

Content Transfer: Getting Data Moved Around the Network

NDIIPP: Preserving Digital Public Television

***DLF Fall Forum
2008***

Joseph G. Pawletko
New York University Libraries

PDPTV Overview:

■ Preserving Digital Public Television (PDPTV)

- preserving 20+ hours of 'born digital' Public Television content in prototype repository
- working groups researching IP & copyright, sustainability, metadata, ...

■ Multiple Sites involved in PDPTV

- Library of Congress
- WNET
- WGBH
- PBS
- NYU

Terminology:

you probably know these already...

■ **SIPs:**

- Submission Information Packages

■ **AIPs:**

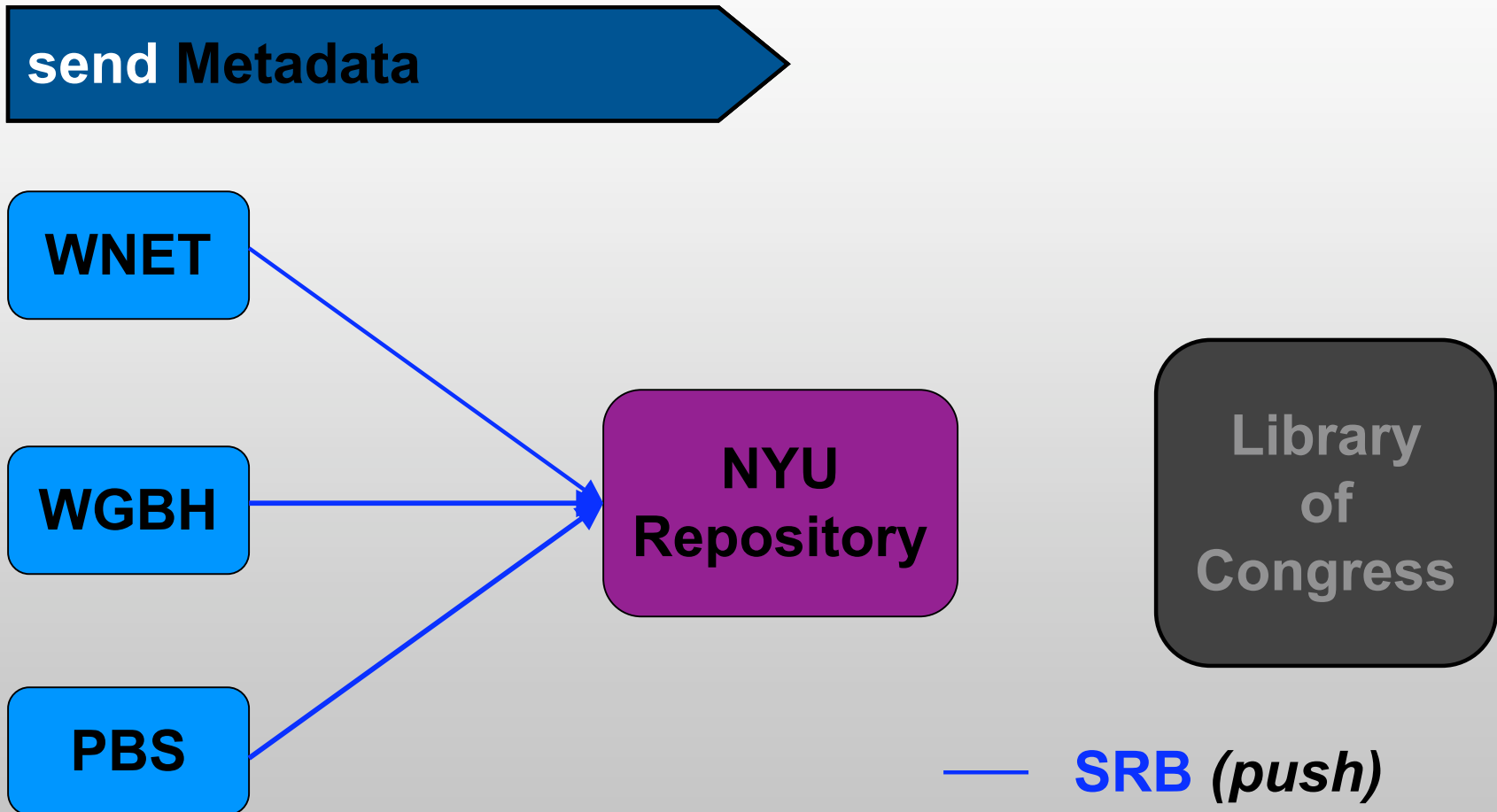
- Archival Information Packages

■ **DIPs:**

- Dissemination Information Packages

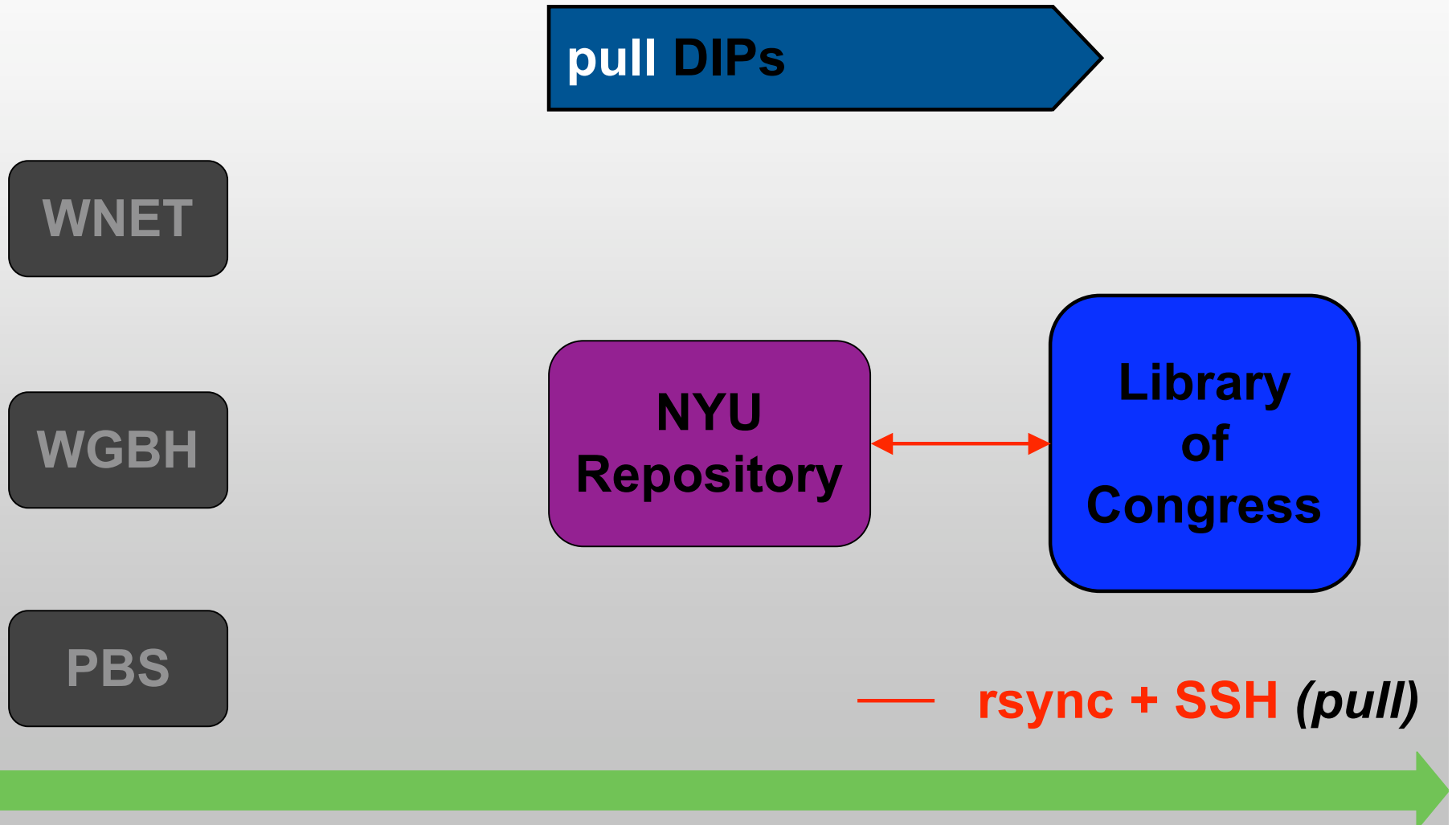
Infrastructure and Data Flow:

aggregation phase



Infrastructure and Data Flow:

dissemination phase



LC DIP Generation:

■ Generated 'Bags' for LC:

- DIPs conform to the **BagIt** specification
- wrote scripts that operate on SRB 'collections' to generate Bags
- generated both SHA1 and MD5 manifests
- leveraged SRB Scommmands in conjunction with *NIX symlinks to create Bags that point into SRB Vault
- froze SRB collections during LC Bag pull

PDPTV Data Transferred to LC:

■ 24 hours of content, 31 bags:

- One **Frontline** episode
