ARCITECTA™

www.arcitecta.com

Managing Your Digital Assets

THE CROWN JEWELS

Jason Lohrey

Chief Technology Officer
Arcitecta Pty. Ltd.

22nd February 2011

ARCITECTA™

A “Digital Asset” .. (*aka*: “crown jewel”)..

A Digital Asset is the intellectual property of an entity (individual, organization, etc.) embodied in digitized form.

The same intellectual property may exist in different forms as multiple digital assets.

ARCITECTA™

Regardless of Size, Some Assets are Costly to Acquire

Must be accessible for a long time (perhaps decades, or centuries).

Quality must be assured and retrievable in original state at any time in the future.

Must have redundant copies.

Need open standards for exchange.

May be repurposed in ways not previously envisaged.

ARCITECTA™

Forms of Digital Embodiment

"Unstructured":

- Files

"Structured":

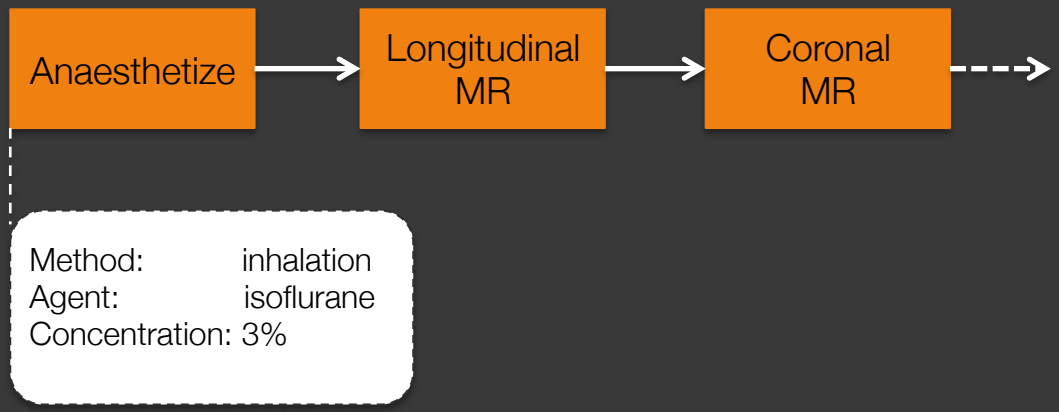
- Tuples in a Relational Database
- Objects in an Object Database

Digital Asset Management

Science and Research

Reproducible and accountable science requires clearly recording, maintaining and making accessible:

- Research methods
- Measurements
- Acquisitions
- Context (e.g. software)



^Traditional

File Systems are Dead

They Will be Supplanted by "Information Systems"..

File Systems are Dead

The File

- A file is simply a *container* for information
- The “information” itself is the intellectual property of an entity
- In addition to the primary purpose, the file content is also important for:
  - Provenance
  - Discovery
  - Access management
  - Life-cycle management

File Systems are Dead

What Can We Do With a File System?

A file system is a *hierarchy* of files, in which we can:

- Provide a name
- Open, read, write, delete
- Attach extended attributes

File Systems are Dead

What Can We Do With a File System?

We can organize information using the following mechanisms:

- Location (meaning inferred from the name)
- Soft-links give multiple axes

```
/data/project1/
            format_X/
                a.x
            format_Y/
                a.y
                b.z
```

File names usually don't change.

File Systems are Dead

What Can We Do With a File System?

Additional information is often stored in separate files:

```
/data/project1/
            format_X/
                a.x
                a.x.xml
            format_Y/
                a.y
                a.y.txt
                b.z
```

File Systems are Dead

What Can't We Do (easily) With a File System?

A typical file system does not allow:

- Searching by file content
- Searching by multiple contexts
- Attaching (complex) information not supported by the file format itself
- Repurposing information in different contexts
- Semantic operations
- …

ARCITECTA™

Structured Databases

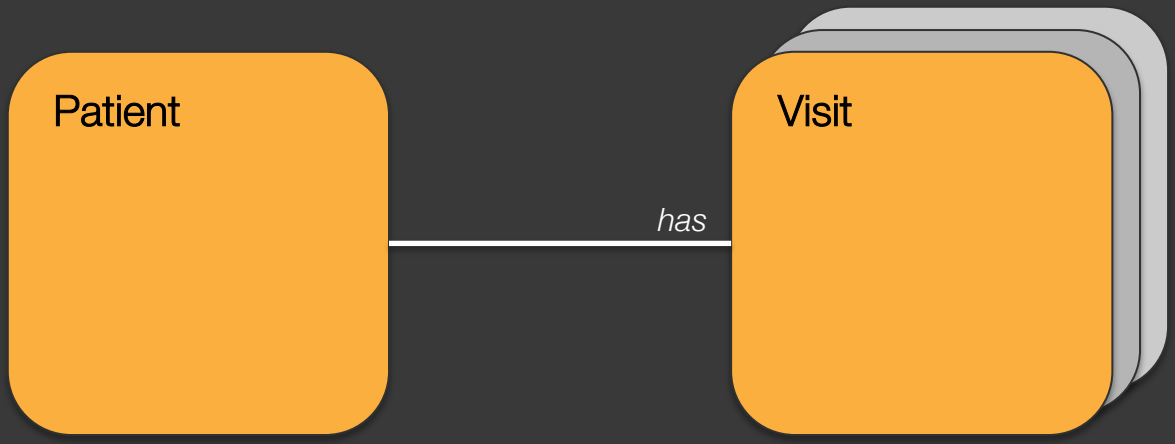
Will be Supplanted by "Information Systems"..

Structured Databases

The Structured Database

A structured database is:

- A set of information organized into an agreed structure
- There are often relationships between the structures



Structured Databases

What Can We Do With a Structured Database?

We can organize information into agreed structures and:

- Create, update and delete structures
- Perform “structured” searches

```
select a.x, b.y from table_a a join table_b b where
a.id = b.id;
```

Structured Databases

What Can't We Do (easily) With a Structured Database?

Typically, cannot:

- Perform "unstructured" searches
- Store and manage "unstructured" data
- Arbitrarily manipulate/transform the data

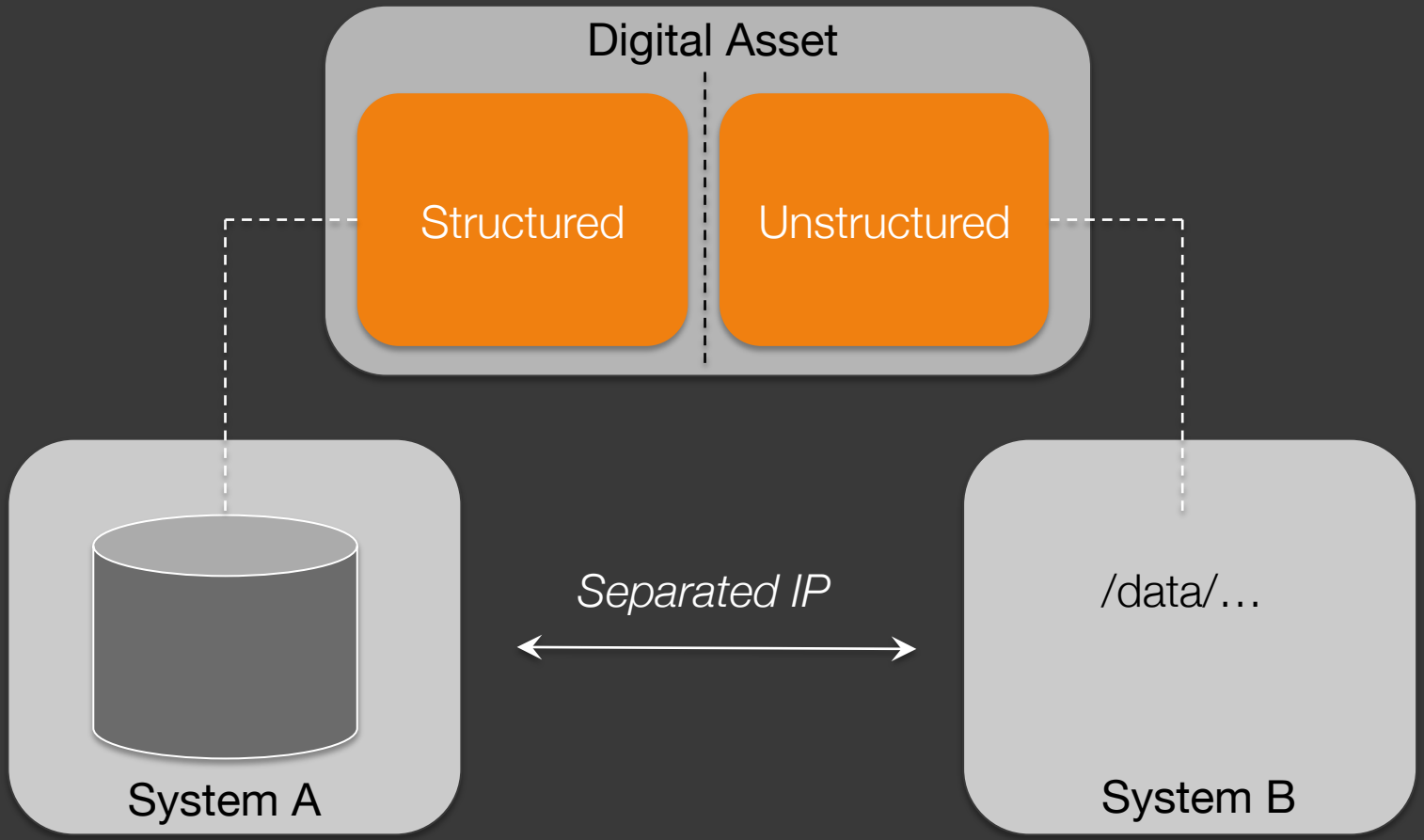
Even structured data presents problems - Relational Databases cannot effectively store large amounts of complex and arbitrarily structured information without significant performance degradation.

Typical Approaches

Aggregated Systems ..

Typical Approaches

Separation of Intellectual Property



ARCITECTA™

Mediaflux™

OPERATING SYSTEM FOR META+DATA

aka: LiveArc™

(SGI Branding)

Mediaflux™

Operating System

Mediaflux™ is a digital asset and information management platform:

- Treats structured and unstructured data as an atomic object
- Extensible Service-Oriented (SOA) architecture
- Written in Java (platform neutral)

Mediaflux™

Objectives

To:

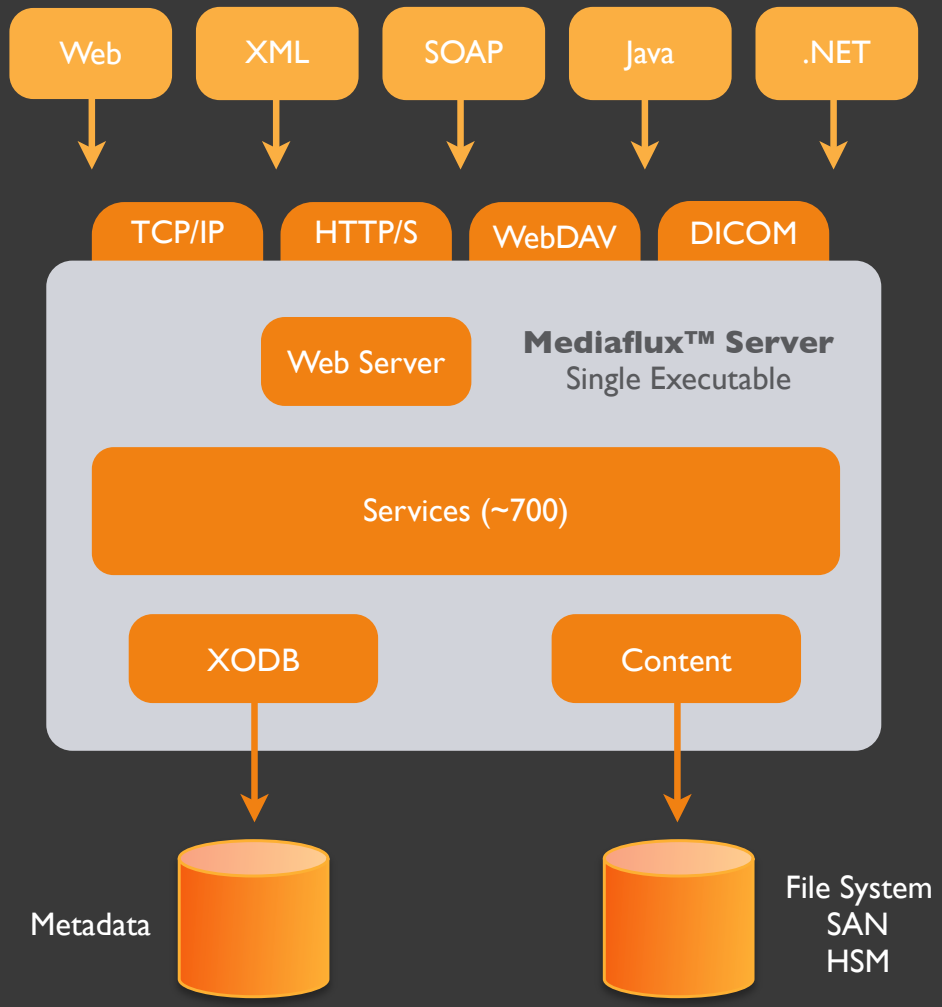
- Manage any type of complex information
- Hide whether requested data literal or computed
- Hide the location of information (logical addressing)
- Make it easy to evolve ontologies (and support multiple ontologies)
- Scale linearly for any given set of evaluants
- Unify information management

