitable economic ups and downs. What about the U.S. government? If they impose requirements on scholarly organizations like the ASP, it is reasonable to expect they would provide funding to make it happen. The ASP is an international organization, however, with authors and readers around the globe. We shouldn't expect the U.S. citizen to pay for international open access, nor would we expect government funding to necessarily be stable over the long term. None of these possibilities seems to be as equitable as the current system.

Our printed books are still very much in demand, and we must balance quality against publication cost. The ASP exists to serve the astronomical community with high-quality publications to increase the impact of conferences and the work of individual authors. At present, we feel that providing open access to Conference Series volumes and PASP journal articles will result in financial damage that would prevent the ASP from achieving its mission. There are some important ways we can move towards the goal of increasing impact, however, without adopting a fully open access model. First, we will do everything we can to increase the quality and timeliness of our publications. We have taken several initiatives to do just that, as described in Sect. 5. Of prime importance is reducing the time between the conference and publication of the proceedings. If we can significantly reduce this time – to 6 months, on average – we will be able to increase the impact of the publications. We have also reduced the embargo period for PASP and ASPCS from 3 to 2 years. We hope that reducing the proprietary period while increasing publication speed will satisfy authors, subscribers, and the many readers around the world, while still maintaining the revenue needed to allow the ASP to fulfill its mission.

References

Huchra, J. 2010, "President's Column," AAS Newsletter 150, 1

Jensen, Joseph B., Moody, J. Ward, and Barnes, J. 2010, in Library and Information Services in Astronomy VI, 21st Century Astronomy Librarianship: From New Ideas to Action, eds. Isaksson, E., Lagerstrom, J., Bawdekar. N., and Holl, A., ASP Conference Series, 433, 174

Joseph, H. 2007, in Library Information and Services in Astronomy V: Common Challenges, Uncommon Solutions, eds. Ricketts, S., Birdie, C., and Isaksson, E., ASP Conference Series, 377, 121

Marvel, K. B. 2010, "From the Executive Office," AAS Newsletter 150, 2

Rich, M. 2010, "Publishers Win A Bout in E-Book Price Fight," The New York Times, Feb. 9, 2010, C1

U.S. Office of Science and and Technology Policy, 2010, "Report and Recommendations from the Scholarly Publishing Roundtable" (http://www.aau.edu/WorkArea/showcontent.aspx?id=10044)

Willinsky, J. 2006, The Access Principle: the Case for Open Access to Research and Scholarship (Cambridge, Massachusetts: The MIT Press)

8

Astronomical Publication Rates in the US, UK, and Europe

Helmut A. Abt

Kitt Peak National Observatory, National Optical Astronomy Observatory,
P.O. Box 26732, Tucson, AZ 85726-6732, USA
abt@noao.edu

Summary. I explored the growth of astronomical research in the US, UK, and four productive European countries to see if there has been any leveling off. I counted pages in the four major astronomical journals, and corrected for format changes and for contributions from other countries. The four European countries were France, Germany, Italy, and the Netherlands. In each area there has been no reduction in output. However, the data show that the UK lags behind the US by 10 ± 1 years and the four European countries lag the US by 12 ± 1 years.

1 Introduction

Since the 1930s the astronomical publication rate in the US has grown by 8.8% per year, but the population has increased by only 1.1% per year. At what stage will the publication rate be reduced to the rate of population increase?

In astronomy most (55%) of the research papers have appeared in just four journals that have impact factors >2.0, namely Astronomy & Astrophysics (A&A), Astronomical Journal (AJ), the Astrophysical Journal (ApJ), and the Monthly Notices of the Royal Astronomical Society (MNRAS). Also, that has been true for 40 years, so we can easily trace the growth of research in Europe, UK, and US during that time. This is much more difficult and uncertain to do in other sciences.

2 The United States

I counted the numbers of pages per year in AJ and ApJ (including the Letters and Suppl.) during each fifth year from 1959 to 2009. I corrected for the changes in format to 1,000-word pages and corrected for foreign input. For the ApJ, Fig. 1 (top curve) shows that the American contribution varied from 90.9% in 1959 to 51.0% in 2009.

A. Accomazzi (ed.), *Future Professional Communication in Astronomy II*,
Astrophysics and Space Science Proceedings 1, DOI 10.1007/978-1-4419-8369-5_8,

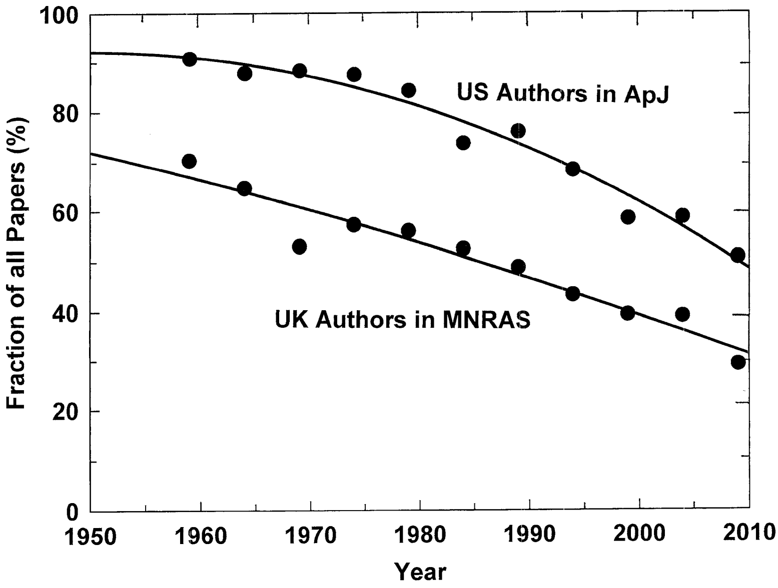


Fig. 1. (*Top*) The fraction of American papers, as judged by the addresses of the first authors, in ApJ as a function of time. The fitted curve is $F = 92.3 - 0.017\ (Y - 1950) - 0.012\ (Y - 1950)^2$ with a cross-correlation of 0.983. (*Bottom*) The fraction of UK papers in MNRAS as a function of time. The fitted curve is $F = 72.2 - 0.538\ (Y - 1950)\ - 0.00235\ (Y - 1950)^2$ with a cross-correlation of 0.965.

Since 2000 much tabular material was published on-line only in American journals. In fact, in 2009 the on-line tables were equivalent to 30.6% of the printed pages. The counting was corrected for that. Long on-line tables are far less frequent in MNRAS and A&A. The ApJ pages, printed plus on-line equivalent, showed a growth of 6.2% per year without any evidence of a leveling off, as will be shown below.

For AJ I collected data in the same way. Figure 2 (top curve) shows the fraction of the AJ published by American authors. The fitted curve ranged from 85.1% in 1959 to 59.1% in 2009. In 2009 the on-line pages were equivalent to 59.1% of the printed pages. The AJ pages, printed plus on-line, showed a growth of 6.7% per year with no evidence of a leveling off.

Figure 3 (left curve) shows the result of combining the ApJ and AJ, and dividing by the appropriate US population in each year. The population ranged from 177,829,628 in 1959 to 305,529,237 in 2009, an increase of 71.8% in 10 years. There is no evidence for a leveling off in pages per million people. I will comment later on the other two curves in Fig. 3.

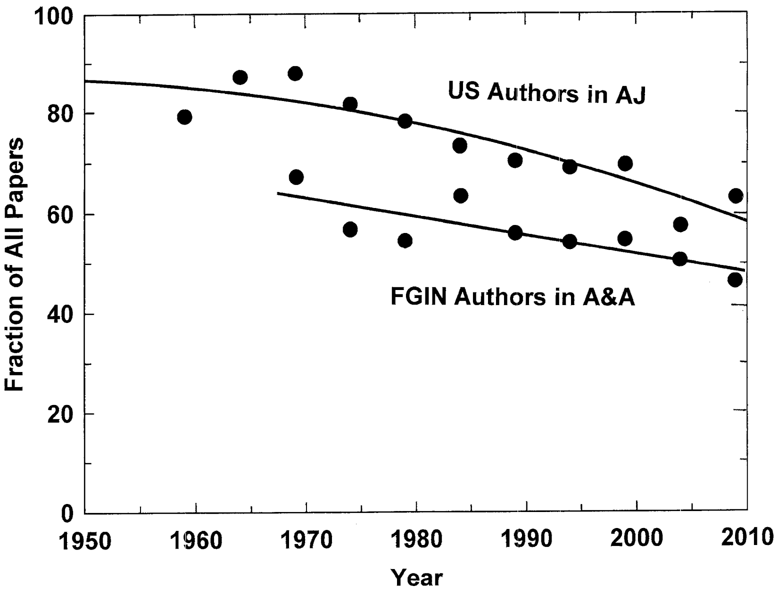


Fig. 2. (*Top*) The fraction of American papers in AJ as a function of time. The fitted curve is F = 86.6 – 0.116 (Y – 1950) – 0.00594 $(Y - 1950)^2$ with a cross-correlation coefficient of 0.915. For FGIN, their contributions in A&A are shown; the fitted curve is 69.8 – 0.335 (Y – 1950) with a cross-correlation coefficient of 0.785.

3 The United Kingdom

Most of the astronomical papers published in England, Scotland, Wales, and Northern Ireland were printed in the MNRAS. Again I corrected for format changes to 1,000-word pages. As shown by the bottom curve in Fig. 1, the UK contributions in MNRAS varied from 72.0% in 1959 to 32.3% in 2009. That means that MNRAS publishes a much larger fraction of foreign papers than the American journals. However, the US was not a major contributor to MNRAS, namely 7.9% in 2009.

The population of UK varied from 51.94 million in 1959 to 61.13 million in 2009, a change of only 17.7%. Dividing the UK 1,000-word pages by the populations gives the results shown in Fig. 3 (middle curve). This too shows no evidence of leveling off, but recent reports (Cleary 2010) state that the UK is withdrawing from most of the large telescope projects in 2010–2012. That suggests that the UK government feels that it has over-extended its commitments to astronomical research and the curve may level off in coming years.

The UK curve in Fig. 3, relative to the US curve, shows a lag of 10 ± 1 years.

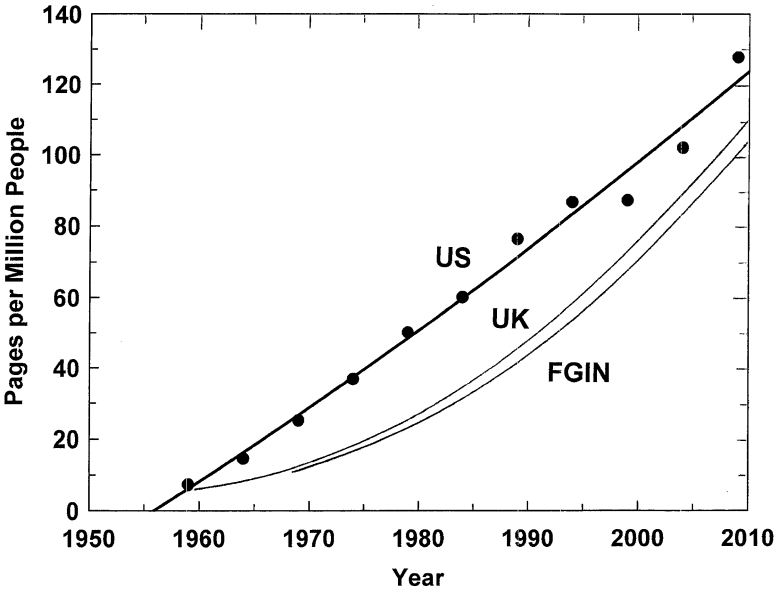


Fig. 3. The left curve shows the combined ApJ and AJ output in American 1,000-word pages, printed plus on-line, per million people in the US. The fitted curve is $-11.2 + 1.88\ (Y-1950) + 0.00618\ (Y-1950)^2$. The middle curve represents the UK output in the same units. The fitted curve is $5.81 - 0.279\ (Y-1950) + 0.0343\ (Y-1950)^2$. For the FGIN output, unlike the other two curves, we combined their contributions in A&A, AJ, ApJ, and MNRAS. The fitted curve is $11.4 - 0.686\ (Y-1950) + 0.0371\ (Y-1950)^2$. The data points for the UK and FGIN curves are not shown to avoid confusion, but they are given in a related paper in press (Abt 2010).

4 Europe

A&A started in 1969 when seven countries terminated their national journals and started to publish "Astronomy & Astrophysics, A European Journal". Participation has now grown to 23 countries, including ones in South America. But some of those countries are still developing major astronomical facilities, such as Bulgaria and Croatia. Rather than include all their populations and their few papers, I counted only four: France, Germany, Italy (after 1972), and Netherlands. These will be abbreviated FGIN. I originally considered including Spain as one of the four major countries in Europe, but I was advised that its large current publication rate started only a decade ago and that it would be better to include the Netherlands.

I counted pages from FGIN and converted them to 1,000-word pages. Then I counted the contributions in A&A from FGIN. Figure 2 shows that they contributed from 63.4% of the A&A pages in 1969 to 50.0% in 2009. Even in 2009 the on-line contributions were trivial. The combined populations of

FGIN ranged from 141.49 million in 1969 to 221.59 million in 2009, a growth of 56.6% in 10 years. The numbers of 1,000-word FGIN pages per million people are only 43% of those of the US to a current value of 59 1,000-word pages per million.

However, many Europeans still publish in UK and US journals. So I counted the numbers of FGIN papers in AJ, ApJ, and MNRAS. Recently the FGIN contributed 6% of the AJ papers, 10% of the ApJ papers, and 20% of the MNRAS papers. When we add the FGIN pages in A&A, AJ, ApJ, and MNRAS, we get the curve shown in Fig. 3. Before comparing that curve with the US curve, we should add the US contributions in A&A (5.4%) and MNRAS (7.9%). After that is done, we find that the FGIN publication rate lags that of the US by 12 ± 1 years. There is no evidence of a leveling off of the European output.

In a recent study (Abt 2007) I found that the numbers of papers in five sciences (astronomy, chemistry, geophysics, mathematics, and physics) depend only on the numbers of scientists in their appropriate societies (e.g. American Astronomical Society, American Chemical Society, etc,). The relations are linear and constant with time. Therefore one way for the UK and FGIN countries to catch up to the US is to employ more astronomers.

References

Abt, H. A. 2007, Scientometrics, 73, 281
Abt, H. A. 2010, PASP, in press
Cleary, D. 2010, Science, 327, 22

9

Fully Digital: Policy and Process Implications for the AAS

Chris Biemesderfer

American Astronomical Society, 2000 Florida Avenue NW, Suite 400, Washington DC 20009, USA
chris.biemesderfer@aas.org

Summary. Over the past two decades, every scholarly publisher has migrated at least the mechanical aspects of their journal publishing so that they utilize digital means. The academy was comfortable with that for a while, but publishers are under increasing pressure to adapt further. At the American Astronomical Society (AAS), we think that means bringing our publishing program to the point of being fully digital, by establishing procedures and policies that regard the digital objects of publication primarily. We have always thought about our electronic journals as databases of digital articles, from which we can publish and syndicate articles one at a time, and we must now put flesh on those bones by developing practices that are consistent with the realities of article at a time publication online. As a learned society that holds the long-term rights to the literature, we have actively taken responsibility for the preservation of the digital assets that constitute our journals, and in so doing we have not forsaken the legacy pre-digital assets. All of us who serve as the long-term stewards of scholarship must begin to evolve into fully digital publishers.

1 Introduction

Scholarly communication has evolved in interesting ways for millennia. From Archimedes writing letters on scrolls to his correspondents, to Neil deGrasse Tyson tweeting to his followers, scholars have discussed not only the issues of their disciplines and the insights of their research, they have (at least recently) discussed how other scholars communicate as well (Odlyzko 2002; Heuer et al. 2008; Renear and Palmer 2009). However, in this paper I won't talk about blogs, wikis, txts, tweets, youtubes, crowd-sourced GPS-enabled climate-sensitive thoroughly-modern mobile droidbots – and all that jazz. I intend to confine my remarks to *formal* communications: information that is published in the things we recognize today as the peer-reviewed journals. It is worth stating plainly that at the AAS, we are primarily concerned with communications that benefit professional astronomers. For the purpose of this paper, that means I won't worry about how we might address audiences other than researchers.

A. Accomazzi (ed.), *Future Professional Communication in Astronomy II*,
Astrophysics and Space Science Proceedings 1, DOI 10.1007/978-1-4419-8369-5_9,

I'm also going to assume that the notion of formal communication as we know it is basically a good thing, because it comprises other good things: peer review is a good thing, consistency is a good thing, permanence (durability, longevity) is a good thing. I acknowledge that there are voices crying out against some of these things – "peer review must die", e.g. – and certainly there is nothing wrong with questioning fundamental tenets. But let me make an observation.

Archimedes wrote to his colleagues in a form that is structurally very much like communications between scholars today. We can read modern translations of Archimedes, and they seem familiar to us; the same is true for works of Euclid or Eratosthenes or Galileo. The information has certainly been durable (although in the case of Archimedes, not without some close calls; Netz and Noel 2007). What is remarkable, however, is how consistent the form is, whether written in parchment scrolls, bound codices, printed books, or LaTeX files. In spite of hyperventilated claims that "digital changes everything", formal communication has survived format changes intact for thousands of years. My point is not to belittle Web 2.0 style communications; my point is not that formal communications are better somehow; my point is that formal communication is not going away.

As a specific instance of formal communication, scholarly journals have a number of merits for the organization of these communications. They provide an initial selection of articles through different editors with different aims and scopes; they offer a range of editorial temperaments; they provide peer review; and they are a platform for the methodical organization, normalization, and preservation of information. In addition, as a publisher of scholarly journals, the AAS offers scholars a sensible approach to intellectual property, and it offers to the research community a rational and sustainable business model. Every one of these things – every single one – is technology-neutral. We've done them all along, we're still doing them, and at the American Astronomical Society, we think we should keep doing them.

