



Neil Clarke
Director, ITS Research Support Services
rev 8 April 2009

Mediaflux – research data management for workgroups

– a digital asset management system (DAMS)

Mediaflux introduction

Mediaflux – positioning

- e-Research aims to deploy a *multiplicity* of *overlapping* ICT tools to meet the *varied* needs of different researchers and disciplines

**Where does Mediaflux sit
within the *continuum* of tools?**

- The *continuum* can be seen in a number of different attributes...

Research data management continua

Personal ← Workgroup → Public

Raw data → data life-cycle → Published results

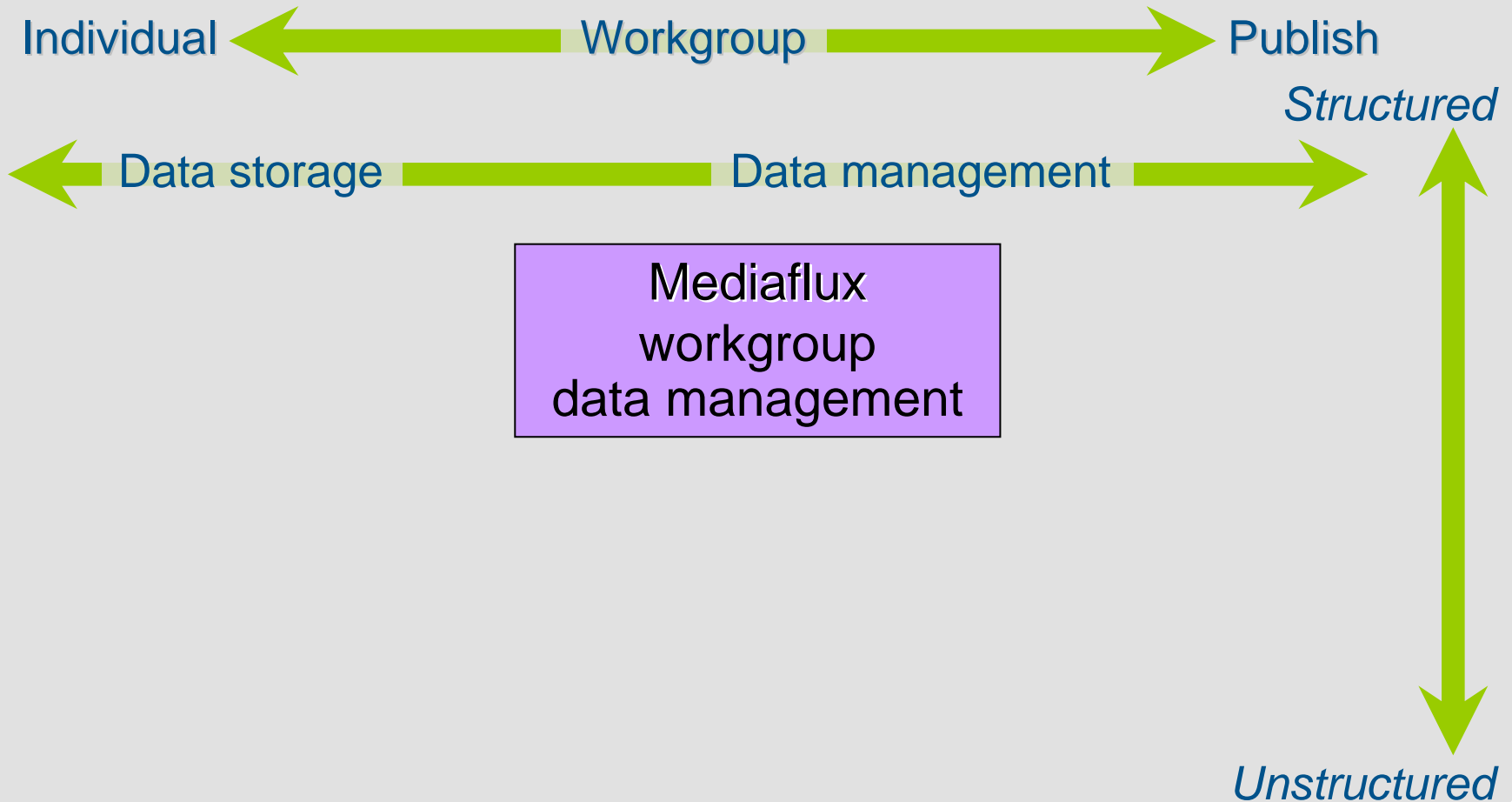
Unstructured data → Structured data

**Where does Mediaflux sit
within the *continuum* of tools?**

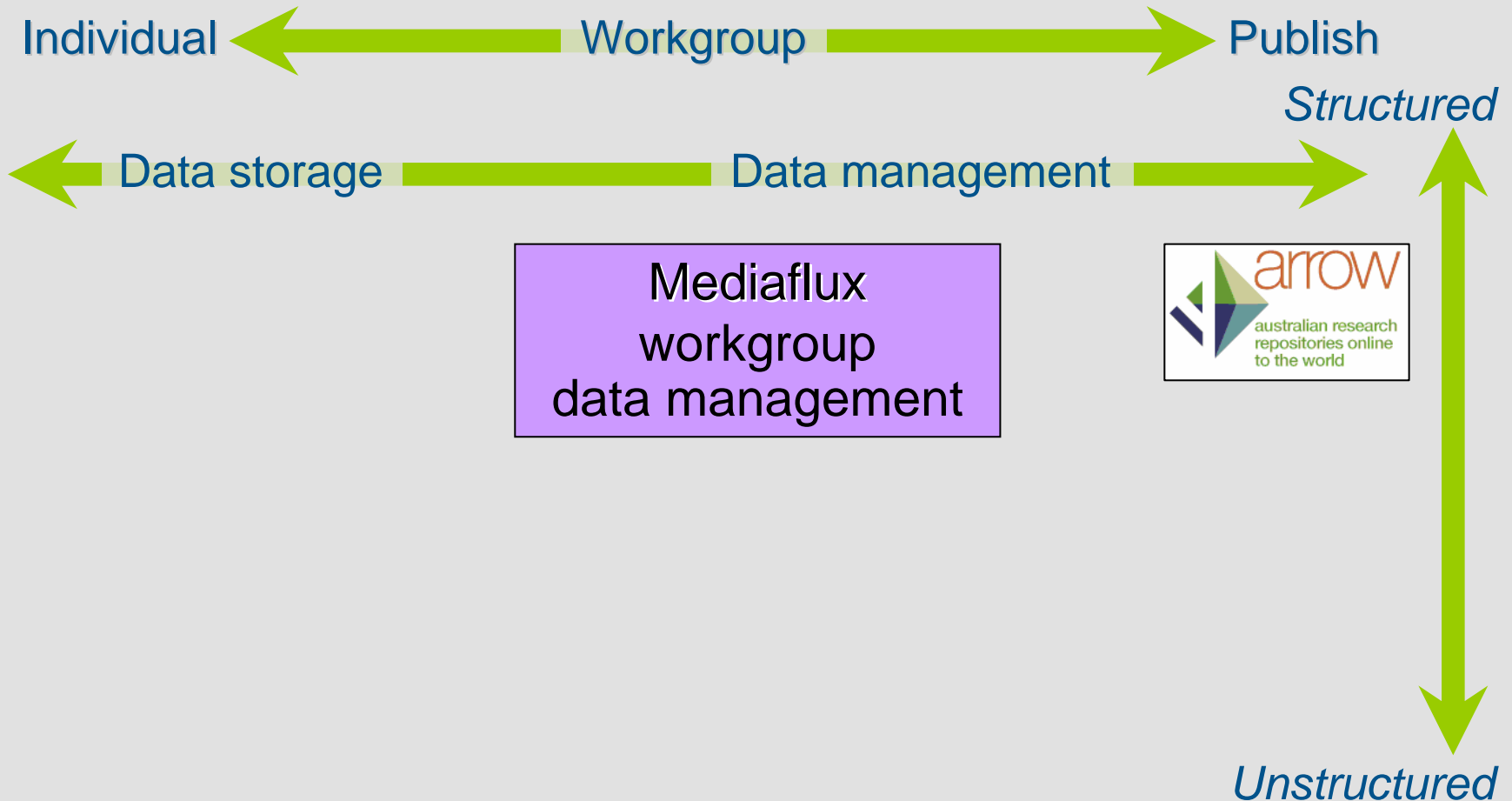
Generic tools → Custom solutions

Institution → Discipline

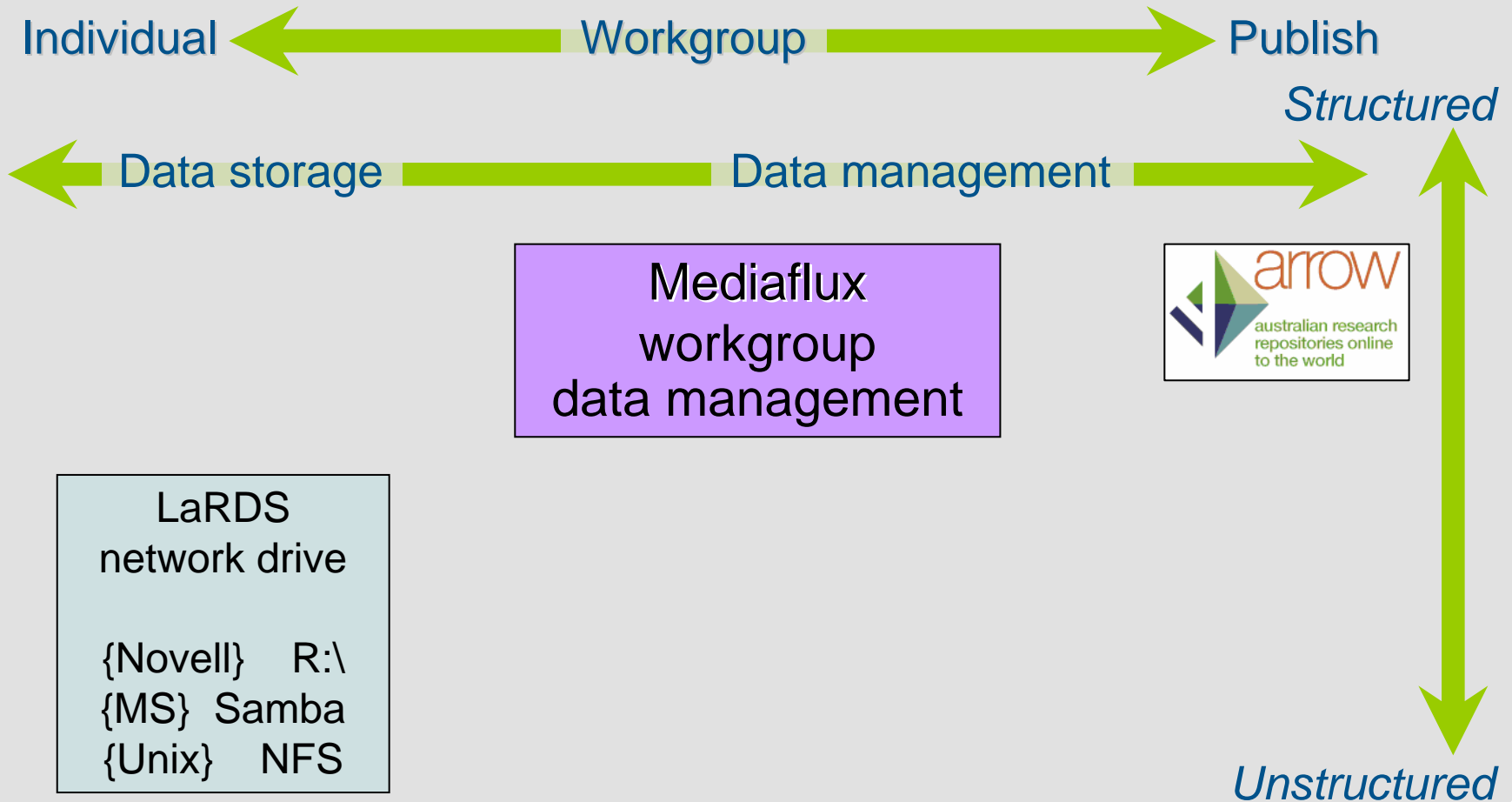
Mediaflux – positioning



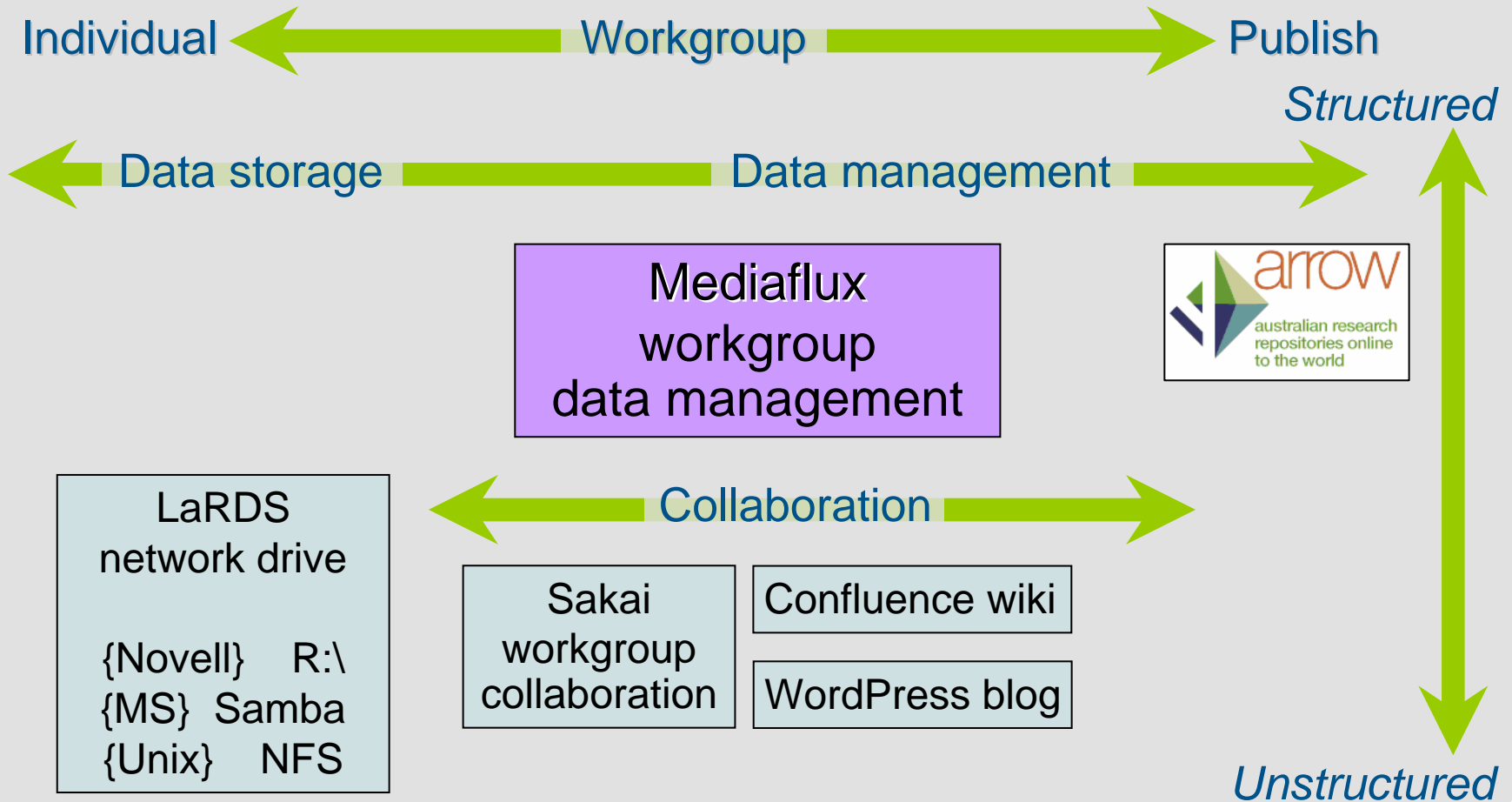
Mediaflux – positioning



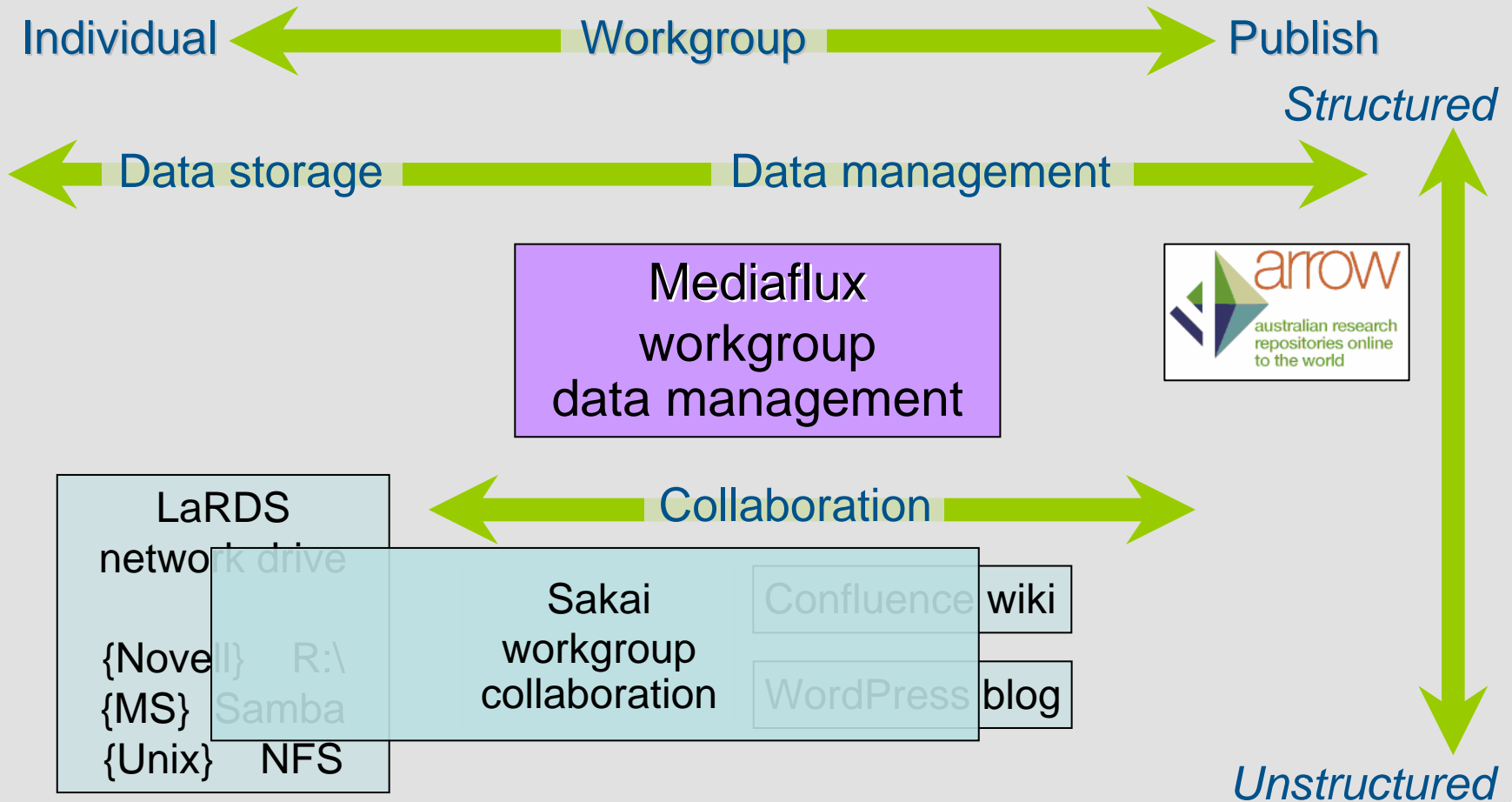
Mediaflux – positioning



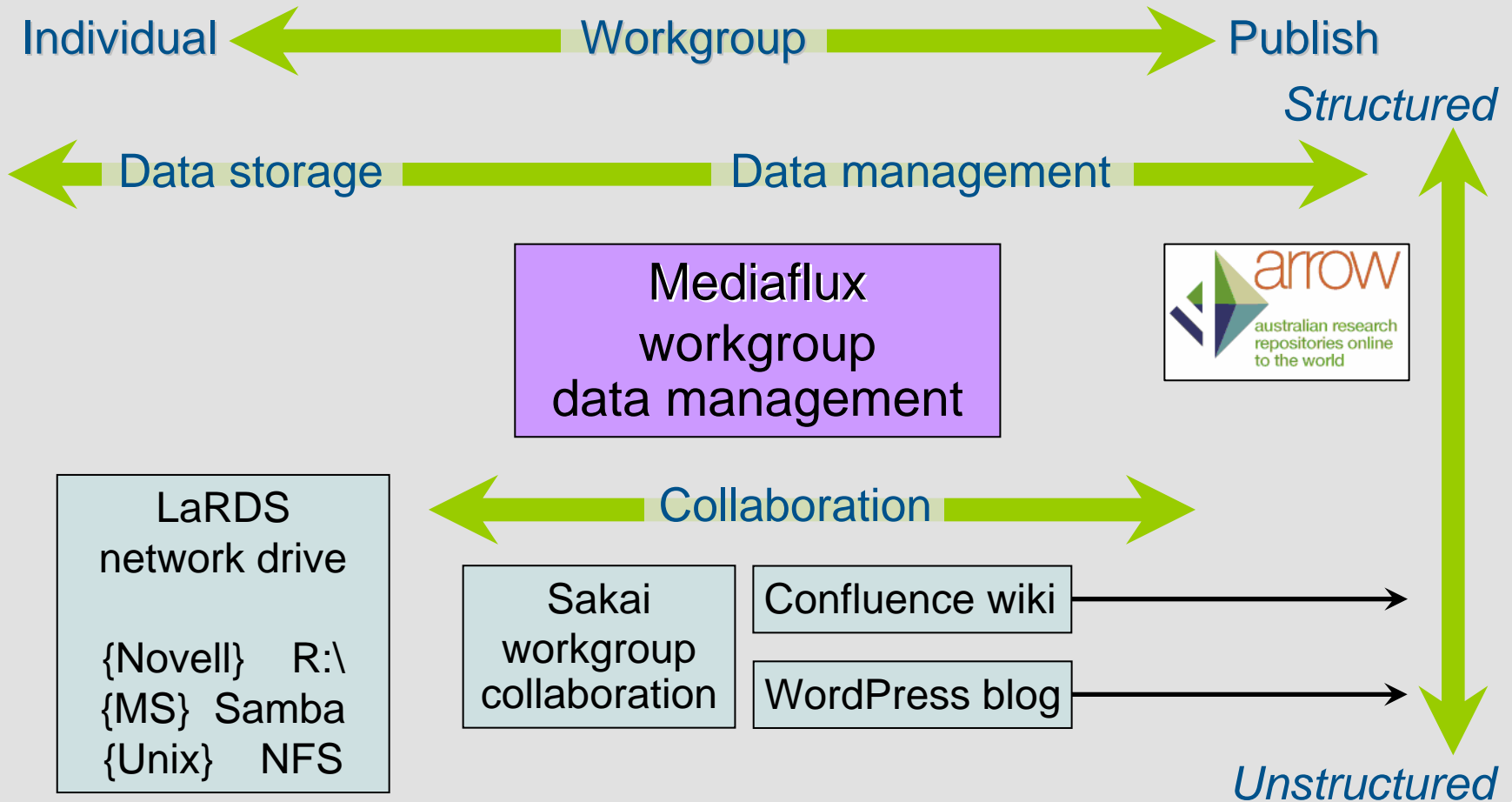
Mediaflux – positioning



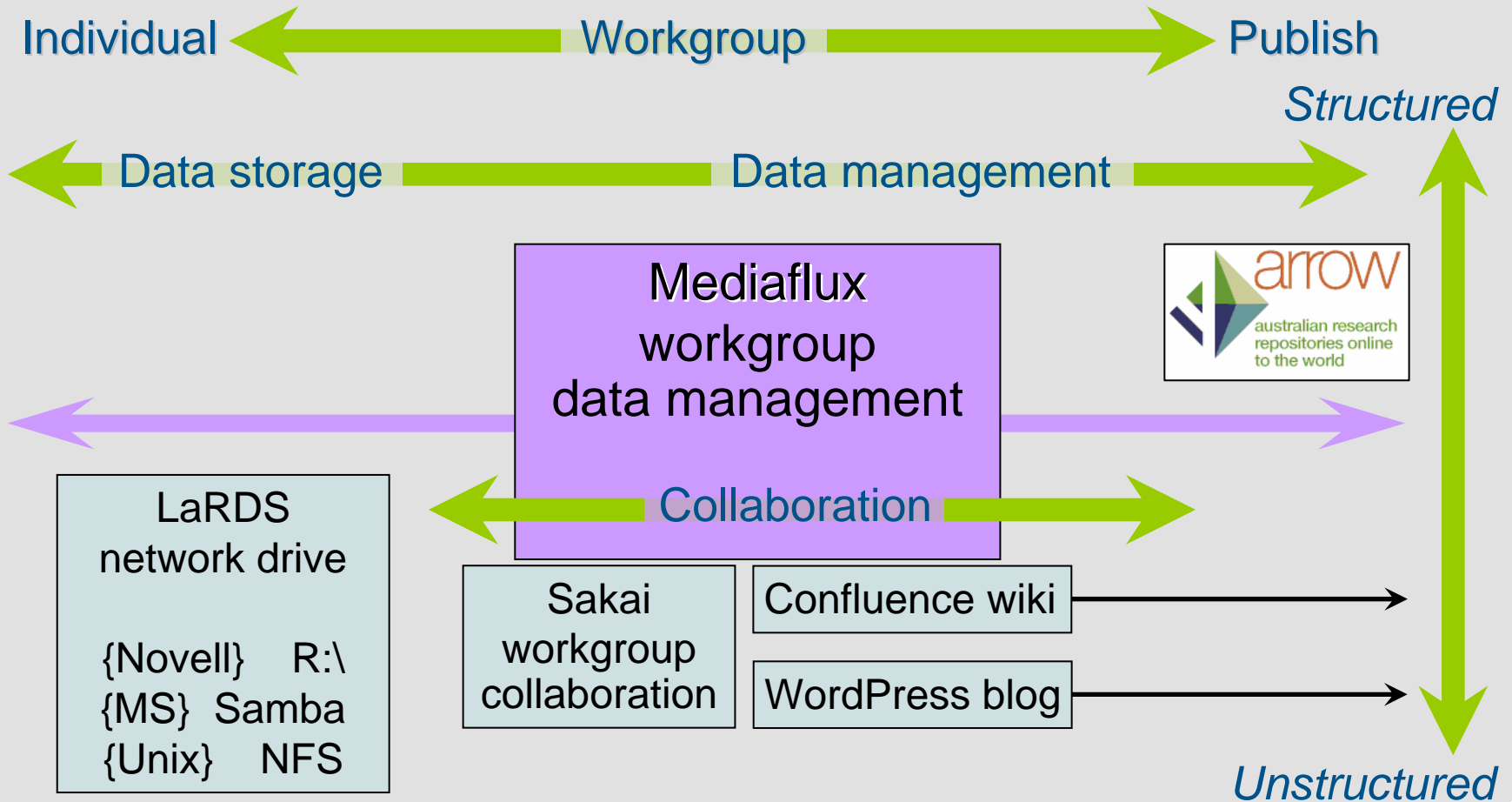
Mediaflux – positioning



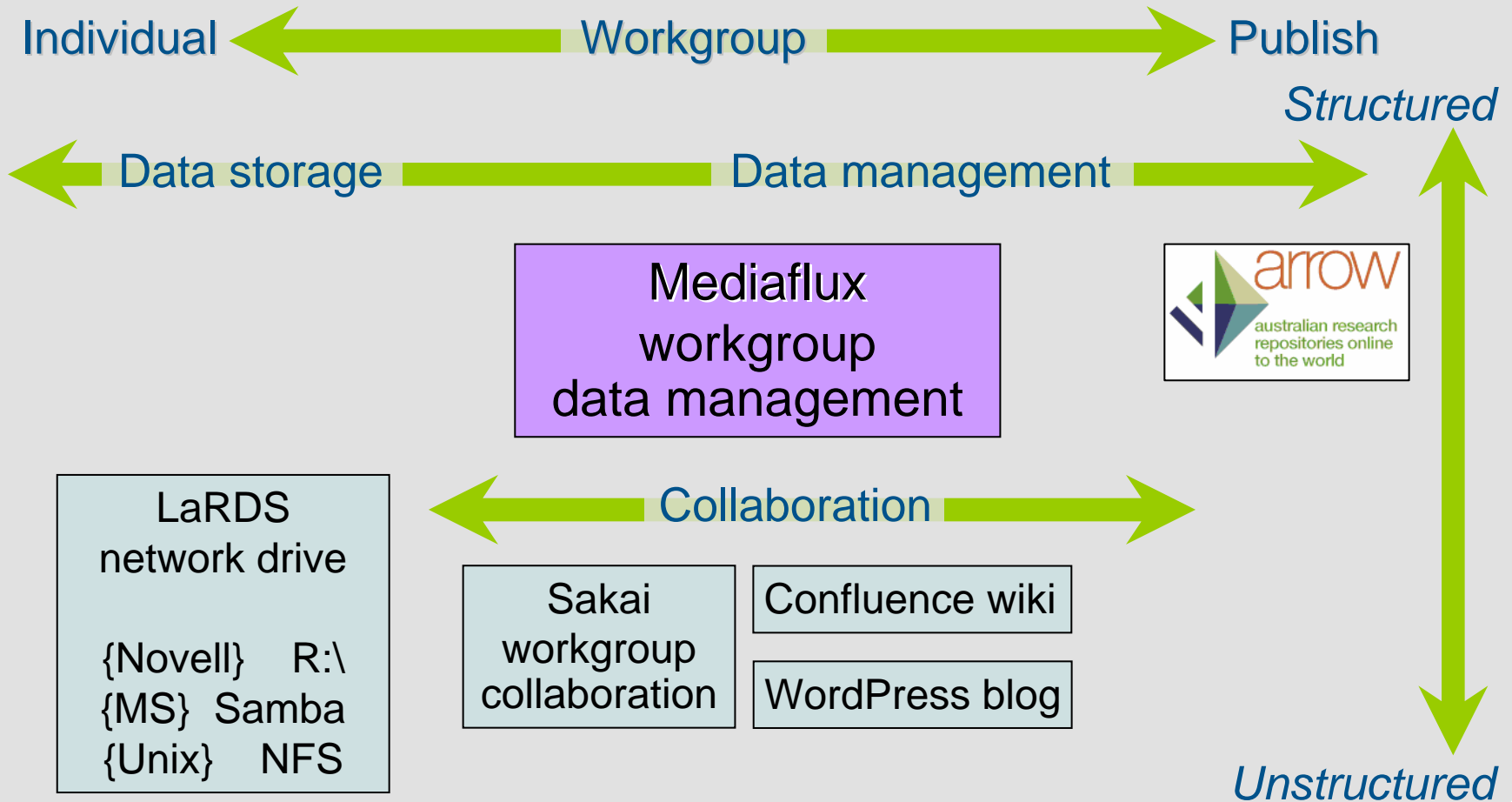
Mediaflux – positioning



