

FROM DATA FILES TO INFORMATION HUBS: BEYOND TECHNOLOGIES AND METHODOLOGIES

A. HECK
Observatoire Astronomique
11, rue de l'Université
F-67000 Strasbourg, France
heck@astro.u-strasbg.fr

Abstract. A few comments are humbly offered in the light of about thirty years of activities linked to information flow in astronomy.

A virtual universe

Let's face it: astronomy is largely a virtual science.

With the exception of experiments carried out *in situ* by solar-system spacecraft, our knowledge of the universe is totally derived from photons reaching us from the outer space. And because of the finite speed of light, we do not observe the objects the way they are, but the way they were when the photons we are collecting actually left them.

What we have thus in our data files is nothing other than a huge and complex virtuality of prior stages, differentiated as a function of the distance in space and time of the various sources. Thus the job of astronomers is to work on that space-time mosaicked virtual universe in order to figure out what is exactly the real universe and to understand the place and rôle of man in it.

As a result of the huge amount of data accumulated, but also by necessity for their extensive international collaborations, astronomers have been pioneering the development of distributed resources, electronic communications and networks coupled to advanced methodologies and technologies often much before they become of common world-wide usage.

But exactly because of this kind of paradox (immense philosophical impact versus inability to interact with the observed subjects), the astronomers have to be specially careful not only of the way data are handled and archived for their own professional usage and for the scientific memory

of the future, but also of the way astronomy information in the broad sense is conveyed to society at large.

At a time when taxpayers (who are nowadays in practice the exclusive source of support for our research) seem to have other priorities (such as environment, health, security, unemployment) than space investigations and cosmological perceptions¹, it is certainly appropriate to keep in mind that a good image (which implies efficient and consistent public relations policies) and top-quality education through all media possible (including ‘simple’ but valuable web sites) remain the best tools to ensure continuing public support of our science. This, in return, will be able to pursue at best its noble mission of building the cosmic context of mankind’s evolution.

The astronomy information flow

As schematically illustrated in Fig. 1 of the Foreword and by the range of matters tackled in this book, the information flow in astronomy is much more complex than what most books and courses on astronomical data are describing – at least if one wishes to encompass the whole elaborate process from sources (cosmic objects) to end (mankind’s knowledge).

The various chapters of this book have detailed many boxes and links of that information flow diagram. It could have been made much more complicated to read, at the limit of legibility, by additional boxes and links, especially by shifting slightly the perspective, such as usage instead of flow. For instance, libraries could be linked to databases, to referees, to expert committees, to amateur astronomers, to the public at large, and so on.

Data centres have not been assigned a box *per se* in that scheme as they can be spread over several existing ones. Even if not fully realized yet, the key rôle of data centres has been recentred over the past decade. While they are now flourishing as distributed resources thanks to the networking of the planet, the advent of these electronic networks might have signalled for ever their destiny as information transit posts (*hubs*) or, as some say a bit disrespectfully, *astrogroceries*. We shall come back to this later on.

It is outside the scope of this book, and of this chapter in particular, to offer a whole treatise on astronomical information. Here are however for the newcomer to the astro-data world a few pointers, again mainly oriented towards data collection, processing, archiving and distributing:

¹It is important to distinguish here two factors, often confused in public surveys: interest (pure interest in astronomy remains high – cf. the survey mentioned by Madsen & West in their chapter of this book) and the priorities set at the time to spend pennies. Fundamental research, including astronomy, benefited indirectly from Cold-War-associated funding. The ‘end-of-Cold-War’ effects predicted by Daniel Goldin (NASA Head) in his talk at the AAS Meeting in Phoenix (Jan. 1992) are already well perceptible (see Heck 2000 or <http://vizier.u-strasbg.fr/~heck/org.htm>).

Borison *et al.* (1996) on modern observing modes; Egret & Albrecht (1995) on a wide selection of databases, archives, data centres and information systems; Kidger *et al.* (1999) on Internet resources; Murtagh & Heck (1987) on multivariate data analysis; Heck & Murtagh (1989) on knowledge-based systems; Heck & Murtagh (1993) on information retrieval; Starck *et al.* (1998) on image processing; Heck (1997a) on electronic publishing; and finally Jaschek (1989) on astronomical data in general, still largely valid.

Series of specialized conferences provide also sources for advanced material: the *Astronomical Data Analysis Software and Systems (ADASS)* conferences (see *e.g.* Mehringer *et al.* 1999); the *Astronomy from Large Databases (ALD)* conferences (see *e.g.* Heck & Murtagh 1992); the *Data Analysis in Astronomy* ('Erice') workshops (see *e.g.* Di Gesù *et al.* 1997); the *Library and Information Services in Astronomy (LISA)* conferences (see *e.g.* Grothkopf *et al.* 1998); the *Converging Computing Methodologies in Astronomy (CCMA)* workshops (see *e.g.* Heck & Murtagh 1996 and Molina *et al.* 1997). Refer also to the numerous references quoted in those works as well as at least two historical colloquiums (*e.g.* Jaschek & Wilkins 1977 and Jaschek & Heintz 1982). Schools have also been occasionally organized nationally and internationally (among these, see *e.g.* Hauck & Sedmak 1985).

The list above is definitely not exhaustive. Because of the quick evolution of information technologies nowadays, and because of the impact they have on the dynamics of the astronomy community, some aspects of the most recent compilations and reviews could become rather quickly outdated. The basic principles will largely remain unchanged though.

The present chapter is mainly offering a few thoughts derived from experience in all these fields, often with references to the other chapters of this book. In no way is there any intention to give lessons to anyone: just some modest experience-sharing exercise from a largely independent observer of, and sometimes actor/user in, the data information world in astronomy during the last three decades – and with some practice of society at large with respect to both the impact of astronomy and communications techniques in the broad sense.