2 Publishing Initiatives at the AAS

The AAS has actively pursued innovation and modernization in its publications. Indeed, the venerable *Astrophysical Journal* (ApJ) itself is sometimes described as having been born in order to communicate the new scientific insights that were forming in the wake of the invention of spectroscopy (Osterbrock 1995). When the ApJ needed to publish more data tables, a Supplement series was created in 1954. When speedier publication of new results was indicated, the Letters were published on a faster schedule (Chandrasekhar 1967). When technology drove us beyond the printed page, the ApJ added videos in 1992, and the AAS began to distribute digital data on CD-ROM in 1993.

The AAS began to take its publications in their entirety into the digital regime with the abstracts for the Society's 180th meeting in Columbus in 1992; by the 182nd meeting in Berkeley in 1993, the abstracts were published on the World Wide Web. The Society had embarked on collaborations, also in 1992, with its publisher at the time, The University of Chicago Press, and with a project called STELAR (Study of Electronic Literature for Astronomical Research) which was associated with the National Space Science Data Center at Goddard Space Flight Center (Warnock et al. 1993). After a period of morphological instability and accretion, STELAR gave way to ADS, the Astrophysics Data System (Kurtz et al. 2000), which had been developing independently in another part of the universe (Cambridge, Massachusetts). The large research journals followed into the digital realm through the mid-1990s (Boyce 1998), and after nearly 20 years, we are preparing once again to update the *Bulletin of the AAS* (which has served as the print home for the meeting abstracts).

The present-day journal publishing initiatives at the AAS have three main strategic drivers: to provide more underlying numerical materials in the journals, to manage the evolution of print, and to adjust the business model accordingly. Since 1990, the Society and its publishing partners have effected the "digitalization" of all the major facets of the publication process: manuscript preparation and submission by authors, peer review, production and delivery, and preservation.

In the near term, we anticipate that enterprise-scale printing will be phased out in the next 2–3 years, as the library subscribers to the journals stop acquiring the print products. We are going to be looking to web-to-print solutions so that customized print products can be specified by the customers themselves, thus allowing the AAS to focus on the larger issues of professional scholarly communication. In the meantime, we are thinking about a business model for the Society that offers only online subscriptions, and we are preparing to charge authors in 2011 based on the quantities of digital material that are submitted, rather than based on the number of typeset pages of the authors' text.

The AAS' interest in providing data sets explicitly dates back to at least 1954, when the ApJ Supplement was born. In the age of computer networks, the research environment has been significantly enriched by direct live connections among resources. We can no longer limit our attention to data "in the journal" or attached explicitly to articles. Now, journal articles can refer to raw data held in archives and data centers, either at the author's initiative or through the addition of query tools in the online article.

3 The View Toward the Future

There are so many exciting things we might do with our journals to enrich formal communication among scholars. We are all interested in connecting our journals' consumers with the full range of resources that support (and may extend) the research they are reading about. We have made some good steps already, but what we have accomplished with our journals to date has been done with a relatively small number of partners, mostly through bilateral collaborations. That's good, but it isn't sustainable. If we want to create a healthy and stable and scalable formal scholarly communication enterprise in the future, we have to put systems in place that utilize standards, because the alternative – having to manage NxN system interfaces through bilateral agreements – is fiscally intractable.

Efforts to deliver more machine-readable data in the journals will resonate with energies now being applied to next-generation web technologies that are intended to facilitate machine interactions: the Semantic Web, or the computable web. We don't have to build the applications for the journals, and at the AAS we don't plan to; we just have to make our content accessible via transparent interfaces so the data can be harvested by users for their preferred applications and services. Even doing so, there are crucial questions regarding data that we obtain with journal articles.

1. How should we obtain it, and what are agreeable formats to get from authors?
2. How can we improve the online presentation, especially for complex data objects? What does a "good" user interface for databases look like?
3. For preservation purposes, is any policy more proactive than byte preservation feasible? Where should this kind of data reside for curation? Should it be the same place as for delivery, or might distinct repositories be preferable?

For some purposes, the underlying data is actually the article text. We have seen an increase in interest in mining the text of the journals themselves as a data set, and we anticipate more of this will happen as people and their software agents become more facile with large corpuses of text. We could argue that many of the rendering enhancements we envision for the online journals qualify as attempts to deliver essential information elements as if they were data. For instance, presenting the math via MathJax can be seen as making the math itself more computable.

Publishers have all done respectable enough jobs of "digitalizing" the major elements of the formal communication process. There is a host of fairly unspectacular tasks left to do, the details in which devils reside. Some of the tasks are obviously complicated, such as making all those live connections to external data resources. Some of them seem straightforward, but aren't. For instance, switching a paginated journal so that it utilizes article numbering sounds easy, but it is fiendishly complicated because it affects bibliographic coordinates and every system that uses them (Chaix 2011).

Much of the work we have to do going forward will be difficult and time-consuming, mostly because it involves so many other parties, parties that we are obliged to work with. It's the price of interconnectedness. The interconnectedness of the present and future holds promise as well as many challenges.

In principle, everything interacts; that's the promise. In practice, everything has to interact; that's the challenge, to be able to find partners and techniques that permit the interconnections to arise in ways so that programs are sustainable.

But what concerns me most is that our view toward the future has to be much larger, much more inclusive and comprehensive, than it is today. We face significant challenges, and after a few decades of evolution of the network, we have to confront a host of external pressures that are not technological, and quite a few of them at the same time.

Some of the external pressures are not new. Publishers of the scholarly literature are concerned with broad dissemination, and we routinely involve secondary providers to aid in discovery and delivery. We have to engage the "new secondaries", which in astronomy means arXiv and ADS for the most part, but in an inclusive and comprehensive world also has to include individual institutional repositories too numerous to count. There are government mandates; open access comes to mind first, but mandatory deposit of digital content is a non-trivial consideration for the national libraries that require it. The scholarly literature is long-lived, and we will always be concerned with conservation and preservation.

In the future, metrics like the impact factor *might* persist for journals, but it seems reasonable that these may decline in favor of metrics that focus on individuals and teams and research organizations. The community is trying hard to bridge that gap, although not with much success I dare say. Nevertheless, I expect these problems will be solved someday, and we will be able to judge the quality of the contributions of organizations or departments or campuses, and of different research teams and coalitions, and maybe even of individual scientists. At that point, journals will no longer proxy for such entities, and the impact factor et al. will not be so interesting. At the same time, journals will instead be favored for things like speed of publication, equanimity of peer review, and the transparency of the journal's databases to processes that compute individual metrics and scores. In the meantime, we will continue to confront statistics that are poorly conceived or unreliable (Adler et al. 2008; Arendt 2010).

Some of the external pressures are unrelated to scholarly publishing. Information overload will get worse before it never gets better. And this evolution in formal communication is happening in a global technological environment that appears to most people to be quite easily switched on, thanks to the effectiveness of advertising and promotion in our modern digital world. As a result, our customers keep asking questions like "How hard can it be?", and our genteel scholarly dialog is drowned out by crowd-sourced cacophony. We will be defending the value propositions of formal communication for a long time.

Let me come back to the idea that our view toward the future has to be more inclusive and comprehensive. In many ways, the world has gotten smaller as globalization has come about. Communication of huge quantities of information is virtually instantaneous thanks to countless kilometers of optical fiber and phalanxes of orbiting satellites. It is possible to travel almost anywhere on the surface of the Earth, quite quickly and inexpensively. These changes have happened fairly rapidly, and in a short span – much less than a lifetime – we have enjoyed a dramatic expansion of our viewpoints. (That larger number of viewpoints and opportunities, by the way, also results in many more sources of information to manage, and more partners to interact with. So in terms of our potential collaborations, the world has also gotten quite a bit bigger.)

However, enjoying expanded viewpoints is not the same as a broadened perspective. Our policies are catching up, but most of our assumptions about the cultural norms that govern the behavior of scholars are rooted in a homogeneous past. The formal communication that has been discussed in this essay is all Western, or at least not any more Eastern than Constantinople. The big small world in which we now communicate challenges us, and thinking only about the West isn't thinking big enough. At the AAS, we are cognizant of the need to adjust our scope and our sensibilities so that we accommodate research and economic engines in Asia, although I think it is fair to say that we do not yet know what all the ramifications of these adjustments will be. However, I do believe that it is a critical part of becoming a fully digital publisher that we are capable of embracing scholarly practices the world around.

We have certainly embraced the technological innovations of the past several decades, the academy has started to adjust attitudes and policies about scholarly communication, and we are beginning to recognize how crucial a broader perspective will be. As fully digital publishers, aided by new technologies and new policies and a broad perspective, we gain the attitudes and the flexibility to meet the present and future challenges of scholarly communication. All of us who serve as the long-term stewards of scholarship need to evolve into fully digital publishers.

I'm grateful for careful readings of drafts by Peter Boyce and Rebecca Jensen; this paper benefits from their suggestions. The AAS has benefited from suggestions and help from all quarters for many years. Recently, I've been able to have stimulating discussions with the AAS editorial and development teams, and in the Executive Office with our Working Group on Communications. Over the years, I've enjoyed fruitful conversations with many innovative colleagues in publishing. And we all have the pleasure of working with a research community full of creativity and new ideas.

References

Adler, R., Ewing, J., and Taylor, P. 2008, Citation Statistics, a report of the International Mathematical Union http://www.mathunion.org/publications/report/citationstatistics0/ (accessed 5 April 2010).

Arendt, J. 2010, Are Article Influence Scores Comparable across Scientific Fields?, Issues in Science and Technology Librarianship, http://www.istl.org/10-winter/refereed2.html (accessed 5 April 2010).

Boyce, P. 1998, Electronic Publishing in Astronomy, in Highlights of Astronomy, volume 11A, 499 (Springer).

Chaix, M-L. 2011, this volume.

Chandrasekhar, S. 1967, ApJ, 148, 1, DOI 10.1086/149125.

Heuer, R-D., Holtkamp, A., and Mele, S. 2008, Information Services & Use, 28, 83, DOI 10.3233/ISU-2008-0570.

Kurtz, M., et al. 2000, Astron. Astrophys. Suppl. Ser., 143, 41, DOI 10.1051/aas:2000170.

Netz, R. and Noel, W. 2007, The Archimedes Codex, Da Capo Press (Philadelphia).

Odlyzko, A. 2002, Learned Publishing, 15, 7, DOI 10.1087/095315102753303634.

Osterbrock, D. 1995, ApJ, 438, 1, DOI 10.1086/175049.

Renear, A. and Palmer, C. 2009, Science, 325, 828, DOI 10.1126/science.1157784.

Warnock A., Van Steenberg M.E., Brotzman L.E., Gass J.E., Kovalsky D., Giovane F. 1993, STELAR: An Experiment in the Electronic Distribution of Astronomical Literature, in ASP Conf. Ser. 52, 137.

10

Open Access: Current Status, AAS Perspectives

Kevin B. Marvel and Chris Biemesderfer

American Astronomical Society, 2000 Florida Avenue NW, Suite 400,
Washington DC 20009, USA
kevin.marvel@aas.org; chris.biemesderfer@aas.org

Summary. Open Access, defined as the free provision of information by science publishers, is not likely to be mandated by law anytime soon in the United States. A collaborative effort, initiated by the House Science Committee, to come to some consensus within the scientific publishing enterprise has resulted in the release of the so-called "Roundtable Recommendations". These will serve as a working model moving forward on fundamental shared starting points for both publishers and authors as well as the Open Access community. The AAS' delayed open access model for publishing is flexible, supportive of our discipline and equitably distributes the cost of publishing to authors and readers. The AAS can support this flexible model because it is not dependent on journal revenues for the support of its member-focused activities.

1 Introduction

It should not come as a surprise that people want access to information for free. People would like to have most things for free, from the proverbial free lunch to registration giveaways at meetings. However, human commerce is based on the concept of exchange. If we get something, we have to give something. The creation of money as a tangible representation of value allowed one to exchange a unit of value instead of one's own personal labor or some other physical good, like chickens. The value of things change over time, driven both by simple supply and demand, as learned in basic economics, but also in more complicated ways, such as through consumer perceptions and technological innovation.

For millennia, preserved information or knowledge, in the form of the written word, has had substantial value. Nearly since writing was invented as a way to preserve, disseminate and consume information, whole buildings were dedicated to the storage and preservation of written, and subsequently printed, information. The buildings were substantial (the library of Alexandria or, today, the Library of Congress) because preserved information, preserved knowledge has substantial value. Yet that value has changed substantially

A. Accomazzi (ed.), *Future Professional Communication in Astronomy II*, Astrophysics and Space Science Proceedings 1, DOI 10.1007/978-1-4419-8369-5_10,

each time technology has enabled a new way to reproduce knowledge and disseminate it. We have moved from scratchings on clay tablets, to writing on scrolls, to writing on stabilized animal skins, to writing on paper to printing. Each change nearly always eradicated the prior technology, while usually decreasing the cost of production. It is nearly impossible to find a handwritten book these days and even personal letters have become a rarity.

We are deep into a new revolution in the preservation and dissemination of knowledge: the digital age. Revolutions are difficult to navigate. Emotions are high. Visionaries claim to know how things will settle out, with some being right, and all the rest being wrong. All we know for sure is that we are in the midst of a substantial change and that we must navigate it carefully as publishers because we impart value to the knowledge we preserve and disseminate. It has yet to be seen if the digital revolution will turn that belief upside down or not.

One movement afoot during this revolution is the so-called 'Open Access' movement or initiative. Open Access means different things to different people, but we like to summarize it in the following way: Open Access is the desire to have the consumer of information or knowledge give nothing in order to access the information or knowledge. This concept stands in stark contrast to the basic definition of human commerce above (give something to get something), but it is not impossible to think that the new technologies available to us, combined with new business models for publishers might allow this goal to be achieved. That is why the American Astronomical Society is not opposed to the Open Access movement. We are however very cautious about rapid changes to our business model, as we feel that the preservation and dissemination of knowledge that we provide is of significant value to our discipline and human knowledge in general.

Hence, we have paid very close attention to recent developments in the United States, where both the Congress and the Administration have taken an interest in codifying the Open Access movement in the form of laws or regulations. Such efforts, if made quickly and without cooperative participation of all parties potentially impacted, could have deleterious results on the preservation and dissemination of astronomical knowledge. The issue is too important to ignore.

This short contribution is meant to provide a synopsis of some recent developments specific to the US government and to detail some positive actions taken by publishers, librarians and consumers of knowledge to sit down together and frame out some general guidelines to help our nation's leaders make sensible decisions.

2 Current Status

A number of different proposals for mandating Open Access have appeared in legislation in Congress since 2000, none of which resulted in an actual law or regulation. In 2004, NIH began an effort called Pub Med Central, to be a central repository for government funded research results twelve months after they appeared in scientific journals. Many publishers, especially non-profit publishers, endorsed this delayed access model and Pub Med Central, as the case that taxpayer funded research should be available to the taxpayers is a solid one. Further, the twelve month delay before posting allowed the revenue earned through subscriptions to remain secure. For without some revenue, the value added to scientific publications by publishers, through peer-review, copyediting, preservation, indexing and all the digital enhancements and linking so necessary in the digital age would not be possible. Subscriptions are an obvious way to pay for these enhancements and there could be others.

In 2009, the US Congress, via the House Science Committee sponsored a roundtable discussion on Open Access, which we will detail below. In December 2009, the US President, via the Office of Science and Technology Policy undertook a call for comments on specific issues surrounding Open Access. This call, the submitted comments and other applicable administration comment and detail is available as of today online[1]. In late January 2010, the roundtable report was issued and is available online[2]. Based on the input the OSTP received, which included the roundtable report and individual comments submitted from publishers and individuals, it is clear that sometime this year (2011) the OSTP will announce a "Proposed Rulemaking" in the Federal Register[3] with the intent of issuing a formal rule on Open Access. More on this below.

3 The Roundtable: Justification

The roundtable effort began based on discussions initiated by interested parties and the House Science Committee. They were initiated because it was felt that there was substantial common ground that should serve as a basis for all parties to begin to work together on the issue of access to scholarly work. Obviously, economic pressures apply to all parties. Readers want things as cheap as possible, libraries have limited budgets to pay subscription fees and the process of peer review and publication along with preservation all cost money as well. It is also clear that scholarly publication is here to stay and plays a vital role in science and scholarship. It is unclear if the revolution we are living through will change the process of scholarly publication, which it could, but for the foreseeable future, scholarly publication as we know it is here to stay.

1 http://www.whitehouse.gov/blog/2010/03/08/public-access-policy-update
2 http://www.aau.edu/policy/scholarly_publishing_roundtable.aspx?id=6894
3 http://www.gpoaccess.gov/fr/

It is also true that disruptive and unsustainable transitions in business models, especially for publishers, could easily disrupt scholarship and the scholarly publishing ecosystem now established. Margins are low, especially for non-profits, and some publishers may not have the financial reserves to survive an abrupt business transition (this is not true for the AAS).

Finally, it is also true that strong reactions and polarization on this issue have prevented discussion of sensible and shared solutions. Since government was stepping forward as a valuable partner, the various communities involved decided to accept the invitation of the House Science Committee to form a roundtable to seek common ground and explore possible future solutions and develop shared recommendations.

4 The Roundtable: Process

The charge to the roundtable, which was officially convened by the Committee on Science and Technology of the US House of Representatives, in coordination with the White House Office of Science and Technology Policy (OSTP) was "to explore and develop an appropriate consensus regarding access to and preservation of federally funded research information that addresses the needs of all interested parties."

