

# Building a large-scale preservation repository based on aDORe

Bill Kehoe & Adam Smith

Cornell University Library

DLF – April 30, 2008

**Cornell University Library**

needs to preserve

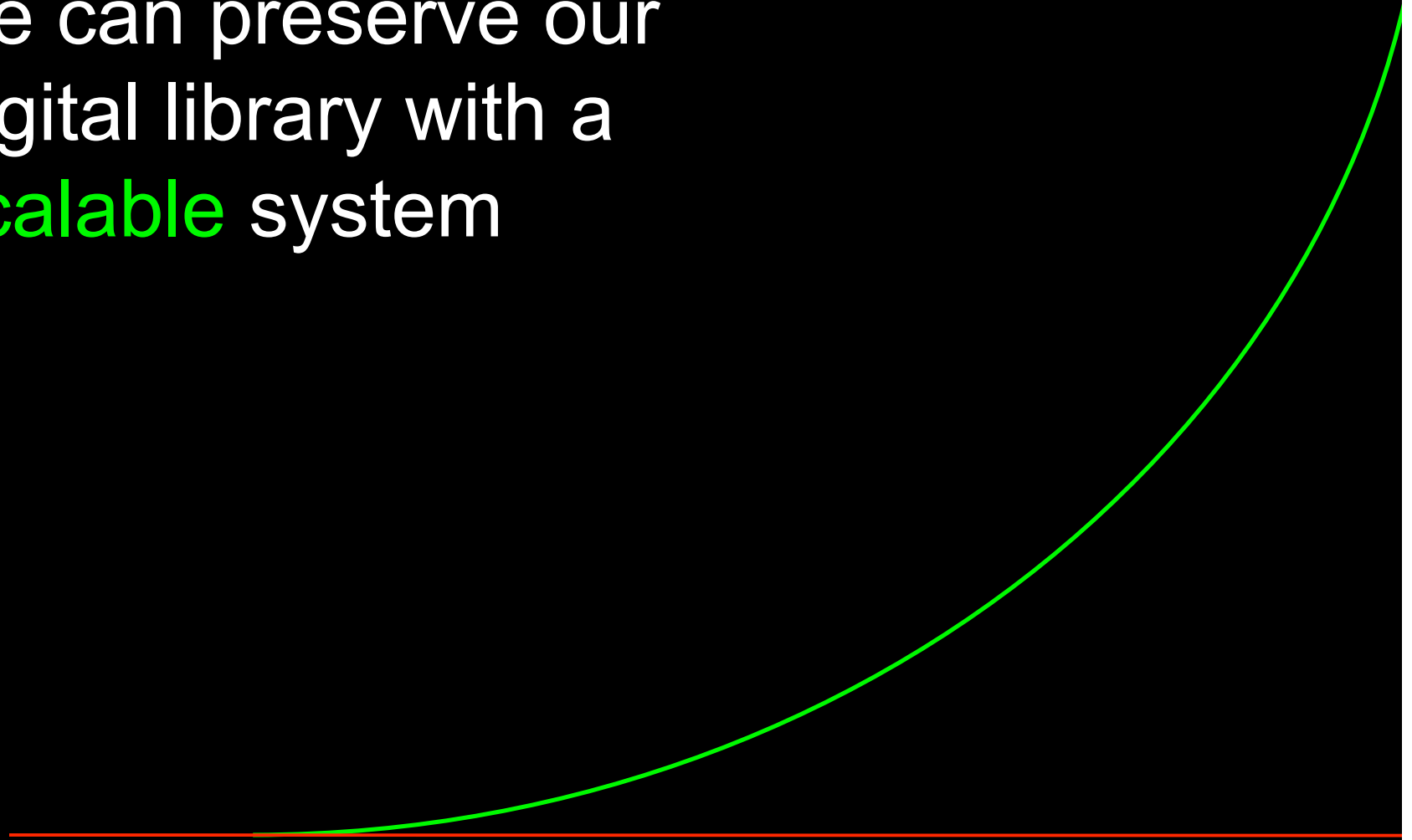
many terabytes of digital assets

anyone building a **large OAIS** faces  
problems on a **new scale**

we risk the loss of expensive  
assets if we manage millions  
of files poorly

£££€€€\$\$\$

we can preserve our  
digital library with a  
**scalable** system



a standards-based system backed  
by **aDORe** will do the job

A 3D rendering of a red book, viewed from a three-quarter perspective. The book is closed, showing the front cover, spine, and back cover. The cover is a solid, vibrant red color. The pages are a light gray color, visible along the edges of the book. The text "1 book" is printed in a clean, white, sans-serif font in the center of the front cover. The background is a solid black color.