Target Problem Spaces

- Keeping track of intellectual property:
  - Structured information
  - “Unstructured” information:
    - Images, video, audio..
    - Documents, software
  - Research
    - Data sets, models, processed data
- Collaboration
- Auditing
- Quality Assurance

ARCITECTA™

MEDIAFLUX ARCHITECTURE



Axis: Operating System

- Everything is a service
- Services can be extended (plugins)
- A service may:
  - Consume or produce information locally or remotely
  - Retrieve or compute information
- Can be scripted
- Callable from any other system
- Can call any other system

Services

ARCITECTA™

Authentication	Asset Management	Triggers	Firewall
Authorization	Dictionaries	Replication	Citeable IDs
Auditing	Classification Schemes	Logging	Events
Automation	Transcoding	WWW	Notification
Scripting	Scheduler	Plugins	E-Mail

Axis: Digital Asset Management

- An asset is the combination of metadata and content
- Metadata and content are independently versioned
- Metadata:
  - Can be easily evolved
  - Automatically extracted from content
  - Manually added
  - Can conform to any schema
  - Can be queried using structured and unstructured (“free text”) searching
- Content:
  - Can be anything
  - Centralized, referenced and distributed heterogeneous storage

ANATOMY OF AN ASSET

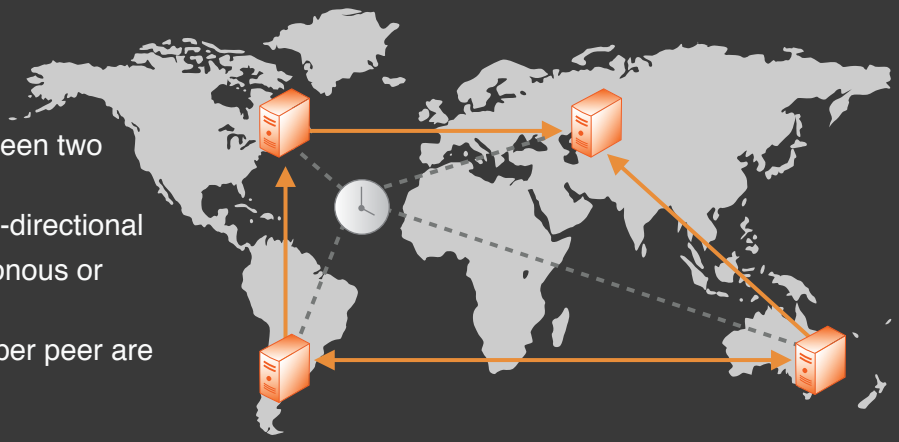
ARCITECTA™

Axis: Distributed Systems

Distributing Data

Replication

+ Data can be replicated between two or more Mediaflux servers
+ Replication can be uni- or bi-directional
+ Replication can be asynchronous or synchronous
+ Multiple replication policies per peer are supported.

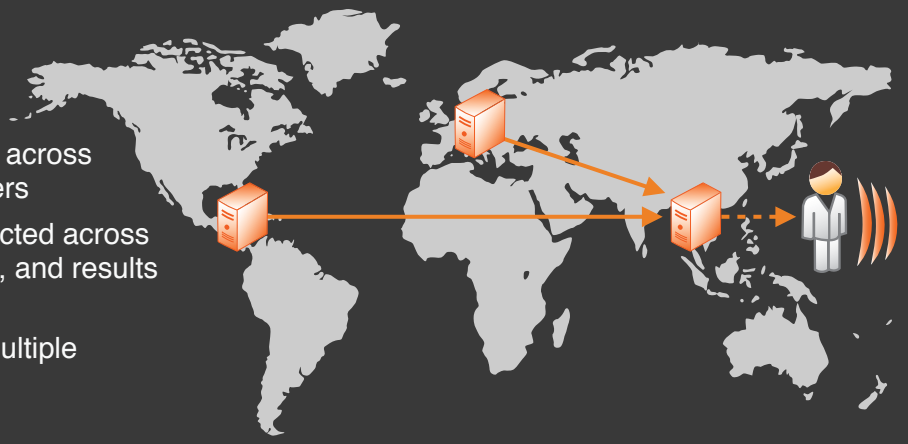


Axis: Distributed Systems

Distributed Services

Federation

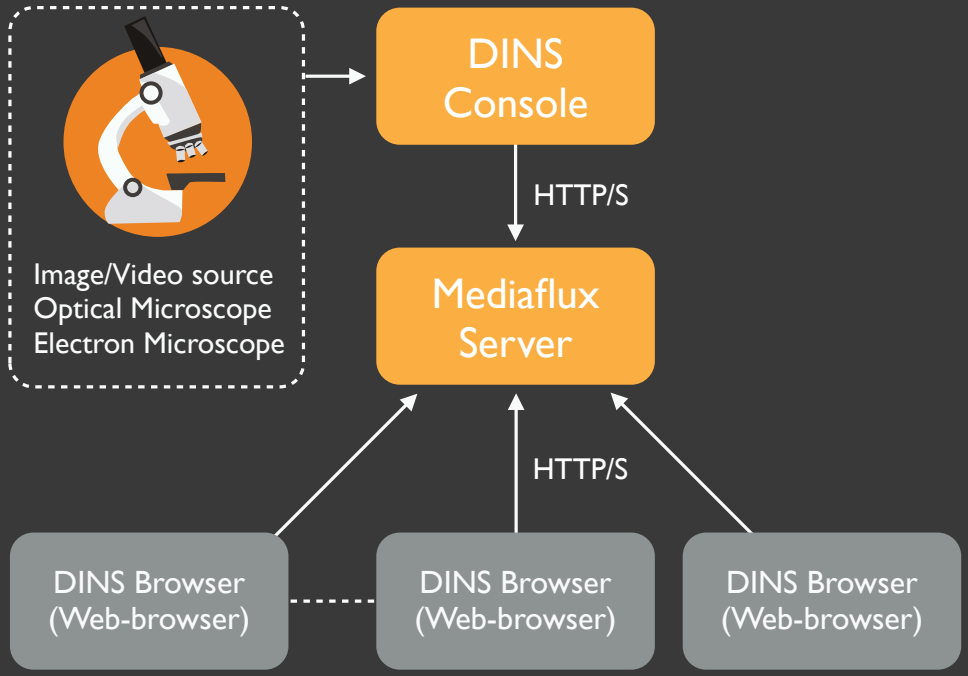
+ Searches may be federated across two or more Mediaflux servers
+ A federated search is conducted across all servers in the Federation, and results consolidated
+ Servers can participate in multiple Federations

