Identifying and retrieving digital objects: A Study of the Handle System

Taarik Hassanmahomed

System and Network Engineering

June 30, 2010

CineGrid.org

CineGrid's Mission:

To build an interdisciplinary community that is focused on the research, development, and demonstration of networked collaborative tools to enable the production, use and exchange of very-high-quality digital media over photonic networks.

AMPAS

- One of the members is AMPAS – the Academy of Motion Picture Arts and Sciences.
- They are just like the rest of the community looking into new way to manage their data explosion.
  - 500 movies every year
  - metadata storage frame by frame
  - over 25 million object per movie

- A candidate for managing this is the Handle System.

Research question

- How can the Handle System help in storage, search, retrieval and preservation of digital content more efficient and reliable within AMPAS/CineGrid in particular?

Overview

- Identifiers
- Metadata
- Handle System
- Applicabilty in CineGrid/AMPAS

Overview

- Identifiers
- Metadata
- Handle System
- Applicabilty in CineGrid/AMPAS

Identifiers

- Namespaces
  - Uniqueness and Persistence
- Locating object
  - Broadcast, Home-base, P2P, Hierarchical
- Scalability
  - Distribution and Replication
- Preservation
  - Loss of meaning and loss of provenance or authenticity

Identifier using metadata

ISAN	ISAN 0000-0000-D07A-0090-Q-0000-0000-X
OpenURL	http://resolver.x.com/cgi?genre=book&isbn=0236218310
PURL	http://purl.oclc.org/oclc/rsch/metadataII

Overview

- Identifiers
- Metadata
- Handle System
- Applicability in CineGrid/AMPAS

Metadata

"An element of metadata describes an information resource, or helps provide access to an information resource .A collection of such metadata elements may describe one or many information resources "

W. Cathro (1997)

- Metadata is an important part in understanding the semantics of digital content
- Power lies in choosing the right set of element.

Metadata catagories

- Metadata itself can be categorised in various ways:
  - General vs. Specialistic
  - Minimalistic vs. Rich
  - Hierarchical vs. Linear
  - Embedded vs. Detached
  - and much more

Metadata schema examples

- Dublic Core,
  - 15 core elements, interoperable
- IEEE LOM,
  - nine categories, sub elements
- MPEG7,
  - Video, audio, generic features, multimedia description
- HTML
  - DESCRIPTION and KEYWORDS
- Conclusion there is no single best solution

Overview

- Identifiers
- Metadata
- Handle System
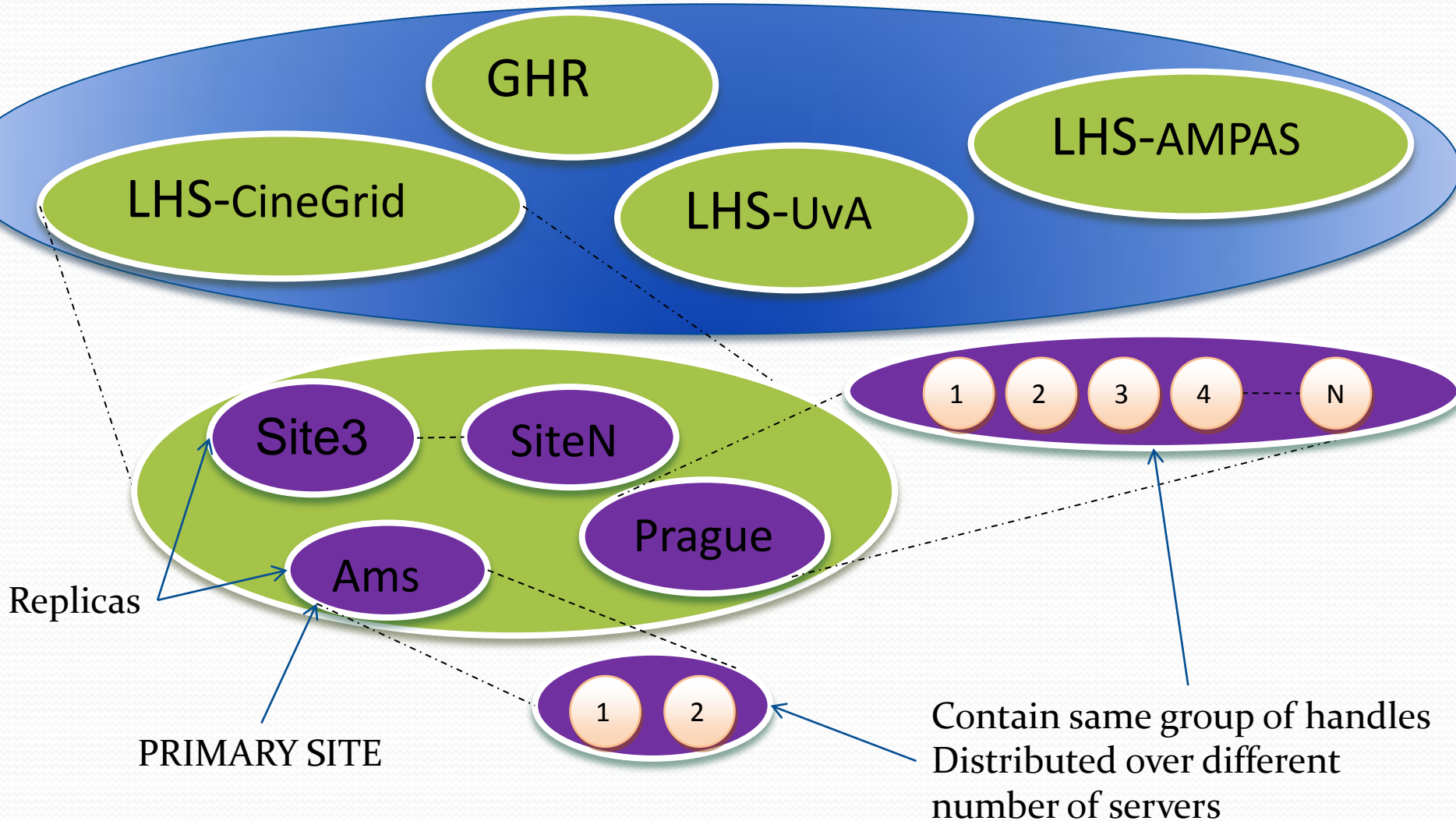- Applicability in CineGrid/AMPAS

