

# A technologist view of the next decade evolution of the librarian work environment

André Schaaff<sup>1,\*</sup> and Marc Wenger<sup>1</sup>

<sup>1</sup> *Université de Strasbourg, CNRS, Observatoire astronomique de Strasbourg, UMR 7550, F-67000 Strasbourg, France*

**Abstract.** The work environment has deeply evolved in recent decades with the generalisation of IT in terms of hardware, online resources and software. Librarians do not escape this movement and their working environment is becoming essentially digital (databases, online publications, Wikis, specialised software, etc.). With the Big Data era, new tools will be available, implementing artificial intelligence, text mining, machine learning, etc. Most of these technologies already exist but they will become widespread and strongly impact our ways of working. The development of social networks that are "business" oriented will also have an increasing influence. In this context, it is interesting to reflect on how the work environment of librarians will evolve. Maintaining interest in the daily work is fundamental and over-automation is not desirable. It is imperative to keep the human-driven factor. We draw on state of the art new technologies which impact their work, and initiate a discussion about how to integrate them while preserving their expertise.

## 1 Introduction

Anticipating what the future will be when it comes to the professional development of people is not easy. One must convince people who are in disbelief and will tell you that this is not possible. It is therefore necessary to argue, and above all not to forget past events, putting them forward to show that what you expect is not always a utopia. To begin with, we draw on an inventory of the main evolutions, in terms of volume and capacity, for data storage, calculation power, and communication and exchanges that have been boosted by the advent of the Internet and the Web. These technological evolutions have gradually built up an extremely elaborate working environment, which has been shaken by the new possibilities offered by increasingly intelligent tools based on machine/deep learning. There is a strong question about the gradual replacement of many trades with intelligent tools that can adapt and produce results that are identical to those of humans, but without them.

---

\*e-mail: [andre.schaaff@astro.unistra.fr](mailto:andre.schaaff@astro.unistra.fr) ORCID: 0000-0002-8790-7048

## **2 Foreword**

### **2.1 Why a decade ?**

A decade is a reasonable duration. It is possible to go further on a purely technical level, but for aspects involving the human factor (acceptance of sometime radical evolutions) it is more complex and going beyond becomes random.

### **2.2 A technologist view ?**

A technologist is often seen as having the following considerations:

- A geek loving science fiction
- A supporter of extreme automation
- The human aspect is not the most important
- Technology always first

We assume to have an asymptomatic approach against these considerations.

## **3 Librarians and documentalists in astronomy**

According to our daily experience of working with librarians and documentalists, these jobs appear to us as having the following main skills:

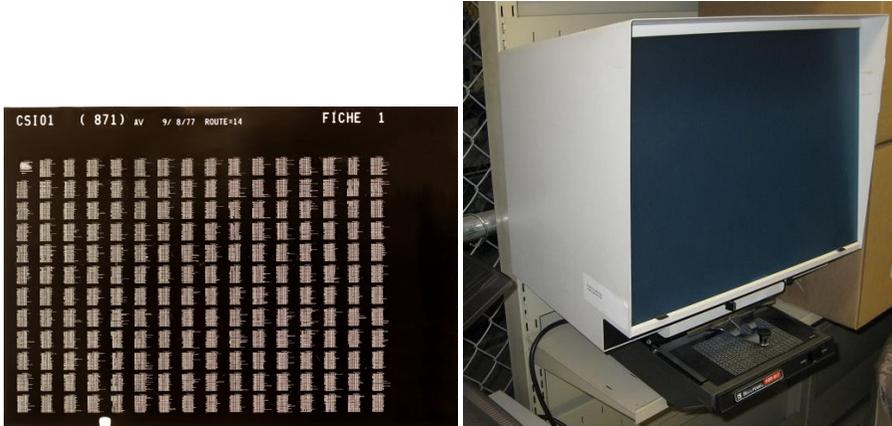
- An intellectual job with a high level of general knowledge
- A sharp knowledge of the scientific field
- An ease with computer tools including scripting
- A strong interaction with scientists and computer engineers [1]
- A strong adaptability to change