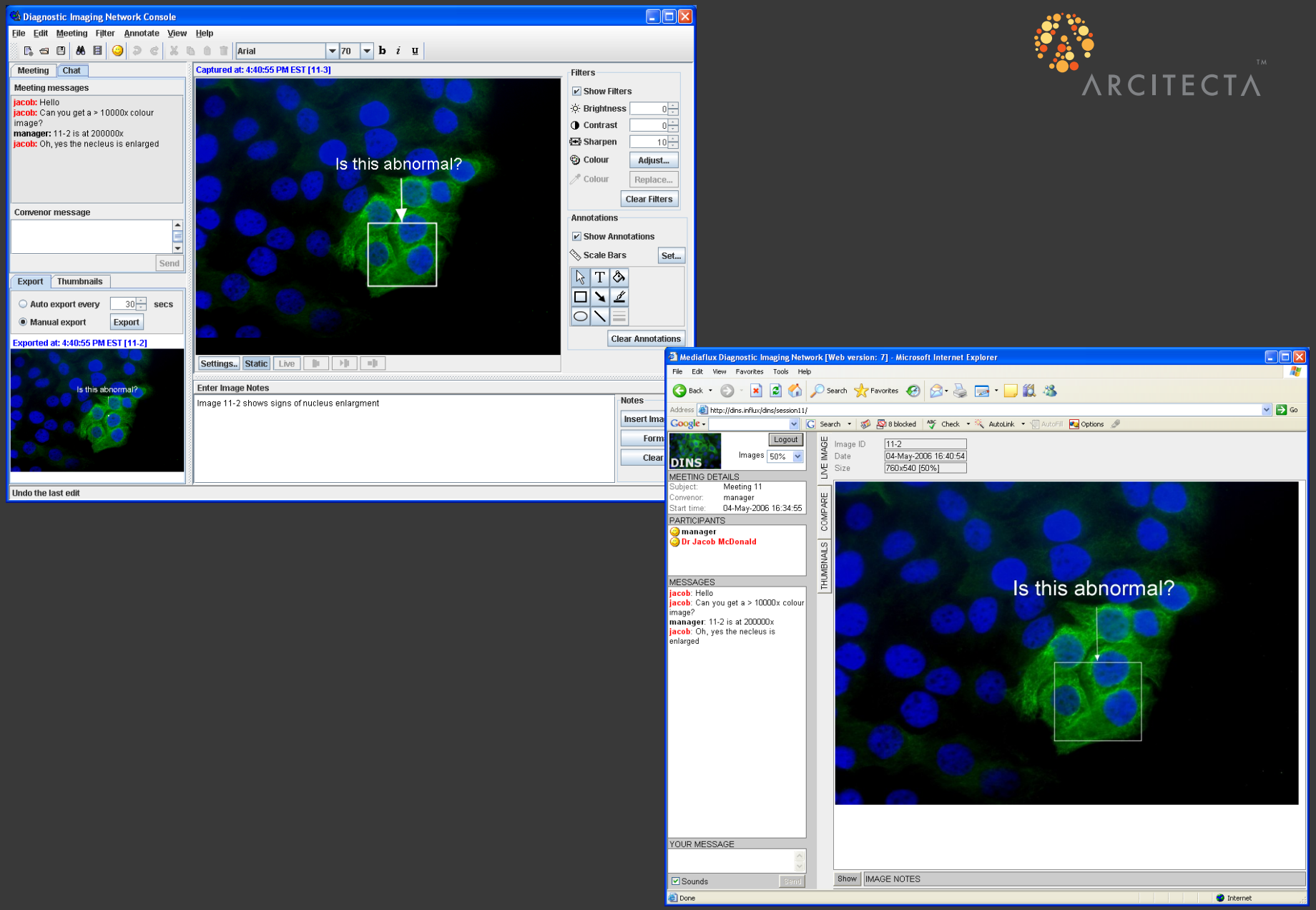


Customer Examples

Applied Mediaflux

CSIRO – AAHL Diagnostic Imaging System

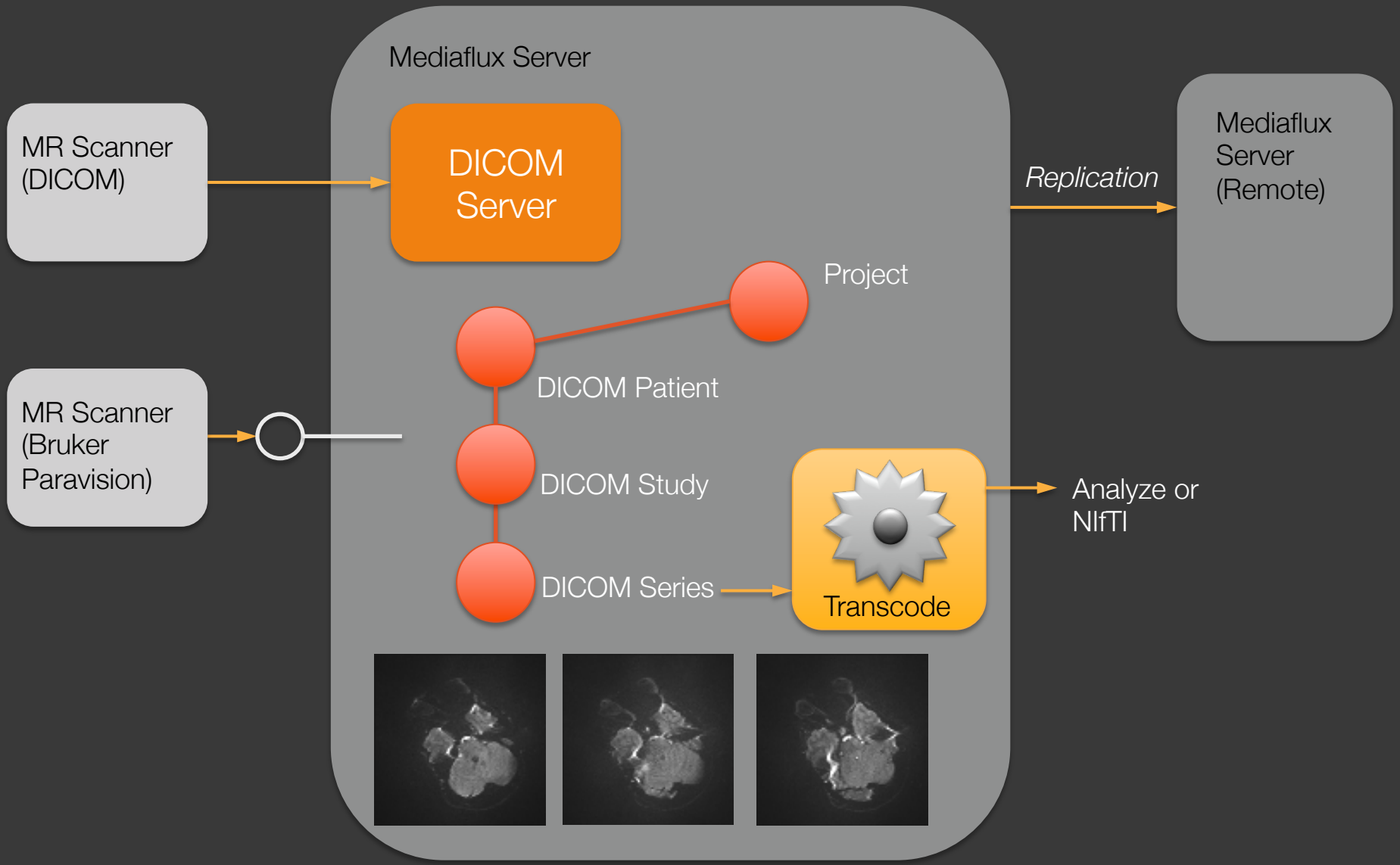


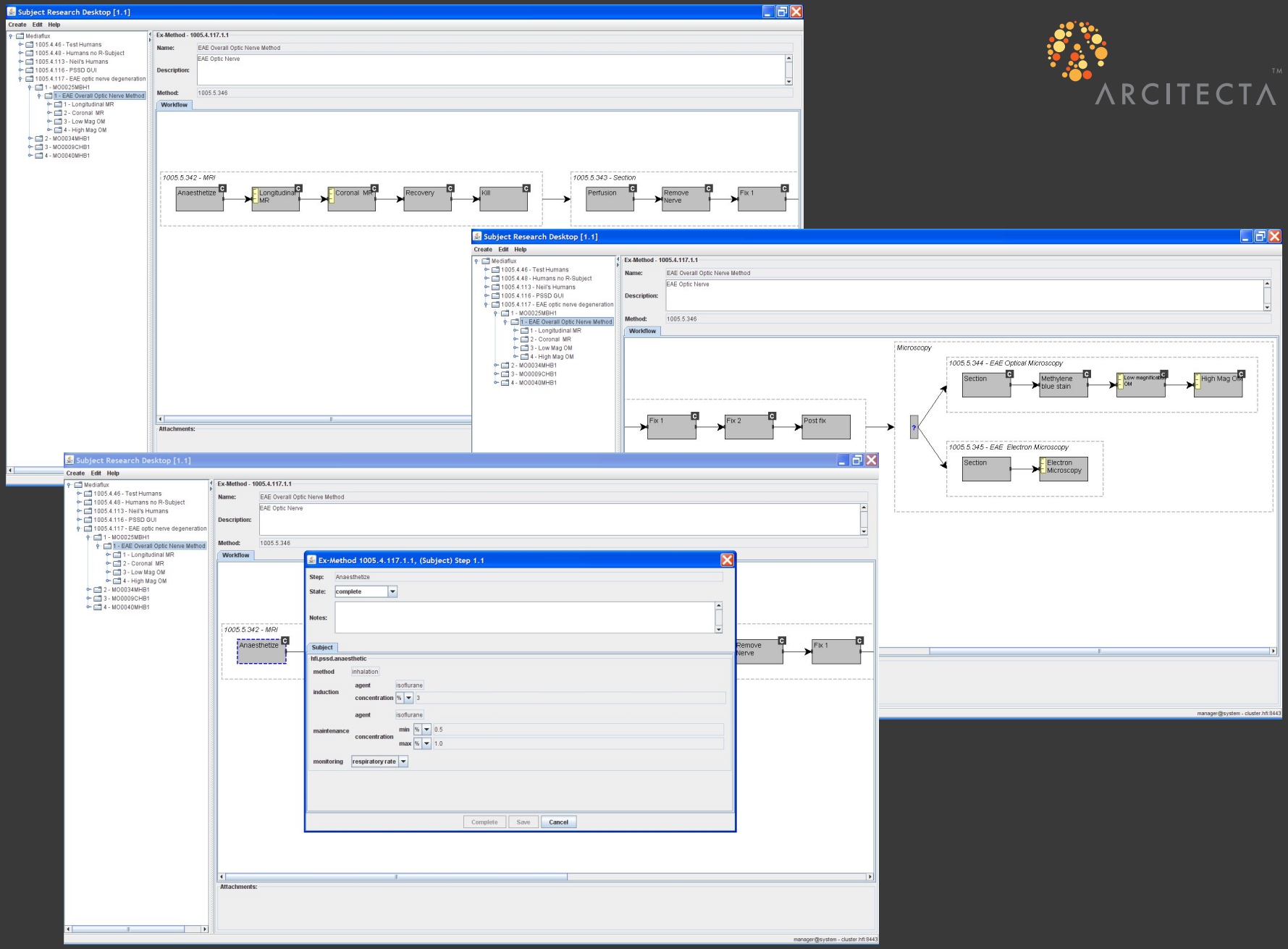

Diagnostic Imaging Network Console
File Edit Meeting Filter Annotate View Help
Meeting Chat
Meeting messages
jacob: Hello
jacob: Can you get a > 10000x colour image?
manager: 11-2 is at 200000x
jacob: Oh, yes the necleus is enlarged
Convenor message
Send
Export Thumbnails
Auto export every 30 secs
Manual export Export
Exported at: 4:40:55 PM EST [11-2]
Is this abnormal?
Captured at: 4:40:55 PM EST [11-3]
Is this abnormal?
Settings.. Static Live
Filters
Show Filters
Brightness
Contrast
Sharpen
Colour Adjust...
Colour Replace...
Clear Filters
Annotations
Show Annotations
Scale Bars Set...
Clear Annotations
Enter Image Notes
Image 11-2 shows signs of nucleus enlargment
Notes
Insert Ima
Form
Clear
Undo the last edit
Mediaflux Diagnostic Imaging Network [Web version: 7] - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites
Address http://dins.influx/dins/session11/
Google Search 8 blocked Check AutoLink AutoFill Options
DINS
Logout
Images 50%
MEETING DETAILS
Subject: Meeting 11
Convenor: manager
Start time: 04-May-2006 16:34:55
PARTICIPANTS
manager
Dr Jacob McDonald
MESSAGES
jacob: Hello
jacob: Can you get a > 10000x colour image?
manager: 11-2 is at 200000x
jacob: Oh, yes the necleus is enlarged
YOUR MESSAGE
Sounds Send
LIVE IMAGE
COMPARE
THUMBNAILS
Image ID 11-2
Date 04-May-2006 16:40:54
Size 760x540 [50%]
Is this abnormal?
Show IMAGE NOTES
Done
Internet