A personal note

If the reader is kind enough to allow a few personal lines, this book is very close to the author's heart because life has been such that, wishing it or not, by taste or by duty, often as a user and sometimes as an actor, he had the opportunity to experience most of the aspects related to contemporary handling of astronomical information. And it is difficult to resist listing them here.

The reader should of course feel free to jump to the next section. However it should be obvious that, if these activities are listed here, this is not to brag about them, but to give more weight to the comments offered in the subsequent sections.

Here is a schematic, roughly chronological and non-exhaustive *aperçu*:

- researcher in both theoretical and observational astrophysics proposing already back in 1969 an SDI system for astronomers²;
- witnessing in an office at Paris Observatory the arrangements for setting up CDS³ of which he was to become the first scientific user and later, for a few years, an active ‘salesman’ of its resources;
- advocating in the mid-seventies advanced information extraction methodologies that would become popular a decade later under various labels;
- being trained in documentation and communication techniques, including acting as a public-relations officer for a professional research institution;
- developing intense public-outreach activities and amateur-astronomer supervision including contacts with all levels of communications media;
- involved in a space experiment (IUE) with preparing proposal calls and assisting selection committees, then training observers, assisting real-time operations, supervising data processing, and managing a large archive⁴;
- developing yellow-page services, first on paper (end of the seventies) and later as electronic resources (see *e.g.* Heck 1997b and the references therein);
- organizing back in 1990 the reflection of the astronomy community on electronic publishing (see *e.g.* Heck 1997a and the references therein);

²SDI = Selective Dissemination of Information. That project was turned down as the boss at that time was instead expecting progress on secular stability in pulsating stars out of the new head in his group ...

³It so happened that the gentleman who accepted to supervise the author’s PhD thesis in 1970 was to become the first CDS Director (CDS = *Centre de Données Stellaires* officially created in 1972 and installed at Strasbourg Observatory; the name changed subsequently to *Centre de Données astronomiques de Strasbourg* in order to encompass non-stellar data while retaining the well-known acronym).

⁴IUE = International Ultraviolet Explorer (NASA/ESA/SERC). The ‘data flow’ that some big observatories have now found necessary to formalize (see *e.g.* Quinn 1996 and <http://www.eso.org/projects/dfs/>) was an everyday notion for all resident astronomers involved in the IUE project and is very familiar to anyone involved in contemporary operations. To many visiting astronomers, the IUE observatories (US and European) were also the first experience of a standardized chain from observation preparation (precise-position constraints, etc.) up to delivery of a processed material with limited-duration proprietary rights.

- currently investigating the interface of the astronomy community with other disciplines and the society at large⁵;

to which one must add the usual activities linked to research, publishing, teaching, thesis supervision, and so on. No special merits as these were/are exciting times!

Concepts and buzzwords

Concepts are sometimes poorly defined in data-related books and occasionally lead to misunderstandings. Without entering so-called ‘Jesuitic’ discussions on some terms (such as data, information, knowledge, ...), it might probably be a good idea for scientists to spend some lines to introduce as precisely and as concisely as possible what is their understanding of the main terms they use, and this especially in new fields where concept forming is at risk of being a bit anarchical at the beginnings.

To complicate the picture, next to scientific concepts, we are also living in a world of buzzwords⁶. These are useful when well introduced and justified. They summarize ideas and projects in an imaginative way and can be excellent vectors to ‘sell’ them to decision makers and takers, to the community, and to the society at large. Some of them might even make it into history.

Their semantic substance must however be representative of what they are labelling and not to be sources of confusion nor, in the worst cases, of deceit. Abuses of language should be avoided and concepts should be appropriately used. The sensationalism introduced by most communications media nowadays should not be followed in science, essentially in terms of excesses flourishing now in the everyday language⁷.

Some big projects announced on web pages are only made of a few links put together in a few minutes, projects for which at least in one case the designer got commended and distinguished. The reader will surely understand if no names are given here and if we rather go to the heart of the message: *substance* (and the work behind it) should always have precedence over the attractiveness of labels and the hype around them.

Virtual observatories?

Speaking of buzzwords, a new one appeared recently in the literature: *virtual observatory*. While highly desirable and commendable, the proposed

⁵See e.g. <http://vizier.u-strasbg.fr/~heck/p-socio.htm> and Heck(2001).

⁶The author has also a training in marketing and advertising where devices, slogans and buzzwords are everyday bread.

⁷The stock exchanges *drop dramatically* when they lose only a fraction of a point. What would one say today for a 1929-style crash?

structures will be quite far from the classical conception of an astronomical observatory devoted to the collection of new data. The label could thus be seriously misleading since additionally a fundamental feature of the actual universe will be disregarded: its omnipresent variability⁸.

For instance, the project known in the US as the *National Virtual Observatory (NVO)*⁹ is basically the aggregation of complementary multi-wavelength surveys. Other projects currently in the air are putting more emphasis on the methodological ways of tackling the existing – and largely dormant – amount of data, not only in astronomy, but also in Earth and environmental sciences¹⁰.

A related project with a less questionable label (only the ‘instrument’ here is virtual) has been launched recently: *AstroVirtel*¹¹ aiming at making accessible the ESO/ST-ECF archive that currently contains more than 7.0 Terabytes of scientific data obtained with the ESA/NASA Hubble Space Telescope (HST) and with several ESO large ground telescopes.

Success stories

Among the many success stories of data centres, an exemplary one is certainly that of the Strasbourg astronomical Data Centre (CDS). Over almost three decades now, it moved from a file holder to an impressive information hub. It is recognized world-wide for the excellence of its work and its products – the kernel of which is the huge database SIMBAD of catalogue synonyms giving access to all individual data from the integrated catalogues, all of this being completed by an object-oriented bibliography linked to the database and by several complementary services.

