

ods for preserving and providing sustained access to electronic records were limited to the simplest forms of digital objects. Even in those areas, proven methods were incapable of being scaled for the expected growth of electronic records. Furthermore, archival science had not responded to the challenge of electronic records sufficiently to provide a sound intellectual foundation for articulating archival policies, strategies, and standards for electronic records [10]. Here, a design that addresses all technical issues reported in the preservation literature is described.¹

WHAT WOULD A PRESERVATION SOLUTION PROVIDE?

What might someone a century from now want of information stored today? Figure 1 suggests users' perspectives and helps illuminate preservation reliability questions. In addition to what content management offerings² and published metadata schema³ already provide, a complete solution would:

- Ensure that a copy of every preserved record survives as long as desired;
- Ensure that authorized consumers can find and use any preserved record as its producers intended, doing so without impact from errors introduced by third parties;
- Ensure that any consumer can decide whether information received is sufficiently trustworthy for his application; and
- Hide technical complexity from end users (both information producers and consumers).

Viable solutions will allow repositories and their clients to use deployed content management software without disruption.

CHALLENGES EXPOSED BY PRIOR WORK

Information in physical books, on other paper media, and in other analog forms cannot be copied without error and always contains accidental information that digital representations can avoid. Per-

fect digital copying is possible, and contributes both to the challenge of preserving digital content and to its solution. Preservation can be viewed as a special case of information interchange—special because information consumers can no longer obtain information producers' responses about missing information or puzzling aspects.

Pervasive Focus on Repositories. Much preservation literature focuses on so-called "trusted digital repositories." Recent articles [9] amplify prior calls for criteria to be used in audits that might lead to public certification that an institution has correctly executed sound preservation practices. However, to execute partly human procedures faithfully over decades would be difficult and expensive. Repository-centric proposals betray problems that call the direction into

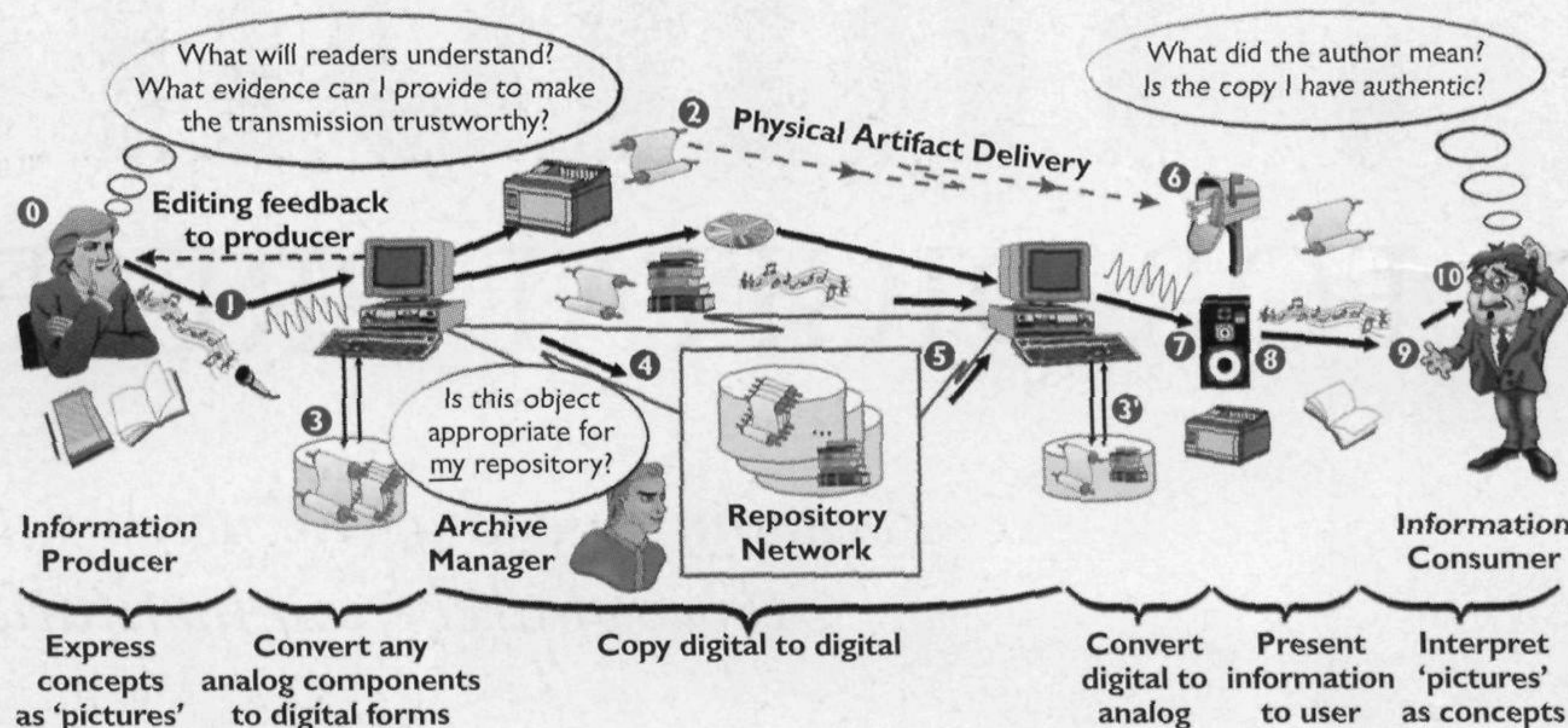


Figure 1. Documentary information interchange and repositories (the object numbering is taken from [5]).

question. Fundamentally, they depend on an unexpressed premise—that exposing an archive's procedures can persuade its clients that its content deliveries will be authentic. Such procedures have not yet been described, much less justified as achieving what their proponents apparently assume. In addition, audits of a digital archive—no matter how frequent—cannot prove that its contents have not been improperly altered by employees or hackers many years before a sensitive record is accessed. Another problem is that the new code needed for digital preservation is likely to be mostly workstation software, not server software, so the people focusing on repositories will find it difficult to design solutions.

The topical literature is replete with epistemological weaknesses. For instance, many of its references to trust are unmodified (unconstrained). Young children trust unconditionally; anyone else who does so is commonly considered childish. The mature formulation has the pattern, "X trusts Y to accomplish some

¹ Designs cited here have been published in *ACM Transactions on Information Systems*.

² Content management is not discussed in this article because archival needs can be satisfied by available software with at most modest and obvious extensions.

³ These include general schema proposed for standardization, such as METS sponsored by the Library of Congress, and many topic- or discipline-specific extensions. An October 2005 Web search for material with "metadata schema" in their titles yielded over 300 hits.