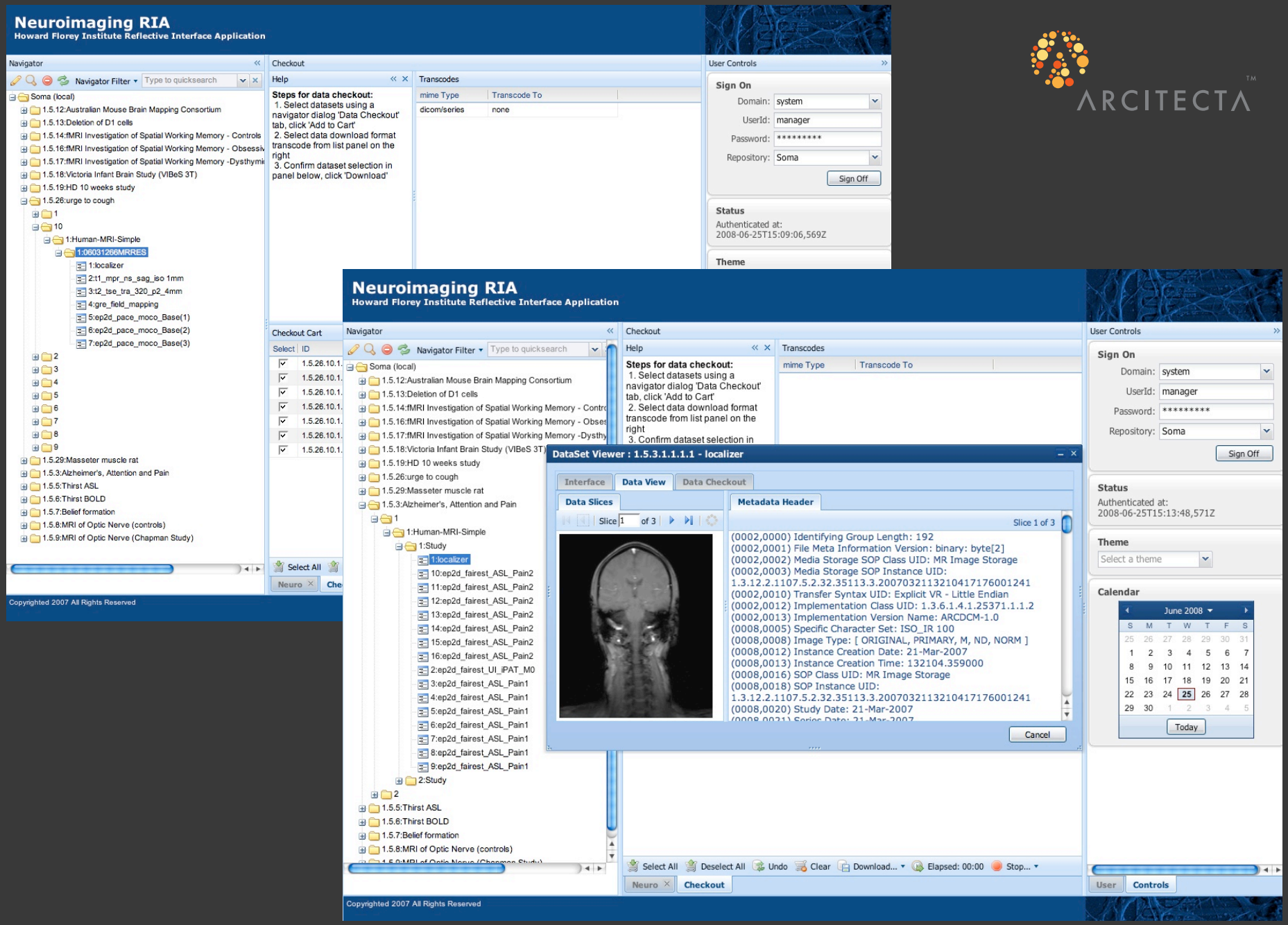
ARCITECTA™

University of Melbourne
Center for Neuroscience

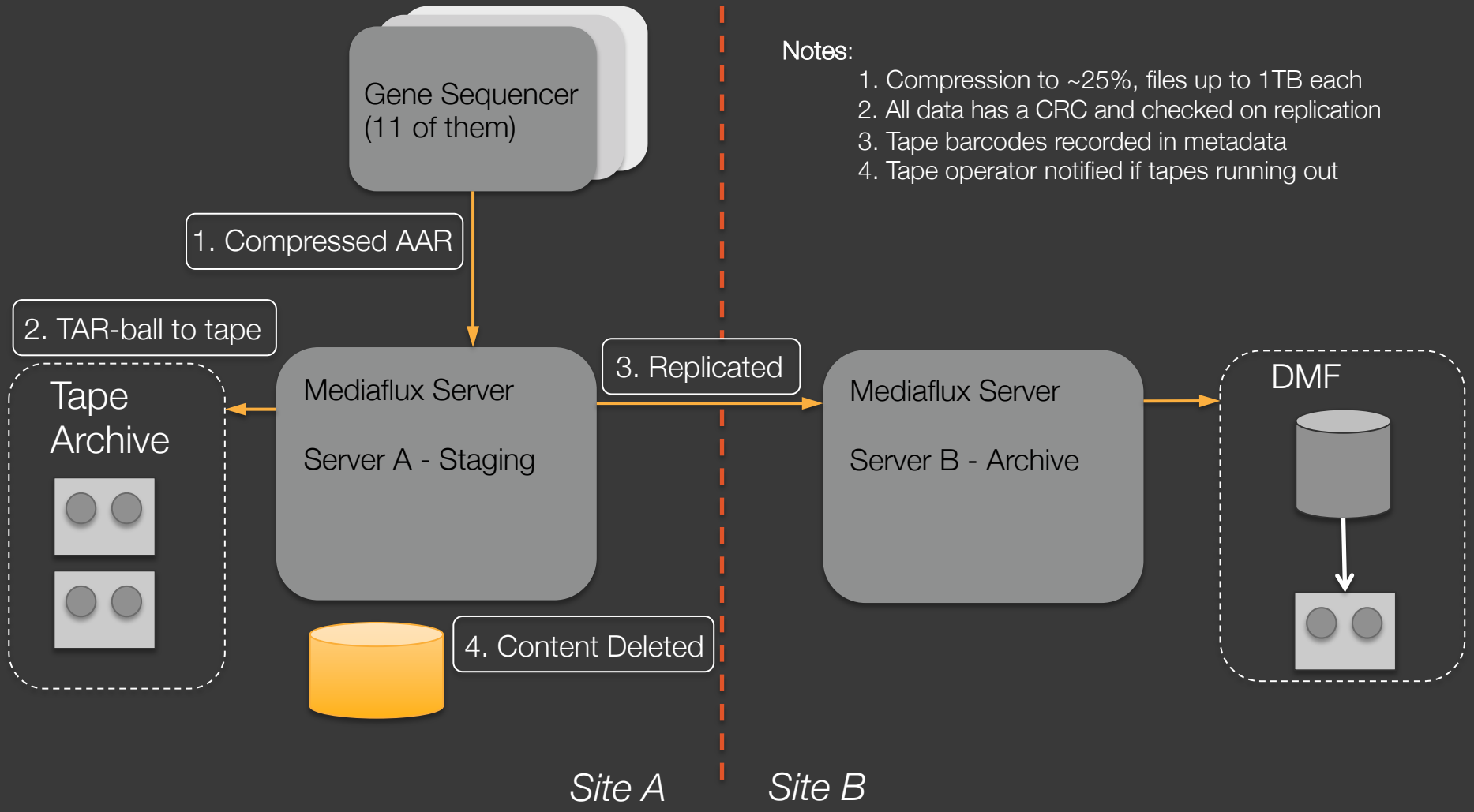
ARCITECTA™




Subject Research Desktop [1.1]
Create Edit Help
Mediaflux
1005.4.46 - Test Humans
1005.4.48 - Humans no R-Subject
1005.4.113 - Neil's Humans
1005.4.116 - PSSD GUI
1005.4.117 - EAE optic nerve degeneration
1 - MO0025MBH1
1 - EAE Overall Optic Nerve Method
1 - Longitudinal MR
2 - Coronal MR
3 - Low Mag OM
4 - High Mag OM
2 - MO0034MHB1
3 - MO0009CHB1
4 - MO0040MHB1
Ex-Method - 1005.4.117.1.1
Name:
EAE Overall Optic Nerve Method
Description:
EAE Optic Nerve
Method:
1005.5.346
Workflow
1005.5.342 - MRI
Anaesthetize
Longitudinal MR
Coronal MR
Recovery
Kill
1005.5.343 - Section
Perfusion
Remove Nerve
Fix 1
Fix 2
Post fix
Microscopy
1005.5.344 - EAE Optical Microscopy
Section
Methylene blue stain
Low magnification OM
High Mag OM
1005.5.345 - EAE Electron Microscopy
Electron Microscopy
Attachments:
Ex-Method 1005.4.117.1.1, (Subject) Step 1.1
Step:
State:
complete
Notes:
Subject
hfi.pssd.anaesthetic
method
inhalation
induction
agent
isoflurane
concentration
maintenance
monitoring
respiratory rate
Complete
Save
Cancel
manager@system - cluster.hfi:8443


Neuroimaging RIA
Howard Florey Institute Reflective Interface Application
Navigator
Checkout
Help
Transcodes
mime Type
Transcode To
dicom/series
none
Steps for data checkout:
1. Select datasets using a navigator dialog 'Data Checkout' tab, click 'Add to Cart'
2. Select data download format transcode from list panel on the right
3. Confirm dataset selection in panel below, click 'Download'
User Controls
Sign On
Domain: system
UserId: manager
Password: *********
Repository: Soma
Sign Off
Status
Authenticated at:
2008-06-25T15:09:06,569Z
Theme
Soma (local)
1.5.12:Australian Mouse Brain Mapping Consortium
1.5.13:Deletion of D1 cells
1.5.14:fMRI Investigation of Spatial Working Memory - Controls
1.5.16:fMRI Investigation of Spatial Working Memory - Obsessiv
1.5.17:fMRI Investigation of Spatial Working Memory -Dysthymi
1.5.18:Victoria Infant Brain Study (VIBeS 3T)
1.5.19:HD 10 weeks study
1.5.26:urge to cough
1
10
1:Human-MRI-Simple
1:06031266MRRES
1:localizer
2:t1_mpr_ns_sag_iso 1mm
3:t2_tse_tra_320_p2_4mm
4:gre_field_mapping
5:ep2d_pace_moco_Base(1)
6:ep2d_pace_moco_Base(2)
7:ep2d_pace_moco_Base(3)
2
3
4
5
6
7
8
9
1.5.29:Masseter muscle rat
1.5.3:Alzheimer's, Attention and Pain
1.5.5:Thirst ASL
1.5.6:Thirst BOLD
1.5.7:Belief formation
1.5.8:MRI of Optic Nerve (controls)
1.5.9:MRI of Optic Nerve (Chapman Study)
Copyrighted 2007 All Rights Reserved
Checkout Cart
Select
ID
Select All
Neuro
Navigator Filter
Type to quicksearch
1.5.26.10.1.
1.5.26.10.1.
1.5.26.10.1.
1.5.26.10.1.
1.5.26.10.1.
1.5.26.10.1.
1.5.26.10.1.
Neuroimaging RIA
Howard Florey Institute Reflective Interface Application
Navigator
Checkout
Help
Transcodes
mime Type
Transcode To
Steps for data checkout:
1. Select datasets using a navigator dialog 'Data Checkout' tab, click 'Add to Cart'
2. Select data download format transcode from list panel on the right
3. Confirm dataset selection in
User Controls
Sign On
Domain: system
UserId: manager
Password: *********
Repository: Soma
Sign Off
Status
Authenticated at:
2008-06-25T15:13:48,571Z
Theme
Select a theme
Calendar
June 2008
S M T W T F S
25 26 27 28 29 30 31
1 2 3 4 5 6 7
8 9 10 11 12 13 14
15 16 17 18 19 20 21
22 23 24 25 26 27 28
29 30 1 2 3 4 5
Today
Soma (local)
1.5.12:Australian Mouse Brain Mapping Consortium
1.5.13:Deletion of D1 cells
1.5.14:fMRI Investigation of Spatial Working Memory - Contro
1.5.16:fMRI Investigation of Spatial Working Memory - Obses
1.5.17:fMRI Investigation of Spatial Working Memory -Dysthy
1.5.18:Victoria Infant Brain Study (VIBeS 3T)
1.5.19:HD 10 weeks study
1.5.26:urge to cough
1.5.29:Masseter muscle rat
1.5.3:Alzheimer's, Attention and Pain
1
1:Human-MRI-Simple
1:Study
1:localizer
10:ep2d_fairest_ASL_Pain2
11:ep2d_fairest_ASL_Pain2
12:ep2d_fairest_ASL_Pain2
13:ep2d_fairest_ASL_Pain2
14:ep2d_fairest_ASL_Pain2
15:ep2d_fairest_ASL_Pain2
16:ep2d_fairest_ASL_Pain2
2:ep2d_fairest_UI_iPAT_M0
3:ep2d_fairest_ASL_Pain1
4:ep2d_fairest_ASL_Pain1
5:ep2d_fairest_ASL_Pain1
6:ep2d_fairest_ASL_Pain1
7:ep2d_fairest_ASL_Pain1
8:ep2d_fairest_ASL_Pain1
9:ep2d_fairest_ASL_Pain1
2:Study
2
1.5.5:Thirst ASL
1.5.6:Thirst BOLD
1.5.7:Belief formation
1.5.8:MRI of Optic Nerve (controls)
DataSet Viewer : 1.5.3.1.1.1.1 - localizer
Interface
Data View
Data Checkout
Data Slices
Slice 1 of 3
Metadata Header
Slice 1 of 3
(0002,0000) Identifying Group Length: 192
(0002,0001) File Meta Information Version: binary: byte[2]
(0002,0002) Media Storage SOP Class UID: MR Image Storage
(0002,0003) Media Storage SOP Instance UID:
1.3.12.2.1107.5.2.32.35113.3.2007032113210417176001241
(0002,0010) Transfer Syntax UID: Explicit VR - Little Endian
(0002,0012) Implementation Class UID: 1.3.6.1.4.1.25371.1.1.2
(0002,0013) Implementation Version Name: ARCDCM-1.0
(0008,0005) Specific Character Set: ISO_IR 100
(0008,0008) Image Type: [ORIGINAL, PRIMARY, M, ND, NORM]
(0008,0012) Instance Creation Date: 21-Mar-2007
(0008,0013) Instance Creation Time: 132104.359000
(0008,0016) SOP Class UID: MR Image Storage
(0008,0018) SOP Instance UID:
1.3.12.2.1107.5.2.32.35113.3.2007032113210417176001241
(0008,0020) Study Date: 21-Mar-2007
Cancel
Select All
Deselect All
Undo
Clear
Download...
Elapsed: 00:00
Stop...
Neuro
Checkout
User
Controls
Copyrighted 2007 All Rights Reserved

University of Queensland
Institute of Molecular Bioscience
Ingestion

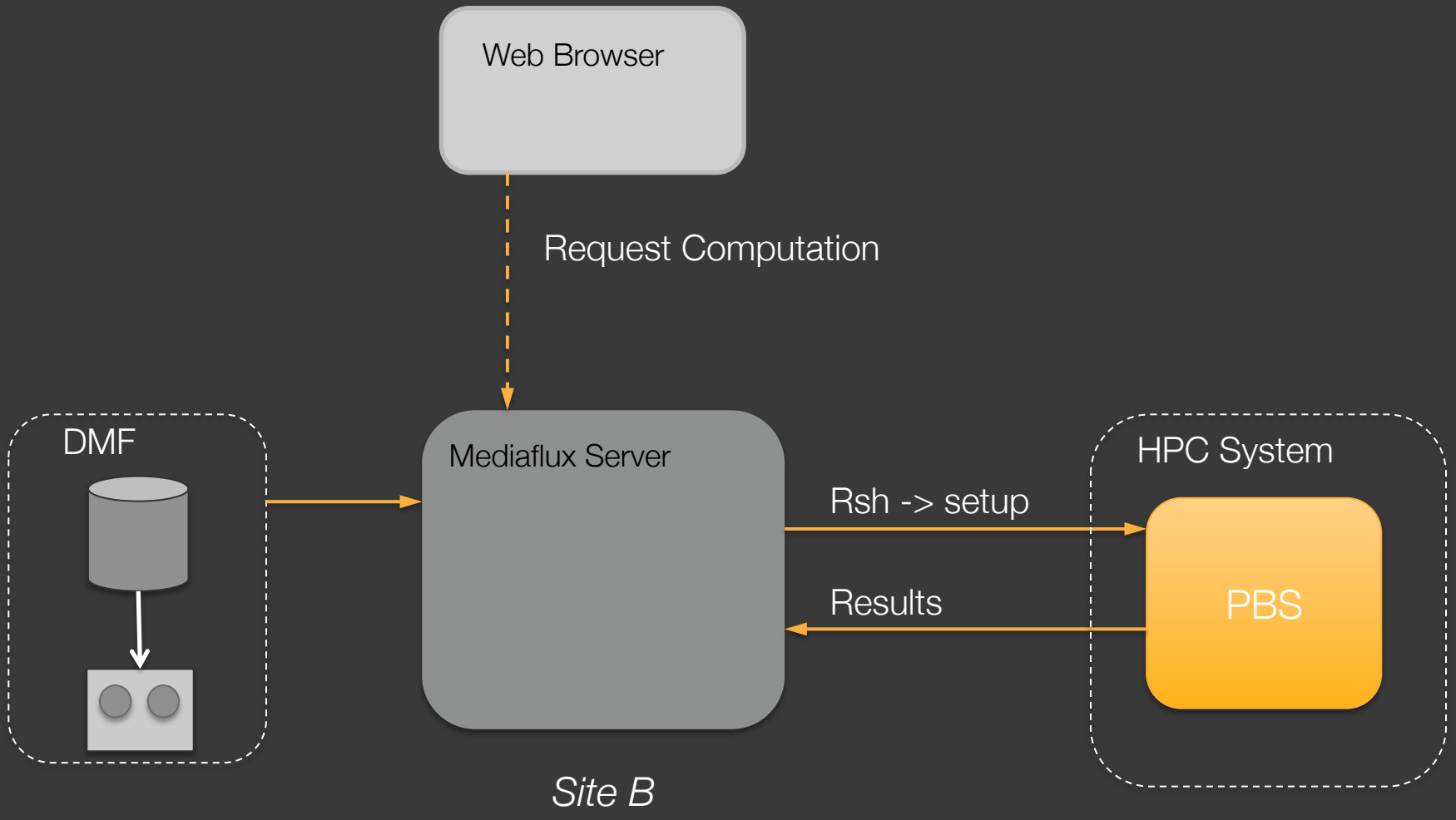


Notes:

1. Compression to ~25%, files up to 1TB each
2. All data has a CRC and checked on replication
3. Tape barcodes recorded in metadata
4. Tape operator notified if tapes running out