The Roundtable was composed of three academic administrators (three provosts and an association executive), three librarians representing academic libraries, two learned society publishers, one commercial publisher and an open access start up publisher along with three researchers from library and information science. The group opted to establish rules of participation that mandated that they participate in the process as knowledgable individuals, not as official representatives of organizations and that they refrain from public disclosure of their deliberations, among other less important guidelines. They further agreed on some simple shared principles:

- Peer review must continue its critical role in maintaining high quality and editorial integrity.
- Adaptable business models will be necessary to sustain the enterprise in an evolving landscape.
- Scholarly and scientific publications can and should be more broadly accessible with improved functionality to a wider public and the research community.
- Sustained archiving and preservation are essential complements to reliable publishing methods.
- The results of research need to be published and maintained in ways that maximize the possibilities for creative reuse and interoperation among sites that host them.

5 Roundtable: Core Recommendations and Reactions

After months of discussion and interaction, the group issued a set of recommendations (available online at the web address presented earlier in this paper) in early January 2010. Although not all participants felt they could sign onto the recommendations, only two participants did not sign onto the set of recommendations. They issued separate statements also available at the AAU-hosted roundtable website.

The core recommendations were:

- Government should develop public access policies to scholarly work funded by government investment.
- The group urged an expeditious, but cautious approach.
- The desired outcome of the development of government policy should be free public access to publicly funded work.
- Public access should happen as soon as possible after publication.
- The group could not develop a specific regulatory or legislative solution.
- As the government proceeds, the process should be consultative, involving all parties.
- The group felt that the policies should be flexible enough to allow field-dependent variation and agency-specific embargo periods.
- Care should be given to defining the 'Version of Record'.
- Any systems developed for providing public access should ensure interoperability

Additionally, the group felt that balance was important. Specifically, it is necessary to validate the need for and potential access to scholarly articles, while preserving the essential functions of scholarly publishing (peer review, preservation, etc.). All stakeholders will have points of view on these matters that need to be considered and all parties should collaboratively participate with the government as a partner.

Reaction so far to the roundtable's recommendations have been measured. Most stakeholders have respect for the roundtable approach, support the general principles and support the recommended consultative process. However, societies and for-profit publishers have reservations about intellectual property and copyright issues as well as trepidation about government intervention in a long-established business. Advocates of a legislative mandate were disappointed because no specific recommendation was provided by the roundtable, though that was not central in their charter.

It is not yet clear what sustained impact the roundtable effort will have, yet it was an important change from the polarized, vocal squabbling that has characterized this debate for several years. As with any revolution, emotions run high, yet it is often the measured voices that prevail. It was important to provide a forum for these voices and we feel the roundtable was a substantial step on the path to a new future.

6 Next Steps

Sometime this year (2011) the Office of Science and Technology Policy will likely issue a notice of a proposed rule-making. After this takes place, there will be 60 days provided for public comment and then 30 additional days for reply comments (and new original comments). Such rule-makings are often preceded by a notice of inquiry. However, it is not clear if the call for comments in January were counted as a notice of inquiry or not, but the general consensus in Washington today is that it did, in fact, count as the notice of inquiry for this probable rule-making. The concept of the rule-making process was designed to allow government to formulate better rules by ensuring that all those impacted by the rule are able to provide their perspective. As any politician knows, often the best solution is a compromise that keeps most people happy. The initial rule-making could be followed by a further proposed rule-making if the submitted comments require a drastic change to the proposed rule. Though this is not common (and the process can repeat again). Eventually the rule will be developed and published in the Federal Register as a 'report and order' (R&O), which stands as a valid regulation unless challenged in court. As a side note, any given rule and order can be amended with additional rules and orders over time, leading us to envision an Escher woodblock print, but let's hope this analogy will not apply to this rule.

Conversations with the OSTP at the highest levels indicate that they have received many serious comments, which they are parsing carefully before proceeding. Given the many other issues confronting the White House, we do not expect any action before June 2011 and would not be surprised if nothing happened until September or even later. Delays in any step of the process would delay the final rule issuance and activation.

7 AAS Perspectives

As of 1 April 2010, the AAS has published approximately 110,000 articles in the AAS journals. Of these, fully 102,000 are available at no cost right now. We do not charge for access to our back content, though we do preserve copyright to ensure the preservation and proper use of our content over time. The AAS also has a delayed open access policy, which limits access for 24 months for its scientific journals and publishes the fully open access Astronomy Education Review. We are an open access publisher already.

We have a complex and robust article provision system, that is augmented by trusted partners. Our electronic articles (e.g. fully digital content) are hosted by our publication vendors, Institute of Physics Publishing (IOP) and the American Institute of Physics (AIP), all of which are linked to by the Astrophysics Data System (ADS) hosted at the Smithsonian Astrophysical Observatory and funded by NASA. Backfile scans, e.g. digital images of printed content, back to volume 1, number 1 are hosted by ADS as well. Digital object identifiers (DOIs) and all applicable metadata have been assigned to all articles and indexed appropriately.

As stated above, we do maintain copyright on all of our articles (not written in full by civil servants as part of their employment). It is our position that free-to-read does not mean free-to-redistribute or reuse. It is not appropriate for third parties to make revenue from the content produced by our authors and published by the Society. By maintaining copyright, the Society can ensure the general use of articles in perpetuity. The estates of authors (or authors themselves) will not be able to prevent or limit access to knowledge published in our journals by withdrawing them from circulation. This is the primary reason the copyright assignment by our authors is required. The Society wishes to ensure, to the fullest extent possible, the long-term access to the knowledge published in its journals.

Although we do require authors to assign copyright to the Society, we grant substantial rights back to the authors. These include the right to grant or refuse permission for third party republication while alive (the Society requires republication requestors to obtain this permission from the authors), the right to use all or part of the article in future and derivative works of their own, the right to make copies of all or part of the article for use for educational or research purposes and the right to place the article, in final published form, on their own or their institution's web pages with a link to the version of record.

The AAS business model is robust and probably unique. We are focused on our key stakeholders, the scholars and authors who perform the research and write the articles and the librarians and readers who want access to our authors' work. We do not use journal proceeds to pay for Society activities, only publication costs and administration. We ensure that the authors pay for peer review, production, archiving and technology development through so-called 'page' charges or author charges. The librarians and readers pay for publication and dissemination and distribution as well as DOI assignment and archiving through subscription fees. We set both author charges and subscription fees annually, based on the forecast costs and publication volume for our journals. It is an iterative process that results in a justified, rational and fair distribution of expense to the interested communities.

However, although we are not against Open Access, we are concerned about substantial and rapid change being forced upon our robust, fair, open and rational business model. We also feel that this distributed model provides a significant robustness to variable effects in both communities. If research dollars become tight, or funds for subscriptions are a challenge to produce, we can adjust to these realities. We have done so in the past and are prepared to do so in the future. If all subscription fees were taken away immediately, our authors would see a substantial increase in the fees they would pay in the new author-pays plan, but the fees would not even come close to doubling. This is because the Society does not use journal proceeds to fund Society programs. This is a deeply held philosophical stance that we are unlikely to give up, specifically because it allows us to publish our journals on behalf of our communities at the lowest possible expense to all.

Further concerns for us include the reality that nearly 60% (roughly) of articles published in our journals are from overseas and mandated US Open Access would set up two classes of author and two classes of papers. This does not advantage scholarship. Finally, although we maintain a talented lobbying staff member to interact with the government to benefit our scientific discipline (following all registration and financial rules governing lobbying our government), we think there is substantial risk to moving to a model where federal funding is relied upon for scholarly publication. Even a cursory study of US government commitment to funding obligations shows substantial variability. One need only mention the superconducting supercollider to set off eye-rolls and exasperated sighs. The US votes on expenditures annually, meaning that funding commitments can change on that timescale as well.

In our submission to the OSTP public comment period, we focused on a few key points. First, a proprietary period, even a short one, acknowledges the value and importance of the scientific publishing process. Second, the version of record should be the version made publicly accessible. Third, any repositories that the government endorses, should be capable of delivering complex digital research reports and have inherent growth capability in this regard as that content changes and evolves. Finally, a twelve month proprietary period is achievable under the AAS model without much financial risk and was implemented in January 2011.

8 Conclusion

A process is underway in the United States that will lead to the imposition of government rules mandating some form of open or public access to published research results funded in whole or part by US taxpayer dollars. The AAS, which has a unique business model that we feel serves our discipline well, is prepared to modify our business model in accordance with government mandate, but we urge caution and an open, collaborative and iterative process to produce the best possible policy. Although we are in the midst of a technological revolution that may very well result in a completely new scientific publishing paradigm, the AAS believes that central components of the publishing process, preservation and peer-review, will always be important to scholarship and human advancement.

A unique roundtable process, sponsored by the US House of Representatives has set a high mark for collegial, constructive discussion of core issues surrounding the Open Access issue. The recommendations from this roundtable will have far reaching impact and the report produced is worthwhile reading.

Looking back historically, we have to wonder how scholarly resources such as the Library of Alexandria were paid for and whether the populace of that small city on the Nile delta demanded open access to the scrolls and the valuable knowledge they contained. Who wouldn't want to learn about and make use of the Pythagorean theorem, for example? we are sure that no matter if they did or didn't, the scribes that authored those scrolls were just as concerned about the preservation of their work as we are with our own journals and we are sure a robust business model paid for their production and preservation. Media may change, but the fundamental value of scholarship and knowledge will be with us forever. At least digital media appear more robust to fire, when archived responsibly with a goal of the permanent expansion of human knowledge.

11

Astronomy Librarian – Quo Vadis?

Jill Lagerstrom[1] and Uta Grothkopf[2]

[1]Space Telescope Science Institute, 3700 San Martin Drive,
Baltimore MD 21218, USA
lagerstrom@stsci.edu
[2]European Southern Observatory, Karl-Schwarzschild-Straße 2,
D-85748 Garching, Germany
esolib@eso.org

Summary. "You don't look like a librarian" is a phrase we often hear in the astronomy department or observatory library. Astronomy librarians are a breed apart, and are taking on new and non-traditional roles as information technology evolves. This talk will explore the future of librarians and librarianship through the lens of some of the recent talks given at the sixth "Libraries and Information Services in Astronomy" conference held in Pune, India in February 2010. We will explore the librarian's universe, illustrating how librarians use new technologies to perform such tasks as bibliometrics, how we are re-fashioning our library spaces in an increasingly digital world and how we are confronting the brave new world of Open Access, to name but a few topics.

1 Introduction

"You don't look like a librarian" is a phrase we hear so often that Ruth Kneale, systems librarian at the National Solar Observatory, decided to write a book with this title (Kneale 2009), after so many years of blogging on the issues surrounding this phrase with frequent regularity.[1] Our popular culture is fascinated by the image of the librarian. An advertisement for a Sony ebook reader states "Sexier than a librarian." Noah Wyle plays a librarian who goes on several Indiana-Jones-type adventures in a series of made-for-TV-movies. And, last but not least, a long time ago in a galaxy far far away, there is even a Jedi librarian with whom Obi Wan Kenobi himself has a mild altercation over the existence of some astronomical data in the archive. Unfortunately, many of these portrayals, paradoxically even the "futuristic" ones, make librarians seem out-of-touch with modern technology and modern culture.

These entertaining portrayals aside, real librarians in astronomy libraries in all corners of the world are doing innovative and critical work to support the

[1] http://desertlibrarian.blogspot.com

A. Accomazzi (ed.), *Future Professional Communication in Astronomy II*,
Astrophysics and Space Science Proceedings 1, DOI 10.1007/978-1-4419-8369-5_11,

missions of the organizations they serve. To explore them, we will take a look at astronomy librarianship through the lens of some of the issues and projects discussed at the recent "Libraries and Information Services in Astronomy (LISA) VI" conference held in Pune, India at the Inter-University Centre for Astronomy and Astrophysics in February 2010. Excellent histories of the LISA conferences by Brenda Corbin and Uta Grothkopf may be found in Volume 7 of the Organizations and Strategies in Astronomy series (Corbin & Grothkopf 2006) and in the proceedings of the fifth LISA conference (Grothkopf & Corbin 2007)[2]. Videos of many of the talks may be seen on the LISA VI website.[3] The proceedings of the conference will be published by the Astronomical Society of the Pacific Conference Series. Finally, we will conclude with further thoughts and observations about the image of librarians today and what it may mean for the future.

2 The LISA VI Conference

Astronomy librarians, appropriately, are a very international and diverse crowd. More than 90 participants from more than 18 countries attended LISA VI. The theme of the conference was "21st Century Astronomy Librarianship: From New Ideas to Action" and the six major topics discussed were: *The Future of Librarianship*, *Metrics*, *Open Access*, *Data Curation and Preservation*, *Virtual Communities* and *Use and Access*. While taking the reader on a tour of the highlights of the "official" themes, we will explore some of the subcurrents running through the presentations that describe the constant adaptation and evolution of astronomy librarians and the diverse ways they serve their local communities. These examples, in no way, represent all that librarians do, but rather a subset of interesting and successful projects.

2.1 The Future of Librarianship

Discussions on the future of librarianship focus on the new and emerging roles that librarians play within the knowledge spheres of their organizations. In addition, they grapple with the role of the print collection in an increasingly electronic environment. Managing both print and online collections simultaneously with a matrix of heterogeneous user expectations and the re-envisioning of the library as a place are two subcurrents that run throughout these discussions.

Advice on managing change, both practical and psychological, was offered by Jane Holmquist, librarian at Princeton University. Departmental science library collections at Princeton have been merged and housed in a new state-of-the-art building designed by Frank Gehry. Jane's situation is becoming

[2] http://www.eso.org/sci/libraries/lisa.html

[3] http://libibm.iucaa.ernet.in/conf/index.php/LISA/conf

more prevalent as universities centralize services. Librarians from the Universities of Oslo and Helsinki are also managing change brought about by similar consolidations within their local academic cultures.

Jill Lagerstrom's poster on "If it's not on the web it must not exist" used the opportunity of a major renovation at the Space Telescope Science Institute to find out how much of the print collection is available online, and how much is indexed by the ADS.

The frontier of "E-science" is being explored by librarians. Lee Pedersen of Brown University chronicled her conversations with faculty about the storage, dissemination, manipulation, protection and publication of the data generated by their research.

Librarians are increasingly involved in "Communicating astronomy with the public." Francesca Brunetti, of the INAF Osservatorio Astrofisico di Arcetri, argued that librarians provide the right mix of skills for this task: scientific authority and accuracy to the public, and the ability to deal with people of different ages, skills and heterogeneous educational backgrounds.

2.2 Metrics

Librarians use publication statistics to analyze the scientific productivity and impact of their organizations, scientific instruments (such as telescopes), groups of people, topics/subjects or even individual scientists. The managers of the organizations that we serve need this information for decision-making purposes and to show funding agencies that money has been put to good use. Fortunately, methods have been developed to construct these metrics efficiently and wisely, and librarians are well versed in their practice. Discussion surrounding this topic deals not only with metrics projects and methods, but also with the shortcomings of information products (journal websites, ADS, Web of Science) when it comes to performing tasks such as these, which need to be done as accurately and systematically as possible.

Christopher Erdmann at the ESO Library has developed two software tools, FUSE and telbib, to manage the workflow specific to constructing bibliographies for citation analysis. FUSE is a full-text searching utility which has features more sophisticated than any publisher's website. Telbib is a content management system that generates next-generation bibliometric statistics.

Eva Isaksson looked at how departmental mergers at the University of Helsinki have affected productivity and impact statistics using the ISI citation database. When an astronomy department merges with a physics department, does this affect the measure of the department's overall performance?

Nishtha Anilkumar examined the types of research used by doctoral students in their theses at the Physical Research Laboratory in Ahmedabad, India. Longitudinal statistics on the use of Open Access materials, non-subscribed journals, and e-prints paint a revealing portrait of the research landscape and how it evolves over time. This type of analysis can reveal gaps in the library's collections.

2.3 Open Access

The Open Access movement promotes the idea that scholarly works (most usually journal articles) should be freely accessible to everyone. Librarians have, in general, been vocal advocates of this idea, which comes in a variety of economic models and paradigms, because we are the ones who get the bills for the journals and often are forced to cancel access to journals as our budgets shrink. According to *Library Journal's* periodicals price survey,[4] the average cost of an astronomy journal has been increasing at a rate of 8–10% each year, certainly far higher than the rate of inflation. Astronomers have historically been early adopters of many aspects of Open Access: they deposit e-prints in astro-ph and the major society journals make articles freely available after two years. They are also accustomed to page charges, whereby the author shares the cost of publishing.

To further complicate matters, the U.S. Office of Science and Technology Policy has recommended that the version of record of federally funded research be made freely available within one year of publication. Terry Mahoney, scientific editor at the Instituto de Astrofísica de Canarias, asked the question "Is Open Access Right for Astronomy?" Of course, the answer is neither a simple "yes" or "no." Terry took a refreshingly critical look at the economic realities of Open Access in the complex and heterogeneous world of astronomical publishing. Joe Jensen, managing editor of the Astronomical Society of the Pacific Conference Series also goes beyond the rhetoric to spell out what Open Access means specifically for society conference proceedings. Terry and Joe both reminded the audience that "there's no such thing as a free lunch" and that there is no one-size-fits-all solution as scholarly publishers confront Open Access.

For M.N. Nagaraj at the Raman Research Institute Library, Open Access is not just about journals. They have created an open digital repository of physics and astronomy theses created at their institution. Before the advent of electronic communication, theses, in particular, were often prohibitively difficult to obtain. A free repository is most welcome. However, these digital repositories don't just "create themselves" and bring with them a host of complex issues, such as determining the best scanning methods, getting permissions and establishing metadata standards.