Fig. 1 shows the various elements constituting today CDS’s hub. It is out of the scope of this chapter to review in detail the history of CDS and its component parts, but interested readers can find more by referring to the CDS web site¹², to the *CDS Information Bulletin* series (now discontinued), as well as to the papers just published as a special issue of *Astronomy & Astrophysics Supplement Series* (Vol. 143, 1, April 2000).

What we wish to stress more particularly here is that CDS’s success is certainly due to the clear-sightedness of its founders and to the extensive network of collaborations set up *ab initio*, but also to the right decisions taken at the right times by its successive managers, to the consistency of the policies followed and, last but not least, to its small but dedicated staff.

⁸And the more we observe it, the more variable it becomes.

⁹See e.g. <http://astro.caltech.edu/nvoconf> and Cheung *al.* (1999).

¹⁰See for instance <http://newb6.u-strasbg.fr/~ccma/vo> for a European project.

¹¹<http://www.stecf.org/astrovirtel/>

¹²<http://cdsweb.u-strasbg.fr/>

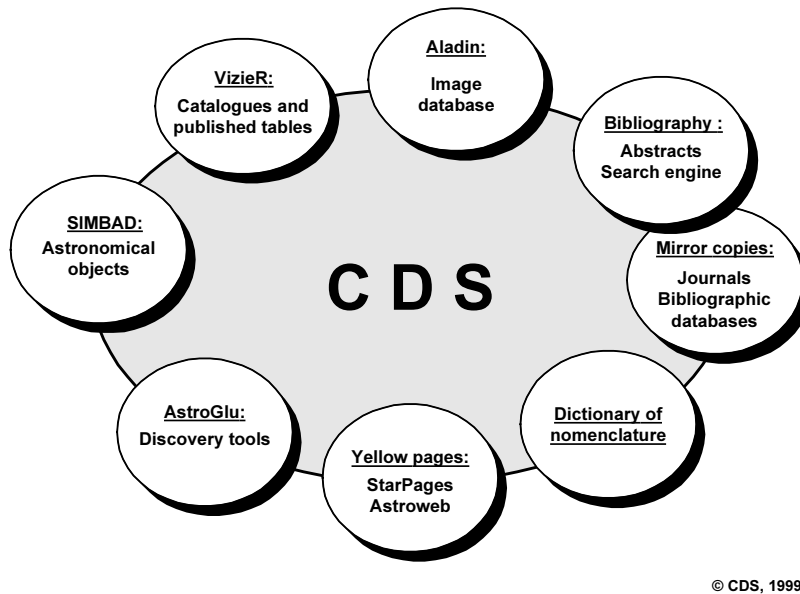


Figure 1. A schematic view of the CDS information hub (courtesy *Centre de Données de Strasbourg, CDS*).

However, if most of the goals set by the founders have been reached – and sometimes largely facilitated by not-so-quickly-expected technologies (such as the electronic networking of the planet), at least one of them did not really come through: astrophysical research directly geared to the data centre. This is worth stressing since such an ambition was explicitly written in the statutes.

Because of the current efficient connectivity, any astronomer able to hook onto the Internet is nowadays getting immediate access to all distributed astronomy resources from his/her office (or his/her home). There is absolutely no difference today whether the users are carrying out their research geographically close or far from the data repositories. It should however be kept in mind that we are not (yet) all equal round the world in terms of Internet access: large gaps are still existing (see for instance Figs. 2-4 reproduced from Heck 2000).

The research, which is also important, that is now proper to data centres is more of a technological and of a methodological nature.

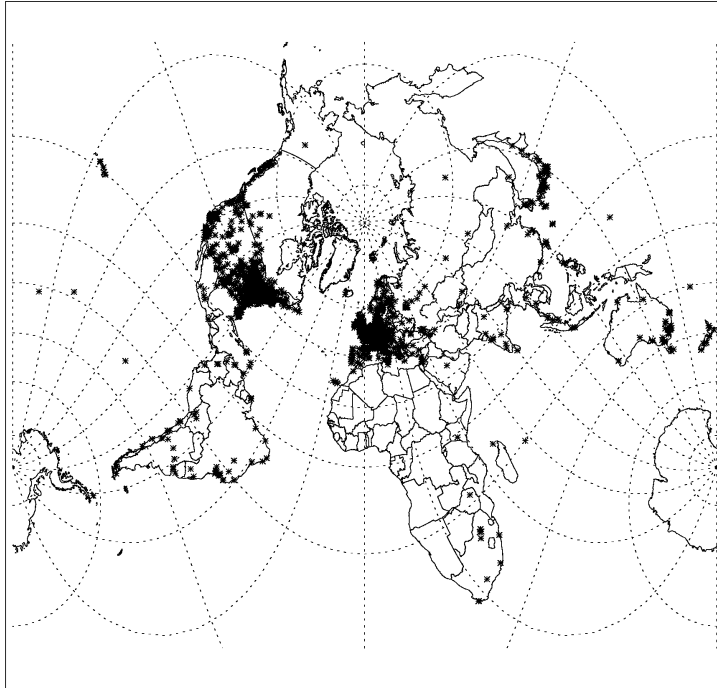


Figure 2. World geographical distribution of astronomy-related organizations (all categories) with an Internet presence (e-mail and/or web pages).

Methodological lessons learned

The world works in small circles and how many times can we observe that each circle tends to become its own finite universe tending to ignore the overall purpose of the whole machinery. Even within a scientifically narrow field such as astronomy (compare to *e.g.* physics or chemistry), it is often difficult for individuals to get enough distance to put their own work in the appropriate perspective and not to lose track that the ultimate aim, conditioning all steps of the information flow, is science and the progress of knowledge.