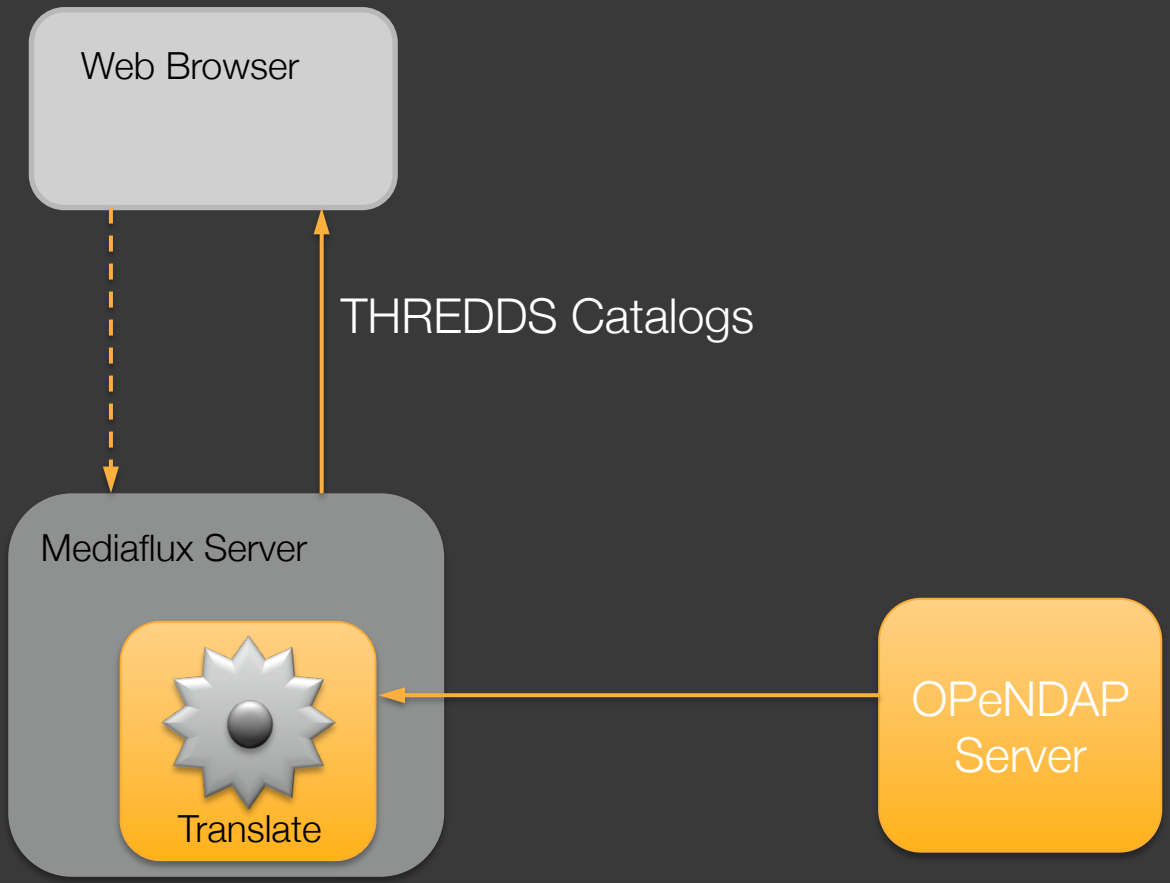
University of Queensland
Institute of Molecular Bioscience
Processing/Computation

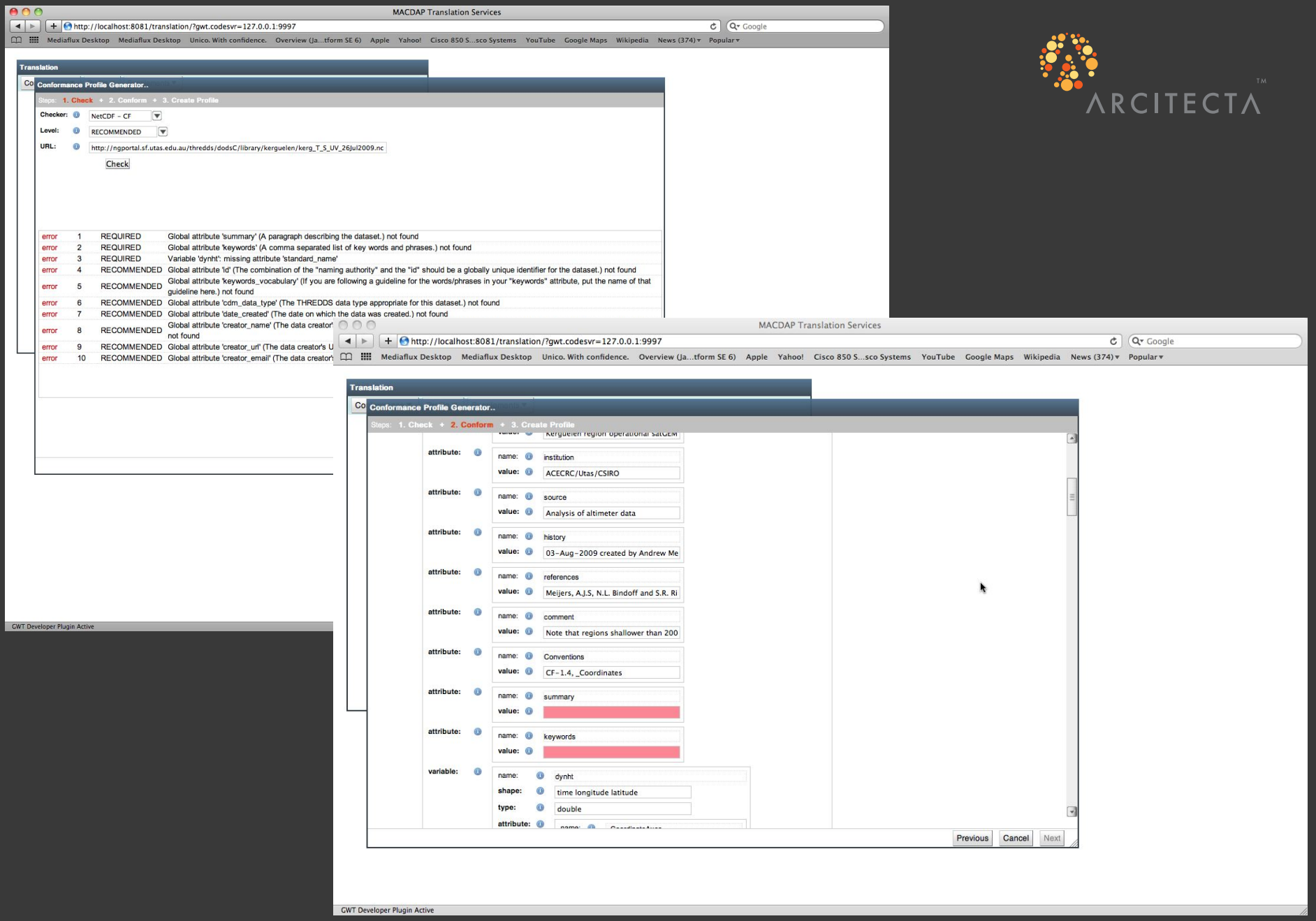
ARCITECTA™



ARCITECTA™

Translation Services



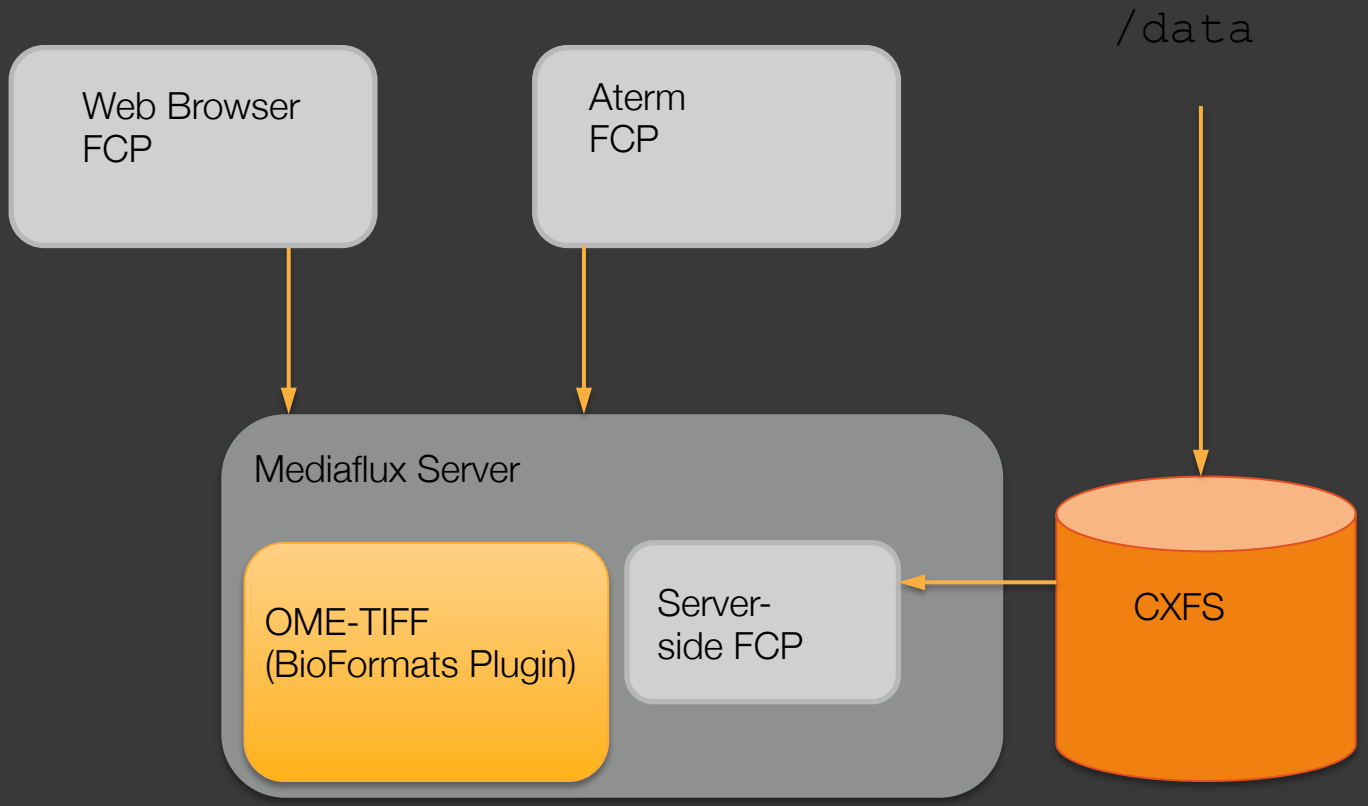

MACDAP Translation Services
http://localhost:8081/translation/?gwt.codesvr=127.0.0.1:9997
Translation
Conformance Profile Generator..
Steps: 1. Check + 2. Conform + 3. Create Profile
Checker: NetCDF - CF
Level: RECOMMENDED
URL: http://ngportal.sf.utas.edu.au/thredds/dodsC/library/kerguelen/kerg_T_S_UV_26Jul2009.nc
Check
error 1 REQUIRED Global attribute 'summary' (A paragraph describing the dataset.) not found
error 2 REQUIRED Global attribute 'keywords' (A comma separated list of key words and phrases.) not found
error 3 REQUIRED Variable 'dynht': missing attribute 'standard_name'
error 4 RECOMMENDED Global attribute 'id' (The combination of the "naming authority" and the "id" should be a globally unique identifier for the dataset.) not found
error 5 RECOMMENDED Global attribute 'keywords_vocabulary' (If you are following a guideline for the words/phrases in your "keywords" attribute, put the name of that guideline here.) not found
error 6 RECOMMENDED Global attribute 'cdm_data_type' (The THREDDS data type appropriate for this dataset.) not found
error 7 RECOMMENDED Global attribute 'date_created' (The date on which the data was created.) not found
error 8 RECOMMENDED Global attribute 'creator_name' (The data creator
not found
error 9 RECOMMENDED Global attribute 'creator_url' (The data creator's U
error 10 RECOMMENDED Global attribute 'creator_email' (The data creator'
GWT Developer Plugin Active
MACDAP Translation Services
http://localhost:8081/translation/?gwt.codesvr=127.0.0.1:9997
Translation
Conformance Profile Generator..
Steps: 1. Check + 2. Conform + 3. Create Profile
attribute: name: institution value: ACECRC/Utas/CSIRO
attribute: name: source value: Analysis of altimeter data
attribute: name: history value: 03-Aug-2009 created by Andrew Me
attribute: name: references value: Meijers, A.J.S, N.L. Bindoff and S.R. Ri
attribute: name: comment value: Note that regions shallower than 200
attribute: name: Conventions value: CF-1.4, _Coordinates
attribute: name: summary value:
attribute: name: keywords value:
variable: name: dynht shape: time longitude latitude type: double attribute:
Previous
Cancel
Next
GWT Developer Plugin Active