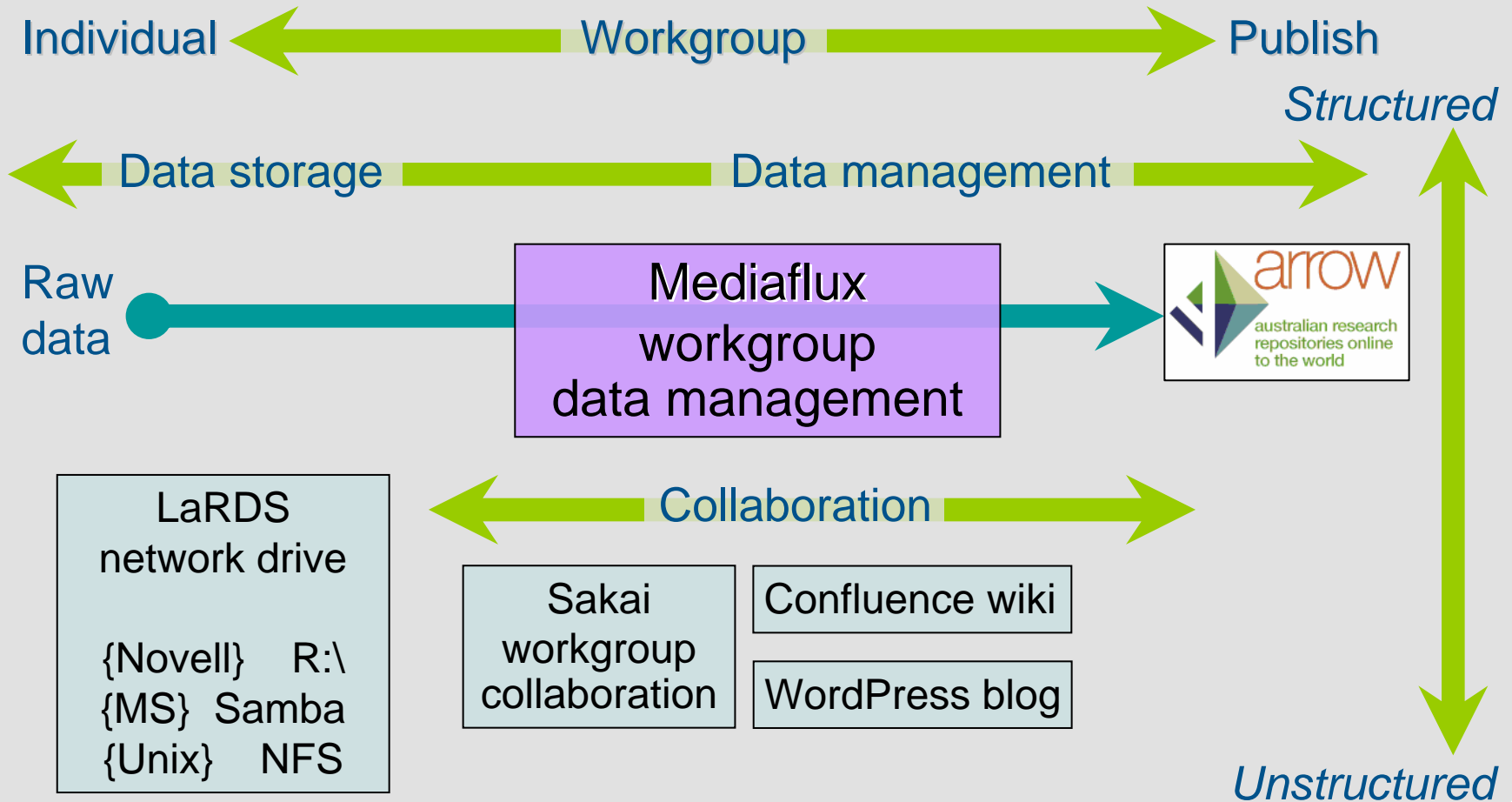
Mediaflux – positioning



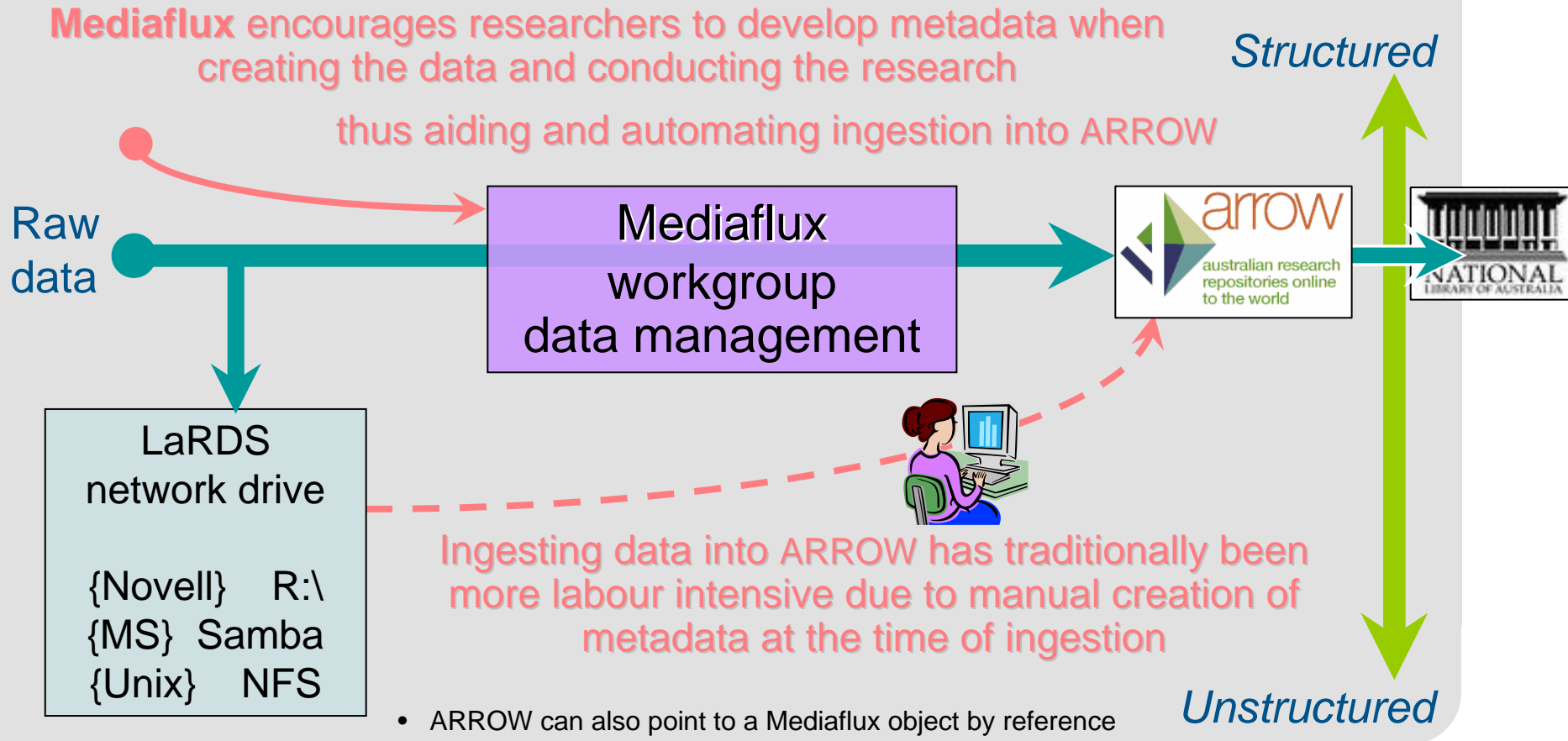
Mediaflux – positioning



Mediaflux – positioning – data life-cycle



Mediaflux – metadata & ingestion into ARROW



Mediaflux – metadata schema

- Mediaflux allows the user to define any arbitrary *metadata schema* for a collection of objects
- by combining any number of existing or new *metadata templates* ('*fragments*') such as:
 - Monash core metadata → minimum for ingestion into ARROW
 - discipline-standard metadata (optional ingestion into ARROW)
 - instrument-specific metadata (optional ingestion into ARROW)
