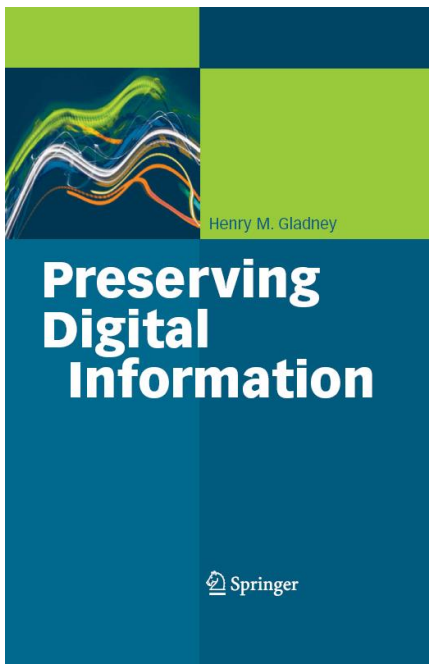
Preserving Digital Information



Springer Verlag, 2007 ISBN 3-540-37886-3

http://home.pacbell.net/hgladney/PDIf.pdf

http://www.springer.com/3-540-37886-3

See also Comm. ACM 49(2), 111-116, 2006.

To Preserve a Digital Work We Must Ensure that:

- Some copy survives as long as wanted
- Any authorized consumer can find a copy
- And **can use it as its producer intended**
- And **has sufficient authenticity evidence**
- No user faces undue technical complexity
- Everything possible is automated
- The embedding infrastructure scales

A Constraint

- ~100 digital content management software offerings exist and many are deployed
- A preservation solution must avoid obligatory changes to:
  - Internet standards and implementations
  - Digital object management infrastructure
  - Widely used document editing tools
- Except for very modest enhancements

Overview of the Book

Part I: Why We Need Long-term Digital Preservation
- – economics, environmental circumstances, state of the art

Part II: Information Object Structure
- – epistemology (theory of knowledge and information sharing)*

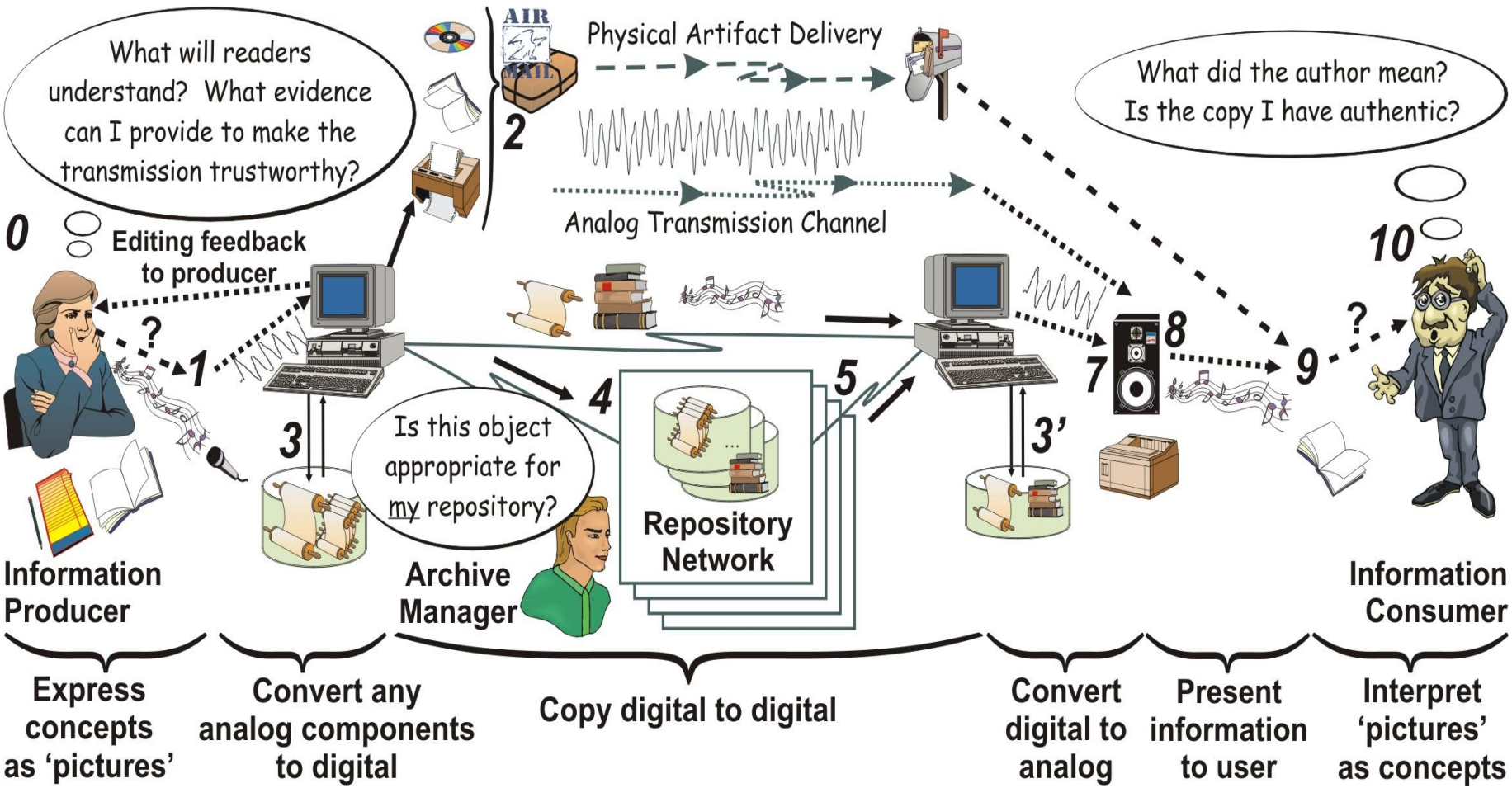
Part III: Distributed Content Management
- – standards, databases, identifiers, digital libraries, …