## **4 Past decades**

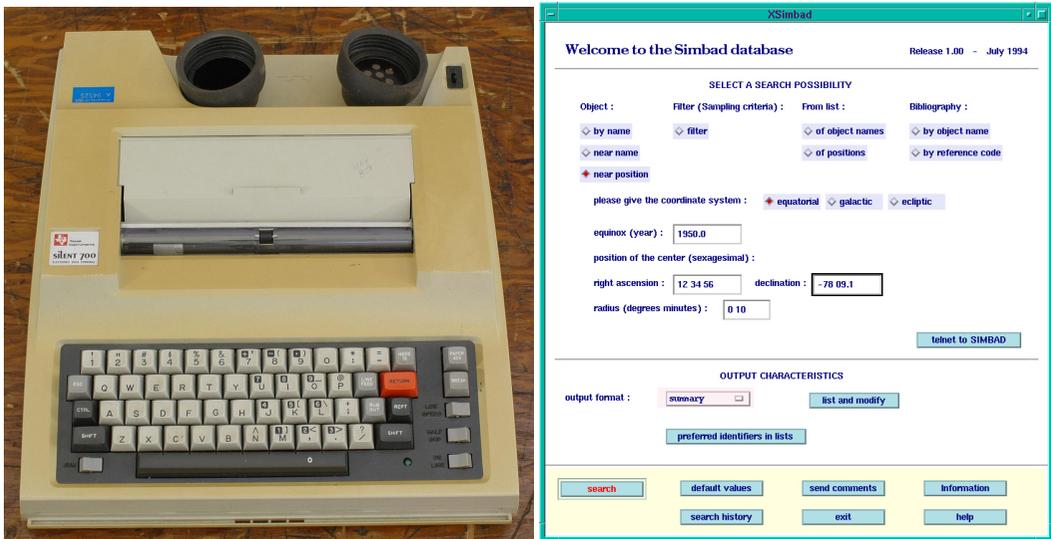
It is essential to review the main developments of the last decades before giving any indications of evolution. Computer science is a relatively young and constantly evolving field. This evolution has for a long time been almost purely technological and closely linked to that of physics and electronics. This has made it possible to miniaturize the components and consequently to increase the performances both in terms of calculation power and storage.

### **4.1 Storage revolution**

In 2017 it seems quite normal to have, in the pocket, a smartphone with 128GB of memory, to store on it many photographs, to listen to hundreds of hours of music and many other things. It is also possible to store 1 TB of data on a USB key that weighs a few grams. At the time of the perforated cards (example: IBM punched card [2] capacity was 80 columns \* 12 rows), the same volume of storage required an entire hangar. Figure 1 gives an example of an astronomical catalogue available on microfiche in 1977. Nowadays the online VizieR service distributes more than 16,000 catalogues.



**Figure 1.** CSI catalogue on microfiche and a screen to read it



**Figure 2.** Accessing Simbad in 1981 and through the Web in 1994

## 4.2 Calculation revolution

The calculation has followed the same curve as the storage. UNIVAC 1 (1951) weighed 7.3 tons and had a calculation power of 0.001905 MIPS (Million Instructions Per Second). A mid-range Smartphone has a computing power of about 100,000 MIPS, and weighs 150 grams.



**Figure 3.** VizieR nowadays, a simple field to query 16,000 catalogues !

### 4.3 Communication and exchange revolution: Internet as a catalyst of the chain reaction... and the Web

The advent of the Internet and especially of the Web was the starting point of a new paradigm concerning the distribution of data. In August 1988, SIMBAD demos were the first to use the Internet in France. Figure 2 shows how Simbad was accessed before and after this date. In 2017, the VizieR Web page (Figure 3) offers to the user the possibility of downloading a part of a catalogue, to cross-match columns from several catalogues, etc. A work around interoperability is ongoing to make possible access to data coming from several surveys [3] through simple Web pages.

### 4.4 The content

We had mainly technical evolutions but technical developments are not sufficient on their own. For example 3D (Movies and TV) does not have enough content and is not seen as really better by the customers. Another example, Immersive 3D has a lot of possibilities, but it is unknown if the content will really follow. The last decades are a cemetery of others, for example object databases, etc.). The content is wanted and its added value.

## 5 The wave

Data is the new Eldorado [4] and this vision is strongly supported by politicians. The existing tools are often not adapted to manage such volume of data. It becomes mandatory to do more, to do it better and to do it quickly.

## 6 Flagships of the next decade

The next decade will be marked by an unprecedented deluge of data and by the return of artificial intelligence.

The keywords of the decade:

- Big Data, Open Data, Linked Data: everywhere
- A.I., Neuronal networks, Deep Learning, Machine Learning, robots, etc.: Chess (1997), Go (2016), Pocker (2017) battles
- "Professional" social networks

## 7 What's going on now ?

### 7.1 What we have

- Large storage and computing facilities
- Fast networks
- A large range of devices to interact, visualize
- A lot of data

We will have more and more intelligent tools able to do an increasing part of our work.

Therefore, is it too late? Will Catherine, Emmanuelle, Evelyne and Mihaela be replaced by Watson (IBM) [5] ?

=> No! [6]

### 7.2 Intelligent tools

Intelligent tools have success rates. Is 96 percent (even 99 percent) acceptable? Human added value lies in the part that aims to reach 100 percent and their work will therefore be increasingly shifted towards expertise. The volume of data and publications is constantly increasing and a few percent can represent a lot of work for humans, and probably not the less interesting.

## 8 So, what

In this part we give a few ideas which should be pillars of their work environment in a near future.

### 8.1 Helping to extract astronomical object names from papers

DJIN2, the new version of DJIN (Figure 4) is becoming an (internal) online tool implementing new technologies like GROBID [7] and is able to capture documentalist's decisions. GROBID uses Machine Learning to improve its recognition of the object names. DJIN2 is an example of a tool which helps documentalists every day and which learns every day to provide a better service.