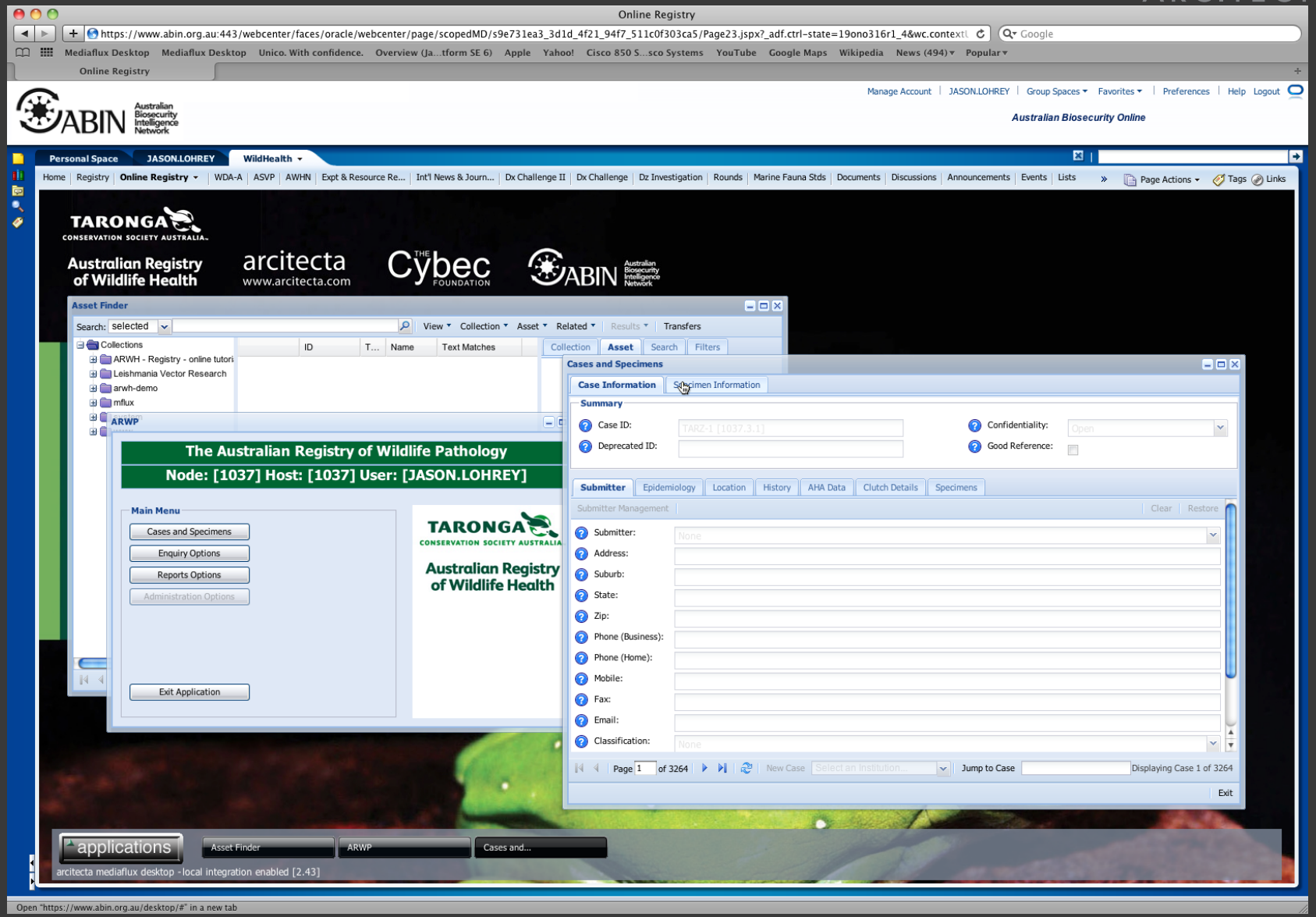
ARCITECTA™

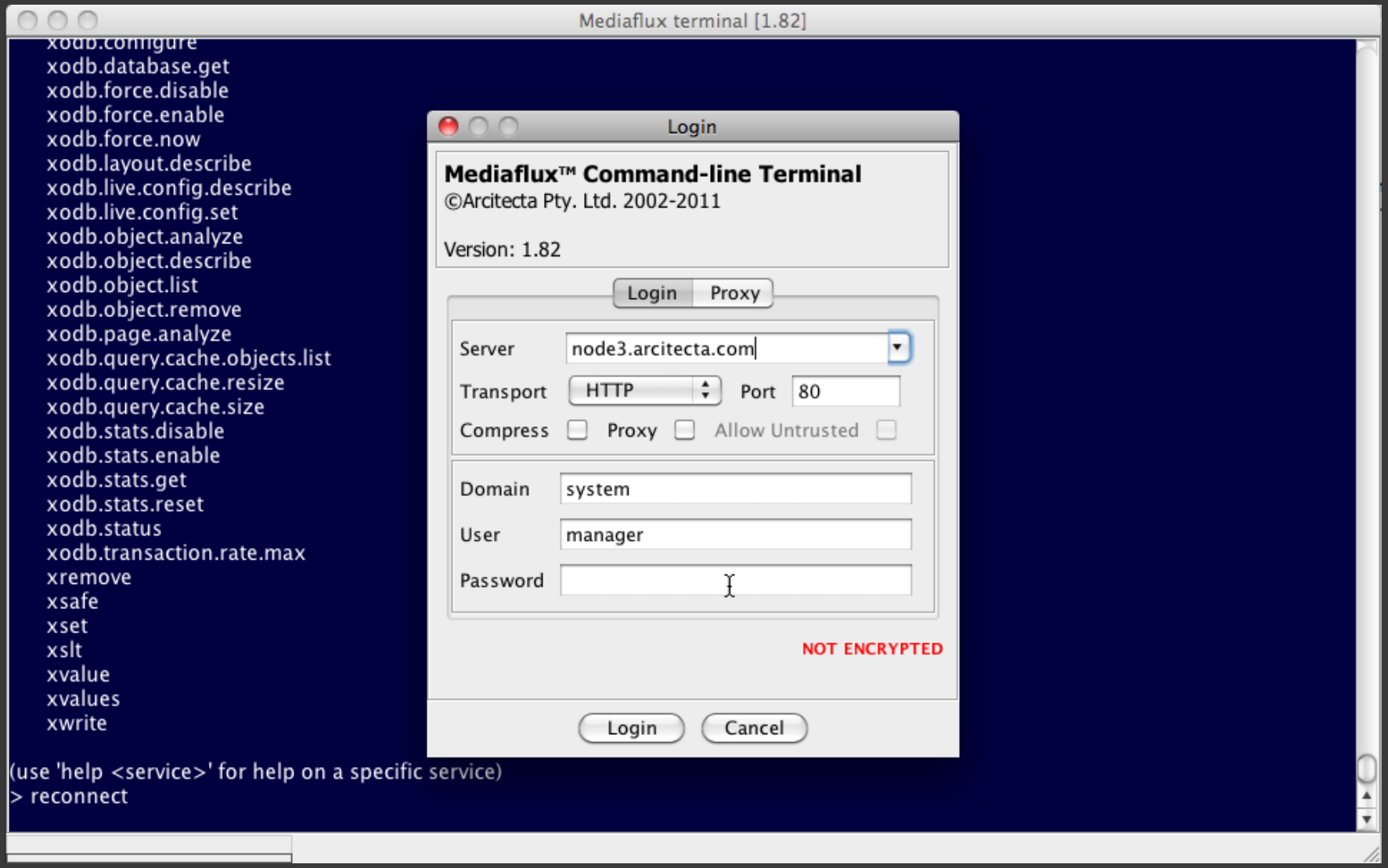
National University Of Singapore

ARCITECTA™

Microscopy Images




Online Registry
Australian Biosecurity Intelligence Network
Australian Biosecurity Online
Personal Space
JASON.LOHREY
WildHealth
TARONGA
CONSERVATION SOCIETY AUSTRALIA
Australian Registry of Wildlife Health
arcitecta
www.arcitecta.com
The Cybec Foundation
Asset Finder
Cases and Specimens
The Australian Registry of Wildlife Pathology
Node: [1037] Host: [1037] User: [JASON.LOHREY]
Main Menu
Cases and Specimens
Enquiry Options
Reports Options
Administration Options
Exit Application
applications
arcitecta mediaflux desktop -local integration enabled [2.43]


Mediaflux terminal [1.82]
xodb.configure
xodb.database.get
xodb.force.disable
xodb.force.enable
xodb.force.now
xodb.layout.describe
xodb.live.config.describe
xodb.live.config.set
xodb.object.analyze
xodb.object.describe
xodb.object.list
xodb.object.remove
xodb.page.analyze
xodb.query.cache.objects.list
xodb.query.cache.resize
xodb.query.cache.size
xodb.stats.disable
xodb.stats.enable
xodb.stats.get
xodb.stats.reset
xodb.status
xodb.transaction.rate.max
xremove
xsafe
xset
xslt
xvalue
xvalues
xwrite
(use 'help <service>' for help on a specific service)
> reconnect
Login
Mediaflux™ Command-line Terminal
©Arcitecta Pty. Ltd. 2002-2011
Version: 1.82
Login
Proxy
Server
node3.arcitecta.com
Transport
HTTP
Port
80
Compress
Proxy
Allow Untrusted
Domain
system
User
manager
Password
NOT ENCRYPTED
Login
Cancel

“Under the Hood”

Some of the Building Blocks ..

Ingesting

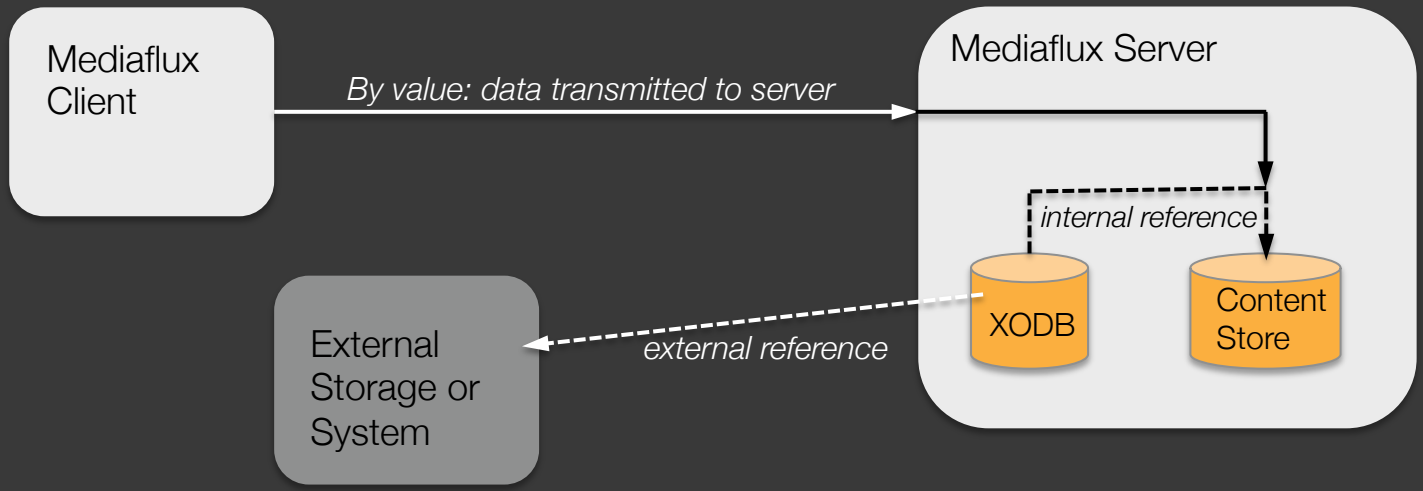
Methods for Getting Assets "In"

Ingest by "value":

- From a Mediaflux Client:
  - Browser (background applet, parallel I/O, AAR packaging)
  - Aterm
- From a custom application (Java, .NET)

Ingest by "reference":

- Data remains in place (server must have visibility to data)



Egesting

Methods for Getting Assets "Out"

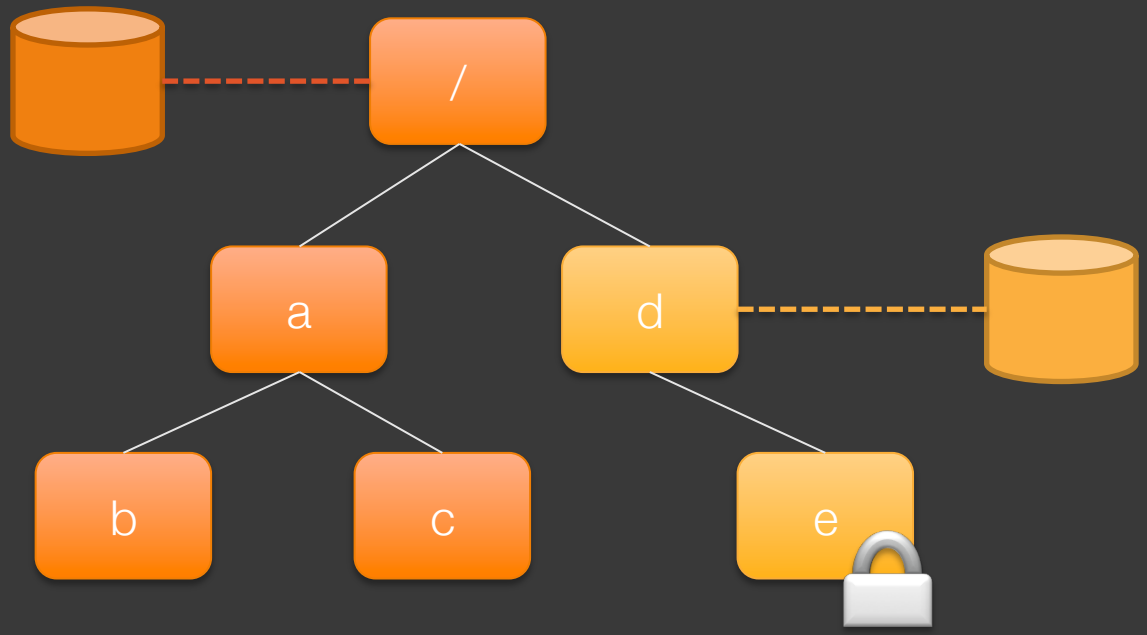
Egest by "value":

- By a Mediaflux Client:
  - Browser (background applet, parallel I/O, AAR unpacking)
  - Aterm
- From a custom application (Java, .NET)

The client need not know where the data comes from (in a federation)

Namespaces

- Assets are stored in a namespace
- Namespaces have associated storage (data stores)
- Access controls and default metadata (templates)



Data Stores

- The place where asset content is stored
- Extensible
- Support for:
  - Database (XODB)
  - File System
  - DMF (HSM)

Metadata

Information about.. information

- Information about an asset (that may not be in the content)
- Stored as binary XML
- Automatically generated:
  - Revision history (audit trail: who, when, what)
  - Plug-in meta data generation
- User generated
- Validated against a Mediaflux document definition
- Server-side manipulation & merging
- Unstructured and/or structured querying

