

Data Observatory: Preservation and Services

Gregory N. Jansen
Digital Curation Innovation Center
UMD iSchool
dcic.umd.edu

Storage Architectures, LOC, September 9, 2015

Objective: Data Observatory

- Big Data Training Sets - develop and provide access
- Test NCSA Data Services - simulate web-scale user load
- Teach Digital Curation - student experience and projects

Components

- 72 Million objects
 - Earth observation data
 - Federal and community archives
 - Digitized archival maps & images
 - Twitter collections
 - Growing research data sets...
- Indigo: Industry partners
Archive Analytics and NetApp
- NSF/DIBBs grant w/NCSA
- NCSA data migration (DAP)
- NCSA feature extraction (DTS)

Data Access Proxy (DAP)

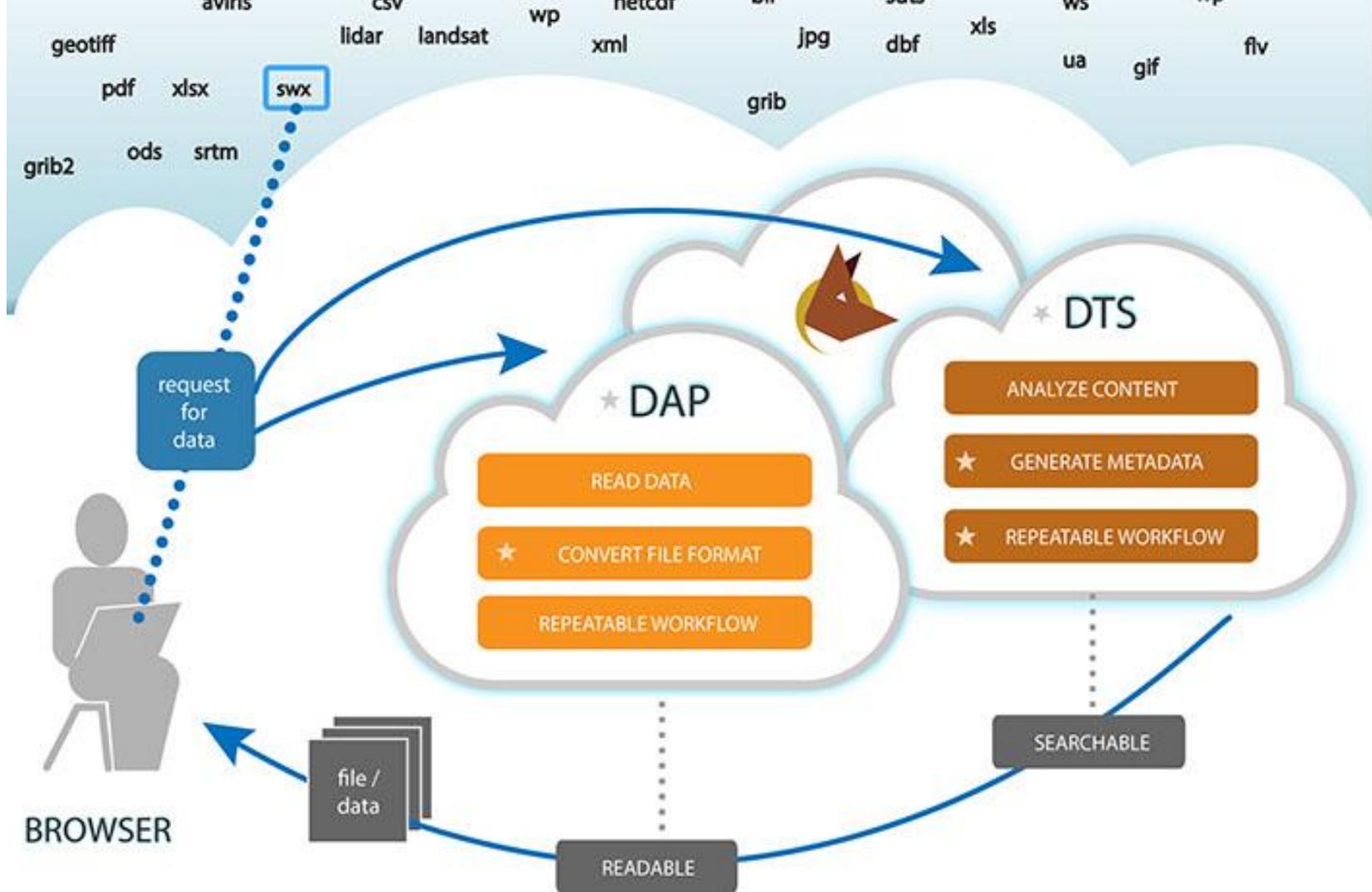
Data Tiling Service (DTS)

NCSA

“Brown Dog”

- “DNS for data”, i.e. web-scale data services
- Unlock the data in your files
- Format migration (DAP)
- Feature extraction and coordination (DTS)
- Strategy of reuse of existing software, “software mutt”

<http://browndog.ncsa.illinois.edu/>



Indigo

Obligatory Requirements Slide

- Scalability
- Redundancy
- Performance
- Audit Trail
- Access Control
- User Metadata
- API

The Unique Challenges...

How to provision big data sets for
iterative, compute-based research?

How to integrate extracted or enhanced data with original objects?

How to be preservation storage and also
a platform for active research?

Archive Analytics Indigo
&
Brown Dog

Questions?