2.4 Data Curation and Preservation

The curation and preservation of data raise many unanswered questions. What is the shelf life of digital data? What does it cost to make data free? How can we cooperate and collaborate better to answer these questions?

Copyright is perhaps the most complex issue for archivists who want to make the electronic documents in their collections freely available and use the

[4] http://www.libraryjournal.com/article/CA6547086.html

web to promote their institutions. Christina Birdie at the Indian Institute of Astrophysics is confronting the challenge of balancing the need to disseminate information freely with the rights of authors under the law. Unpublished works are particularly difficult to manage.

According to Andras Holl, librarians at small observatories must be involved in the team responsible for constructing, managing and ensuring access to local data archives. He outlines a plan with guidelines for the Konkoly Observatory to establish such a data archive.

2.5 Virtual Communities

Virtual tools facilitate communication and collaboration across global networks. Librarians make use of these tools to build communities.

Leila Fernandez participates in a virtual community to engage the public with astronomy at the University of Toronto. An interactive chat facility combined with live viewing of telescope images enable a virtual community to participate in astronomical activities to which they would not normally have access. Participants from the southern hemisphere are able to view portions of the sky not normally visible to them. Groups such as bilingual people and First Nations Communities are targeted participants in this outreach program.

Social networking is best known to us as Facebook and Twitter. However, a host of book social network sites, such as LibraryThing and Shelfari, have also emerged. Francesca Martines, of the INAF Osservatorio Astronomico di Palermo, evaluated these with a librarian's bibliographic eye and contemplated their use in the astronomy library. Librarians often evangelize interesting new technologies to their local users.

Alberto Accomazzi, ADS Project Manager, sees the ADS as a sort of virtual community, a collection of heterogeneous knowledge products interlinked with the ADS as a node. In order to take advantage of powerful web architectures, though, we need to structure our data accordingly. Alberto calls on us to do this so that we can fully collaborate, share and uncover networks of communication.

2.6 Use and Access

Librarians study their communities and attempt to understand their local users' specific needs. A good librarian is familiar with the culture not just of the specialized field she supports, but of her local institution as well. She is also familiar with the marketplace of new and emerging information products. Librarians seek the best products for their users with the intent of using library funds effectively and saving the user time.

Electronic books are an example where librarians are waiting for the future to happen. Molly White, at the University of Texas at Austin, surveyed her astronomy faculty to reveal their attitudes toward the ebooks available to them through the library. When asked if they prefer ebooks to print books, the responses were mixed. The answer was a resounding “sort of.”

Open source library catalogs have entered the marketplace. Uta Grothkopf, of the ESO Library, has become frustrated because commercial proprietary library catalog software is constantly behind the technological times. Investigating open-source alternatives gives her greater flexibility and allows her to configure these tools to provide her users with twenty-first century research capabilities. Uta evaluated the features of various open source catalogs and shared her experience with her endeavour to adopt an open source product in a small library.

Hemant Kumar Sahu of the Inter-University Centre for Astronomy and Astrophysics researched the specific information-seeking behaviors of astronomers and astrophysicists by using a questionnaire to learn about their use of books, online catalogs, print materials and online journals. With the gathered statistics, Hemant can paint a picture of the impact of Information and Communication Technology on the information-seeking behavior of Indian astronomers and astrophysicists. This picture can be used to inform strategic planning initiatives for the library’s infrastructure.

3 Conclusions

3.1 Subcurrents

While there were six official themes to the talks at the LISA VI conference, many subcurrents can be found throughout such as:

- Playing active roles in supporting the communication of astronomy to the public
- Demonstrating the value of the intellectual output of their organizations through bibliometrics
- Pro-actively making local knowledge products freely available
- Becoming more user-centered and less library-centered
- Refocusing on the library as a place
- Exploring the frontier of e-science and data archives

3.2 What’s in a Name?

Will librarians remain librarians? Last year, our trade organization, the Special Libraries Association, voted on whether to change its name to the “Association of Strategic Knowledge Professionals.” The discussion surrounding this topic was very charged, indeed. The leadership of the organization was concerned that the term “librarian” was a liability. While there were many members who were not necessarily attached to the name, which has been in use for 100 years, the result was that the name change was voted down.

The word "libraries" has received some attention in the educational sphere as well. One-third of library schools have become "iSchools." A new type of knowledge worker that has grown out of librarianship is called the "informationist." These new professionals, who primarily provide research assistance, came into existence in the context of clinical care and biomedical research. Informationists receive specialized training in their subject areas that goes beyond "on-the-job" learning. Programs in informatics are popping up in universities with library curricula and have expanded beyond the biomedical to include fields such as chemistry and even music. "Embedded" librarianship describes a set of practices librarians are adopting to get out of their offices and out from behind their desks. Since more patrons are using online information, we do not see them as often and this is an effective way of making sure our services are visible. As librarians' roles evolve, along with our names in some cases, our traditional values of preservation and access to information remain.

3.3 This Book Is Overdue

Marilyn Johnson, in her recent book *This Book is Overdue! How Librarians and Cybrarians Can Save Us All* (Johnson 2010) makes the case that, despite ongoing technological advances, there will always be a need for human help. Librarians must continue to be advocates for their users, or, as a reviewer of this book puts it, "pragmatic idealists who fuse the tools of the digital age with their love for the written word and the enduring values of free speech, open access, and scout-badge-quality assistance to anyone in need."[5]

3.4 Where Are We Going?

The title of this paper asks of the astronomy librarian *"Quo Vadis?"* which in Latin means "Where are you going?" The answer, that we hope has become apparent from the many subcurrents running through the talks at the LISA VI conference, is "we go where the information goes" and "we go where our users go." Our good faith, however, may not be enough. Terry Mahoney's Declaration concerning the evolving role of libraries in research centres warns us of the "increasing invisibility of research libraries vis-à-vis recent accelerated changes in publishing and reader-access technology" and calls on librarians to "adopt a more proactive stance in making their contribution known to the research communities they serve" (Mahoney 2007). Hopefully, recent efforts to combat the image of the librarian as antiquated and out-of-touch will help us to demonstrate that we are vital parts of the organizations we serve.

[5] http://www.harpercollins.com/books/9780061431609

References

Corbin, B.G. & Grothkopf, U. 2006, LISA – The Library and Information Services in Astronomy conferences, in: *Organizations and Strategies in Astronomy (OSA)*, Vol. 7, A. Heck (ed.), Springer, Dordrecht, 285-306.

Grothkopf, U & Corbin, B.G. 2007, LISA – How We Got Where We Are Now (Closing Remarks) in: *Library and Information Services in Astronomy V*, S. Ricketts, C. Birdie E. Isaksson and C. Birdie (eds.), Astronomical Society of the Pacific, San Francisco, ASP Conference Series, 377, 25-26.

Johnson, M. 2010, *This book is overdue! : how librarians and cybrarians can save us all*, Harper, New York.

Kneale, R. 2009, *You dont look like a librarian : shattering stereotypes and creating positive new images in the Internet age*, Information Today, Inc., Medford, NJ.

Mahoney, T.J. 2007, Declaration concerning the evolving role of libraries in research centres, *The Observatory*, 127, 401-402.

12

Telescope Bibliometrics 101

Uta Grothkopf[1] and Jill Lagerstrom[2]

[1]European Southern Observatory, Karl-Schwarzschild-Straße 2,
D-85748 Garching, Germany
esolib@eso.org
[2]Space Telescope Science Institute, 3700 San Martin Drive, Baltimore,
MD 21218, USA
lagerstrom@stsci.edu

Summary. During recent years, bibliometric studies have become increasingly important in evaluating individual scientists, institutes, and entire observatories. In astronomy, often librarians are involved in maintaining publication databases and compiling statistics for their institutions. In this paper, we present a look behind the scenes to understand who is interested in bibliometric statistics, which methodologies astronomy librarians apply, and what kind of features next-generation bibliographies may include.

1 Introduction

There are many ways to assess research output, for instance by investigating how many research grants have been received, when and where research has been presented at conferences, or how many students graduated under the supervision of specific researchers. Another common tool is bibliometric studies, i.e., using metrics to measure productivity and impact through publications and citations.

Bibliometric studies have quite some history; studies go at least back to the 1960s when the Science Citation Index was first issued. A large number of papers have been published on this topic in general and more specifically in the context of astronomy.

In this paper, we will focus on the following aspects:

- Bibliometric studies – what are they and who is interested?
- Linking publications and data – how does this happen, and where can interested users get access?
- Telescope bibliographies – who compiles them, and how? What are the current tools and methodologies?

Finally, we will take a look ahead to see what might be coming next.

A. Accomazzi (ed.), *Future Professional Communication in Astronomy II*, Astrophysics and Space Science Proceedings 1, DOI 10.1007/978-1-4419-8369-5_12,

2 Typical Bibliometric Measures

Many bibliometric studies use one or several of the following measures. However, they all have some advantages as well as some disadvantages.

– *Number of publications*: measures productivity, but does not report anything about the impact
– *Number of citations*: gives information about the impact, but can be inflated due to many reasons, for instance incorrect or incomplete citations, as well as biased citing behavior (authors citing well-known authors rather than young, unknown ones)
– *Mean or median cites per paper*: this measure is better suited for comparisons of scientists or facilities that have been active for different numbers of years, but it seems to reward low productivity
– *'High-impact papers'*: this metric has been introduced in astronomy by Juan P. Madrid, then at the STScI, a few years ago. Basically, Madrid used the ADS to retrieve information about the 200 most-cited papers in a given publication year, identified those that were based on observational data and calculated the impact of facilities (telescopes, observatories) of each paper. The results per paper were added, and the facilities ranked by contribution to this set of *Top 200* papers. The drawback of this method is that so-called hot topics are favored and can outnumber all other facilities for instance in a year of data release. The method is also time-consuming and, to a certain extent, subjective as the contribution percentage is assigned by the bibliometric researcher
– *h-index*: the h-index is meant to combine metrics for productivity and impact (Hirsch 2005); h itself is not suited for comparisons as it does not contain information about the number of years of operation. For comparisons, another value, the so-called m parameter, also described by Hirsch, should be used.

All measures have to be applied with greatest care and never in an isolated way as they only shed light on a very limited area of performance of research output. If used for comparisons, several metrics should be applied in parallel to get a more complete picture.

3 Users of Bibliometric Studies

A wide range of groups are interested in bibliometrics, including instrument scientists, management of observatories, governing bodies and funding agencies who want to

– evaluate the performance of telescopes and instruments
– measure the scientific output from observing programs
– define guidelines for future facilities

- compare in-house facilities with other observatories and telescopes
- interconnect resources, for instance in the context of Virtual Observatory projects

Establishing telescope bibliographies is the prerequisite to bibliometric studies in astronomy; it closes the loop from (a) observing programs carried out by astronomers, (b) data stored in an archive, (c) scientific papers which use the data, and (d) records in telescope bibliography databases that connect the papers with the data through observing programs (program IDs). Archives can be searched by program ID, and all papers published so far that use the data these programs generated can be listed. This procedure assures a maximum return of science benefits from observing proposals, for instance by making archived data easily available to other interested researchers.

Telescope bibliographies can be accessed in various ways. Firstly, a listing has been compiled by the ESO Library and is available on the web[1]; it provides links to the databases of many large observatories. The user interfaces of telescope bibliographies of all major observatories provide the option to search by observation or data set number. If these numbers are noted in the bibliographies' records, a listing of all published articles using the data is just a click away.

For example, in the case of ESO, the public interface allows to search for specific programs, or for papers that use data from certain instruments. The result will be a list of papers for which the facilities that were used are shown, as well as the program IDs. A click on the program ID will take users to the observing schedule from where more information on the proposals can be accessed. The data can be requested, or other papers using the same program can be found.

Another way of finding lists of papers that use observational data from specific observatories is to use the ADS *Filters* section which can be found in the lower part of the main ADS search screen.[2] The option *"All of the following groups"* as well as at least one of the listed facilities need to be selected. The records contained in the results set will all have *D* (for data) links which link back to the underlying observations.

4 Compiling Telescope Bibliographies

In order to compile such literature lists with links to data, or telescope bibliographies, certain prerequisites are necessary (Fig. 1).

Compilers need to have access to the literature. This can be in the form of print versions of scientific journals, or electronic access. The latter can be established by either pulling over PDFs to screen them locally, or by

[1] http://www.eso.org/libraries/publicationlists.html
[2] http://adsabs.harvard.edu/abstract_service.html

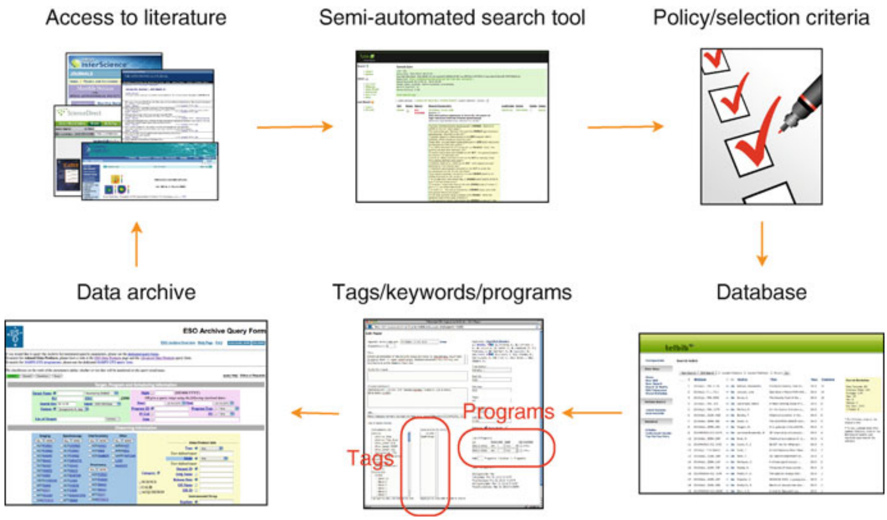


Fig. 1. Telescope bibliographies close the loop from published literature to data located in archives, and back.

searching full-texts at the publishers' sites. Ideally, a semi-automated search tool is available that helps compilers to pinpoint relevant articles. This will be explained in more detail below. If such a search tool is not in place, PDFs have to be searched one-by-one, or the paper version has to be inspected visually.

Most importantly, a policy needs to be in place that determines which papers shall be included in the telescope bibliography, and which are to be excluded. The topic of selection criteria can be handled quite differently among large observatories. Once relevant articles are identified, a database has to be established to host records with bibliographic information about the published papers.

Observatories will have their own individual set of tags and keywords which are assigned to records to describe the content of the papers as well as the facilities that generated the data. Specific information, such as program IDs, will be added at this stage, too. These identifiers link to the observatories' archive, and from there back to the published articles.

5 Staffing Situation, Journals Screened, and Search Strategies

An additional very important "ingredient" needed to establish and maintain telescope bibliographies is staff. In many observatories, librarians are involved in the process, or are even the main person in charge.

In early 2010, Jill Lagerstrom, librarian at STScI, conducted a survey among 16 large observatories (Lagerstrom 2010). A total of 14 institutions replied, namely CFHT, Chandra, FUSE/Galex/IUE, ESO, Gemini, HST/STScI, Isaac Newton Group, Keck, NOAO, NRAO, SDSS, Spitzer, Subaru, and XMM Newton. The survey revealed that large observatories involve between one and four staff members to compile their bibliographies, even though not all of them full-time.

The survey also focused on the journals which are searched by these institutions. Respondents provided a combined list of more than 30 journals which are screened, but surprisingly, only four are searched by all survey participants: *A&A*, *AJ*, *ApJ* and *Supplement*, as well as *MNRAS*.

The full-text can be accessed using one or several of the following methods:

- *Print journals*: screening the full-text by using the print version of journals provides access to the entire text of papers including footnotes, figures, captions, etc. However, it depends on the skills and capabilities of the compilers to detect all instances of relevant keywords and facilities in the text to gather a complete list of published articles
- *E-journals (PDFs)*: this methodology implies retrieving the PDF format of articles from the publishers' web sites and searching them locally. Like with method 1, all sections of papers are screened.
- *Journal web sites*: many publishers provide an option to search full-texts of articles directly at their web site. While this functionality can indeed be very useful, it is usually not immediately understandable for users whether this feature actually searches the entire text, or whether certain sections are not indexed and can therefore not be searched properly.
- *ADS abstracts*: at present (May 2010), ADS provides full-text searches only for historic literature; access to recent publications includes only title, abstract, and selected footnotes. If compilers of telescope bibliographies rely on ADS abstract searches for their work, they must be aware that almost inevitably they will miss many relevant papers.
- *Author self-reporting*: some compilers use reports from authors as an additional or even exclusive means to find papers. It must be noted that, if applied as the sole method for collecting articles, relying on author self-reports will reveal only a fraction of all relevant papers; it is therefore the least recommended way of compiling telescope bibliographies.

The survey also investigated whether compilers of bibliographies use similar search statements, and how elaborate and comprehensive they are in their attempts to identify papers. Not surprisingly, all respondents stated that they search for their observatory's telescopes, typically both the abbreviation as well as the full name. Some are more specific and include instrument names, as well as surveys, archives, or specific science programs (e.g., GOODS). For some institutions, it can be useful to search for geographic locations (Green Bank, Paranal, etc.) or even for concepts (e.g., X-ray).

The final survey question aimed at the biggest challenge compilers face in gathering papers. The majority of respondents stated that their biggest problem is authors who do not provide sufficient details about the source of their data, so that the programs that generated the data cannot be traced without doubt. Others replied that they face problems with restrictive publishers when they access large numbers of articles at their web sites. Many compilers feel that the amount of work time they can dedicate to their respective telescope bibliography is not sufficient in order to complete an exhaustive search. The idea of having a central search option for full-texts at the ADS was ventilated; this would be more time and resources-efficient than duplicating retrieval efforts locally at each observatory, as is currently done.