How many times did we have to remind software people that their packages should be designed as tools to be used by ‘plain’ scientists and not as challenges for computer freaks? Database people tend sometimes to gather data together for the sake of it, with secondary interest for the science that can be extracted from those compilations. Examples could be multiplied.

The success of the database-related (*e.g.* ALD) and software-related (*e.g.* ADASS) conferences is the best indication that these fields are lively,

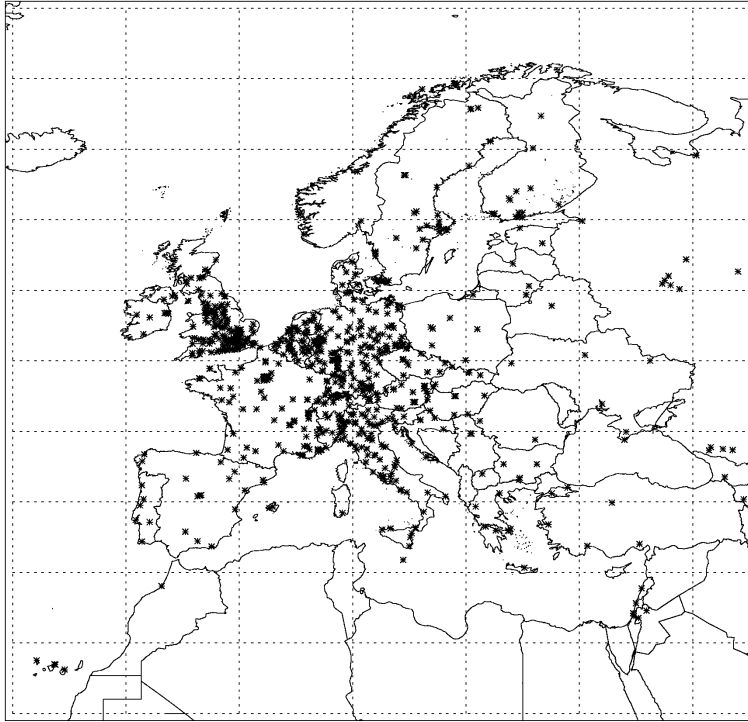


Figure 3. West European geographical distribution of astronomy-related organizations (all categories) with an Internet presence (e-mail and/or web pages).

but, as astronomers, we would certainly like to see more pointers towards significant scientific advancements at such gatherings, even if they are not their main purposes. Along the line of the comments made by Benvenuti (1988) at the ALD-I conference, it is appropriate to recognize that one sees too many technicalities at such events and not enough astrophysics: too many data collections *per se*, too many prototypes, too many demonstrations of feasibility without new results, too many school examples and illustrations, in other words too many openings of doors already open.

Take for instance the Kohonen self-organizing maps with which there is a kind of fashion these days. One of them appears on a CDS web page¹³. Personally I love to see such illustrations, but should we not ask also ourselves whether they are bringing anything to the progress of astrophysics?

Those beautiful graphical summaries are excellent to demonstrate to students or to beginners the proximities of fields or research works in mul-

¹³See <http://vizier.u-strasbg.fr/cgi-bin/VizieR> and another one as Fig. 5 of Murtagh's chapter in this book.

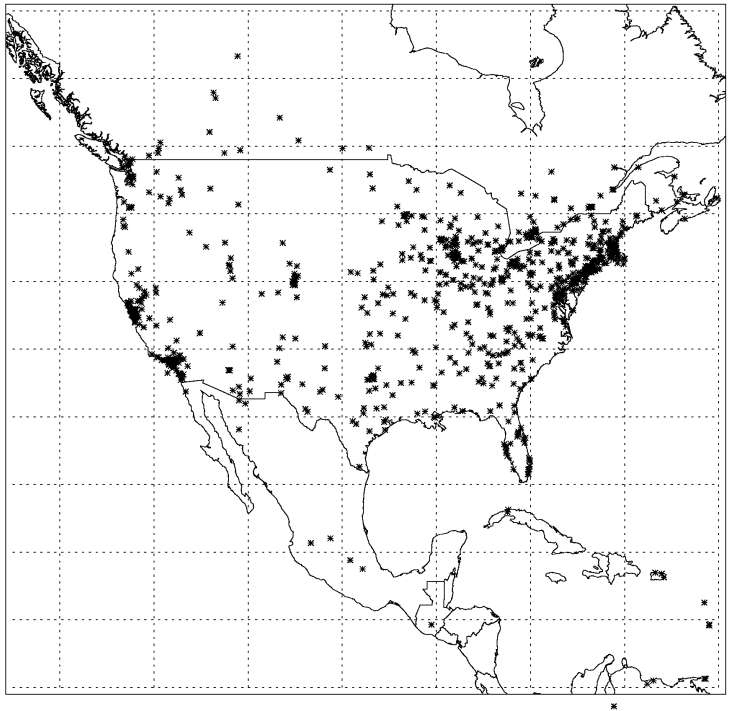


Figure 4. North American geographical distribution of astronomy-related organizations (all categories) with an Internet presence (e-mail and/or web pages).

tidimensional spaces, but should not we put to shame a scientist who would ignore who else is working in the same area or what are the astrophysical relationships of the objects he/she is studying – *i.e.* his/her working context? In other words, that person would not learn anything from such illustrations.

Perhaps after all such methodologies find their real usefulness in other disciplines where the underlying physics is less developed than in our field. And there are instances too where they can lead to weird conclusions.

The message of this is that each development of new methodologies, software packages, impressive web sites, and so on, should be pursued not as an end *per se*, but according to what they really and efficiently bring to the progress of astrophysics and knowledge.