Part IV: Digital Object Architecture for the Long Term
- – bit-string saving, durable evidence, durable encoding

Part V: Assessment and the Future
- – solution properties, cost estimates, software needed

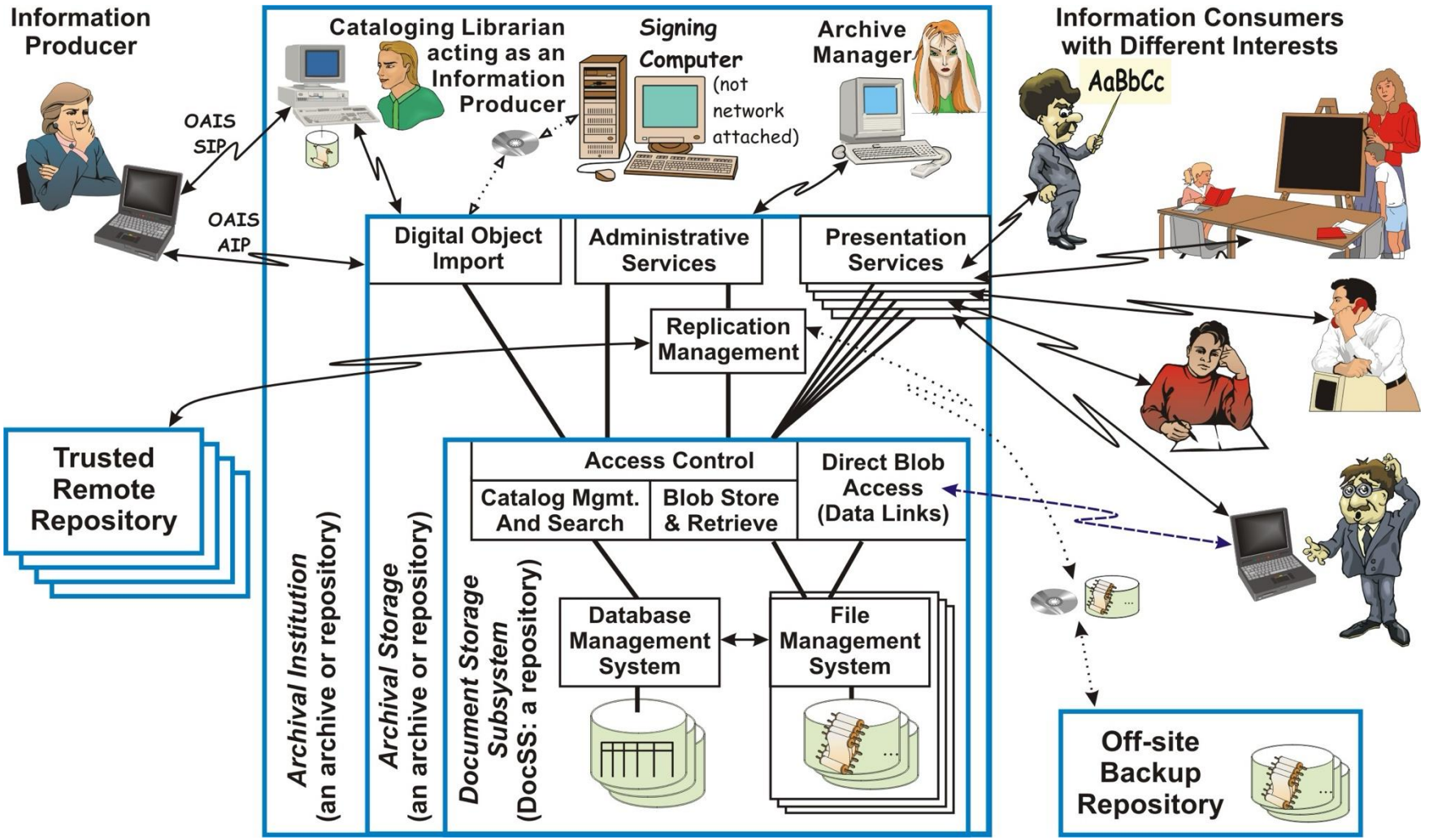
* Needed because of prominent confusions in the preservation literature, e.g., in the *Variable Media Initiative* <http://www.variablemedia.net/>