It became evident that the most difficult part is to define a policy to clearly govern which papers fulfill the criteria for inclusion, and which do not, as well as consistency in applying this policy. Recently, a group has been formed, initiated by the compilers of telescope bibliographies at the Chandra Archive, STScI, and ESO, to provide best practices for maintaining bibliographies, to develop recommendations for using cross-facility bibliometrics, and to share solutions to common problems.[3]

6 Tools of the Trade: FUSE and Telbib

While compilers of bibliographies in the past had to rely on their own abilities to retrieve relevant papers by visually scanning the literature, lately some tools have become available to support them in their work. One of these tools is the ESO Full-Text Search tool (FUSE) which has been developed by the ESO Library. Based on an ADS search, FUSE pulls over PDFs from publishers, converts them into text, and searches for keywords and text strings chosen by the individual observatory. If keywords are detected, they are shown in context (two lines of text) for inspection by the compilers. The excerpts often reveal immediately whether the highlighted paper should be investigated in more detail, or whether the keywords are used in a context that is not relevant for the compilation of the telescope bibliography (Fig. 2).

It is important to note that FUSE can only be used as a help application in order to spot potential candidates for inclusion in telescope bibliographies. By no means does it replace the human (intellectual) investigation that is necessary in order to determine whether an observatory's selection criteria for inclusion in the bibliography are actually met by the paper.

Another important tool is telbib, a content management software also developed by the ESO Library. It is used to establish and maintain a database of records with bibliographic information about papers pertaining to a telescope bibliography, to store additional metadata (e.g., tags to describe observing facilities and programs), and to generate reports and statistics. The librarians' user interface provides access to records through a large variety of search

[3] http://groups.google.com/group/astrobib/

Fig. 2. Screenshot of the ESO Full-Text Search tool (FUSE).

criteria (Fig. 3). A public user interface is also available, even though with less detailed query features.[4] Using telbib, bibliography compilers can import bibliographic information and other metadata from the ADS, including author affiliations and number of citations. Observatory-defined tags and keywords can then be added, as well as specific information about observing programs, modes (e.g., visitor or service mode) and types (normal, large, guaranteed time observing types, etc.).

Telbib is also used to generate reports and statistics either on demand or regularly, for instance the "Basic ESO Statistics" document[5] that is available from the Libraries' home page.[6] It provides information on ESO publication and citation statistics at large as well as for specific instruments, identifies the "ESO Top 20" papers, and looks at ESO publications in comparison with other large observatories.

A more detailed description of FUSE and telbib including their development and features can be found in Erdmann & Grothkopf (2010).

[4] http://www.eso.org/libraries/telbib.html
[5] http://www.eso.org/sci/libraries/edocs/ESO/ESOstats.pdf
[6] http://www.eso.org/libraries/

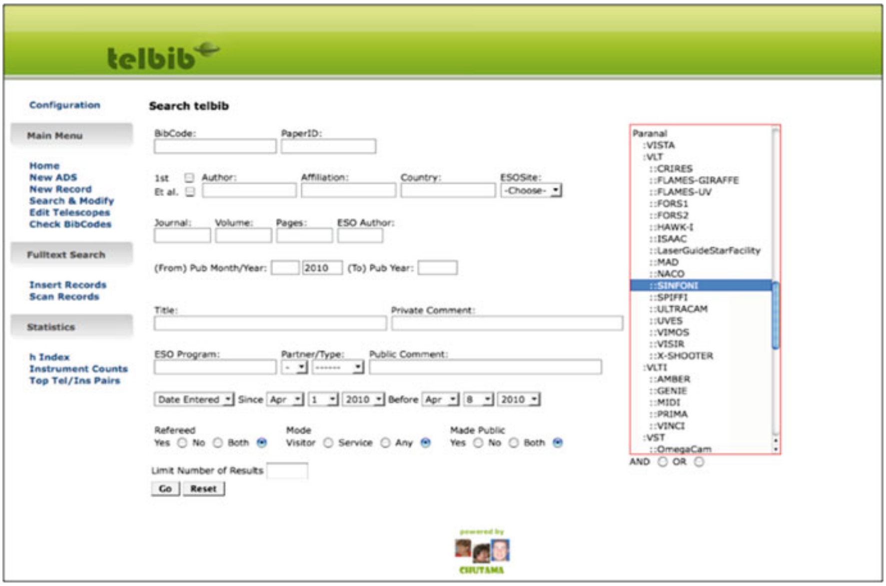


Fig. 3. Screenshot of the ESO telbib tool.

FUSE is currently in use at ESO, STScI, Gemini, Subaru, and the Carnegie Observatories. The libraries of SAAO (South African Astronomical Observatory) in Cape Town, South Africa, and IUCAA (Inter-University Centre for Astronomy and Astrophysics) in Pune, India, are in the process of installing the software. Because of this obvious duplication of efforts by conducting searches locally, recently the idea of creating a central search facility at the ADS has been discussed. Compilers of bibliographies could then run tailored searches at the ADS without the need to download PDFs locally for inspection. The search feature would only highlight potentially relevant matches by showing brief excerpts of text; access to the full-texts would still be governed by the publishers through subscriptions.

7 Beyond Citations

As already mentioned, studying citations to measure scholarly output is problematic, despite their wide-spread use for this purpose. Lists of citations are often incomplete, references are incorrect, various abbreviations for the same journal are used in parallel and are sometimes not attributed to the correct publication, and the typical citing behavior often favors well-known researchers instead of referencing younger, unknown authors.

Are there other, possibly better ways of evaluating scientific impact? Compilers of telescope bibliographies are experimenting with other measures. Some of them are described below.

It is well known that the ADS provides information about citations for each paper in their database; these are limited to those citing papers that also reside in the ADS database, hence they are not entirely complete. However, in the area of astronomy this is a minor disadvantage as the ADS coverage of papers is extremely large.

Few users are aware that ADS also gathers information about so-called Reads. Reads are defined as those instances where users access more detailed information than is available in the standard brief results list displayed after an ADS search. Hence, a Read can mean that a user looked at the abstract, the list of references, or the citing papers. Also any time the electronic version of an article is accessed, be it in PDF or in HTML format, counts as a Read.

The number of Reads typically differs considerably from the number of citations an article has gathered. More importantly, the distribution of Reads versus citations as a function of years reveals that many papers which are not cited frequently anymore still are used regularly by the community.

An even more telling statistic about actual usage could be obtained by limiting Reads to the actual request of full-texts, i.e., PDF and HTML downloads. Such numbers are traced by the ADS, but in contrast to Reads, they cannot be retrieved through the ADS user interface. However, the ADS team is always helpful in providing such statistics on request.

Another project that operates along the same line of thought is Citebase.[7] Citebase is a citation index that "harvests pre- and post-prints (mostly author self-archived)" from arXiv and other repositories and lists them together with papers that cite them. In addition, they provide information about downloads by country, by date and by organization. Unfortunately, they are still in experimental stage, and users are cautioned not to use the results for academic evaluation.

A newcomer in the area of evaluating research impact is the idea to look at social networking platforms and forum discussions. In astronomy CosmoCoffee[8] might be one such discussion group that reveals the community's interest in specific papers. Registered users refer to manuscripts posted on arXiv and suggest discussion among their peers. In order to track impact, bibliography compilers could track which arXiv preprints are discussed so that these papers can be linked to the final versions once they are published, and the number of comments and discussion threads specific papers generate can be traced.

8 Ongoing Projects

In order to provide even more informative reports and to anticipate future interests of management and funding organizations, it is desirable to import further metadata from the ADS, for instance subject terms and keywords (to gain

[7] http://www.citebase.org/

[8] http://www.cosmocoffee.info/

knowledge about the specific subject areas in which users of observational data publish), full information of citing papers (in order to eliminate self-citations), all other available links (e.g., Digital Object Identifiers (DOI), eprint IDs to establish links between eprints and published papers), as well as the author gender. The latter might be of interest to researchers involved in topics such as "Women in Astronomy"; identifying an author's gender may be feasible in future with the help of initiatives like ORCID (Open Researcher and Contributor ID)[9] that are hoped to solve the name ambiguity problem in scholarly research.

It is also intended to establish additional links, for instance between telescope bibliography records and press releases that feature specific papers, as well as record detailed information about the observing dates so that the delay from data acquisition to the publication can be computed.

9 Conclusion

The task of compiling telescope bibliographies has evolved considerably during recent years, providing ever more detailed and sophisticated information about the papers contained in the databases of large observatories. This trend will continue in the future as management and funding agencies rely increasingly on bibliometric studies in order to evaluate research output. It is important that compilers collaborate and exchange ideas and solutions to common problems in order to develop best practices and recommendations for establishing, maintaining, and using telescope bibliographies in a standardized, reproducible manner.

Acknowledgement. The ESO Telescope Bibliography (telbib) and Full-Text Search tool (FUSE) use NASA's Astrophysics Data System (ADS) Abstract Services. The authors wish to thank the ADS team for their excellent service to the entire community.

References

Erdmann, C. & Grothkopf, U., 2010: Next generation bibliometrics and the evolution of the ESO Telescope Bibliography, in: *Library and Information Services in Astronomy VI*, Isaksson, E., Lagerstrom, J., Holl, A. and Bawdekar, N. (eds.), Astronomical Society of the Pacific, San Francisco, ASP Conference Series **433**, p. 81

Hirsch, J.E., 2005: An index to quantify an individual's scientific research output, *Proc. Nat. Acad. Sci.*, **102**, 16569, DOI 10.1073/pnas.0507655102

Lagerstrom, J, 2010: Comparison of methods for creating telescope bibliographies, in: *Library and Information Services in Astronomy VI*, Isaksson, E., Lagerstrom, J., Holl, A. and Bawdekar, N. (eds.), Astronomical Society of the Pacific, San Francisco, ASP Conference Series **433**, p. 89

[9] http://www.orcid.org/

13

Progress on the 'Declaration Concerning the Evolving Rôle of Libraries in Research Centres'

Terence J. Mahoney

Instituto de Astrofísica de Canarias, C/ Vía Láctea s/n, E-38200 La Laguna, Tenerife, Spain
tjm@iac.es

Summary. I outline the goals of the 'Declaration concerning the Evolving Rôle of Libraries in Research Centres' and describe progress in obtaining its official adoption as IAU policy.

> The Observatory had two great libraries: one contained the printed word, and the other the photographic records of the stars and their spectra. I applied myself first to the former, and it is a course that I recommend to every student.
>
> Cecilia Payne-Gaposchkin, *The Dyer's Hand*

1 Libraries in Astronomical Research Centres

Times have certainly changed since Payne-Gaposchkin made her recommendation to young researchers, but her advice to learn what libraries have to offer still holds good today. Gone are the days when library tables groaned under the weight of opened volumes of abstracts and journals, as astronomers gleaned results and bibliographical information in the preparation of their articles: today's researchers prefer to access the literature directly via their laptops. Far fewer research workers frequent libraries now than was the case before the advent of instant Internet access to journals. The rôle of the research centre library has also evolved to adapt to the new technology; however, that evolution has largely been a silent and invisible one, to such an extent that some researchers, unaware of the librarian's part in bringing digital literature to their laptops, have begun to wonder whether a space-occupying and resource-consuming library is really necessary at all.

The way in which library services in astronomy have evolved is well represented in the six published LISA volumes (Wilkins & Stevens-Rayburn

A. Accomazzi (ed.), *Future Professional Communication in Astronomy II*, Astrophysics and Space Science Proceedings 1, DOI 10.1007/978-1-4419-8369-5_13,

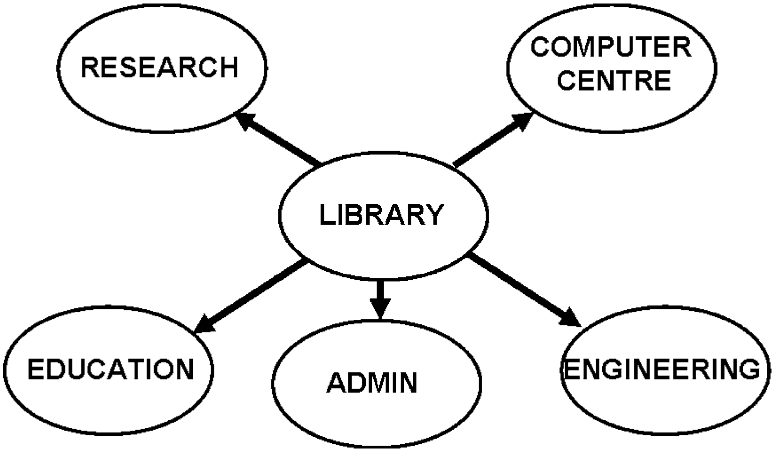


Fig. 1. Libraries in research centres.

1989; Murtagh et al. 1995; Grothkopf et al. 1998; Corbin et al. 2003; Ricketts et al. 2007; Isaksson et al., in the press). These volumes document how the computer has changed from a peripheral to a central tool of the astronomical librarian. Indeed, the technology has now reached such a level of development that some see the library becoming entirely digital. But what then, should be done with all those printed volumes. Do we need them? Can't we just ditch the printed copy altogether?

It was to address our concern 'at the increasing invisibility of research libraries vis-à-vis recent accelerated changes in publishing and reader-access technology' that a group of astronomical librarians and myself, while attending FPCA I (Heck & Houziaux 2007),[1] decided to work on a declaration that would explain the new rôle played by modern research libraries (Mahoney 2007).

In a typical large research establishment the library may be seen as a bibliographical and information resource for all other departments (Fig. 1). Traditionally, the library has provided journals, magazines, textbooks, monographs, theses, conference proceedings and observatory publications in printed form. Print is still (at the time of writing) important, but the librarian's accession tasks now include the purchase of licences for electronic journals and ensuring access to online databases and bibliographical resources such as ADS (Fig. 2).

[1] What became the 'Declaration Concerning the Evolving Rôle of Libraries in Research Centres' was drafted subseqent to FPCA-I but was not part of the proceedings of that meeting.

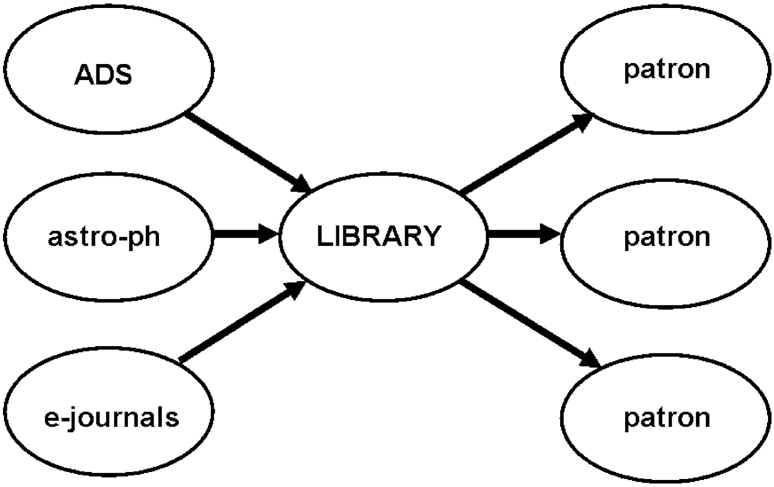


Fig. 2. The virtual library.

2 The Modern Librarian's Tasks

Today's librarians must offer a combination of traditional and new skills, including:

- Providing up-to-date coverage of subjects of interest to the centre by the acquisition of new publications
- Providing information to patrons
- Playing a part in ensuring continuity of the astronomical record from antiquity to the present both in print and on line
- Providing access to scholarly information resources in all media and formats
- Identifying, retrieving, organizing, evaluating, repackaging, filtering and providing electronic access to digital information sources
- Maintaining access through license and other software purchase
- Adapting the librarian's expert knowledge of information resources with astronomer's specialized subject knowledge
- Enhancing access through the librarian's ability to identify and link patrons to their required information needs.

It should be stressed here that a fair number of astronomical research centres house priceless historical records in the form of observing logs, article and book manuscripts, photographic plates and historical instruments. We are surely all mindful of this at this meeting, which is being held on the same site where, among others, Henrietta Swan Leavitt, Annie Jump Cannon, Antonia C. Maury, Harlow Shapley and Cecilia Payne-Gaposchkin spent all or most of their working lives.[2] Even if records and instruments are eventually housed in museums and archives away from the observatory, the library has to play an important part cataloguing and keeping records.

A further point to note is that not all astronomical literature is available online and there does not seem to be any prospect that this situation will change in the near future. It is therefore vital that printed copies of journals and documents are retained in sufficient numbers for the purposes of both maintaining as complete a bibliographical coverage as possible and ensuring the prolonged existence of the astronomical record for future generations of researchers.[3]

3 'Not Fade Away...'

The Declaration urges librarians to become more proactive in promoting their services to research centres. Generating and sharing information requires specialized knowledge of how to locate, filter, organize and provide access to that information: librarians, as trained information scientists, combine all these skills, yet, as already mentioned, application of those skills can pass unperceived by researchers accessing astro-ph and ADS on their laptops. Librarians therefore need to make researchers aware of their contribution to the smooth running of research centres. Since librarians are now being asked to provide a more wide-ranging service than was traditionally the case, library space and staffing levels need to be maintained. It is also important to continue to preserve comfortable and pleasant reading areas and recent journals sections (a healthy antidote to laptop autism).

The Declaration closes by noting that the main astronomical research journals (*ApJ*, *AJ*, *MNRAS* and *A&A*) have tripled in size and price over the past 10 years. It calls on funding bodies to take this increased essential expenditure into account and it urges research centres to distribute its financial resources in a way to meet these costs while recognizing the relative importance within the institution of the research library.

