

No thing is preserved in Digital Preservation

A preservation-as-communication approach towards modeling information preservation

Simone Sacchi sacchi1@illinois.edu

Center for Informatics Research in Science and Scholarship — Graduate School of Library and Information Science
University of Illinois at Urbana-Champaign

When we look at the kinds of things that participate in the representation of information in digital form, it appears that preservation as traditionally conceived does not literally apply. No thing appears to be literally preserved in digital preservation. If digital preservation is not about preserving things, what is it about, exactly?

Introduction

In DP we routinely talk about “preserving” or “maintaining” digital information, content, objects, material, documents, data, etc.

This rhetoric is derived from “physical” preservation, the assumptions being:

- There is an *object* of some kind
- This object is prone to decay or can be destroyed
- Preservation efforts must ensure its persistence

Does preservation, in this sense, **really** apply to the digital domain?

The notion of preserving information

Consider the notion of preserving information:

- extensively adopted in the literature, documentation, etc.
- In general, it is used to characterize the ultimate goal in digital preservation.

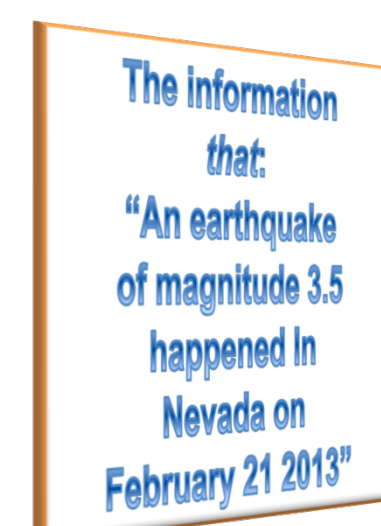
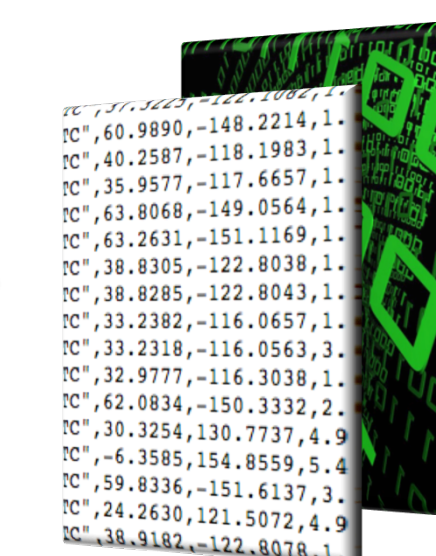
What **exactly** do we mean by “preserving information”?

What exactly is preserved in digital preservation?



NOT physical objects per se...
And in fact media migration is often required

NOT symbol structures...
Symbol structures cannot deteriorate or be destroyed, they persistence is ensured by their nature of repeatable abstracta.



NOT content...
Information, propositional content, cannot strictly speaking decay or be destroyed, it is an abstraction itself!

No thing is literally preserved in digital preservation!!!

If digital preservation is not about preserving things...

...what is it about, **exactly**?
...and how should it be **modeled**?

A preservation-as-communication approach

Most events in the lifecycle of a digital resource can be understood as **communication events** because they *sustain the flow of information* from the creator of a resource to its potential users.

Agents participating in these events establish **relationships** between content, symbol structures, and physical objects:

- Symbol structures might express content
- Symbol structures might encode other symbol structures

In virtue of these relationships, **types** acquire **roles**

- Symbol structures become bit sequences, data, text, etc.
- Propositional content becomes communicated information

These roles are **contingent** upon *socio-technical agreements* — what we called Interpretive Frames [1].