The real slot of electronic publishing

Do not be mistaken: no astronomy journal is practicing electronic publishing in the full sense of the medium. What they do is to put on line digitized

files still basically along the lines of linear structure of a document on paper, since the first thing most users of such ‘electronic’ journals want is ... to print the ‘papers’! And what we still have largely is a system equivalent to TV bulletin news zooming on newspapers (on paper) or showing people reading magazines.

A fully electronic resource would make use of all the degrees of freedom of the medium, such as the hypertextual structure, the colors, the sound and motion, the applets and whatever might come next.

The only deviations from the linearity in a classical article are the footnotes and the bibliographical references. The present chapter has actually been first conceived as a hypertextual document, then constrained into a linear sequence – with still quite a few footnotes.

Certainly some e-peculiarities have been introduced in electronizing the journals, such as advanced integration in databases and forward referencing, not to forget the possibility of shipping quickly the papers forth and back between authors, editors and referees, as well as exactly that flexibility for quick download from the web sites.

The real difficulties in scientific electronic publishing is not to make papers available electronically with whatever e-gimmicks and degrees of freedom. The true points to be solved are the following ones:

- how do we make compatible fluidity *and* validation of scientific information?
- are *all* expert bodies and evaluation committees ready to include in their reviews fully electronic documents (*i.e.* with all the possibilities mentioned above)?

Another frequently encountered misconception relates to the exact rôle of the electronic medium compared to paper. Some were actually talking of a U-turn, dropping the latter for the former, instead of simply thinking in terms of *diversified publishing* since the electronic medium and the paper are not competing against each other, but rather complementing each other¹⁴ (such as television with newspapers and magazines).

Scientific electronic publishing can probably expect natural progress from (at least) two areas: on one hand, cellular telecommunications including advanced voice recognition and all-web capabilities (this is solved in principle already) and, on the other hand, neurobiology with direct plugging of our own neural network into the technological communications loop (perhaps not so far away). This would induce dramatic changes in the rôle of publishing that goes already much beyond the mere transfer of knowledge

¹⁴See for instance <http://vizier.u-strasbg.fr/~heck/cern.htm> and the references quoted therein.

in the dynamics of our communities (by being a key factor in *recognition* and *evaluation* procedures).

Green literature has never really been identified as an important issue in astronomy, perhaps because of the small size of the community and the rather fast publishing procedures (compare to other disciplines).

However, and with all the consideration and appreciation due to preprint servers such as the LALN one¹⁵ (see *e.g.* Ginsparg 1996), one must recognize that the system is somehow heavy and little time-efficient (files need often to be compressed, possibly to be uuencoded, plus ftp-ed and/or e-mailed before the paper is up and available).

Again for all its value at the time it was set up, such a system could certainly be simplified nowadays by taking advantage of the web structure and by pointing to papers residing at the authors' sites. The maintenance would be lighter (especially in case of paper upgrade) and the validation procedures could remain very similar to the current ones. An abstract and a bibliographical reference could always be included in the main database together with the paper URL.

As detailed in Grothkopf's chapter in this book, librarians are more and more involved in the whole evolution of the astronomy-related information environment. This increasing participation should be encouraged with adequate financial and material means, not only to libraries, but also for librarians themselves in order they get optimally trained and organized.

Quality and automation

To the best of our experience, those two fellows are not quite good friends yet and this is said at the risk of being hated by colleagues developing automatic procedures for information extraction¹⁶.

To give a couple of examples (no names!), there have been attempts of automatic spectral classification and other bibliographical studies ruined because the initial samples were polluted by that unavoidable fringe of data that cannot be sorted out but by the human brain still as of today or because of some carelessness in the way the data were labelled in the original logs. We should also remember that astronomy does not quite work statistically (for instance, a star is expected to have a specific spectral type, not a chance percentage of being of such or such class).

Our own experiences with artificial intelligence and knowledge-based systems showed that, with currently available technology and problematics, it would take much too much time to capture the corresponding human expertise and to put it into a computer system. All the discourses heard

¹⁵<http://xxx.lanl.gov/>

¹⁶Will that be an opportunity for them to prove the statement is wrong?

on interviewing people before they retire or they die are very attractive theoretically. In practice, if that expertise is complex, it might take another life to record it. This does not mean we should not try to develop better ways to capture people's expertise, but we might have to wait for significant advances in neurobiology again to progress decisively in that direction.

As a general rule, the quality of results can at best reflect the quality of the input material when dealing with data-related methodologies (in other words, garbage in can only lead to garbage out). With the tools we have today, increase of quality is still the result of time and human sweat.

Beware also what is sometimes said on automated methodologies. Spontaneous generation of knowledge does not exist as no methodology will ever reveal knowledge that is not already somehow in the data.

With the massive amount of information available nowadays thanks to the networking of the planet and in particular via the web, there are endless discussions on the virtues, properties and techniques of data mining. The problem today is not so much accessing information, but to identify good-quality information, then information relevant to the matter under consideration, and finally information exactly on target.

Additionally, as Gell-Mann (1997) recently expressed it, with the digital age producing an "immense sea of data that threatens to drown humanity", people need to adapt how they think so that true knowledge can be distilled from the deluge. "We hear, in this dawn of the so-called information age, a great deal of talk about the explosion of information and new methods for its dissemination. It is important to realize, however, that most of what is disseminated is misinformation, badly organized information or irrelevant information. How can we establish a reward system such that many competing but skillful processors of information, acting as intermediaries, will arise to interpret for us this mass of unorganized, partially false material?" What is true for the web is also valid much more generally.

According to our experience, as of today, validated information¹⁷ cannot be secured automatically. Take for instance those two yellow-page services that we contributed to set up: on the one hand the *StarPages*¹⁸ (see e.g. Heck 1997b and the references therein) and on the other hand *AstroWeb*¹⁹ (Jackson *et al.* 1994). The resources are factually compared at <http://vizier.u-strasbg.fr/~heck/awsp.htm> .