[2] See Shapley (1969) and Payne-Gaposchkin (in Haramandunis 1984) for captivating first-hand accounts of life in Harvard College Observatory in the first half of the twentieth century. See also Johnson (2005) for a brief account of the life of Henrietta Swan Leavitt.

[3] I have dealt with the issue of the need to maintain sufficient printed copies of journals and books elsewhere (Mahoney 2007).

4 Progress on the Declaration

Many astronomical librarians have now endorsed the Declaration (there were 15 signatures on the published version in *The Observatory*; Mahoney 2007). The Declaration has also appeared in the *Society for the History of Astronomy Newsletter* (Mahoney et al. 2008) and has been translated into Italian by Monica Marra of the *Biblioteca dell'Osservatorio Astrofisico di Arcetri* (Mahoney 2008). A Spanish translation in now in preparation.

The Declaration has also been reported in *NRAO Library Updates* (Nunemaker 2008), *Physics and Astronomy Library News* (Johnson 2008). As part of plans to get the Declaration adopted as official IAU policy, it was reported (van der Hucht 2009) to Commission 5 (Documentation & Astronomical Data) of Division XII (Union-Wide Activities) and a formal proposal will be made to have the Declaration accepted at the XXVIII General Assembly of the IAU in Beijing in 2012.

References

Corbin, B. G., Bryson, E. P. & Wolf, M. (eds) 2003, *Library and Information Services in Astronomy IV (LISA IV)* (Washington, DC: USNO)

Grothkopf, U, Andernach, H., Stevens-Rayburn, S. & Gomez, M. (eds) 1998, ASP Conf. Ser., 153, *Library and Information Services in Astronomy III (LISA III)* (San Francisco: ASP)

Haramandunis, K. (ed.) 1984, *Cecilia Payne-Gaposchkin: An Autobiography and Other Recollections* (Cambridge: CUP)

Heck, A. & Houziaux, L. (eds) 2007, *Future Professional Communication in Astronomy* (Brussels: Académie Royale de Belgique)

Isaksson, E. (ed.) 2010, *ASP Conf. Ser., Library and Information Services in Astronomy VI: 21st Century Astronomy* (San Francisco: ASP)

Johnson, G. 2005, *Miss Leavitt's Stars* (New York: Atlas)

Johnson, L. 2008/4/16, *Physics and Astronomy Library News* (http://blog.lib.umn.edu/ljohnsto/physics/scholarly-communication/)

Mahoney, T. J. 2007, in ASP Conf. Ser., 377, *Library and information Services in Astronomy V: Common Challenges, Uncommon Solutions*, ed. S. Ricketts et al. (San Francisco: ASP), 389

Mahoney, T. J. 2007, *Obs* **127**, 401

Mahoney, T. J. (transl. M. Marra) 2008, *Giornale di Astronomia* **3**, 22

Mahoney, T. J. et al. 2008, *Soc. Hist. Astron. Newsletter* **16**, 28

Murtagh, F., Grothkopf, U. & Albrecht, M. (eds) 1995, *Library and Information Services in Astronomy II (LISA II)*, *Vistas* **39 (2)**

Nunemaker, M. 2008/4/18, *NRAO Library Updates* (http://www.nrao.edu/library/NRAOLibraryUpdates.html)

Ricketts, S., Birdie, C. & Isaksson, E. (eds) 2007, ASP Conf. Ser., 377, *Library and Information Services in Astronomy V: Common Challenges, Uncommon Solutions* (San Francisco: ASP)

Shapley, H. 1969, *Through Rugged Ways to the Stars* (New York: Scribner's)
van der Hucht, K. (ed.) 2009, *Reports on Astronomy 2006-2009 (IAU XXVIIA): IAU Trans. XXVIIA* (Cambridge: CUP), p. 364
Wilkins, G. A. & Stevens-Rayburn, S. (eds) 1989, IAU Colloq. 110, *Library and Information Services in Astronomy* (Washington, DC: USNO)

14

Finding Your Literature Match – A Recommender System

Edwin A. Henneken, Michael J. Kurtz, Alberto Accomazzi, Carolyn Grant, Donna Thompson, Elizabeth Bohlen, Giovanni Di Milia, Jay Luker, and Stephen S. Murray

Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA 02138, USA
{ehenneken,mkurtz,aaccomazzi,cgrant,dthompson,ebohlen,gdimilia,jluker, smurray}@cfa.harvard.edu

Summary. The universe of potentially interesting, searchable literature is expanding continuously. Besides the normal expansion, there is an additional influx of literature because of interdisciplinary boundaries becoming more and more diffuse. Hence, the need for accurate, efficient and intelligent search tools is bigger than ever. Even with a sophisticated search engine, looking for information can still result in overwhelming results. An overload of information has the intrinsic danger of scaring visitors away, and any organization, for-profit or not-for-profit, in the business of providing scholarly information wants to capture and keep the attention of its target audience. Publishers and search engine engineers alike will benefit from a service that is able to provide visitors with recommendations that closely meet their interests. Providing visitors with special deals, new options and highlights may be interesting to a certain degree, but what makes more sense (especially from a commercial point of view) than to let visitors do most of the work by the mere action of making choices? Hiring psychics is not an option, so a technological solution is needed to recommend items that a visitor is likely to be looking for. In this presentation we will introduce such a solution and argue that it is practically feasible to incorporate this approach into a useful addition to any information retrieval system with enough usage.

1 Introduction

Authors publish because they want to transfer information. An essential ingredient for this transfer is being able to find this information. Of course, it is also in the interest of a publisher or somebody running a search engine service to have efficient and intelligent tools for information discovery. As our Literature Universe is expanding rapidly, finding your way in this deluge of information can be a daunting task. Looking just at the SAO/NASA Astrophysics Data System (ADS) holdings as of April 19, 2010, there are 1,730,210

A. Accomazzi (ed.), *Future Professional Communication in Astronomy II*, Astrophysics and Space Science Proceedings 1, DOI 10.1007/978-1-4419-8369-5_14,

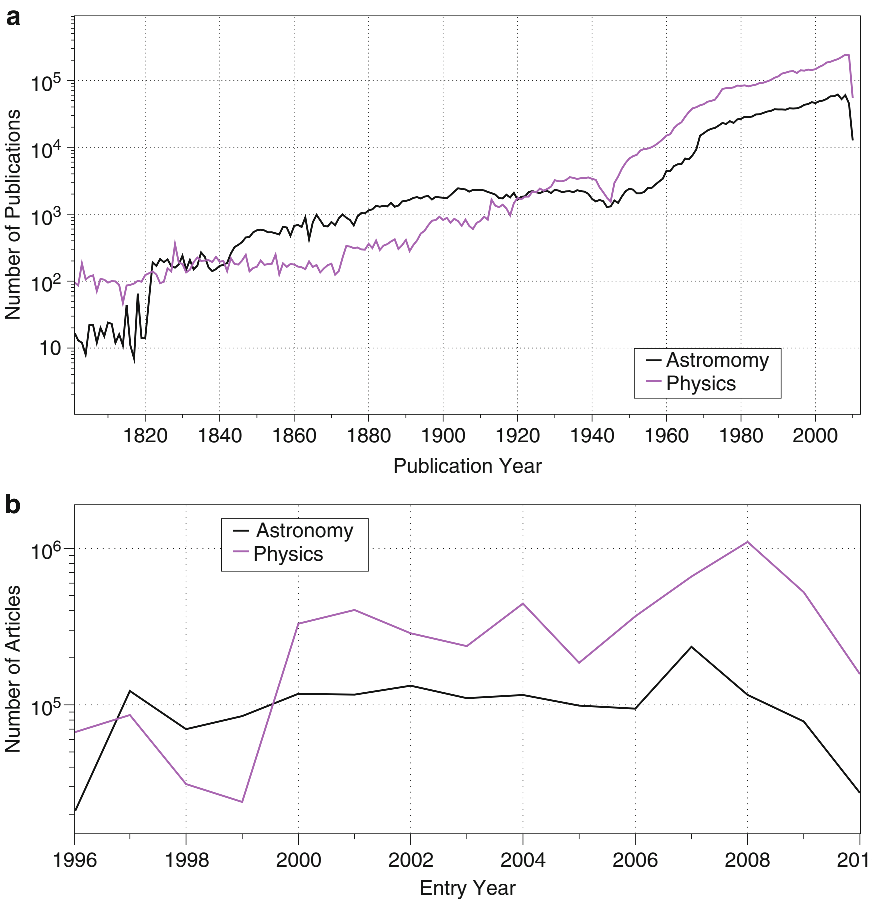


Fig. 1. *Top*: Records in the ADS as function of publication year. *Bottom*: Number of records entered in the ADS.

records in the astronomy database, and 5,437,973 in the physics database, distributed over publication years as shown in Fig. 1 (top). In astronomy, as in other fields, the Literature Universe expands more rapidly because of dissolving boundaries with other fields. Astronomers are publishing in journals and citing articles from journals that had little or no astronomy content not too long ago. Figure 1 (bottom) shows how many records were entered in the ADS during the period 1996–2010, in both the astronomy and physics database. About 100,000 astronomy records have been added each year, since 2000, with the exception of 2007 when ADS received a large amount of meta data from Springer and ARI (the Astronomisches Rechen-Insitut in Heidelberg, Germany). The large spike for physics in 2008 is due to the ingest of CrossRef meta data.

How do you find what you are looking for and more importantly, information you could not have found using the normal information discovery model?


myADS Personal Notification Service
for Edwin Henneken
Wed Apr 7 15:50:46 2010
Physics database

ADS Main Queries
Astronomy
Physics
arXiv e-prints
FAQ
What's new

Current Tables of Contents
European Physical Journal B
Geophysical Research Letters
International Journal of Modern Physics A
Journal of Geophysical Research A (Space Physics)
Journal of Geophysical Research C (Oceans)
Journal of Geophysical Research D (Atmospheres)
Journal of Geophysical Research E (Planets)
Journal of Geophysical Research F (Earth Surface)

HENNEKEN, EDWIN - Citations: 123 (total 162)
2010JGRF..11501009G: Gardner,+: A review of snow and ice albedo and the development of a new physically based broadband albedo parameterization
2010ThApC..99..115K: Kruk,+: Downward longwave radiation estimates for clear and all-sky conditions in the Sertãozinho region of São Paulo, Brazil
2009IJCli..29.2309V: van den Broeke,+: Surface layer climate and turbulent exchange in the ablation zone of the west Greenland ice sheet
2009TCry....3...75S: Sedlar,+: Testing longwave radiation parameterizations under clear and overcast skies at Storglaciären, Sweden
2008TCry....2..179V: van den Broeke,+: Partitioning of melt energy and meltwater fluxes in the ablation zone of the west Greenland ice sheet

"SOLAR WIND", etc - Recent Papers
2010GeoRL..3703107E: Edberg,+: Pumping out the atmosphere of Mars through solar wind pressure pulses
2010PhPl...17b2105N: Nielson,+: Numerical modeling of Large Plasma Device Alfvén wave experiments using AstroGK

Favorite Authors - Recent Papers
2010JGP....60..664S: Stevenson: Geometry of infinite dimensional Grassmannians and the Mickelsson--Rajeev cocycle

"IMPACT PHENOMENA", etc - Recent Papers
2010JGRE..11503003H: Hiesinger,+: Ages and stratigraphy of lunar mare basalts in Mare Frigoris and other nearside maria based on crater size-frequency distribution measurements
2010JGRE..11500D14S: Skok,+: Spectrally distinct ejecta in Syrtis Major, Mars: Evidence for environmental change at the Hesperian-Amazonian boundary
2010SolED...2...69K: Klokocník,+: Candidates for multiple impact craters: popigai and chicxulub as seen by EGM08, a global 5'×5' gravitational model
2009MinPe..96..141D: Dill,+: The ``Donauplatin'': source rock analysis and origin of a distal fluvial Au-PGE placer in Central Europe
2010PhRvA..81c2901V: Villalba,+: Light-induced atomic desorption and diffusion of Rb from porous alumina

"IMPACT PHENOMENA", etc -

Fig. 2. The myADS weekly newsletter.

When you have some prior information (like author names and/or subject keywords), you can use your favorite search engine and apply that information as filters. There are also more sophisticated services like myADS (as part of your ADS account, see Fig. 2 for an example and e.g. Henneken et al. (2007) for more information), that do intelligent filtering for you and provide you with customized suggestions. Alternatively, you can ask somebody you consider to be an expert. This aspect emphasizes that "finding" essentially is a bi-directional process. Would it not be nice to have an electronic process that tries to mimic this type of discovery? It is exactly this type of information discovery that recommender systems have been designed for.

2 Recommender Systems

Recommender systems can be characterized in the following way. Recommender systems for literature recommendation...

– are a technological proxy for a social process

- are a way of suggesting like or similar articles to a user-specific way of thinking
- try to automate aspects of a completely different information discovery model where people try to find other people considered to be experts and ask them to suggest related articles

In other words, the main goal of a literature recommender system is to help visitors find information (in the form of articles) that was previously unknown to them.

What are the key elements needed to build such a recommender system? The most important ingredient is a "proximity concept". You want to be able to say that two articles are related because they are "closer together" than articles that are less similar. You also want to be able to say that an article is of interest to a person because of its proximity to that person. The following approach will allow you to do just that:

- build a "space" in which documents and persons can be placed
- determine a document clustering within this space ("thematic map")

How do we build such a space? Assigning labels to documents will allow us to associate a "topic vector" with each document. This will allow us to assign labels to persons as well ("interest vector"), using the documents they read. Placing persons in this document space can be used in essentially two different ways: use this information to provide personalized recommendations or use usage patterns ("reads") of expert users as proxies for making recommendations to other users ("collaborative filtering"). As far as the labels themselves are concerned, there are various sources we can distill them from. The most straightforward approach is to use keywords for these labels. One drawback that comes to mind immediately, is the fact that there are no keywords available for historical literature. However, keywords are an excellent labeling agent for current and recent literature.

Figure 3 shows a highly simplified representation of that document space, but it explains the general idea. Imagine a two-dimensional space where one axis represents a topic ranging from galactic to extra-galactic astronomy, and where the other ranges from experimental/observational to theoretical. In this space, a paper titled "Gravitational Physics of Stellar and Galactic Systems" would get placed towards the upper right because its content is mostly about theory, with an emphasis on galactic astronomy. A paper titled "Topological Defects in Cosmology" would end up towards the upper left, because it is purely theoretical and about the extra-galactic astronomy. A person working in the field of observational/experimental extra-galactic astronomy will most likely read mostly papers related to this subject, and therefore get placed in the lower left region of this space. A clustering is a document grouping that is super-imposed upon this space, which groups together documents that are about similar subjects. As a result, this clustering defines a "thematic map". As mentioned, this is a highly simplified example. In reality the space

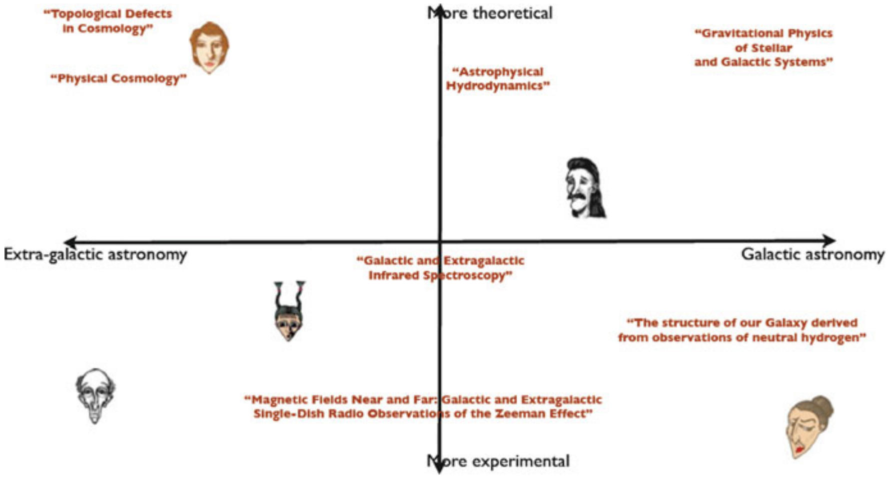


Fig. 3. A simplistic representation of topic space.

has many dimensions (100–200), and these cannot be named as intuitively as "level of theoretical content". However, the naming of various directions in this "topic space" is not something we need to worry about. The document clustering is the tool that we will be working with. Now that we have established this "topic space", together with the document clustering (defining the "thematic map"), how do we use this information to find recommendations?

Knowing to which cluster a new article has been assigned will allow us to find papers that are the closest to this article within the cluster. The first couple of papers in this list can be used as a first recommendation. The more interesting recommendations, however, arise when we combine the information we have about the cluster with usage information. The body of usage information is rather specific: it consists of usage information for "frequent visitors". People who read between 80 and 300 articles in a period of 6 months seems like a reasonable definition for the group of "frequent visitors". We assume that this group of frequent visitors represents either professional scientists or people active in the field in another capacity. People who visit less frequently are not good proxies because they are most likely incidental readers.

To find recommendations from usage data, we determine a group of papers, say 40, that are the closest to the new article within the cluster to which it was assigned. On the basis of this knowledge, we can recommend the following papers:

– the paper read the most directly before reading a paper from that group of 40
– the paper read the most directly after reading a paper from that group of 40

– the most read paper by the persons who read a paper from the group of 40
– the most recent paper in the top 100 of the most also-read list from point 3
– the paper which cites the most papers in that group of 40

where the last recommendation includes citation statistics in the recommendation. There are more recommendations than the ones mentioned above, but these 5 give a good impression of what is possible.

