AN UPDATE ON CHRONOPOLIS

AND THE

EMERGING DIGITAL PRESERVATION INFRASTRUCTURE IN THE UNITED STATES

2014 PRDLA Meeting, Macau

Brian E. C. Schottlaender
The Audrey Geisel University Librarian

The Library
UC SAN DIEGO

PREVIEW

- **Brief reminder of "last time"**

- **Three generations of digital preservation**

- **Evolution of the digital preservation ecosystem**

- **Chronopolis as microcosm**

Chronopolis™:
Federated Digital Preservation Environment Using Data Grid Technologies

2006 PRDLA/PNC/ECAI Meeting
Seoul, Korea

Brian E. C. Schottlaender
University Librarian, UCSD

"LAST TIME, I ... "

... described an information universe ...

- ... that was rapidly becoming *all-digital.*

- ... whose digital content was very fragile, more fragile than analog.

- ... in which there was little widespread support for digital preservation.

... Tick ... Tick ... Tick ...

The Library
UC SAN DIEGO

EIGHT YEARS AGO ...

... we were very much in the

First Generation of Digital Preservation.

The Library
UC SAN DIEGO

FIRST GENERATION: PRE-2009

Characterized by:

- **Focus on a single problem or solution**

- **Focus on immediate solutions**

- **Extramural funding**

- **No clear long-term business model**

The Library
UC SAN DIEGO

FIRST-GENERATION "ACTORS"

- **Chronopolis**
  - **Focused on preservation technologies**
  - **Funded by grants**
- **HathiTrust**
  - **Focused on digitized print resources**
  - **Funded by memberships**
- **MetaArchive**
  - **Focused on collaborative models**
  - **Funded by grants & memberships**
- **Portico**
  - **Focused on electronic journals**
  - **Funded by customers**

PROBLEMS W GEN 1

- **Narrowly focused**
- **Challenging for users**
- **Dependent on the largesse of funding agencies (e.g., Mellon), larger institutions (e.g., LC's NDIIPP) and/or complicated funding models (e.g., Hathi and Portico)**
- **Struggled with "long-term" in "long-term preservation"**
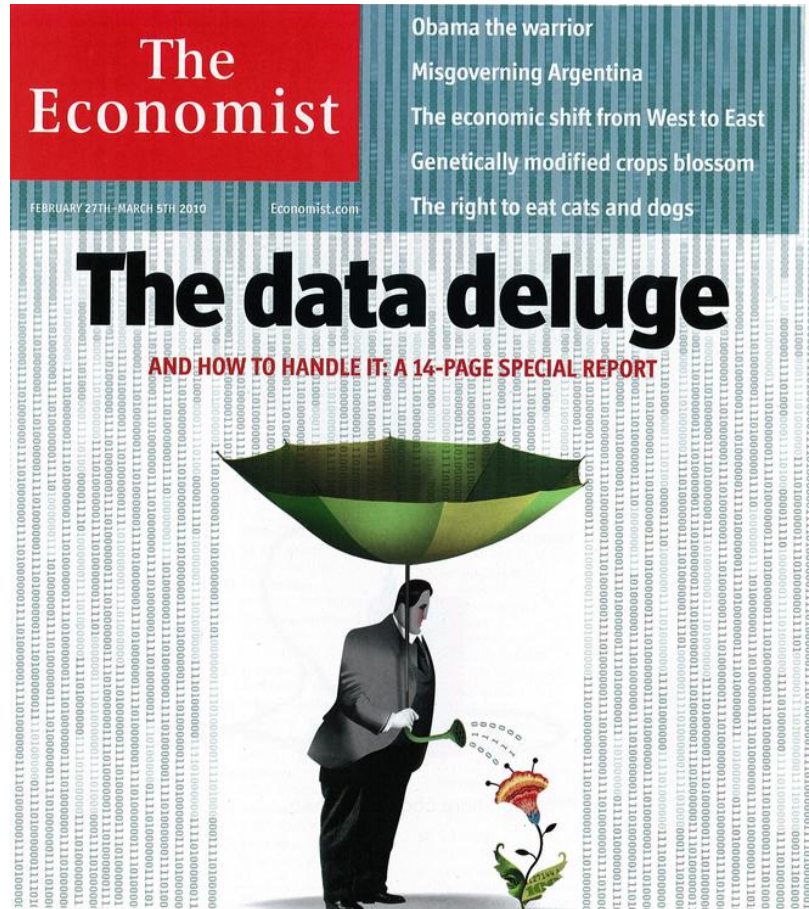
The Library
UC SAN DIEGO

SECOND GENERATION: 2010-2014

Characterized by:

- **Broadening of scope and focus**

- **Increase in types of content preserved**

- **Increase in customers/members/participants**

- **Still, no clear long-term business model**

The Library
UC SAN DIEGO

(And plus also, the world woke up.)

SECOND-GENERATION "ACTORS"

- **Chronopolis**
  - **Moved to a customer (i.e., charging) model**
- **CLOCKSS (Controlled LOCKSS)**
  - **Joint venture between academic publishers and research libraries**
- **HathiTrust**
  - **Created services and research tools with corpus**
- **MetaArchive**
  - **Greatly widened collaborative, even outside US**
- **Portico**
  - **First digital preservation service to be TRAC certified**

The Library
UC SAN DIEGO

PROBLEMS W GEN 2

- **Had begun to explore linking services, but still remained silos**

- **Had a hard time explaining high cost of preservation relative to basic storage**

- **Still, struggled with "long-term" in "long-term preservation"**

  But now we'd scared everyone with the seriousness of the problem!