Handle System

- **CineGrid.AMPAS\NiceMovie4K**

Handle System properties

- **Globally** unique identifiers
- Handle name **persistence**
- Multiple **instances** and **attributes** of an object
- **Extensible** namespace
- **Distributed** service model
- **Secured** name service
- **Distributed administration** service

HS Architecture

GHR : Global Handle Service
LHR : Local Handle Service

GHR

LHS-CineGrid

LHS-UvA

LHS-AMPAS

Site3

SiteN

Prague

Ams

Replicas

PRIMARY SITE

1   2   3   4   N

1   2

Contain same group of handles
Distributed over different
number of servers

Overview

- Identifiers
- Metadata
- Handle System
- Applicability in CineGrid/AMPAS

Handle metadata access

- Web proxy
  - GHS proxy at hdl.handle.net
  - local LHS proxy (port 8000)
  - Any other non local LHS proxy
- Java tool from handle.net
  - Admin tool
  - Java command line tool
- Custom client Java/C application
  - OpenHandle (open source)
    - http://code.google.com/p/openhandle/

Web proxy resolution

- "wget" handles with the web proxy hdl.handle.net
  - Recursive resolution
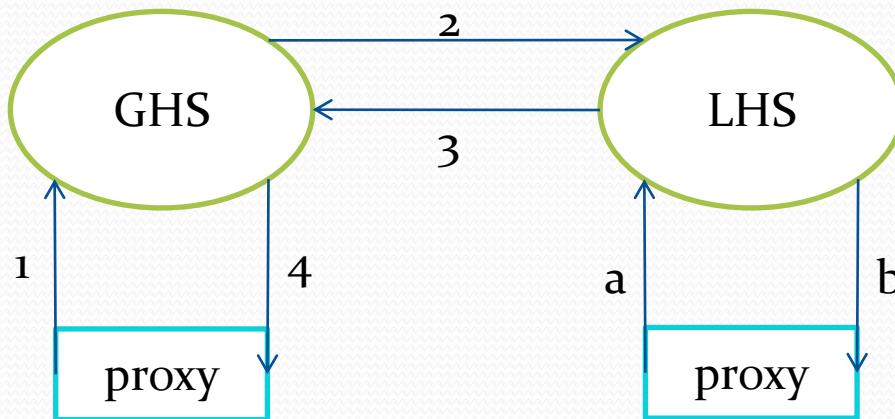
1 request handle resolution
2 request specific handle
from primary of mirror
3 return handle
4 generate HTML responds

a request handle resolution
b generate HTML responds

hdl.handle.net	time req-resp	Ping
glow.handle.net	0.2945 seconds!	0.0975 seconds
Local Proxy	0.0103 seconds	same server