3 Example

As an example, take a recent e-print from arXiv, which, besides the arXiv subject headings, does not come with keywords. Assuming that the bibliography of the e-print contains papers with keywords, we use its bibliography to place the e-print in the document space. Once we know its position in this space, we can establish to which cluster this document belongs.

Michael Kurtz (Kurtz et al. 2009) shows an example of the recommender system applied to astronomy. The creation of the topic space for astronomy included an additional step of creating a normalized keyword set, because the set of documents used to create the topic space did not share a unique keyword system. The normalized keyword set is essentially a set of identifiers to which the keywords are mapped, using a number of filters and translation rules. Here we want to show an example taken from the physics literature. It was more straightforward to create a topic space for physics because of a large body of documents sharing a unique keyword system: the subject codes from the Physics and Astronomy Classification System (PACS). The topic space was constructed from articles that appeared in the *Physical Review* (A,B,C,D,E and Letters), *Physics Letters* (A,B), *Reviews of Modern Physics*, *Nuclear Physics* (A,B), the *Journal of Chemical Physics* and *Applied Physics Letters*, during the period 1995 through 2009. The resulting data set consisted of 144,110 articles. These articles were clustered into 100 clusters containing about 1,000 articles each, on average, using the clustering toolkit "CLUTO" (Karypis 2002) and the statistical toolkit "R" (R Development Core Team 2008). For usage data we used the anonymized arXiv usage logs for the period of July through December of 2009.

The example below takes a recent e-print (arXiv:1004.1856), titled "The Hubble Constant" by Wendy Freedman and Barry Madore. The bibliography of this e-print allowed us to assign it to a cluster. Based on this cluster assignment and the usage data (see above), the overview of "Related Articles" was constructed as shown in Fig. 4. All related articles deal with the structure of the Universe and aspects that are related to it, both observational and theoretical. Just from its title "An excess of cosmic ray electrons at energies of 300–800 GeV", you might wonder why the Nature article (Chang et al. 2008) is among the recommendations. But at closer inspection we see that, among others, the following subjects are discussed in its contents: supernova

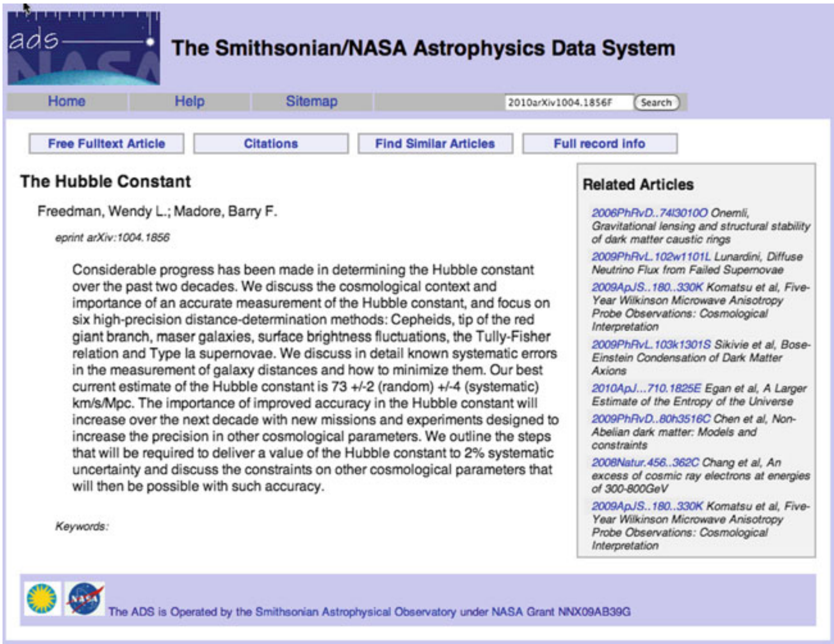


Fig. 4. Example of Literature Recommendation. The section of "Related Articles" has been constructed using PACS identifiers and anonymized usage logs.

remnants, dark matter particles and the Wilkinson Microwave Anisotropy Probe (WMAP). These are highly relevant to cosmology in general and the structure of the Universe is particular (see e.g. Spergel et al. 2003). The authors conclude "If the Kaluza-Klein annihilation explanation proves to be correct, this will necessitate a fuller investigation of such multi-dimensional spaces, with potentially important implications for our understanding of the Universe", a cosmological statement you might not have expected by just looking at the article title. In this sense the recommender system is not unlike an oracle from Ancient Greece, known for their often cryptic answers. The point we are making here, is the following: the title of an article, and even its abstract, only contain a superficial representation of the article content. Indexing the full article text will obviously result in the best representation of its content, but at the cost of significantly more processing and complexity. One would assume that the key aspects of an article are accompanied with citations of relevant literature, which should make the article bibliography an accurate representation of its content. If keywords classify articles with sufficient accuracy, keywords should therefore be sufficient to base a recommender system on. This illustrates another point with respect to the Nature article we just discussed: because our example was based on the physics literature, using PACS identifiers, all articles from e.g. *The Astrophysical Journal* in the bibliography did not participate in the analysis, because they use a different keyword system. Therefore, a recommender system should use labels, in our case keywords, that represent the whole body of literature it is intended for. This means for fields like cosmology and cosmic ray physics, for example, that our recommender system should use both the PACS identifiers and the normalized keyword set, mentioned earlier, in conjunction for optimal accuracy.

4 Concluding Remarks and Discussion

The technique used to build the recommender system has been around for quite a while. As early as 1934, Louis Thurstone wrote his paper "Vectors of the Mind" (Thurstone 1934) which addressed the problem of "classifying the temperaments and personality types". In this paper he extended and generalized Spearmans method of factor analysis into what is called the "Centroid Method" and which became the basis for modern factor analysis. Peter Ossorio (1965) used and built on this technique to develop what he called a "Classification Space", which he characterized as "a Euclidean model for mapping subject matter similarity within a given subject matter domain" (Ossorio 1965). Michael Kurtz (1993) applied this "Classification Space" technique to obtain a new type of search method. Where the construction of the "Classification Space" in the application by Ossorio relied on data input by human subject matter experts, the method proposed by Michael Kurtz builds the space from a set of classified data. Our recommender system is a direct extension of the "Statistical Factor Space" described in the appendix "Statistical Factor Spaces in the Astrophysical Data System" of this paper by Michael Kurtz.

We still need to answer the following questions: How stable is the recommender system? What is the best way to update the system? Do we want to allow cluster overlap when creating the thematic map in the topic space? Do keywords really describe the document universe with sufficient accuracy? What do we do with the body of literature for which no keywords are available? Can we optimize the system so that we can provide recommendations "real-time"?

In order to analyze the stability of the recommender system, we need to explore the dependence of its results on assumptions and simplifications that lie at the basis of the system. For example, we clustered the topic space into 100 clusters of about 1,000 articles on average per cluster. The number of papers (40) we selected for also-read analysis was arbitrary. If we vary these numbers, the results will undoubtedly vary. The main question is: do the results stay relevant?

The Literature Universe keeps expanding, and therefore the topic space needs to be updated on a regular basis. Do we have to recalculate it from scratch, every time, or is there a smart way to update it incrementally? It is probably a bit of both.

We probably should allow articles to be members of multiple clusters. For example, there are papers that are interesting to some people because they discuss e.g. an observational technique or instrument, and interesting to others because they provide information on a particular astronomical object.

Whether keywords really describe the document universe with sufficient accuracy is directly related to the question of whether a keyword system is sufficiently detailed to classify articles. We assume the latter is true. Having said this, we do realize that a keyword system can never be static because of developments within a field and because of diffusing boundaries with other fields. We use the keywords provided by the publishers, so the scope and the evolution of the keyword spectrum is out of our hands. It also means that a recommender system based on publisher-provided keywords has one obvious vulnerability: if a major publisher would decide to stop using keywords (e.g. PACS identifiers), it would pose a significant problem.

One way to build a recommender system without the use of keywords is to use Latent Semantic Indexing. This would also take care of the body of literature for which we do not have keywords, for example historical literature. It is probably not sufficient to index just article abstracts, because of the reason we mentioned in the previous section. However, if we were to index the abstract of an article together with the abstracts of the articles in its bibliography, the level of accuracy would probably be acceptable. This approach is more noisy, but there is enough redundancy in the data to find correlations of a strength similar to the ones we find when using keywords.

The part of the recommender system that needs the most computing power can be pre-computed: the topic space and the resulting clustering. The "real-time" part consists of assigning a document to a cluster and essentially intersecting a subset of articles from that cluster with usage data. These data sets are small enough to be processed "real-time", so we feel that providing visitors with recommendations "on the fly" is an achievable goal.

We conclude that a recommender system is feasible for any information retrieval system containing sufficient meta data to classify its records and with a large enough set of regularly returning visitors, providing either implicit ("click stream") or explicit feedback. Prototype results seem to indicate that using a keyword system for article labeling provides a meaningful literature recommender system. We realize that a different approach will be necessary to include records without keywords.

Acknowledgement. The ADS is Operated by the Smithsonian Astrophysical Observatory under NASA Grant NNX09AB39G. The ADS gratefully acknowledges the support of the arXiv team.

References

Chang, J., et al. 2008, Nature, 456, 362

Henneken, E., Kurtz, M. J., Eichhorn, G., Accomazzi, A., Grant, C. S., Thompson, D., Bohlen, E., & Murray, S. S. 2007, Library and Information Services in Astronomy V, 377, 106

Karypis, G. 2002, Cluto. http://glaros.dtc.umn.edu/gkhome/views/cluto/

Kurtz, M. J. 1993, Intelligent Information Retrieval: The Case of Astronomy and Related Space Sciences, 182, 21

Kurtz, M. J., Accomazzi, A., Henneken, E., Di Milia, G., & Grant, C. S. 2009, arXiv:0912.5235

Ossorio, P. G. 1965, J. Multivariate Behavioral Research, 2, 479

R Development Core Team (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Spergel, D. N., et al. 2003, ApJS, 148, 175

Thurstone, L. L. 1934, Psychological Review, 41, 1

15

Linking Literature and Data: Status Report and Future Efforts

Alberto Accomazzi

Harvard-Smithsonian Center for Astrophysics, 60 Garden Street,
Cambridge, MA 02138, USA
aaccomazzi@cfa.harvard.edu

Summary. In the current era of data-intensive science, it is increasingly important for researchers to be able to have access to published results, the supporting data, and the processes used to produce them. Six years ago, recognizing this need, the American Astronomical Society and the Astrophysics Data Centers Executive Committee (ADEC) sponsored an effort to facilitate the annotation and linking of datasets during the publishing process, with limited success. I will review the status of this effort and describe a new, more general one now being considered in the context of the Virtual Astronomical Observatory.

1 Introduction

Links between papers in ADS and data products hosted by astronomy archives have existed since 1995. These links have been created and curated by librarians and archivists as part of the data center's effort to collect information about the scientific use of the data being hosted by the archive. The links provide more than just useful connections between bibliographic records and observations that allow users to access related material. They represent part of the scientific artifacts created during the research lifecycle of an astronomer, and as such are needed to fully document and describe the research activity itself (Accomazzi 2010).

One obvious benefit which comes from maintaining such links is the ease with which one can generate metrics about the scientific impact of the observations in the form of published papers or citations. Thus, in a highly competitive scientific discipline such as astronomy, maintaining the complete record of paper-data connections has become an accepted way to evaluate a project, mission, or even an entire research field (Grothkopf and Lagerstrom 2010; Trimble and Ceja 2010). The metadata collected during the creation of links maintained by ADS and its collaborators have so far been limited to some very basic information about the location of the resources linked together. Typically these are simple mappings of ADS's bibliographic identifiers (*bibcodes*) and URLs pointing to one or more particular data product(s)

A. Accomazzi (ed.), *Future Professional Communication in Astronomy II*,
Astrophysics and Space Science Proceedings 1, DOI 10.1007/978-1-4419-8369-5_15,

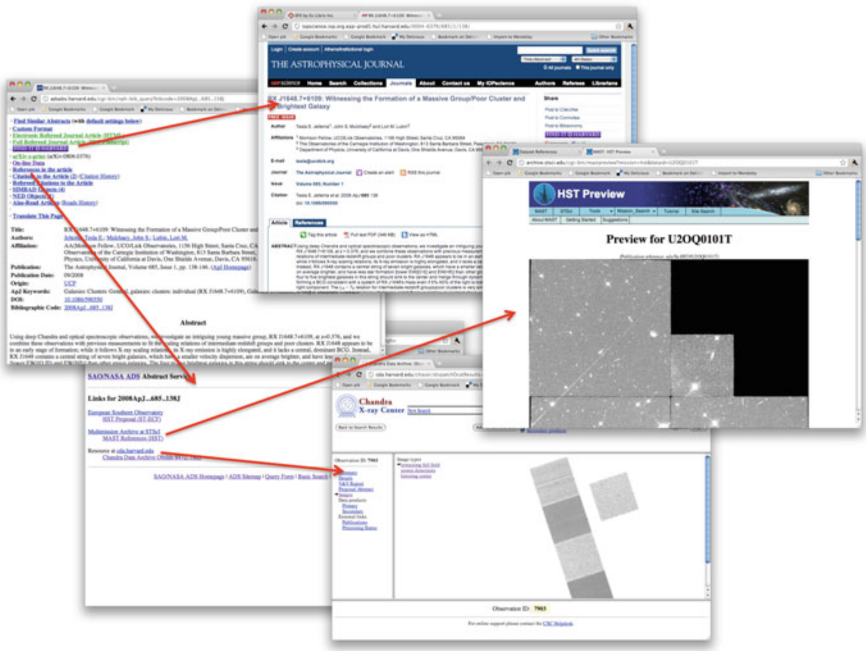


Fig. 1. Links between an ADS record, the full-text manuscript hosted by a publisher, and data products available from MAST and Chandra.

hosted by an archive. The ADS record will of course also have a link to the published paper itself, acting as a "bridge" between the manuscript and the data described therein. Figure 1 shows the connection between a paper and data products available from the Multimission Archive at Space Telescope (MAST) and the Chandra X-Ray Archive.

2 Development of Dataset Linking Infrastructure

In 2002, it became apparent that the methodology adopted to create, maintain and share these linkages could be improved. Thus, in 2003, the NASA Astrophysics Data Centers Executive Council (ADEC) and the AAS journals issued guidelines aimed at improving the situation (Eichhorn 2004). This new effort was aimed at addressing four separate issues in the management of these links: their curation, naming, resolution and persistence.

The creation of links to data products has been a time-consuming activity usually carried out by a librarian or archivist. Rots et al. (2004) describe the effort required to perform this activity, which typically consists of scanning the literature to identify which papers mention one or more data products from a particular archive, and then link those papers with the relevant datasets.

In 2004, in order to facilitate this activity, and in coordination with the ADEC proposal, *the Astrophysical Journal* introduced the capability for authors to properly tag the datasets analyzed in the paper. This introduced a mechanism to formally “cite” data in a way similar to how scientists cite other papers. According to this plan, citations to data products would be vetted by both editors and referees during the manuscript editorial process, and links would be created to the corresponding data products as part of the process which generates the online HTML version of the paper. The correlation between a paper and the datasets referenced therein would then be propagated back to the ADS and the participating data centers via metadata exchange.

The implementation of this linking proposal would not only benefit end-users, but would potentially provide significant savings in the curation efforts of archivists and librarians, who could now harvest these linkages directly from ADS, thus reducing the need for the manual scanning of the literature.

In order to properly cite the datasets in the literature, the ADEC and AAS adopted a standard way to uniquely identify data resources based on the IVOA Identifier standard (Plante et al. 2006). The proposed system of nomenclature (Accomazzi and Eichhorn 2004) provided a standard for dataset identifiers which featured some important properties. Among them: uniqueness (one resource corresponds to a single identifier), and persistence (identifiers do not change even when data products are migrated to a different archive). The identifiers were designed to support the naming of resources with a broad range of granularity and included a “public” prefix identifying the archive or mission that generated the dataset as well as a “private” key identifying the data item within a specific collection.

In order to ensure the proper use and persistence of links to datasets, the ADEC charged the ADS with the task of setting up a verification and resolution service for dataset identifiers. In this role, the ADS would act as the registration authority on behalf of the community, creating the infrastructure necessary to enable the dataset linking. During copy-editing of a paper, the editors would use an automated tool provided by ADS to verify that a particular dataset identifier is known and can be resolved to an online resource. Upon successful verification, the identifier would be incorporated into the paper with a link to a resolution service provided by ADS (rather than a simple link to the current URL for the resource). This model provides a level of redirection which can be used to properly track a dataset if and when it moves from one archive to another, and allows the resolver to provide options should multiple versions of a data product be available. A complete description of this implementation can be found in Accomazzi et al. (2007). Elements of this architecture are similar to the Digital Object Identifier standard used for the persistent linking of scholarly publications, which is discussed in Sect. 4. However, this system was designed to be fully managed by the astronomical community requiring a minimal level of effort for institutional buy-in.

3 Current Status

Six years have passed since the introduction of the dataset linking infrastructure, so now is a good time to take stock of this effort. From a system design point of view, some of the features that made the implementation of this system attractive have, in retrospect, proven to be obstacles to its long-term success. Chief among all problems with the registration of dataset identifiers has been enforcing their persistence. Since data products ultimately reside within archives that participate in the ADEC but which are run independently of each other, the implementation and maintenance of services that provide access to the data is left to the archives themselves. Given that the thrust of this effort is completely voluntary, there is no contract or reward system which can be leveraged to enforce the long-term resolution of and access to a particular dataset. Experience shows that unless requirements for the preservation of these linking services become part of the archive operations, a simple system upgrade is enough to break valuable links to dataset resources. As an example, over 200 dataset identifiers which were published in a 2004 ApJ Supplement special issue on the Spitzer Space Telescope are no longer resolvable due to a change in the Spitzer Science Center interface.

