NDIIPP Architecture and the Archive Ingest & Handling Test

DLF

Clay.Shirky@NYU.edu

November 7, 2005

Goals of Test

- Evaluate Systems
- Look for Areas of Interoperability
- Encapsulate Institution-specific Goals
- Generalize Interfaces
- Provide View Towards Federation

Archive Ingest & Handling Test

Goal: Give a moderately complex archive to several institutions and have them:

- Describe it
- Mark it up
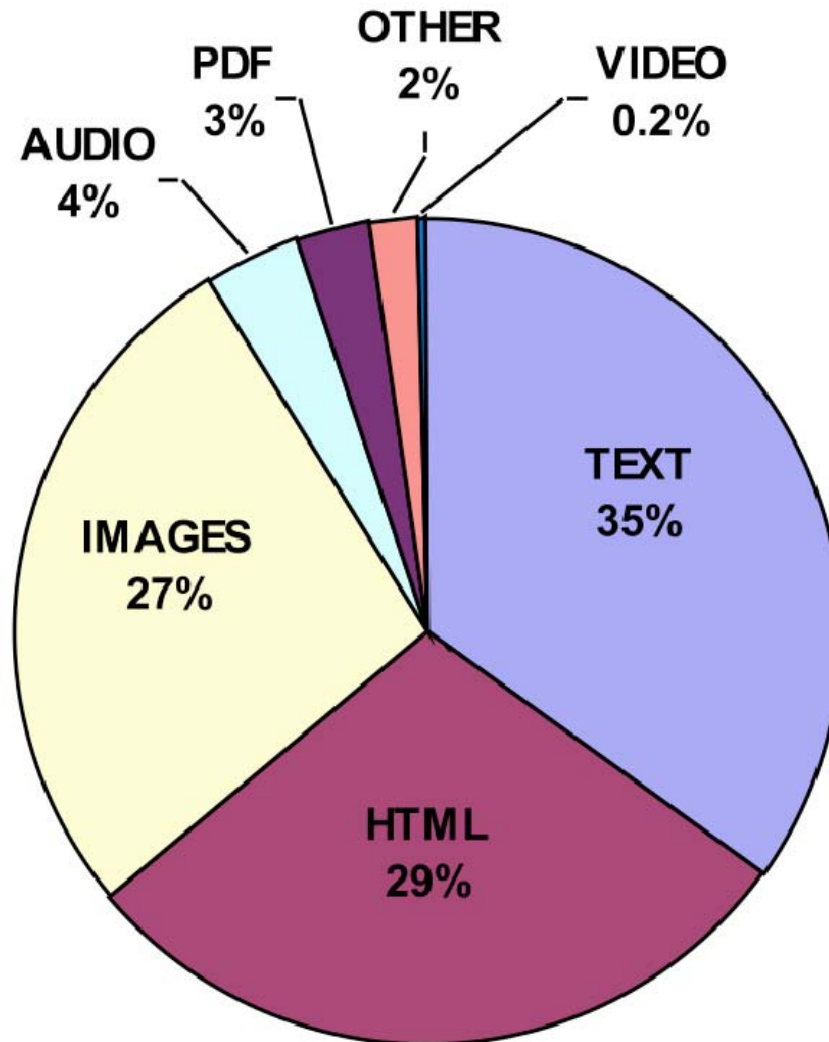- Ingest it
- Transform it
- Share it

GMU 9/11 Archive

- Physically small (~12Gb)
- Conceptually large (~57K files, many types)
- Messy (Amateur contributions, inconsistent naming)
- No solid meta-data
- No access to original sources
- As inconsistent as real life

Inconsistent

	GMU Document	GMU DB	GMU TMD	LC Inspection
Size	12GB			12GB
File Count	57,442	57,492	57,540	57,540
"Collections"	2,105	171		
"Sub-Collections"		1,934		
"Contributors"		17,504		170

Imbalanced Content

Great Breadth of Contribution