Java tool resolution

- Direct request with the Java command line
  - Iterative resolution

GHS*   LHS

2   3

1   client   4
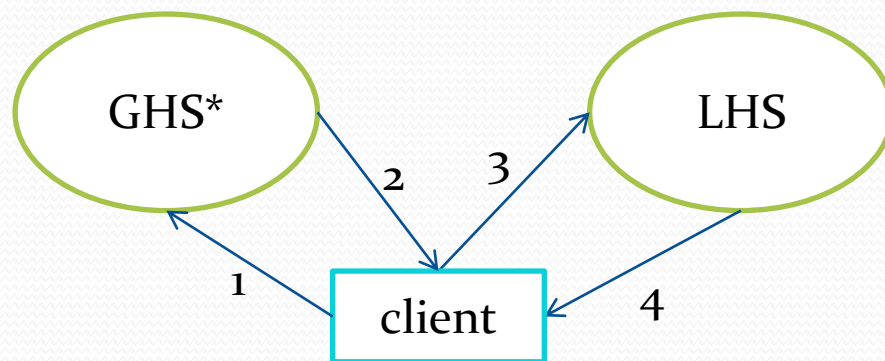
1 request handle resolution from random server
2 random reply with address of primary or mirror
3 request handle resolution
4 respond with handle data

time req-resp	GHS	location	ISP	Ping
0.6102 seconds!	glow.handle.net	US, United States	PSI	0.0963 seconds
0.6251 seconds!	macmini1.handle.net	Reston, VA 20191	CNRI	0.0978 seconds
0.5997 seconds!	hercules.handle.net	Reston, VA 20191	CNRI	0.0989 seconds
0.5826seconds!	Crossref.org	Lynnfield, MA 0194	Verizon	0.0987 seconds
0.7634 seconds!	China	Beijing, 22	CNIC	0.1103 seconds

Failover test

- Turning of the primary to see how many times handles from the primary are requested.
  - The command line tool (50% of the time)

action	GHS	LHS	LHS	time req-resp
primary down	Random	primary (3x)	mirror	6.5711 seconds!

  - The web proxy (10% of the time)

action	GHS	time req-resp
primary down	glow.handle.net	10.2958 seconds!

Resolution recommendations

- There is just one recommendation and that is to use a local caching server which is one of the optional component of the Handle System
  - It caches handle data, service information of any LHS and allows re-use of information obtained from earlier queries, reducing traffic between Handle System clients and servers.

Handle metadata access

- Custom client Java/C application
  - OpenHandle v0.21 (open source)
    - http://code.google.com/p/openhandle/