¹⁷Be it refereed papers, homogenized and/or critically analyzed elements of a database, or others.

¹⁸A set of three databases reachable at <http://vizier.u-strasbg.fr/starworlds.html> (organizations), <http://vizier.u-strasbg.fr/starheads.html> (astronomers), and <http://vizier.u-strasbg.fr/starbits.html> (acronyms).

¹⁹<http://cdsweb.u-strasbg.fr/astroweb.html>

While *AstroWeb* is almost exclusively based on automated procedures registering URLs together with brief descriptions and verifying the corresponding links do not die out, the *StarPages* rely on a painstaking daily manual maintenance and various postal procedures for updating, authenticating and validating the information published. *AstroWeb*'s quality of entries is heterogeneous, sometimes questionable and definitely non-exhaustive. Little if anything is checked beyond the aliveness of the links²⁰. Its mirror sites are sometimes down or plagued by updating problems that may remain unnoticed for quite some time.

Even on the level of the quantity, automation does not perform better since the *StarPages* include currently (amongst much more comprehensive information) more than 11,000 URLs while the URL-only-oriented *AstroWeb* does not even reach 3,000 URLs.

Need for prospective

Believe it or not, but not so long ago, the space agencies had no post-mission plans for what to do with the data collected by their spacecraft. Their commitment was to select the best ones, to supervise their manufacturing, to launch them and, in the best of the cases, to operate them.

The scientific community or rather the principal investigators (PIs) were largely responsible for what was happening to the data gathered by the various experiments. In that respect, the policy was essentially a transposition of what was happening at ground observatories: once the observing runs were completed, the astronomers were going home with 'their' data. But how many plates ended in drawers and closets, and were never looked at? How many tapes never saw a drive again?

Fortunately things changed with the cost of the missions, with the observing pressure on instruments (percentage of requested over available time) and with the general pressure from the community in parallel with the digitization of most data that became easy to duplicate.

Now no space experiment worth its name is launched without *ad hoc* provisions for archiving and disseminating its data to the community at large after some possible proprietary periods in favor of PIs. One can appreciate how complex such tasks have become through the chapters by Hanisch and by Cheung & Leisawitz in this book.

On a more general level, Thomas J. Watson declared when opening a new IBM laboratory in 1932: "There is no business in the world that can hope to move forward if it does not keep abreast of the time, look into the future and study the probable demand of the future." (Berghel 1999).

²⁰But this is not because a web site is alive that it is validated: being able to turn on the engine of a car does not mean the car is in good shape.

Predicting is definitely a difficult game, but few would disagree with the need for sound and accurate technology forecasting in any organization that seeks to remain competitive. There is thus a need for anticipating appropriate information policies in projects. In that respect, the sections on ‘lessons learned’ appearing in many chapters of this book are precious references.

As we have seen earlier, summarizing information, *i.e.* reducing it to its most essential points, is a real issue nowadays (see *e.g.* Endres-Niggemeyer 1998). As pointed out by Albrecht in his chapter, there is a definite need to innovate in terms of knowledge processing and transfer. Refer also in this respect to Böcker (1998) and to various special issues of the *Communications of the ACM*.

Education and communication

No astronomer will question the need to increase astronomy teaching at all levels. The field has been well covered by Percy and by Norton *et al.* in this book. The difficulties encountered in some countries with lobbies such as the creationism-related ones should not be underestimated, but they should also be carefully treated to avoid that actions taken backfire through some propaganda in favor of such groups.

In the same vein, astronomers should also be encouraged to become more involved in organizations fighting pseudo-sciences such as the *Committee for the Scientific Investigations of Claims of the Paranormal (CSI-COP)*²¹ and its world-wide affiliates. Amateur astronomers can also be usefully involved in such actions.

Amateur astronomers are generally classified in two categories: the active and the armchair amateur astronomers. While the latter ones have generally a passive interest in astronomy (reading magazines, attending lectures, and so on), the former ones carry out some observing, often with their own instruments, and such activities can be useful to professional astronomy. The chapter by Mattei & Waagen in this book beautifully exemplifies how a well-organized and hard-working organization can efficiently contribute to the gathering of data and thus to the expansion of cosmic knowledge.

Inversely the sharing of knowledge with more passive amateur astronomers and with the society at large has many facets. Professional astronomers should also be encouraged to share their work more often with the open world. Beyond the world-wide audience of a journal such as *Sky &*

²¹<http://www.csicop.org/>

*Telescope*²², there are many national journals which deserve more attention from our community.

Public observatories and planetariums (on the latter, see the chapter by Petersen & Petersen in this book) are other outlets where professional astronomers should be seen more often.

As already pointed out earlier when speaking of electronic astronomy, there are still large portions of some of our continents where astronomy is almost inexistent²³ and where help is dramatically needed. Refer for instance to the chapter by Andersen in this book for the various actions undertaken by the *International Astronomical Union (IAU)*²⁴ in this respect.

As amply illustrated in the chapters by Andersen, by Maran *et al.*, and by Madsen & West in this book, astronomers need to learn how to communicate properly and it is true that, in general, this is not part of their education. And narrow-minded, often personally-motivated, initiatives have sometimes ended in disastrous results for our community because outsiders generally believe that we are all talking with the same voice.

The *American Astronomical Society (AAS)*²⁵ has understood this very well, not only by its dedicated news media service (refer to the chapter by Maran *et al.* in this book), but also by opening years ago a bureau in Washington, DC. It is lobbying directly the US Congress while instructing also adequately its membership via its newsletter and its electronic announcements for concerted actions at appropriate times with adequate arguments.