Unfortunately the adoption and use of dataset identifiers in the literature has not been a success story. Citations to datasets began appearing in 2005 and increased in the following 2 years, peaking in 2007, before decreasing in 2008 and finally going down to zero in 2009 (see Fig. 2). The reasons for this

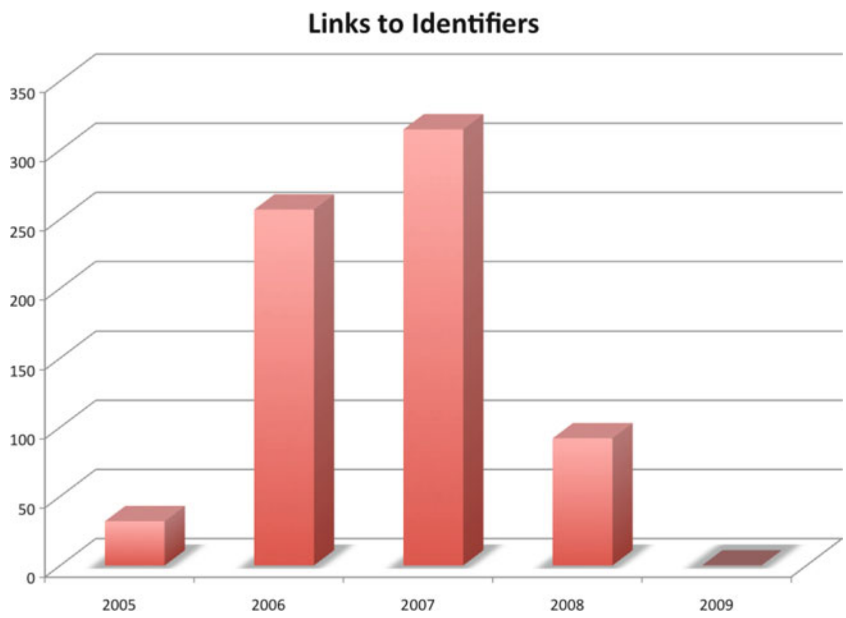


Fig. 2. The number of links to data entered in AJ and ApJ papers as part of the ADEC/AAS dataset linking effort.

reversal are not entirely clear, but can probably be attributed to a variety of factors. First and foremost, even though the ADEC approved a policy encouraging archives and users to make an effort to more widely publish dataset identifiers, anecdotal evidence shows a low level of awareness from scientists of this possibility. Researchers tend to be busy and, unless properly coached by editors and archivists, will easily overlook new demands or stipulations requiring additional work on their part. In addition, data archives don't always make it obvious how a particular dataset (or data file) should be cited in the literature by the scientists who make use of the data in their research. Since astronomers have been accustomed to referring to the data they have used in terms of specific observations or regions of the sky, this general practice is still used. Rather than unambiguously identifying the data using dataset identifiers, astronomers describe how the data can be obtained. While this is a reasonable way for an author to convey the necessary information about the data being studied, it obviously defeats the goal of creating persistent, unambiguous, machine-readable links to the data products. Finally, it was hoped that after the initial adoption of the standard for dataset identifiers within the ADEC more data centers and more journals would follow suit, but this did not materialize. The critical mass and community awareness necessary to make this common practice was never reached.

4 A Way Forward

Despite the lack of adoption from the community of the proposal described above, every scientist, librarian and archivist agrees that preserving data products and publishing links to data in the literature is a worthwhile effort. We believe that at this point in time we should even be more ambitious, and recognize that in fact for our discipline to flourish in the digital era all artifacts related to the research lifecycle need to be available online, and properly interlinked (Accomazzi 2010). Thus, the issue of creating links from the literature to data products can be recast in a wider scope – the preservation and interlinking of digital assets in astronomy. The use of the term "digital assets" in this context refers to artifacts used and generated during the research activity of an astronomer. This includes observing proposals, observations, archival data from surveys and catalogs, observing logs, tables and plots that are published in manuscripts. In short, we advocate capturing all the data and knowledge that has gone into the research activity itself, with the aim of providing a digital environment that can support the repeatability of the research described in a publication. While there are many possible implementations of a digital environment for data preservation, it is clear that any such effort must satisfy a set of principles. Below we identify some of the basic requirements that we believe will need to be addressed in the near future.

4.1 Management of Digital Assets

First, one should consider the issue of nomenclature and persistence of digital assets. Since we can expect that the data referred to in a paper will be hosted on a distributed set of digital repositories, naming and linking standards need to be clearly defined and adopted in order to create persistent links to such resources. The solution proposed by the ADEC was primarily designed to satisfy the requirements of uniqueness and persistence for data already available in well-established archives using community-developed standards. While this approach is technically sound and seemed at the outset to provide the best solution to the problem, new technologies and standards since developed by the digital library community now offer attractive alternatives that should be considered. For the creation and management of unique identifiers, the Handle System is a general purpose distributed information system for the minting and resolution of unique identifiers on the internet. The Digital Object Identifier (DOI) system is an application built upon the Handle system and is widely used by the digital publishing industry. Organizations making use of the DOI system agree to a business model that requires the deposit and active curation of metadata for the digital assets registered in the system, and are subject to fines if found to be in breach of the DOI Foundation rules. This elevated level of commitment provides a certain level of assurance that the digital assets registered in the system will be properly maintained. In addition, the DOI foundation explicitly imposes requirements on its members to ensure the long-term survival of the system. For instance, should one of its members cease operations, the DOI resolution of its content would be transferred to other members of the foundation. When it comes to the preservation of data products, this type of long-term commitment has never been formalized or made explicit by most of the publishers, societies or even astronomical data centers (except for the case of the active NASA missions, whose digital assets are transferred to archival centers at the end of a mission). However, it is exactly the type of commitment we believe is essential for our community to make at this time.

4.2 Archival and Preservation

In order to enable the publication and broader re-use of scientific high-level data products, researchers should be required to upload such data to one or more trusted, community-curated, digital repositories. Not only does this requirement allow repeatability of experiment and analysis, but it promotes a level of transparency and trust that is an important component of the scientific discourse.

While much of the tabular data now published in scientific articles ends up being stored in services such as NED and Vizier, a significant amount of supplementary material does not make it into such archives. To be sure, authors are often encouraged to submit machine-readable versions of the data (or even computer code) as supplementary material submitted to the journal, but the uniformity, longevity, re-usability and discoverability of such material are at this point highly inconsistent and questionable. In addition, no explicit or common migration plan has so far been defined by publishers or learned societies, so future access and curation of these assets is not assured. We believe that it is essential to encourage the deposit of digital assets in a wide range of trusted repositories curated by the community in collaboration with the journals. The kind of material deposited in such repositories will supplement the products which are currently curated by projects such as NED and Vizier, and consist of anything which does not fit in the typical description of a data table or catalog. This may include published images, plots, observing notes, workflows, software, intermediate results, and large data collections. In order for these data products to be useful, it is essential for the user to deposit and for the repository to expose not just raw data but also its related metadata. This should include, at a minimum, a description of the datasets in the sample, a set of applicable keywords, and some notes relating the data in question to the published paper(s) in which they were used.

The need to create a digital infrastructure in support of these activities is increasingly being recognized both in the US and in Europe by funding agencies such as NASA, the NSF, JISC and digital preservation programs are now being defined (Choudhury et al. 2007). In particular, the NSF DataNet Data Conservancy program has established as one of its goals the support of scientific inquiry through the adoption of a comprehensive data curation strategy. Today there are a number of open-source digital repository systems available, some of which have already been deployed by universities and projects involved in preserving digital institutional assets (e.g. Fedora or DSpace). Other initiatives, such as the Dataverse Network (King 2007), provide a scientist-centered framework for storing data products associated with publications and encourage their citation through the use of unique, permanent dataset identifiers based on the Handle system.

5 Discussion

Even though there is general agreement that publishing and citing data is a noble goal and worthwhile effort, the experience of the ADEC data linking effort has shown that it takes sustained community engagement to turn a proposal into a successful activity. When many people and organizations are involved in providing crucial components of such a distributed system, the risk of multiple points of failure becomes significant and can ultimately spell the demise of even the best thought-out technical scenario. Rather than giving up on this worthwhile idea, we should take the failure of adoption as a learning opportunity to devise a more robust system that can not only provide links to existing datasets available from well-established archives, but also provide the capability of storing author-supplied data and metadata related to publications.

In the long run, it is likely that there will be an ecosystem of different repositories and technologies used for the preservation of research products. Some of them will be more focused on actively capturing and curating datasets, as is today done by projects such as Vizier and NED, and others which will provide an infrastructure which can be used by scientists and publishers to self-manage data products associated with published papers. The overarching goal of such systems should be to provide useful services to the astronomy community and guarantee that the data deposited in such repositories will be properly curated and preserved for the foreseeable future. This includes providing a migration path for obsolete data formats, curating and exposing metadata of digital assets in the repository, and providing discovery services to its content. This commitment comes at a cost, which should be shared by the community as part of the effort which funds the infrastructure supporting astronomical research. In addition, it is essential that we promote and encourage policies that foster and facilitate the growth of the digital scholarly environment that the Virtual Observatory has been envisioning. The recent funding of digital preservation frameworks such as the Data Conservancy project suggests that the time has come for the VO to play a major role in the capture and preservation of the astronomy research lifecycle. We look forward for the members of the International Virtual Observatory Alliance to take a pro-active role over the next decade in order to make this vision a reality.

References

Accomazzi, A., & Eichhorn, G. 2004, Astronomical Data Analysis Software and Systems XIII, Astronomical Society of the Pacific Conference Series, 314, 181

Accomazzi, A., Eichhorn, G., & Rots, A. 2007, Astronomical Data Analysis Software and Systems XVI, Astronomical Society of the Pacific Conference Series, 376, 467

Accomazzi, A. 2010, Library and Information Systems in Astronomy VI, Astronomical Society of the Pacific Conference Series, 433, 273

Choudhury, S. et al. 2007, The International Journal of Digital Curation, 2, 20

Eichhorn, G., & Astrophysics Datacenter Executive Committee (ADEC) 2004, Bulletin of the American Astronomical Society, 36, 805

Grothkopf, U, & Lagerstrom, J. 2011, These proceedings, Chapter 12

King, G. 2007, Sociological Methods & Research, 36, 173

Plante, R., Linde, T., Williams, R., & Noddle, K. 2006, http://www.ivoa. net/Documents/latest/IDs.html

Rots, A. H., Winkelman, S. L., Paltani, S., Blecksmith, S. E., & Bright, J. D. 2004, Astronomical Data Analysis Software and Systems (ADASS) XIII, Astronomical Society of the Pacific Conference Series, 314, 605

Trimble, V., & Ceja, J. A. 2010, Astronomische Nachrichten, 331, 338

16

FPCA-II Concluding Remarks

Michael J. Kurtz

Harvard-Smithsonian Center for Astrophysics, 60 Garden Street,
Cambridge, MA 02138, USA
kurtz@cfa.harvard.edu

1 The Times They Are a-Changin'

We live in interesting times. Modern technology is turning much of our world upside-down and inside-out; this is especially true in the realm of communication. This conference, the latest in a long series organized by André Heck and his collaborators, has made it clear where we are, how we got here, and (so far as one can see into the future) where we go from here for the professional communication between astronomers, and with the general public.

The rate of change in the last 20 years has been astounding, centuries old traditions and methods have been replaced in but a twinkling of the historical eye. In the last 20 years:

1990: CDS/Simbad was nearing 10 years old, NED had just been started. There were no electronic journals, but discussions had been started. ADS did not yet exist; nor did arXiv.

2000: All astronomy journals were electronic; arXiv had become the dominant provider of newly published articles; ADS had replaced the print libraries as the source of journal articles.

2010: Journals are beginning to cease production of their paper editions; many young scientists have never used a paper journal. Several initiatives are working toward the direct publication of data.

2 Some Things Stay (Somewhat) Stable

The principal goal of professional communication is just that: professional communication. This is where the new technologies are making the largest impact. As both André and John discussed in their introductory talks, there are other considerations.

Recognition, Validation, Preservation are among the secondary goals of the communication process. The new dissemination techniques have not fully

A. Accomazzi (ed.), *Future Professional Communication in Astronomy II*, Astrophysics and Space Science Proceedings 1, DOI 10.1007/978-1-4419-8369-5_16,

addressed this, a great deal of discussion at this, and other venues, involves these issues.

ArXiv, perhaps, represents the issues most clearly. While the journals continue to fulfill the recognition, validation and preservation roles, arXiv has already taken over the distribution and awareness functions. In money terms the per article cost of arXiv is less than 1% that of the journals and arXiv does not require an army of referees to work for free. The timeliness of the publications is perhaps the greatest "cost." Are the benefits of the referee required revisions worth the 6 month delay?

Does the hiring and promotion process require the validation and recognition step of publication in a well-cited refereed journal? We cannot all be Grisha Perelman.

Publishers and journals are working to provide better services; refereed data and rich semantic tagging seem two promising directions.

3 The End of Paper – Implications

The millennia long reign of paper as the medium of communication is clearly at its end. While authors and publishers have more or less been able to adapt to the change from paper based to electronic printing, the organization most affected by the change is the library.

Libraries were formed as local information centers. With the advent of wide area networking, local information stores are much less needed. Of the 10,000 or so arXiv papers per month read at Harvard, none is stored in Cambridge; 97% of the more than 12,000 journal article downloads per month originating at Harvard, and via ADS, also come from sources physically outside Cambridge.

As printed paper changes from active research tool to historical artifact, the physical nature of libraries will change. Many libraries will find more compelling uses for their physical space than as centers for archival preservation.

The intellectual contributions of librarians will also change. Till now a substantial portion of a librarian's responsibility has been choosing what items to have in the library. This has now evolved into which electronic licenses to purchase. As the funding models for electronic publishing change from the paper/library based model, in particular as Open Access models become dominant, this responsibility will become less important.

In the past couple of decades new, electronic libraries have been created with fully globally distributed clientele; CDS, ADS, MAST and the Chandra and ESO archives are examples. These new libraries are developing joint operational capabilities through the standards and protocols being developed by the virtual observatory projects, IVOA, VAO, EuroVO, etc.

Librarians are now taking the lead in developing and maintaining productivity measures for individuals, projects, and facilities. They are increasing acting as creators of meta-data, organizing data-bases of individuals, observations, instruments, and many other, often abstract, topics.

Librarians are likely to be the glue which holds our vast digital resources together.

4 New Forms of Scholarship

Perhaps the most obvious "new" form of scholarly output is the data-base. This is not exactly new; catalogs and atlases have long been important scholarly works. The Sloan Digital Sky Survey is a clear descendant in an ancient process which would include the 150 year old Bonner Durchmusterung and the 50 year old Palomar Observatory Sky Survey.

Recognition, Validation, and Preservation are all serious issues for modern scientific data projects.

Often the key finding of a research project can be shown in a single graph or chart. Normally these would be included in a journal article, and after a significant delay. Modern technology makes it quite possible to publish single atoms of information, perhaps as blog entries. Will things like this happen?

For me the most exciting new development in scientific communication is the introduction of provenance or work-flow systems. These are basically publishing systems which allow one to encode and share exactly how a result was obtained, in such a manner that portions can be extracted and reused on different inputs. Twenty something years ago Alberto Accomazzi and I wrote down the steps necessary to get calibrated photometry of galaxies from glass copies of the POSS. There were more than 150 (sometimes quite complex) steps, with a large number of data dependent decision points and looping branches. It would have been nice to be able to share this, certainly at the time we published the redshift survey based on the photometry.

The most important first use of provenance systems in astronomy will likely be to publish the details of the pipeline processing systems for our largest and best used instruments.

5 Coming Attractions – FPCA-III

In a few years many of us will likely meet again to discuss how things are and how things have changed. Much of what has been presented at this meeting will still be of current interest, but for some topics we will have moved on.

The journal of the future will be well underway in a few years, dense interlinking and interactive graphics are high on the list of possible improvements.

Progress, perhaps mediated by government access mandates, will have been made in developing new funding mechanisms for scholarly communication. Perhaps some will see a final stable solution for the future (perhaps not).

The role of the journals in the vetting and archiving of the data used in articles will continue to be defined.

Librarians will continue to re-define their roles and responsibilities.

ADS will have yet another nifty new feature.

Future Professional Communication in Astronomy II

(Cambridge, Massachusetts, 13–14 April 2010)

Alberto Accomazzi

The present volume gathers together the talks presented at the second colloquium on the *Future Professional Communication in Astronomy (FPCA II)*, held at Harvard-Smithsonian Center for Astrophysics (Cambridge, MA) on 13–14 April 2010.

This meeting provided a forum for editors, publishers, scientists, librarians and officers of learned societies to discuss the future of the field. The program included talks from leading researchers and practitioners and drew a crowd of approximately 50 attendees from 10 countries.

These proceedings contain contributions from invited and contributed talks from leaders in the field, touching on a number of topics. Among them:

- The role of disciplinary repositories such as ADS and arXiv in astronomy and the physical sciences;
- Current status and future of Open Access Publishing models and their impact on astronomy and astrophysics publishing;
- Emerging trends in scientific article publishing: semantic annotations, multimedia content, links to data products hosted by astrophysics archives;
- Novel approaches to the evaluation of facilities and projects based on bibliometric indicators;
- Impact of Government mandates, Privacy laws, and Intellectual Property Rights on the evolving digital publishing environment in astronomy;
- Communicating astronomy to the public: the experience of the International Year of Astronomy 2009.

A. Accomazzi (ed.), *Future Professional Communication in Astronomy II*,
Astrophysics and Space Science Proceedings 1, DOI 10.1007/978-1-4419-8369-5,