- Repopulate a copy of the Cinegrid Amsterdam portal with handles

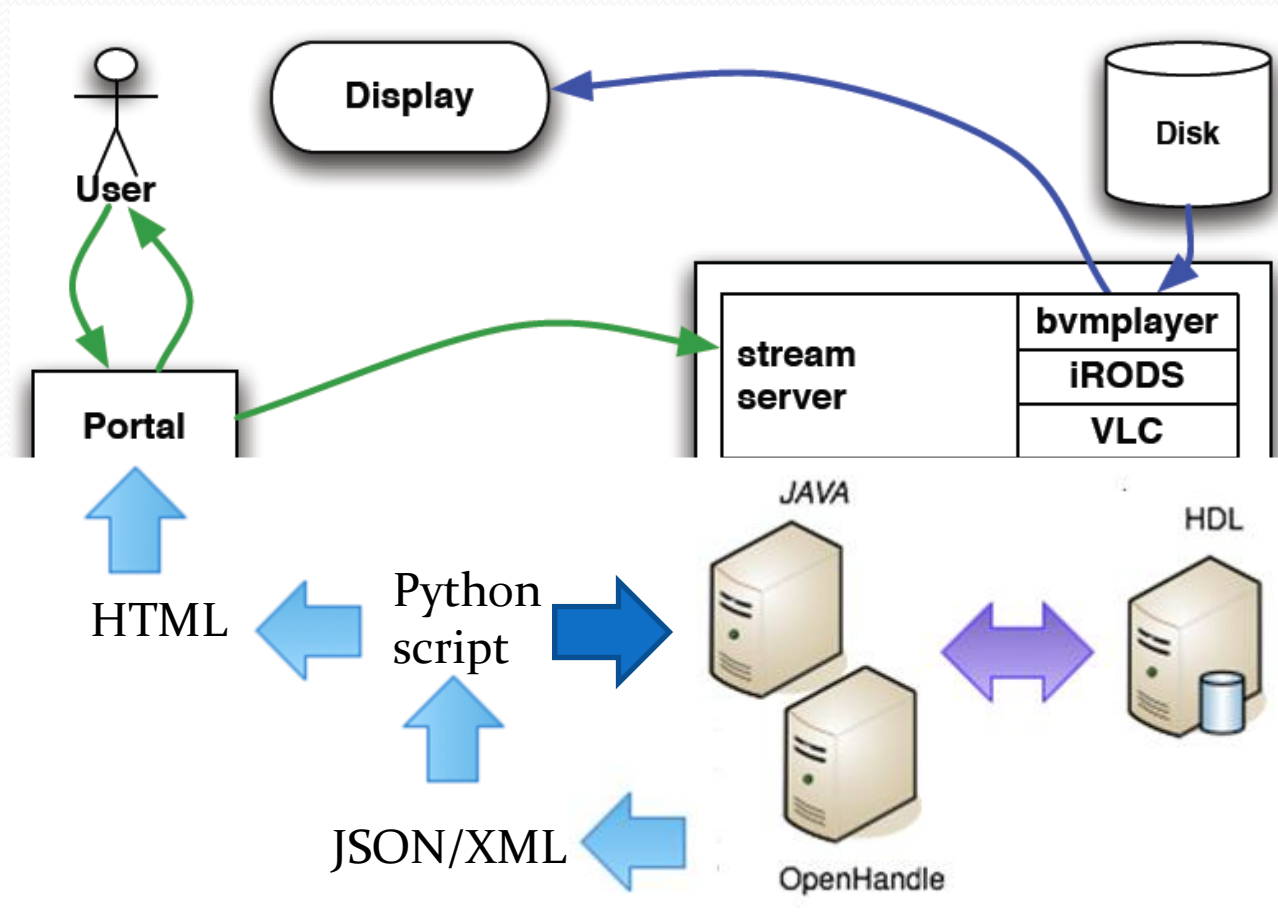# CineGrid Amsterdam Portal

Architecture

Architecture
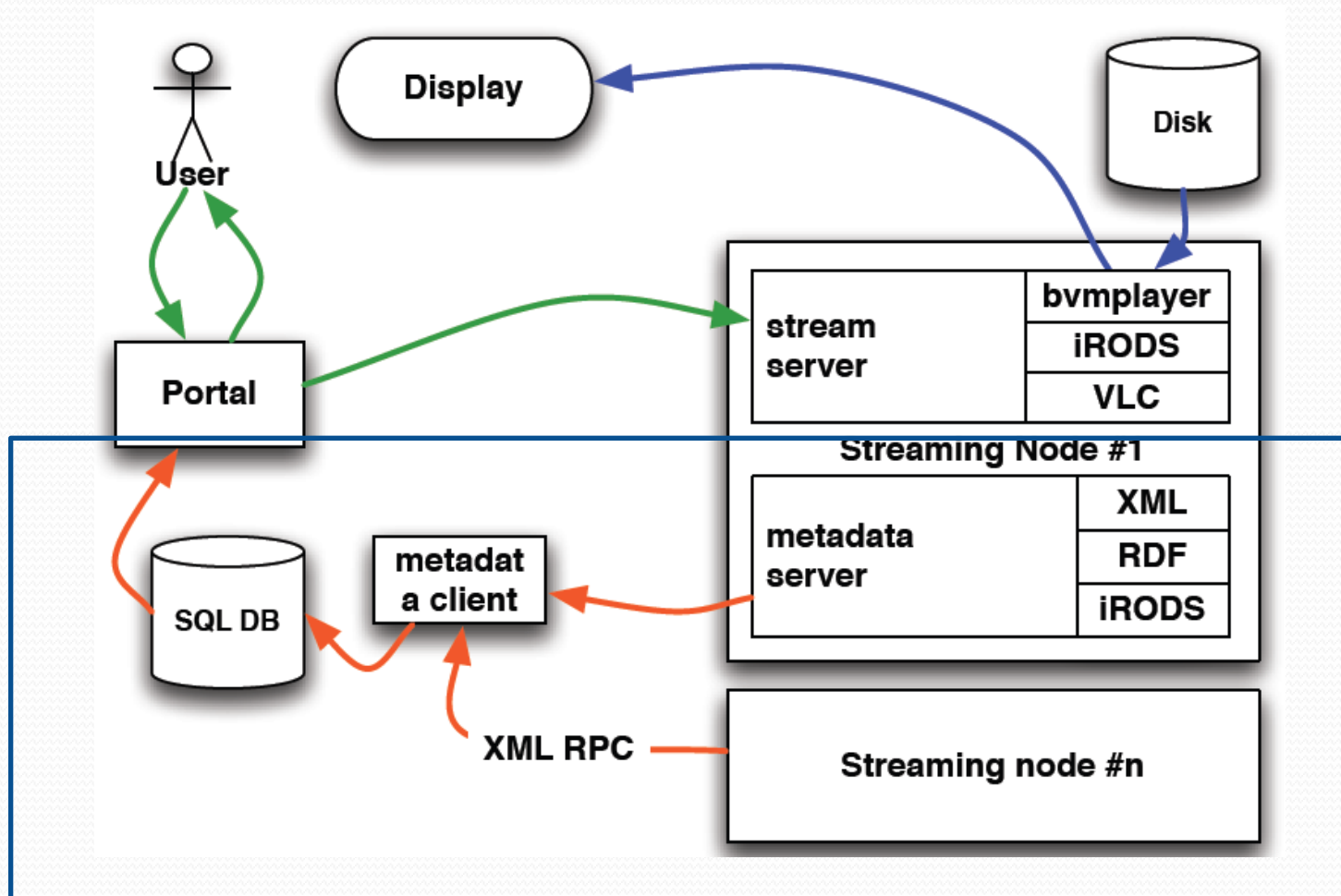
New architecture

Example of Population

Template used:

Author: []
Created: ago
Tags:
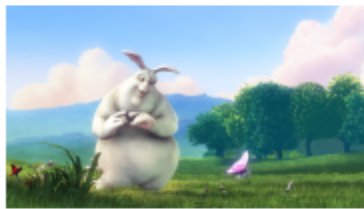**Preview images:** [4k tiff] [4k jpeg] [4k small] [4k normal]

Description:

Codec	Format	Location	Size	Framerate	Duration	Filename
	x		bytes	frames	minutes	

After choosing handle:

Big Buck Bunny

Author: [Blender Foundation]
Created: 1970-01-01 01:33:28 ago
Tags:
Preview images: [4k tiff] [4k jpeg] [4k small] [4k normal]

Description:

(c) copyright Blender Foundation | http://www.bigbuckbunny.org

Codec	Format	Location	Size	Framerate	Duration	Filename
mpeg4	1920x1080	cgdevil	1135094932 bytes	30 frames	3600 minutes	big_buck_bunny_1080p.mp4v

Selecting multiple handles

Big Buck Bunny

Author: [Blender Foundation]
Created: 1970-01-01 01:33:28 ago
Tags:
Preview images: [4k tiff] [4k jpeg] [4k small] [4k normal]

Description:

(c) copyright Blender Foundation | http://www.bigbuckbunny.org

Codec	Format	Location	Size	Framerate	Duration	Filename
mpeg4	1920x1080	cgdevil	1135094932 bytes	30 frames	3600 minutes	big_buck_bunny_1080p.mp4v

7 Bridges

Author: [CineGrid]
Created: 1970-01-01 01:33:27 ago
Tags:
Preview images: [4k tiff] [4k jpeg] [4k small] [4k normal]

Description:

(c) A boat ride on the canals of Amsterdam.

Codec	Format	Location	Size	Framerate	Duration	Filename