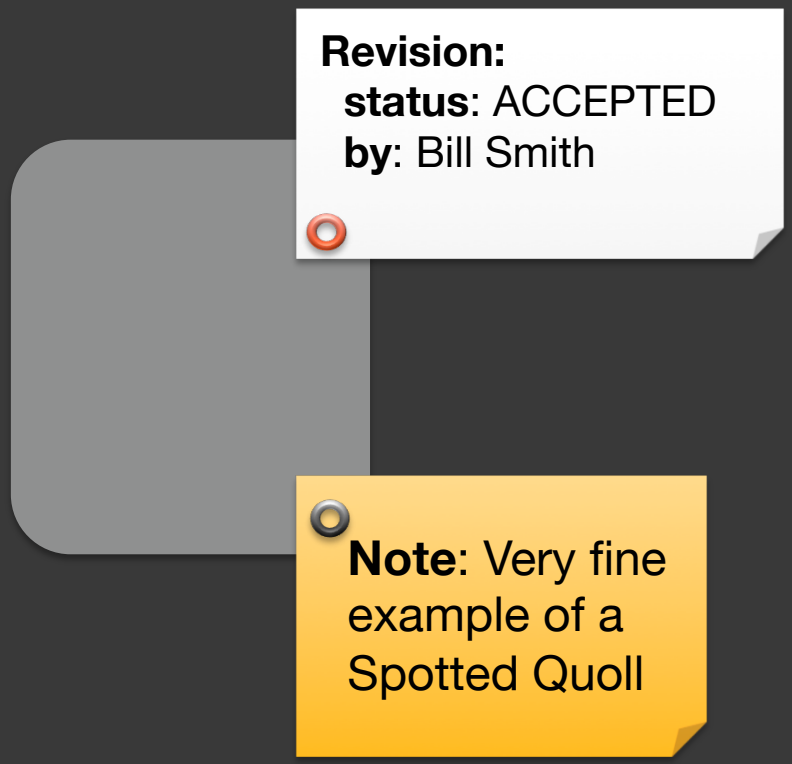
Applying Metadata

At Any Time..

- Metadata simply “attached” as required to an asset
- Can attach any number of different fragments of metadata

```
asset.query :where \
xpath(revision/status)='ACCEPTED'
```

```
asset.query :where \
text contains 'spotted quoll'
```



Evolving Metadata

At Any Time

- Who knows what metadata might be required in 2/5/10 .. years?
- Can evolve by:
  - Adding to existing definitions
  - Defining and “attaching” new definitions
- Evolve incrementally or by direct update

Revision:
status: enumeration {
ACCEPTED,
REJECTED
}, max-occurs 1

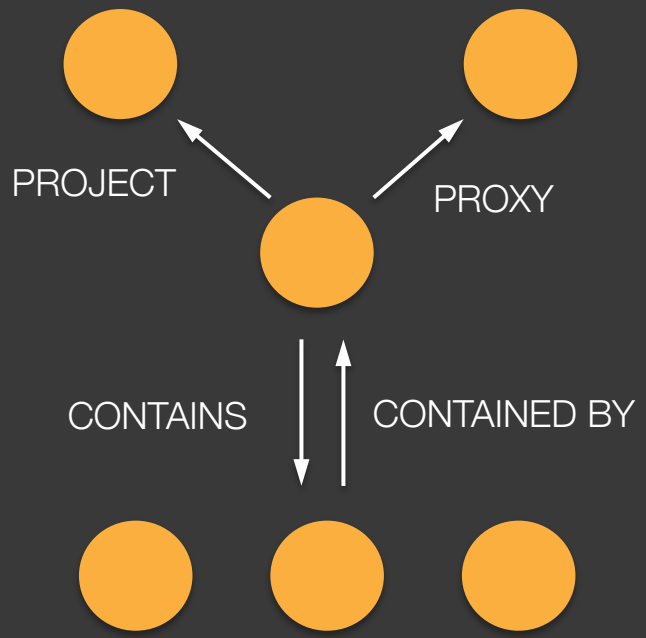
by: String, max-occurs 1

Revision:
status: enumeration {
ACCEPTED,
REJECTED
}, max-occurs 1

by: String, max-occurs 1
publish-date: Date, max-occurs 1

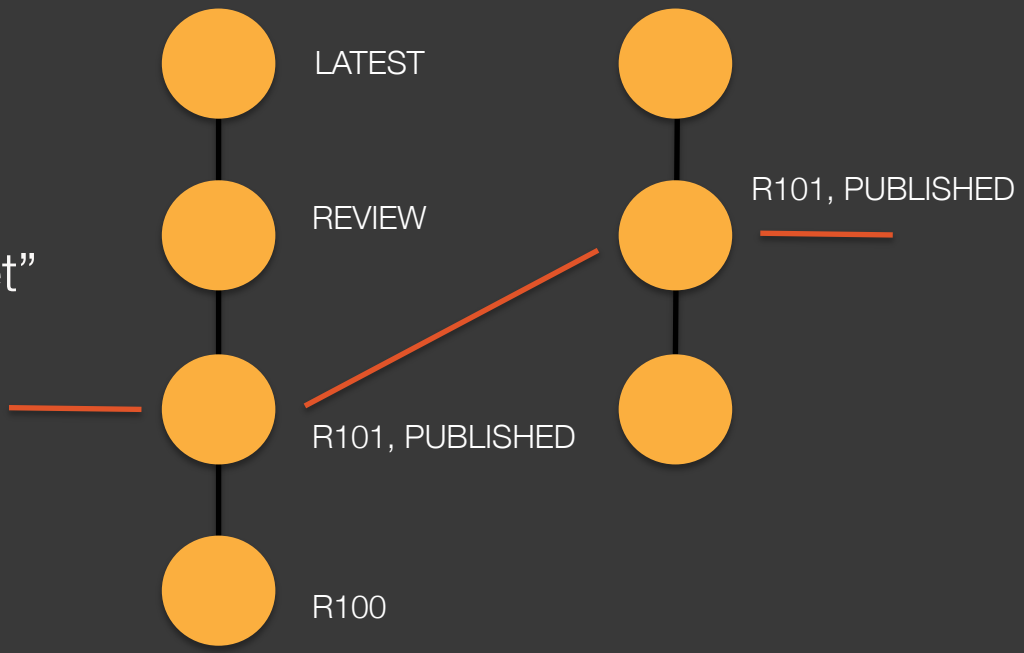
Relationships

- Established between assets
- Arbitrary definition:
  - Name
  - Description
  - Uni- or bi-directional
  - Cardinality of each arc
  - “Contains” or “associative” binding



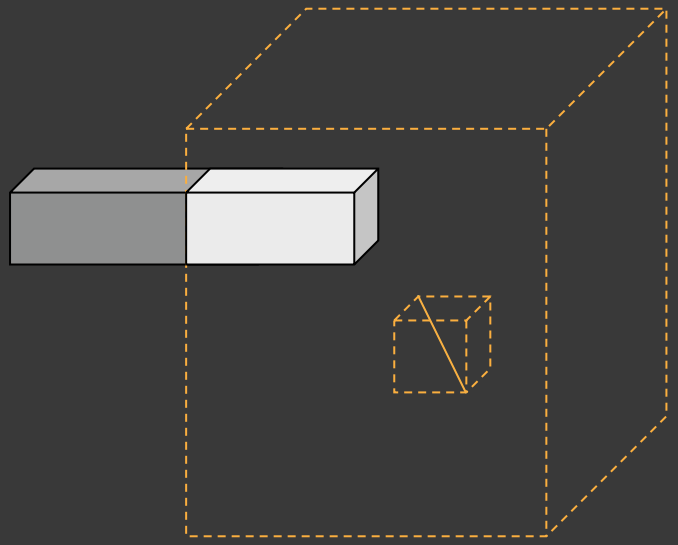
Labels

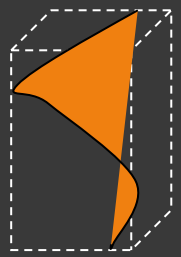
- Applied to asset versions
- Arbitrary definition:
  - Name
  - Cardinality
- Used to create a “matched set”



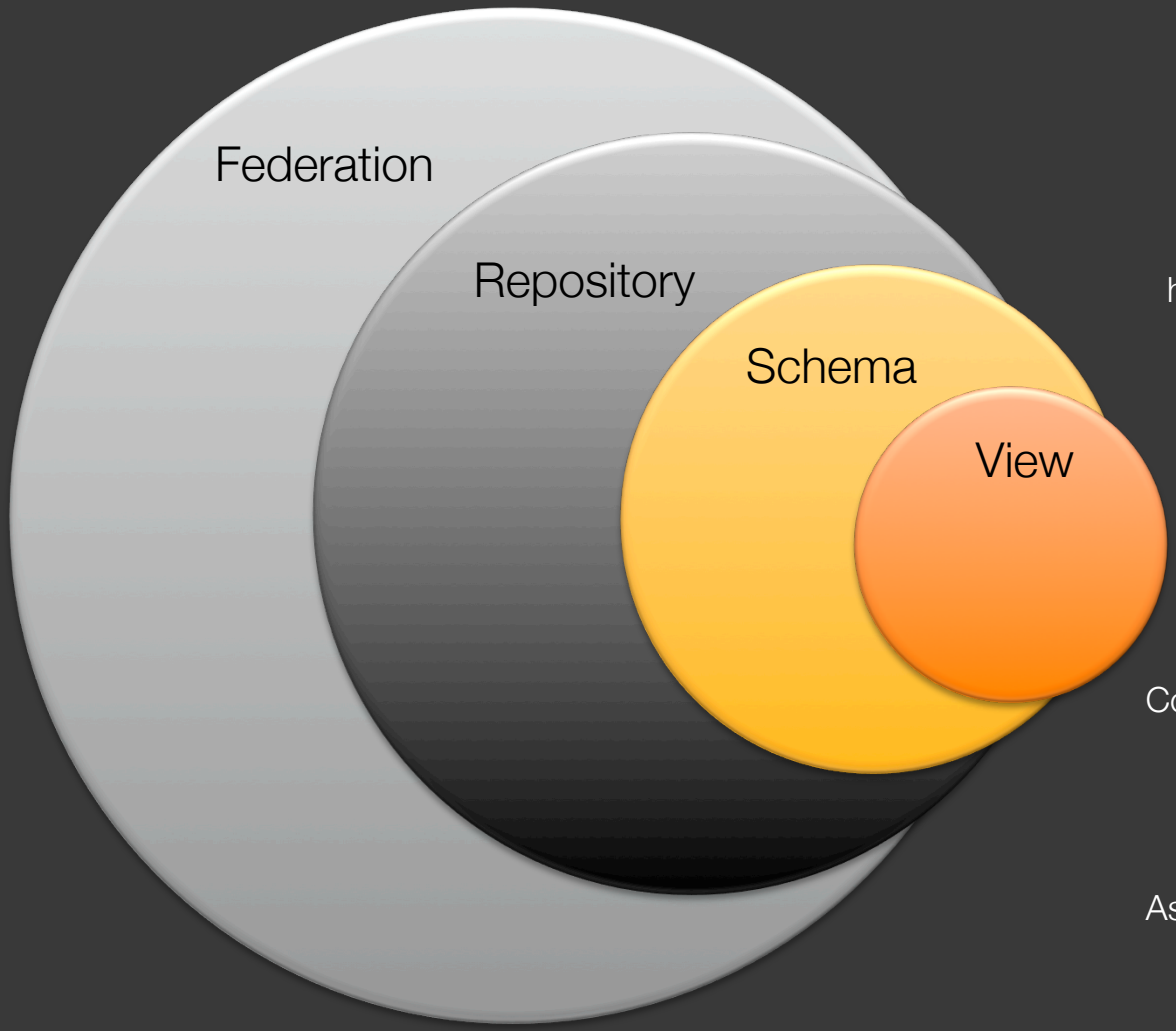
Finding Things

- Unstructured queries:
  - Meta text, content text, name text, annotation text, or all
- Structured queries:
  - Metadata specific, type, time, relationships
  - Locks, labels, collections, etc.
- Spatial queries:
  - Inside, Outside, Intersects
  - < N-D search criteria
  - Search by arbitrary polyline/polygon later
- Restricted to what you have access to
- Classification schemes
- Text suggestions





Views



http://public.abc.com

Constrained to:

- Namespace
- Labels
- Arbitrary Query

Associated with:

- Domain and/or user
- HTTP host or URL

Authentication

Identity Management

- Triplet: Domain/User/Password
  - Local
  - External (LDAP, Kerberos, Active Directory)
- SSO:
  - Oracle

