

Digital Mathematics Libraries: The Good, the Bad, the Ugly

Thierry Bouche

Université de Grenoble I & CNRS,
Institut Fourier (UMR 5582) & Cellule Mathdoc (UMS 5638),
BP 74, 38402 St-Martin-d'Hères Cedex, France
thierry.bouche@ujf-grenoble.fr
<http://www-fourier.ujf-grenoble.fr/~bouche/>

Abstract. The mathematicians' Digital mathematics library (DML), which is not to be confused with libraries of mathematical objects represented in some digital format, is the generous idea that all mathematics ever published should end up in digital form so that it would be more easily referenced, accessible, usable. This concept was formulated at the very beginning of this century, and yielded a lot of international activity that culminated around years 2002–2005. While it is estimated that a substantial part of the existing math literature is already available in some digital format, nothing looking like *one* digital mathematics library has emerged, but a multiplicity of competing electronic offers, with unique standards, features, business models, access policies, etc.—even though the contents themselves overlap somewhat, while leaving wide areas untouched. The millenium's appealing idea has become a new Tower of Babel.

It is not obvious how much of the traditional library functions we should give up while going digital. The point of view shared by many mathematicians is that we should be able to find a reasonable archiving policy fitting all stakeholders, allowing to translate the essential features of the past library system—which is the central infrastructure of all math departments worldwide—in the digital paradigm, while enhancing overall performances thanks to dedicated information technology.

The vision of this library is rather straightforward: a third party to the academic publishing system, preserving, indexing, and keeping current its digital collections through a distributed network of partners curating the physical holdings, and a centralised access facility making use of innovative mining and interlinking techniques for easy navigation and discovery.

However, the fragmentation level is so high that the hope of a unique portal providing seamless access to everything relevant to mathematical research seems now completely out of reach. Nevertheless, we have lessons to learn from each one of the already numerous projects running. One of them is that there are too many items to deal with, and too many different initial choices over metadata sets and formats: it won't be possible to find a nontrivial greatest common divisor coping with everything already available, and manual upgrading is highly improbable.

This is where future management techniques for loosely formalised mathematical knowledge could provide a new impetus by at last enabling a minimum set of features across projects borders through automated procedures. We can imagine e.g. math-aware OCR on scanned pages, concurrently with interpreters of electronic sources of born digital texts, both producing searchable full texts in a compatible semistructured format. The challenge is ultimately to take advantage of the high formalisation of mathematical texts rather than merely ignoring it!

With these considerations in mind, the talk will focus on achievements, limitations, and failures of existing digital mathematics libraries, taking the NUMDAM¹ and CEDRAM² programs as principal examples, hence the speaker himself as principal target. . .

¹ <http://www.numdam.org>

² <http://www.cedram.org>