Each decade in the US too, an *Astronomy and Astrophysics Survey Committee (AASC)*²⁶ surveys the field of space- and ground-based astronomy and astrophysics, recommending priorities for the most important new initiatives. The publication of the next report (addressing the decade 2000-2010) is expected in Fall 2000.

We are still waiting for such undertakings in Europe, even at the national levels as political lobbying is largely left to individual initiatives and to short-sighted personal promotions and political connections.

However fascinating it can be, the communication process needs to be carefully planned: the formulation of a message (*i.e.* an information set), its conveyance, and its reception by targets who will each perceive it differently. In a scientific context, the matter is not only to deal with 'true' information (*i.e.* authenticated, verified and validated), but also for each

²²<http://www.skypub.com/>

²³See <http://vizier.u-strasbg.fr/~heck/org.htm> or the maps reproduced in Heck (2000).

²⁴<http://www.iau.org/>

²⁵<http://www.aas.org/>

²⁶<http://www.nas.edu/bpa/projects/astrosurvey/>

scientist to get the recognition he/she deserves among his/her peers, as well as for a scientific community to position itself adequately compared to other disciplines and to society at large. And in astronomy, as already mentioned, we are not only ‘selling’ products (our research results) or ourselves, but also the fundamental understanding of mankind’s position in the universe.

Innovations and assertive attitudes, in other words creativity, towards society at large should probably be put more often into practice, for instance when it comes to countering the problematic practice of selling stars by offering instead cosmic objects for adoption while educating people adequately (see *e.g.* Heck 1997c).

Conclusions and final comments

First of all, never forget: it all comes down to a few photons reaching us from outer space. Astronomy has thus to rely on the ingenuity of instrumentalists to conceive and design a whole range of observing tools exploiting optimally the latest technologies and the most sensitive detectors to obtain the most relevant and most varied information allowing progress of astronomical knowledge.

It is a truism to say that the last decades have seen dramatic changes in the way information in general is handled. Some of us surely remember how (not so long ago) we were still using logarithmic tables, mechanical typewriters, speaking to colleagues over noisy phone lines (sometimes hard to get and not uncommonly breaking down) and how we were dependent, to work and publish, on what we call nowadays ‘snail mail’. Astronomy has been no stranger to that evolution in parallel with instrumentation developments and panchromatic integration of wavelength ranges. Radio-astronomers and space scientists are no longer separate branches of the family.

It is also true that, over the last three decades, we went from the (almost) individualistic way of dealing with astronomical data to the current institutionalized handling of astronomical information with teams becoming larger and larger and involving specialists (astronomers or others) in instrumental technology, telecommunications, computing, image processing, knowledge extraction, electronic publishing, and so on.

Now we can also (in principle) cross-check the latest p/reprints on the *astro-ph* server or check out the ADS abstracts and papers²⁷ from the middle of nowhere through a PDA phone facility – and a large part of this book has been put together via the Internet link of a ship sailing international waters down under.

²⁷ADS = NASA Astrophysics Data System (<http://adsabs.harvard.edu/>).

But have we gone through a revolution as some claimed maybe a bit too hastily or merely through natural evolution of technologies and methodologies? The few comments in this chapter have pointed out a number of serious shortcomings and language abuses compared to the actual situation. The real revolution might still be to come, with the direct connection of our brains to a huge knowledge base, somehow foreseen in Gibson's (1986) *Neuromancer*.

What is sure however, as technology leaders agreed at the ACM97 50th anniversary conference in San Jose, CA, change is the only sure thing about computing and communication technologies in the next decade. As always in the past, astronomers will certainly be among the first eager users of the latest technologies whatever they will be – and we shall let future generations answer the question above as they will have the advantage to look at us with some historical perspective.

Maybe a very final word of caution: in spite of the marvellous potentialities offered by the technological and methodological advances, one must refrain, even within a specific scientific discipline, from engulfing enormous amounts of energy and manpower in oversized endeavours with questionable return.

The history of European astronomy has this pitfall illustrated by the ambitious project of the *Carte du Ciel* that mobilized during critical decades the resources of various countries and delayed the flourishing of astrophysics in Europe while it was happily taking off on the other side of the Atlantic. Therefore to avoid the repetition of such unfortunate episodes, it would be advisable that mammoth projects be undertaken with *ad hoc* world-wide collaboration and with a dynamical management avoiding dispersion and waste of expertis and resources.

Acknowledgements

It is a real pleasure to acknowledge here the impact of numerous readings and enlightening conversations with many colleagues involved in astrophysics, communications, computing, information science, library management, communications media, publishing, sociology of science, and so on.

References

1. Benvenuti, P. 1988, Introduction, in *Astronomy from Large Databases – Scientific Objectives and Methodological Approaches*, Eds. F. Murtagh & A. Heck, *ESO Conf. & Workshop Proc.* **28**, v-vi
2. Berghel, H. 1999, The Cost of Having Analog Executives in a Digital World, *Comm. Assoc. Comp. Machinery* **42** (Nov. 1999) 11