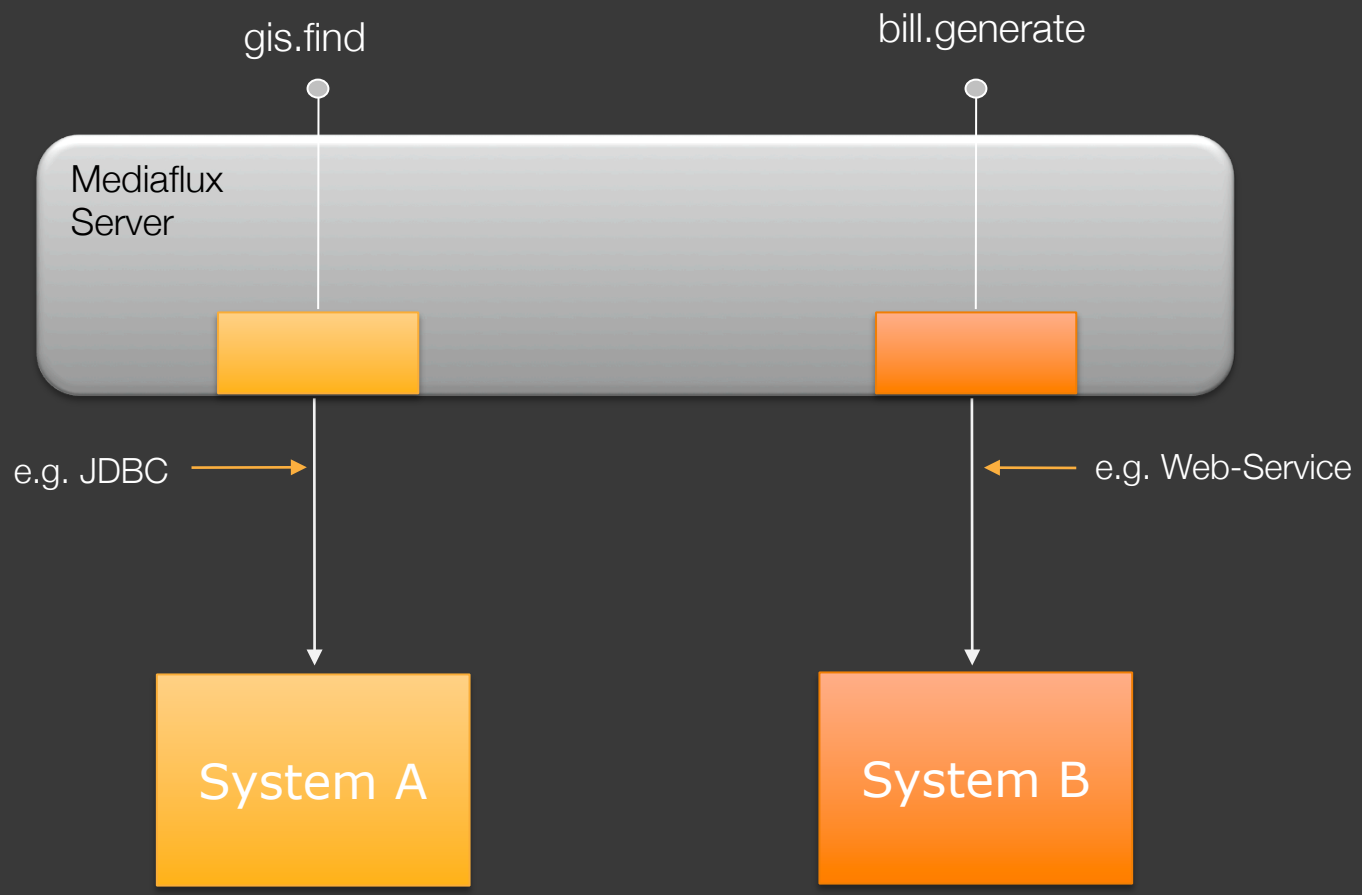
Authorization

Access Control

- "Actor" based (Domain, User, Role, View, Plugin, …)
- Hierarchical
- Control (access, update, publish, delete, …) for:
  - Namespaces
  - Assets
  - Services
  - Metadata
  - Schemas
  - Authentication domains
  - Views
  - Etc.

Integrating With Other Systems

Adding Services



AAR

Arcitecta Archive Format

- Up to 2^{63} byte total file size
- Up to 2^{63} bytes per entry
- Uncompressed, or Deflate compression
- Streamed or random access
- Created/Read by:
  - Aterm
  - Web-clients
  - Stand-alone AAR.jar tool

File System Compiler

Pattern Matcher for Files

File Systems are full of “patterns”. The file system compiler uses a profile to interpret the structure of the file system and package at ingestion time.

```
profile test.profile.two {
     construct text.file {
       match {
            file extension "txt"
         }

         logical type "application/gene-sequence"
       encapsulation type "text/plain"

         consume no
     }

     construct spatial.image {
       match {
            file name ignore case "spatialimage.jpg"
         }

       encapsulate as archive level 0

         logical type "application/my-type"
         encapsulation type "image/jpg"

         consume yes
     }
```

```
  construct esri.shape {
      match {
        group {
         file extension "shp"
         file extension "dbf"
         file extension "shx"
         optional file extension "prj"
          }
      }

      logical type "application/esri-shape"

   consume no
  }

construct GIS_FILES {
    match {
       file name "GIS_FILES" and
       directory contains construct esri.shape
    }

    encapsulate as archive
    logical type "application/gis"

    consume yes
}
```

ARCITECTA™

Hierarchical Content Store

SGI Data Migration Facility (DMF)

DMF

Higher Order Semantics

Mediaflux™ has specific support for DMF:

- Retrieving status to:
  - Inform end-users of potential delays
- Migrating data between tiers

These capabilities are exposed via services:

```
> asset.content.status :id 1
    :asset -id "1"
        :state "online”
> asset.content.migrate :id 1 :destination offline
```

DMF

Higher Order Semantics

These services can be combined with a query:

- To utilize arbitrary metadata to control migration

```
asset.query :where \
"geoshape intersects rectangle [(16,143),(16.56,144.89)]" \
:action pipe \
:service -name asset.content.migrate
< :destination online >
```

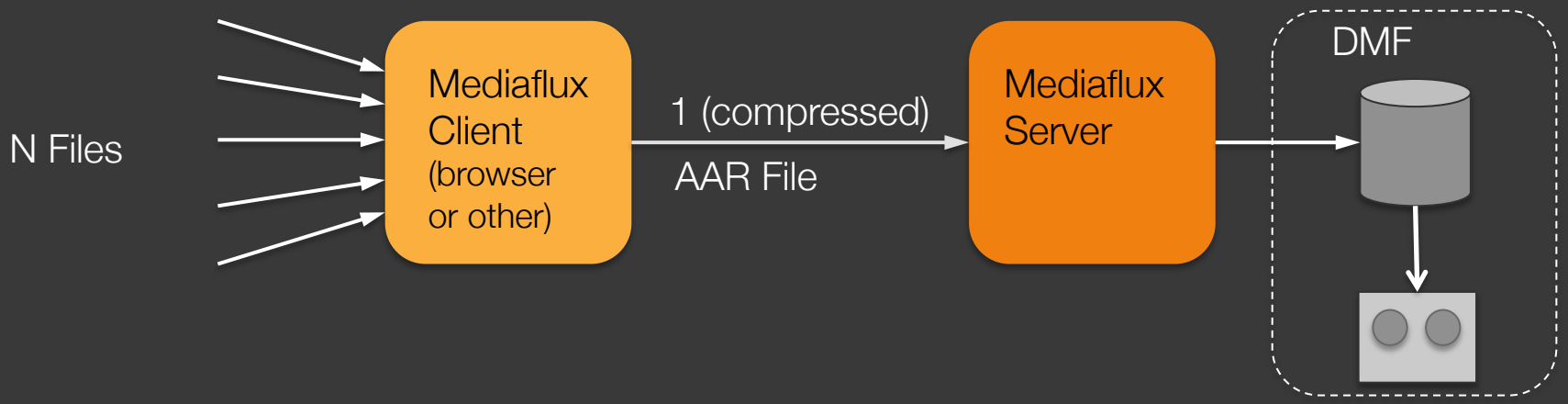
This allows migration policies based on *any* metadata (including information extracted from the content).

As with all Mediaflux services, these can be scripted and automated with the scheduler.

DMF
File Coalescing

Mediaflux can coalesce many files into a single Arcitecta Archive (AAR), and automatically re-inflate on extraction. This:

- Significantly reduces the number (and potentially size) of files managed by DMF
- Significantly reduces the number of files transmitted versus other mechanisms such as Windows Explorer, etc.



ARCITECTA™

The Future

The Return of The File System..

File Systems are Dead?

Not Quite..

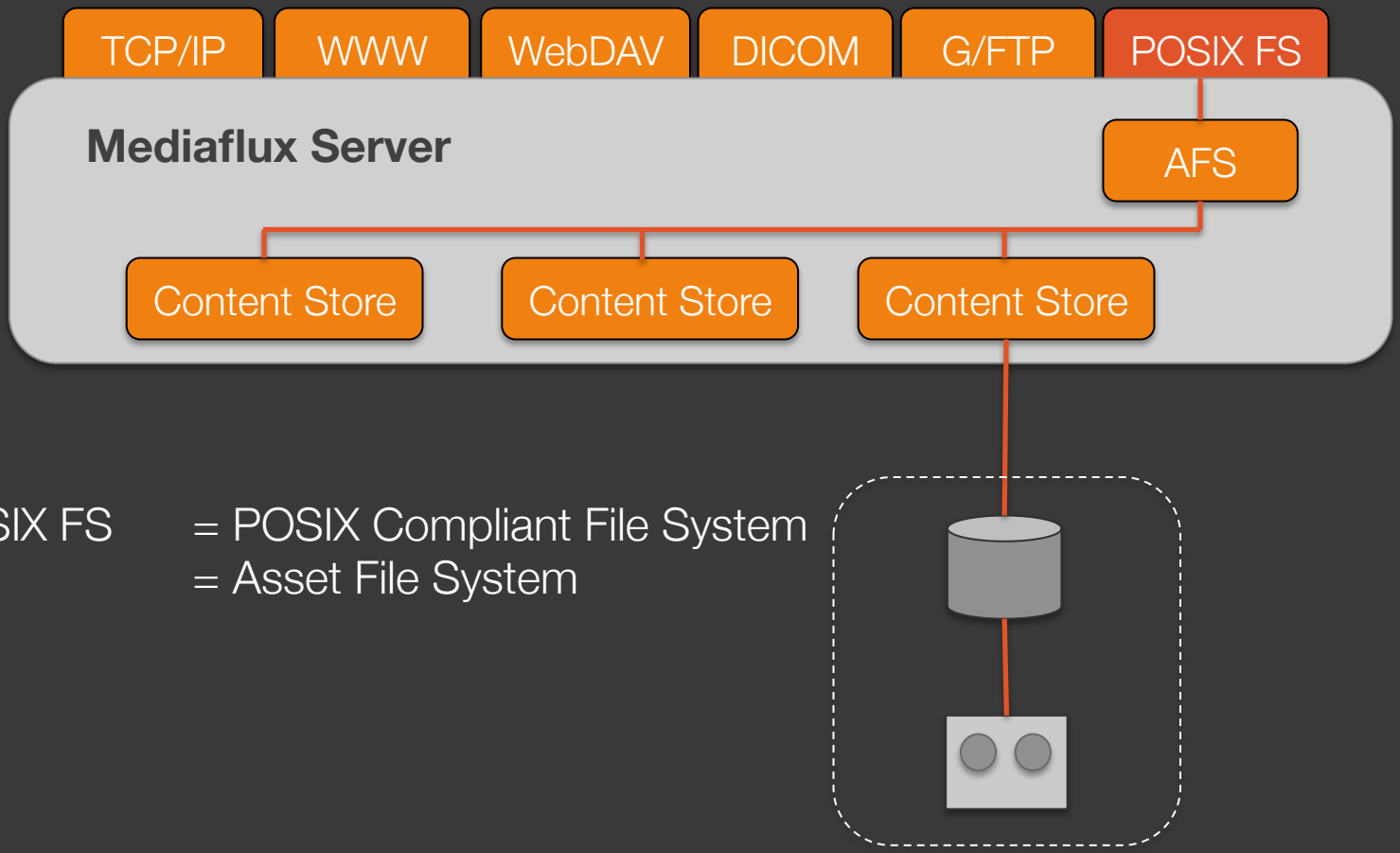
Arcitecta and the SGI DMF and CXFS teams are working closely to create an “archive file system” with Mediaflux and DMF.

Mediaflux will provide a POSIX compliant file system interface:

- Allowing data to be written and read directly to/from Mediaflux using standard applications without the need for subsequent ingestion
- Mediaflux can directly drive DMF based on arbitrary metadata based policies. E.g.:
  - Data is immediately migrated, but must remain on disk until analyzed
  - Scheduled retrieval
  - Etc.
- Operations are journalled, removing the need for many file-system scanning operations, significantly improving scalability.

File Systems are Dead?

Not Quite..



POSIX FS = POSIX Compliant File System
AFS = Asset File System

File Systems are Dead?

The Underlying Storage..

Can be:

- A regular file system
- A clustered file system (CXFS)
- A HSM managed file system (DMF)

File Systems are Dead?

Now For the (really) Fun Stuff..

The Mediaflux File System will:

- Be able to change “shape” depending on:
  - Your authority
  - View configuration (e.g. only PUBLISHED data)
- Provide data in transcoded form. E.g.
  - Provide low resolution JPEG images (even through the original data is high resolution JPEG 2000)
  - Display metadata in JP2064 format in virtual XML files
- If configured to, transparently pack/unpack data to/from archives for replication and HSM optimization.

ARCITECTA™

The Future

Better Archive Format..

AAR

Parallel Compression/Decompression

Arcitecta is working on the next generation AAR format:

- Multi-threaded compression/decompression
- Tables of Contents generation for sequentially written archives

This will:

- Significantly reduce the compression and ingestion times for multi-terra byte files (e.g. 10-20TB per gene sequence file)
- Ensure all archives are random-access making it easier to selectively extract individual files.

Mediaflux™

OPERATING SYSTEM FOR META+DATA

www.arcitecta.com