THIRD GENERATION: POST-2014

Characterized by:

- **Uniting mature systems**

- **Interaction of diverse architectures**

- **Focus on long-term business model**

*We are now at the dawn of
the third generation.*

The Library
UC SAN DIEGO

THIRD-GENERATION "ACTORS"

All previous actors, plus:

- **CRL's Trustworthy Repositories Audit & Certification (TRAC)**
  - **Portico, Hathi, Chronopolis, and CLOCKSS all certified**

- **Educopia's Distributed Digital Preservation (DDP)**
  - **Focus on aligning preservation actions taken in extant distributed systems**

- **Aligning National Approaches to Digital Preservation (ANADP)**
  - **Focus on international alignment (Organizational, Technical , Legal, Standards, Economic, Educational)**

- **MetaArchive's "Chronicles in Preservation" Project**
  - **Focus on cooperative solutions for digitized & born-digital newspapers**

- **Digital Preservation Network (DPN)**
  - **Largest preservation network ever conceived**

The Library
UC SAN DIEGO

PROBLEMS W GEN 3

- **We're going to build it, but will they come?**

- **More, and bigger, commercial actors have entered the space. Will they take over?**

  - **Amazon Glacier, Google Cloud, Microsoft Azure Etc...**

- **If so, what is OUR value proposition?**

- **Ask us about our 20-year business model in 10 years ...**

GEN 1 PRESERVATION STACK

Unfederated Access Repositories

Libraries/IRs

Research Collections

Publishers

Unlinked Preservation Repositories
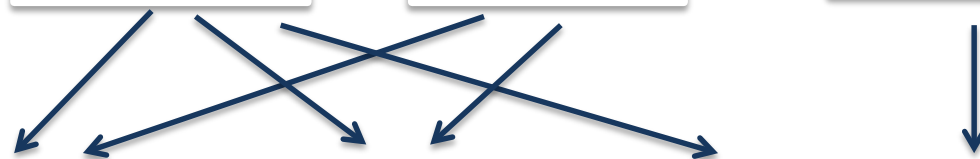
Chronopolis

MetaArchive

HathiTrust

Portico

Code

iRODS

LOCKSS

Isilon

CLOCKSS

Slide derived from James Hilton, w permission

Gen 3 Preservation Stack

Discovery

DPLA	SHARE	Etc...	PRL

Publishers

Unfederated Access Repositories

Libraries/ IRs	Shoah Visual Archive	Sloan Digital Sky Survey	Etc...

Preservation Repositories Linked to DPN

AP Trust	Chrono	Hathi Trust	SDR	TACC

Thumb Drives Etc...

DPN Backbone

Dark, diverse replication with audit and succession

Data From proprietary systems

Code

iRODS	LOCKSS	DSpace	Fedora	Vivo

Slide derived from James Hilton, w permission

Characterized by:

- **Focus on a single problem or solution**

- **Focus on immediate solutions**

- **Extramural funding**

- **No clear long-term business model**

The Library
UC SAN DIEGO

CHRONOPOLIS AS MICROCOSM 1

- **First generation:**
  - **Funded by NDIIPP**
  - **Focused on technologies for preservation**
  - **Focused on "behind-the-scenes" work**
  - **Didn't worry much about business model**
  - **Existed to satisfy NDIIPP's needs**
  - **20Tb of content under active management**

The Library
UC SAN DIEGO

SECOND GENERATION: 2010-2014

Characterized by:

- **Broadening of scope and focus**

- **Increase in types of content preserved**

- **Increase in customers/members/participants**

- **Still, no clear long-term business model**

The Library
UC SAN DIEGO

CHRONOPOLIS AS MICROCOSM 2

- **Second generation:**
  - **Moved into a per-terabyte charging model**
  - **Worked to build an infrastructure that could satisfy large number of customers**
  - **Still focused on "behind the scenes" work**
  - **Worried a lot about business model**
  - **50Tb of content under active management**

The Library
UC SAN DIEGO

THIRD GENERATION: POST-2014

Characterized by:

- **Uniting mature systems**

- **Interaction of diverse architectures**

- **Focus on long-term business model**

We are now at the dawn of the third generation.

The Library
UC SAN DIEGO

CHRONOPOLIS AS MICROCOSM 3

- **Third generation:**
  - **TRAC certified**
  - **Tighter integration with campus curation needs**
  - **Tighter integration with DuraCloud for a more robust system**
  - **Tighter integration into DPN for total preservation solution**
  - **200Tb of content under active management**

The Library
UC SAN DIEGO

CHRONOPOLIS: ON THE HORIZON

- **Complete integration with DuraCloud to offer an end-to-end preservation solution**

- **Complete integration with DPN, particularly in role of "low-entry-bar" solution**

- **Implement targeted solution for campus research data (Research Data Library)**

- **Leverage all of the above to create a stable, decades-long preservation platform**

The Library
UC SAN DIEGO

THANK YOU

QUESTIONS?

Brian E. C. Schottlaender
UC San Diego

BECS@ucsd.edu

The Library
UC SAN DIEGO