Information Flow Model



Model helps us to focus on process and to avoid epistemological confusions (e.g., Thibodeau 2002, as discussed on book's p.100, and Duranti 2004, as discussed on p.104)

Digital Repository Model

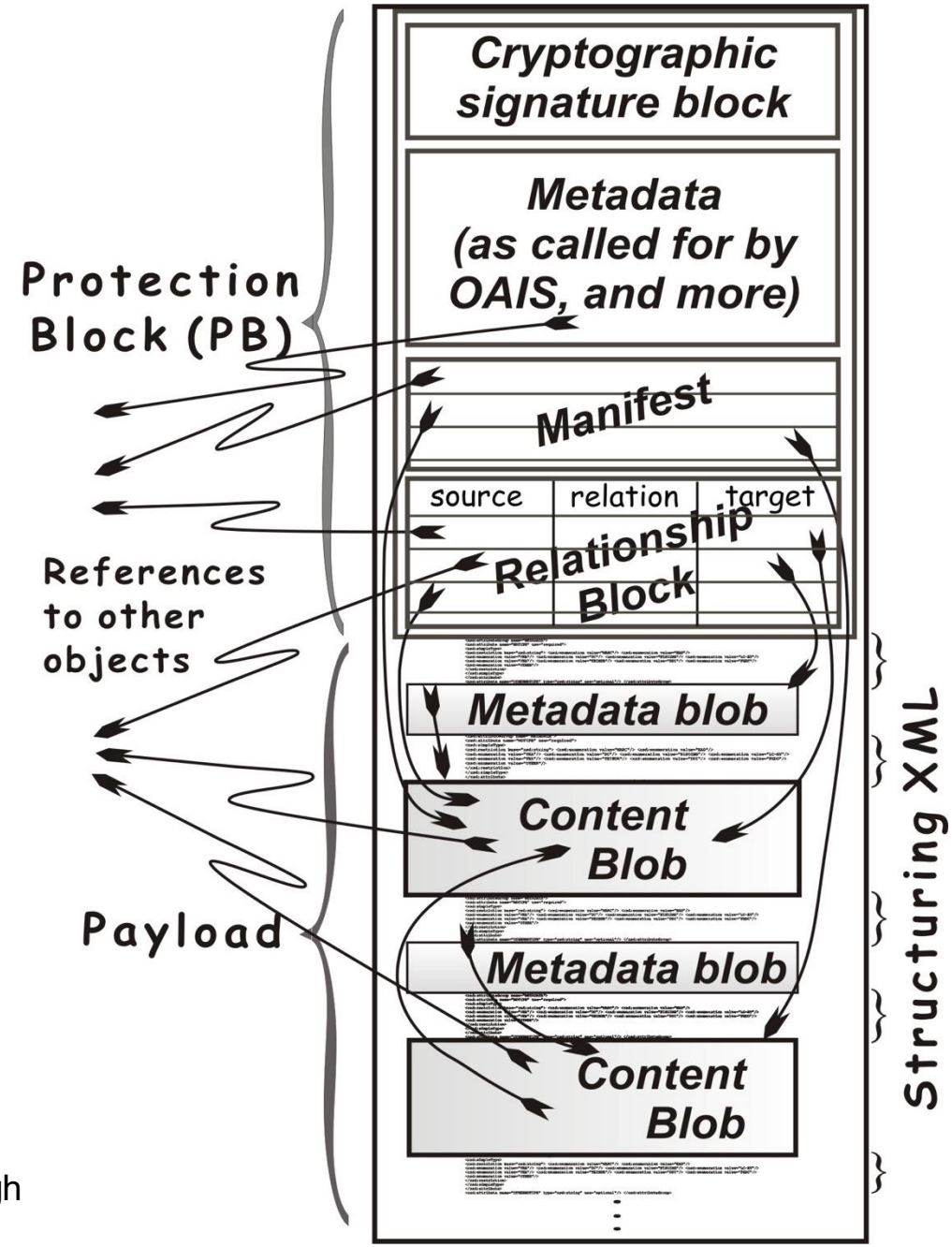


Every document requires context

Trustworthy Digital Object Model

Contents tightly bound and sealed

External pointers bound to signatures



Status and Next Steps

- We claim an “in principle” solution that is complete, economical, and optimal (i.e., the constraints identified are met)
- We invite specific criticisms of this claim
- We know what software is missing, and that it is almost all PC software.
- Providing this software would be relatively inexpensive

How does this relate to other work?

Different Questions ⇒ Different Answers

How can we preserve digital collections?

How should we manage preservation today?

How can we provide digital library services?

How should we manage electronic records?

How should we serve Web users?

Preserving Digital Information

http://home.pacbell.net/hgladney/PDIf.pdf

http://www.springer.com/3-540-37886-3

Comments to hgladney@gmail.com