1 book

The image features a black background densely populated with small, cyan-colored circles. Each circle has a thin white outline. The circles are scattered across the entire frame, creating a textured, stippled effect. In the center-left area, the number '1600' is displayed in a large, bold, yellow font. The number is slightly offset from the center, with the '1' being the tallest and the '0's being shorter and wider. The overall composition is simple and graphic.

1600



1 file

(1 pixel)

about 1 million files

(625 books)

how can we manage so many files?

we can bundle them  
in an aDORe archive

a METS *wrapper* describes and transports the digital object

storing the metadata



the  
XMLtape

storing the datastreams



the  
ARCfile

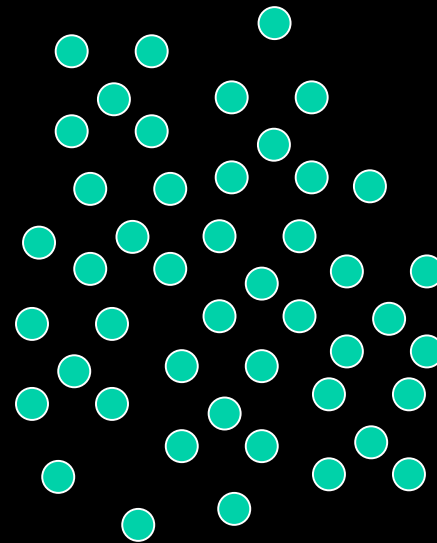
how can we retrieve the files?

with OAI and  
OpenURL requests

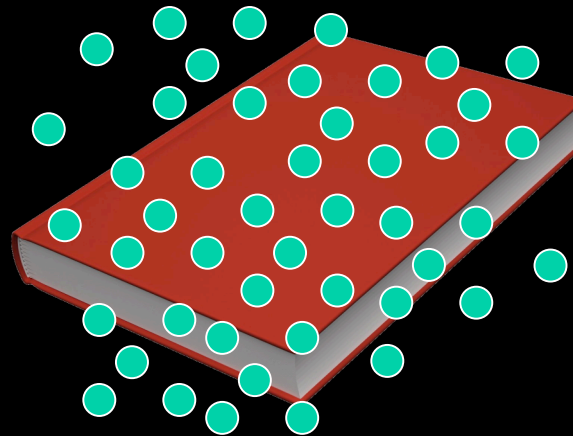
retrieving the metadata

< XML />

retrieving individual files



reconstituting the original complex  
object



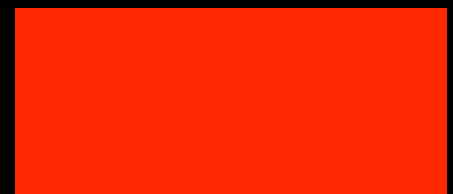
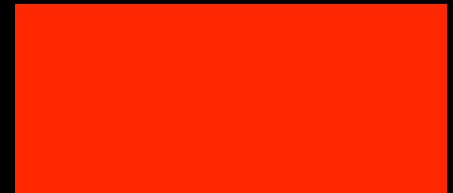
how does the system grow?