3. Böcker, H.D. 1998, Riding The Knowledge Waves of the Centuries to Come, in *The Impact of Electronic Publishing on the Academic Community*, Ed. I. Butterworth, Portland Press, London, 176-180
4. Boroson, T., Davies, J. & Robson, I. (Eds) 1996, *New Observing Modes for the Next Century* Astron. Soc. Pacific, San Francisco, xxiv + 262 pp. (ISBN 1-886733-08-2)
5. Cheung, C.Y., Borne K.D., Shaya, E.J. & Blackwell, J.H. 1999, Using Online Databases to Estimate the Galaxy Interaction Rate: Looking Forward to a National Virtual Observatory, *Bull. Amer. Astron. Soc.* **31**, 1492
6. Di Gesù, V., Duff, M.J.B., Heck, A., Maccarone, M.C., Scarsi, L. & Zimmermann, H.U. (Eds.) 1997, *Proc. Fifth Workshop on Data Analysis in Astronomy*, World Scientific, Singapore, viii + 422 pp. (ISBN 981-02-3171-7)
7. Egret, D. & Albrecht, M.A. (Eds.) 1995, *Information & On-line Data in Astronomy*, *Astrophys. Sp. Sc. Lib.* **203**, Kluwer Acad. Publ., Dordrecht, vi + 292 pp. (ISBN 0-7923-3659-3)
8. Endres-Niggemeyer, B. (Ed.) 1998, *Summarizing Information*, Springer-Verlag, Heidelberg, xii + 376 pp. (ISBN 3-540-63735-4)
9. Gell-Mann, M. 1997, *Lecture to ACM97*, San Jose, CA
10. Gibson, W. 1986, *Neuromancer*, Grafton, London, 318 pp. (ISBN 0-586-06645-4)
11. Ginsparg, P. 1996, *Winners and Losers in the Global Research Village*, in *Electronic Publishing in Science*, eds. D. Shaw & H. Moore, ICSU Press & UNESCO, Paris, 83-88
12. Grothkopf, U., Andernach, H., Stevens-Rayburn, S. & Gomez, M. (Eds.) 1998, *Library and Information Services in Astronomy III*, *Astron. Soc. Pacific Conf. Series* **153**, xxxii + 320 pp. (ISBN 1-886733-73-2)
13. Hauck, B. & Sedmak, G. (Eds.) 1985, *Data Handling in Astronomy and Astrophysics*, *Mem. Soc. Astron. Ital.* **56**, 249-570
14. Heck, A. (Ed.) 1997a, *Electronic Publishing for Physics and Astronomy*, *Astrophys. Sp. Sc. Lib.* **224**, Kluwer Acad. Publ., Dordrecht, xiii + 250 pp. (ISBN 0-7923-4820-6)
15. Heck, A. 1997b, *Electronic Yellow-Page Services: The Star*s Family as an Example of Diversified Publishing*, *Astrophys. Sp. Sc.* **247**, 211-220
16. Heck, A. 1977c, *An Alternative to Selling Stars*, *Sky & Telescope* **March 1997**, 8
17. Heck, A. 2000, *Characteristics of Astronomy-Related Organizations*, *Astrophys. Sp. Sc.*, in press
18. Heck, A. (Ed.) 2001, *Organizations and Strategies in Astronomy and Related Activities*, Kluwer Acad. Publ., Dordrecht, in press
19. Heck, A. & Murtagh, F. (Eds) 1989, *Knowledge-Based Systems in Astronomy*, *Lect. Notes Phys.* **329**, Springer-Verlag, Heidelberg, iv + 280 pp. (ISBN 3-540-51044-3)
20. Heck, A. & Murtagh, F. (Eds.) 1992, *Astronomy from Large Databases II*, *ESO Conf. & Workshop Proc.* **43**, x + 534 pp. (ISBN 3-923524-47-1)
21. Heck, A. & Murtagh, F. (Eds.) 1993, *Intelligent Information Retrieval: The Case of Astronomy and Related Space Sciences*, *Astrophys. Sp. Sc. Lib.* **182**, Kluwer Acad. Publ., Dordrecht, ii + 214 pp. (ISBN 0-7923-2295-9)
22. Heck, A. & Murtagh, F. (Eds.) 1996, *Strategies and Techniques of Information for Astronomy*, *Vistas Astron.* **40**, 361-452
23. Jackson, R., Wells, D., Adorf, H.M., Egret, D., Heck, A., Koekemoer, A. & Murtagh, F. 1994, *AstroWeb – A database of links to astronomy resources (announcement of a database)*, *Astron. Astrophys. Suppl.* **108**, 235-236
24. Jaschek, C. 1989, *Data in Astronomy*, Cambridge Univ. Press, Cambridge, xii + 198 pp. (ISBN 0-521-34094-2)
25. Jaschek, C. & Heintz, W. (Eds.) 1982, *Automated Data Retrieval in Astronomy*, *Astrophys. Sp. Sc. Lib.* **97**, Kluwer Acad. Publ., Dordrecht, xx + 324 pp. (ISBN 90-277-1435-5)
26. Jaschek, C. & Wilkins, G.A. (Eds.) 1977, *Compilation, Critical Evaluation and Distribution of Stellar Data*, *Astrophys. Sp. Sc. Lib.* **64**, Kluwer Acad. Publ., Dordrecht,

- xiv + 316 pp. (ISBN 90-277-0792-8)
27. Kidger, M.R., Pérez-Fournon, I. & Sánchez, F. (Eds.) 1999, Internet Resources for Professional Astronomy, Cambridge Univ. Press, Cambridge, viii + 328 pp. (ISBN 0-521-66308-3)
 28. Mehringer, D.M., Plante, R.L. & Roberts, D.A. (Eds.) 1999, Astronomical Data Analysis Software and Systems VIII, *Astron. Soc. Pacific Conf. Series* **172**, xl + 520 pp. (ISBN 1-886733-94-5)
 29. Molina, R., Murtagh, F. & Heck, A. (Eds.) 1997, From Information Fusion to Data Mining, *Vistas Astron.* **41**, 327-461
 30. Murtagh, F. & Heck, A. 1987, Multivariate Data Analysis, *Astrophys. Sp. Sc. Lib.* **131**, Kluwer Acad. Publ., Dordrecht, xvi + 210 pp. (ISBN 90-277-2425-3)
 31. Quinn, P. 1996, The ESO Data Management Division, *ESO Messenger* **84**, 30-33
 32. Starck, J.L., Murtagh, F. & Bijaoui, A. 1998, Image Processing and Data Analysis – The Multiscale Approach, Cambridge Univ. Press, x + 288 pp.