mpeg4	1920x1080	cgdevil	214748364 bytes	30 frames	138 minutes	hollandfestival07.7bridges.1080p.30fps.dxt.bmv

Handle system metadata

Handle:INDEX	#	TYPE	HANDLE DATA
10677/7_Bridges	1	TITLE	7 Bridges
	2	AUTHOR	CineGrid
	3	DESC	(c) A boat ride on the canals of Amsterdam.
	4	CREATED	"1970-01-01 01:33:27"
	…	…	…
	25	URL	http://cinegrid.uvalight.nl/portal
	26	IMAGES	http://cinegrid.uvalight.nl/images/bridge.p

CineGrid Handle Access

- CineGrid community consist of various members all over the globe and therefore content all over the world.
- Handles are not hard linked to any digital content
- Use the whole prefix: CineGrid\LHS
  - Use a member site as the primary, other member can then choose to become a mirror or leave the replication to the primary.
- Make use of subprefixes: CineGrid.AMPAS\LHS
  - Every member can be a primary and have administration and storage near by.

Conclusion

- Handle system is fairly good for storage, retrieval and preservation of metadata and the location of digital content.

- combination of the LHS and GHS with all its characteristics ensures that digital content can continue to grow and still be available without much loss in performance.

- Still it can be seen as a form of middleware, which require application like OpenHandle to get its full potential.

More info?

- Handle System website "http://www.handle.net/"
- RFC3650 "http://www.handle.net/rfc/rfc3650.html"
- RFC3651 "http://www.handle.net/rfc/rfc3651.html"
- CineGrid.org website "http://www.cinegrid.org/"
- CineGrid Ams website "http://cinegrid.uvalight.nl/portal/"
- AMPAS "http://www.oscars.org/"

Questions?