by **federating**  
multiple  
repositories

the **aDORe federation** software  
presents a **façade**

**discovery**  
and **access**

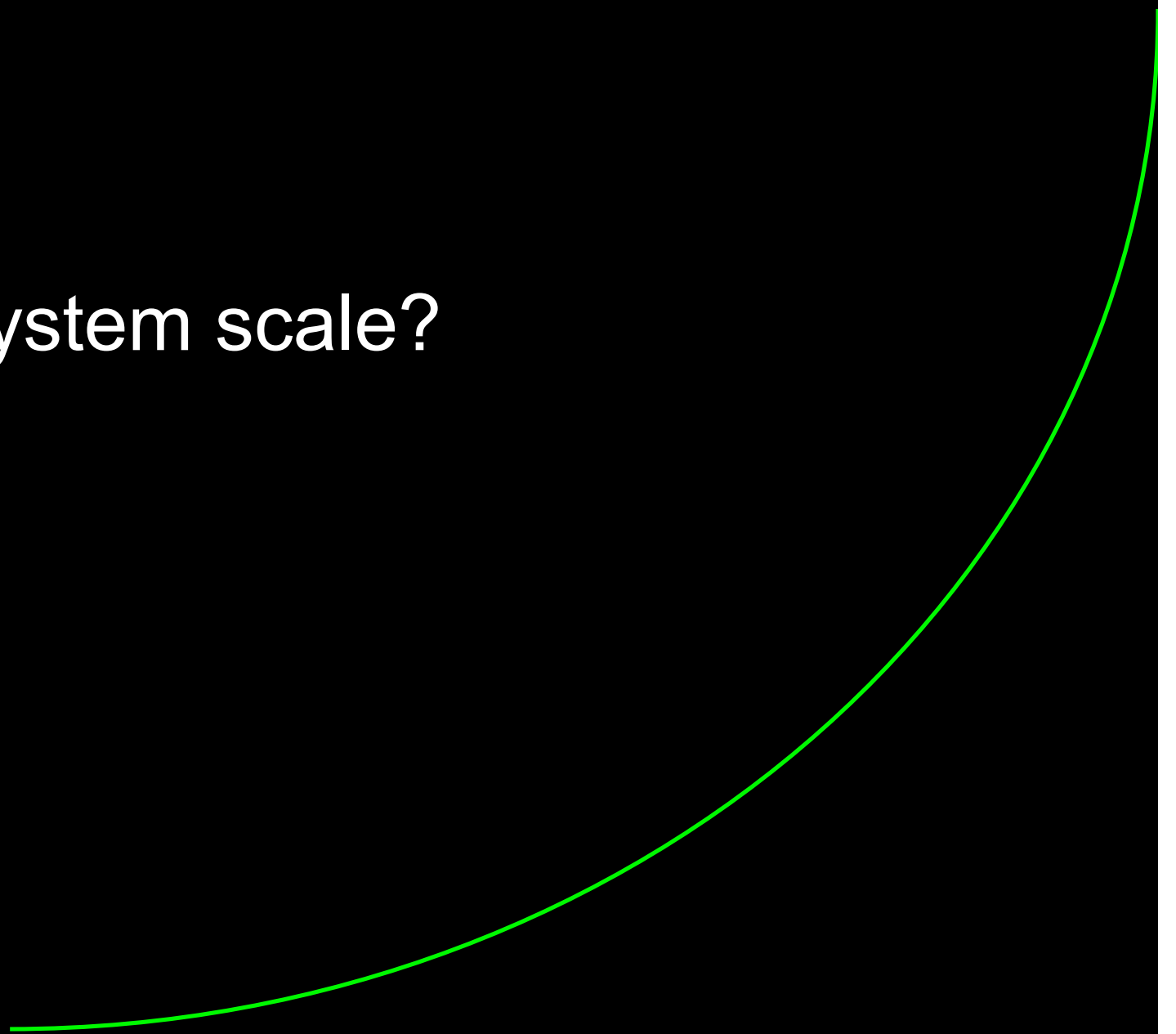
a simple data model for **all**  
repositories



finding anything

everything is a  
web resource

will the system scale?



we risk the loss of our assets,  
our time,  
our money,  
our reputations,  
our self-esteem...

a set of **federated aDORe archives**  
will allow our system to **grow huge**

# the aDORe Federation architecture

a façade  
that federates  
all repositories?

aDORe at CUL:

wrk1@cornell.edu

aDORe at LANL:

[http://african.lanl.gov/aDORe/projects/  
adoreArchive/](http://african.lanl.gov/aDORe/projects/adoreArchive/)

[http://african.lanl.gov/aDORe/projects/  
adoreFederation/](http://african.lanl.gov/aDORe/projects/adoreFederation/)