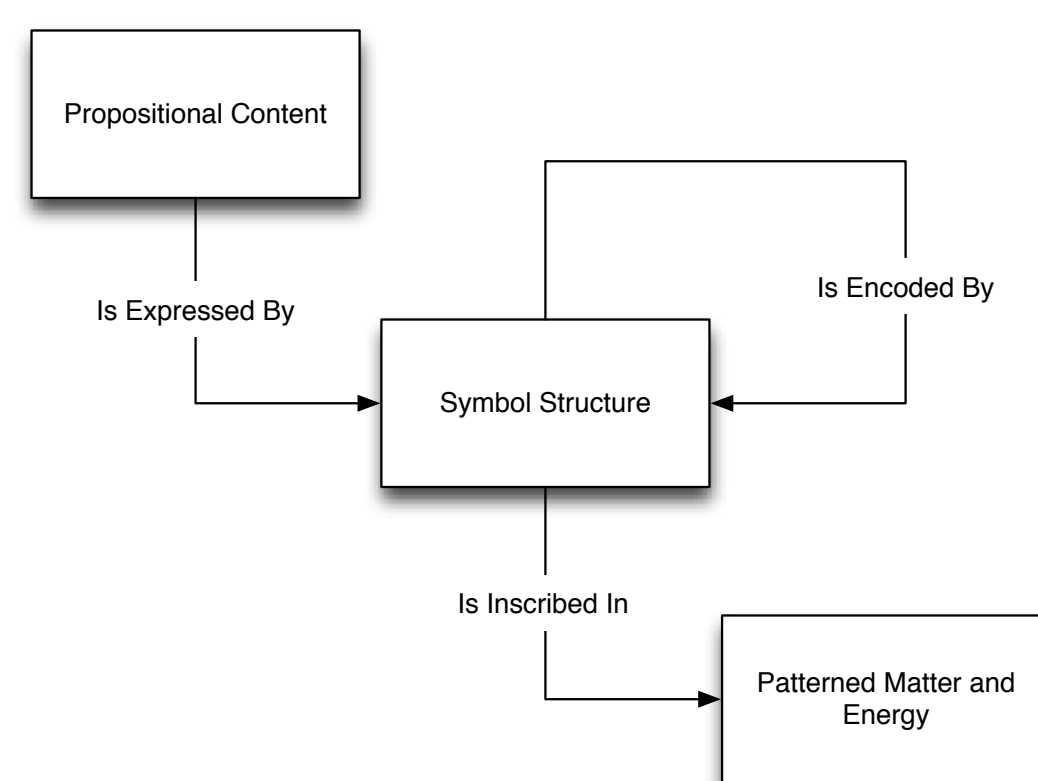
This **contingency** is the problem we face in DP

Digital preservation is a form of communication itself

It sustains the transmission of information by re-mapping these roles when technology changes

The Basic Representation Model (BRM)

The **Basic Representation Model** [2,3,4] represents the *types of things* participating in the representation of information in digital form.



- **Propositional Content**
 - The information carried by a digital resource, its meaning, etc.
- **Symbol structures**
 - Bit streams, character sequences, tables of numerals, etc.
- **Physical objects**
 - Hardware devices, components, etc.

References

- [1] Dubin, D., Wickett, K. M., & Sacchi, S. (2011). Content, Format, and Interpretation. In B. T. Usdin (Ed.), *Proceedings of Balisage: The Markup Conference 2011* (Vol. 7). Montréal, Canada.
- [2] Sacchi, S., Wickett, K. M., Renear, A. H., & Dubin, D. S. (2011). A Framework for Applying the Concept of Significant Properties to Datasets. In *Proceedings of ASIS&T 2011: The 74th Annual Meeting of the American Society for Information Science and Technology* (Vol. 48). New Orleans, LA.
- [3] Sandore, B., & Unsworth, J. (2010). ECHO DEpository — Phase 2: 2008–2010 Final Report of Project Activities (pp. 30–37). University of Illinois at Urbana-Champaign.
- [4] Wickett, K. M., Sacchi, S., Dubin, D. S., & Renear, A. H. (2012). Identifying Content and Levels of Representation in Scientific Data. In *Proceedings of ASIS&T 2012: The 75th Annual Meeting of the American Society for Information Science and Technology* (Vol. 49). Baltimore, MD.

Acknowledgments

I wish to thank my dissertation committee: Allen H. Renear, Jerome P. McDonough, Carole L. Palmer and Adam Farquhar. This research is being carried out at the Center for Informatics Research in Science and Scholarship (CIRSS) at the University of Illinois at Urbana-Champaign and grounded in previous work by Dave Dubin at NDIPP and work conducted by the Data Concepts research group under the Data Conservancy, a multi-institutional NSF funded project (OCI/ITR-DataNet 0830976).