- One **American Experience** episode
- One **Antiques Roadshow** episode
- Twelve **Nature** episodes
- Sixteen **New York Voices** episodes

■ > 900 GB of data

■ Average of > 29 GB per Bag

Generalized Data Transfer and 'Bags':

■ Plan to use 'Bags' for general content transfer

- NYU digitizing stations ---> NYU storage servers
- outside vendors ---> NYU
- Producer SIPs ---> NYU
- NYU ---> other repositories
- ...

■ Developing scripts to automate processes

- source-node script 'bags' content
- target-node script validates bag
- still discussing how to handle receipts
- plan to use BagIt's '[package-info.txt](#)' to facilitate content processing at target node

Acknowledgements:

■ **Library of Congress**

- Carl Fleischhauer, Jane Mandelbaum, Brian Vargas, Michelle Gallinger

■ **New York University**

- Dean Carol Mandel, Dr. David Ackerman, Professor Howard Besser, Kara van Malssen, Rasan Rasch, Grace Yang, Unni Pillai

■ **Thirteen/WNET**

- Ken Devine, Nan Rubin, Jonathan Marmor, Ryan Weston, Winter Shanck, Dave Rice

■ **WGBH**

- Dave MacCarn, Leah Weisse

■ **PBS**

- Bea Morse, Glenn Clatworthy, Wendy Allen, Irene Taylor, Michael Hunt

Wrap up...

Questions?

Thank you for your time!

Reference:

BagIt Specification:

<http://www.cdlib.org/inside/diglib/bagit/bagitspec.html>