

Preface

This compilation of documents from the Seminar organised by ICSTI (The International Council for Scientific and Technical Information), in association with ICSU (The International Council for Science), CODATA (the ICSU Committee on Data for Science and Technology), and with the kind support of UNESCO, is the latest outcome from a programme of work which started in 1996.

ICSTI's interest in digital preservation dates back to 1996 when as a result of the ICSU Press/UNESCO Electronic Publishing Conference in Paris, ICSTI took on digital preservation as a major focus for its activities. Digital preservation was the theme of ICSTI's technical program in 1997. In 1998, it funded a major study, conducted by Gail Hodge, on the State of the Art and Practice of digital archiving. The focus was not just on archiving electronic journals, but on data, technical reports, and images. Over 30 organizations were surveyed or interviewed, and several organizations (including the scientific data community) were highlighted as presenting state of the art contributions to digital archiving and preservation. In 2000, ICSTI and ICSU Press co-sponsored a conference in Paris that brought together participants in the major projects underway at that time.

Through the years, ICSTI has sought to break down the barriers between projects, stakeholder groups and national interests. While every implementation must address local constraints and opportunities, there is a need to think globally and to consider the broader context for science and technology in today's world.

In the few years since the 1996 Conference the publication of Scientific and Technical Information has changed almost beyond recognition. The process of change was one of the driving forces behind the Conference but not many of those who were present then could have foreseen the rate of change.

The transfer to digital production is almost complete, it is inconceivable that any scientist would not present the results of his or her work in anything but digital form, even if it is in parallel to a physical copy delivered to a publisher. Even the concept of publishing has changed, scientific results are produced for comment on web sites, circulated via email and lodged on "pre-print" servers, before being formally submitted for "publication".

The changes have given rise to much debate in the scientific community – but this debate is not the subject of this compilation, another consequence of the changes is.

Preservation or archiving or, to use a more modern term, "permanent availability" of the record of science as represented by scientific and technical information is one of the processes which is dramatically affected by the change to an all digital world. While once the printed page was the basic medium and was archived by libraries today the record exists in a number of locations and in many cases is not printed. Preservation is now a matter of identification of the item in the first place, the location is not always so obvious; then ascertaining the preservation policy of the "keeper" – who may not be the owner; ensuring that there will be preservation or archiving and last but not least keeping track of the technology so that the item can be read in the future, despite obsolescence or other technology changes.

Based on the decision about the stages of the research process that warrant preservation, various types and formats of research output must be managed for future access. A variety of formats must be accommodated from bit streams to image formats, to pdf and word processing formats. As research output

becomes more complex, archiving and preservation strategies must accommodate ever more complex formats such as multi-media, where the presentation device and content are tightly coupled.

There exists also a debate on what should be archived. Consideration needs to be given to presumed value to future users, practicality and present costs amongst other issues. In the discussions at the Seminar of which this compilation is the outcome there was a complete spectrum of opinions from “all must be kept” through a process of identifying research which has been updated or changed and archiving the latest results to keeping material of “value” although who should decide the value was not clear. In addition the requirements of historians, archivists and other specialisations need to be considered. Finally there are a series of issues surrounding the identification of trends, developments or effects, which may necessitate long term preservation and availability of data.

Since a variety of research outputs may have to be archived at various stages in the research and communication process, it is unlikely that one stakeholder group will be able to take responsibility for the entire process. From authors to learned societies, to commercial publishers, to national and academic libraries and trusted third parties, multiple stakeholder groups will be involved. Unfortunately, these roles are still to be defined or redefined for the digital environment. Organizations that never considered themselves to be in the preservation business are now being faced with these issues.

Ultimately, an infrastructure for long-term preservation and future access to scientific and technical information that incorporates all the important information regardless of the point in the research process at which it was created, the type and format of the research output, or the place of stewardship, is needed. Achieving this level of access requires integration across research output types (journals, conference proceedings, data, books, etc.), across formats and across stakeholder groups.

Shared standards and best practices and common understandings in certain basic infrastructure areas are needed. In the debate on what should be preserved formal criteria for categorizing and rating products will need to be developed, if one accepts the proposition that not everything can nor should be preserved permanently for future access. Distinctions can be made, for example, according to type of content and its likelihood of change. Databases are updated, journal articles typically are not. Web site content may change daily and few if any would argue that permanent access should be guaranteed to all previous iterations. But it is entirely reasonable that permanent access to the site itself should be guaranteed. The criteria themselves may be institution specific, or adopted by consensus amongst many collaborators.

The papers in the following pages cover the status of the different interests, the situation of the different projects underway or completed and the viewpoints of various technologies which will be important in the preservation process.

As such the compilation can only be seen as a snapshot of the situation at a point in time, early 2002. ICSTI will continue to pay attention to the topic, as behoves a body with the interests of the scientific information process at its centre.

The Editors would like to thank all the authors who have been kind enough to provide copies of their presentations for this publication. We would also like to thank IOS Press for their interest and support, especially Fred Drissen for his patience in dealing with the inevitable delays and Einar Frederikson for his support.

Barry Mahon and Elliot Siegel
Editors, July 2002