### 8.2 Virtual Research Environment

A VRE is not only a nice user interface to access tools and services, it must be seen as an environment providing a well designed and customisable user interface with a large range of tools, data services (both to store and access), computing facilities, and community networking. Many experiments are carried out all over the World; VRE4EIC, EVER-EST (H2020) in Europe, Science gateways in the USA, and Virtual Laboratories in Australia, etc.). A VRE is not only a technical solution for a better working environment but also a way to collaborate, share, preserve, etc. The Research Data Alliance is working in this framework through an Interest Group which aim to bring people from several communities together to exchange knowledge (technical problems to solve, interoperability between VREs, common efforts to enable, etc.). A VRE for librarians / documentalists should be experimented with.

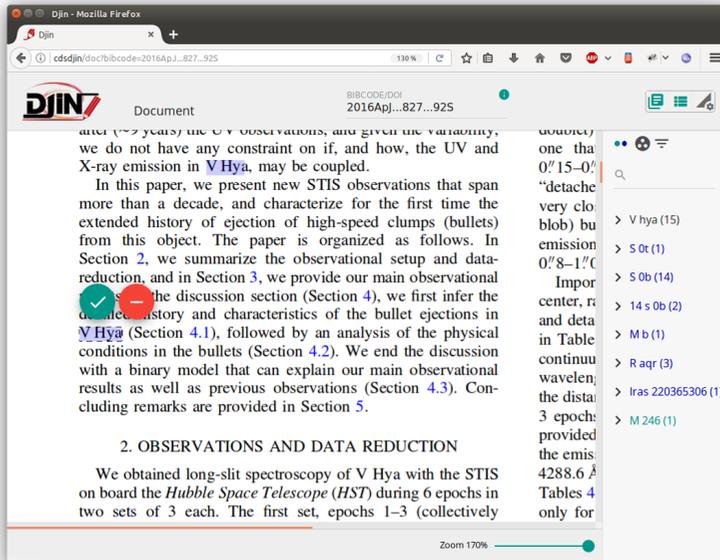


Figure 4. DJIN2, the next generation of DJIN

### 8.3 Professional social network

Facebook is developing Workplace [8], a social network for professionals. A social network for astronomy librarians and documentalists should develop for the exchange of information and good practices. Astronomy is a relatively small community and this will help to facilitate. It is a way for "working together" at a larger scale.

### 8.4 Collaborative tools

A good way to develop expertise is to exchange information with other people in the same domain. A number of collaborative tools are available like Wikis, and source code repositories (GitHub, Gitlab, etc.). Sharing and preserving both the data and the process are key actions.

## 9 Conclusion

Astronomy librarians and documentalists will themselves as flagships of the coming decade. They will work with tools capturing a part of their expertise, which is also a way to preserve their expertise. It will give them more time to focus on the complex parts and the opportunity to strengthen their expertise. A key point is certainly to develop collaborations with other people of the same domain through professional social networks. Astronomy is a relatively small area and this should not be a problem. Their added value will remain precious. This will remain true on the condition that they accompany the development of the new tools.

They have to coach "Watson" to help them!

## References

- [1] E. Perret, T. Boch, F. Bonnarel, C. Bot, M. Buga, M. Brouty, C. Bruneau, C. Brunet, L. Cambrésy, S. Derrière et al., *Working Together at CDS: The Symbiosis Between Astronomers, Documentalists, and IT Specialists*, in *Open Science at the Frontiers of Librarianship*, edited by A. Holl, S. Lesteven, D. Dietrich, A. Gasperini (2015), Vol. 492 of *Astronomical Society of the Pacific Conference Series*, p. 13
- [2] IBM, *The IBM Punched card*, <http://www-03.ibm.com/ibm/history/ibm100/us/en/icons/punchcard/> (2017), [Online; accessed 31-July-2017]
- [3] A. Schaaff, *Data in Astronomy: From the Pipeline to the Virtual Observatory* (Springer Berlin Heidelberg, Berlin, Heidelberg, 2007), pp. 52–62, ISBN 978-3-540-77010-7, [https://doi.org/10.1007/978-3-540-77010-7\\_6](https://doi.org/10.1007/978-3-540-77010-7_6)
- [4] E. U., *How Europe can gain from the rising tide of scientific data* (European Commission, 2010), [http://ec.europa.eu/information\\_society/newsroom/cf/document.cfm?action=display&doc\\_id=707](http://ec.europa.eu/information_society/newsroom/cf/document.cfm?action=display&doc_id=707)
- [5] IBM, *IBM Watson*, <https://www.ibm.com/watson/> (2017), [Online; accessed 31-July-2017]
- [6] H. Rachlin, *The Behavior Analyst* **35**, 1 (2012)
- [7] P. Lopez, *GROBID: Combining Automatic Bibliographic Data Recognition and Term Extraction for Scholarship Publications* (Springer Berlin Heidelberg, Berlin, Heidelberg, 2009), pp. 473–474, ISBN 978-3-642-04346-8, [https://doi.org/10.1007/978-3-642-04346-8\\_62](https://doi.org/10.1007/978-3-642-04346-8_62)
- [8] Facebook, *Workplace*, <https://www.facebook.com/workplace> (2017), [Online; accessed 31-July-2017]