 - researcher-defined metadata (optional ingestion into ARROW)
- The researcher can progressively refine the schema even while in use
- *the Monash core metadata covers necessary basics like creation/disposal date/time, author, etc.*
- *the Monash core is a sub-set of Dublin core, as determined by the Monash Library*

Mediaflux – generic and custom solutions



- **Generic: Mediaflux Desktop**

Addresses the question:

“ How can I organize my data better ? ”

- **Custom: Mediaflux Toolkit**

Build custom solutions for specific research apps

- **also 3rd party access:**

Data mining & access using open standards: XML

Mediaflux Desktop – viewing screens

- Use the viewing screens to *create, organize, view and search* collections of digital objects
 - Collection tree and object list
like Windows Explorer
 - Geo-spatial viewer
Google Maps for objects with geographical coordinates
- and under development:
 - Text explorer
contextualizes objects within a text-based narrative
 - M-eNotes
MS-Word based front-end, primarily intended as a Laboratory Information Management System (LIMS) or text-based Document Management System
 - Collaboration tools *for real-time collaboration*

Mediaflux – data types & viewing screens

- Mediaflux can handle all data types including:
 - **Images** (including nD imaging)
 - **Audio**
 - **Video**, time-domain imaging
 - **Simulations**, graphics and animation
 - **Document** management
 - text files, PDFs, html files, etc.
 - any **other** file types
 - spreadsheets, tables, proprietary formats, etc.
- Mediaflux has players to display many common image, audio, video and document file formats
 - including **annotation** of elements *within* objects

Mediaflux Desktop – control screens

- Use the control screens to perform the following functions
 - Metadata library
to create, modify or find suitable metadata templates
 - Meta mapper
to import and export objects and metadata to other systems
 - Access controls
to create users, assign group memberships and access rights to view, change and contribute to sets of collections

Summary: What is Mediaflux ?

- **Mediaflux** is a general purpose Digital Asset Management System (DAMS), specifically designed for research data
- Mediaflux comprises:
 - Mediaflux Desktop – web GUI application
 - Mediaflux Toolkit – for developing custom applications
 - Mediaflux XODB – XML object database engine
- *Mediaflux was developed by Arcitecta and supplied to Monash by SGI as prime contractor*
- *Mediaflux Desktop was developed by Arcitecta to a specific Monash specification*

What can Mediaflux store ?

- Digital research data of any type: audio, video, imaging, documents (PDF, Word,...) , numerical data (spreadsheets), text, or any other file types
- Metadata (information about the data) in any standard or custom metadata schema
 - You can create your own metadata schema
 - You can progressively modify metadata schemas as needs or standards evolve
- All data are stored in their original native formats
- All metadata are stored in open XML standard format

What can you do with Mediaflux ?

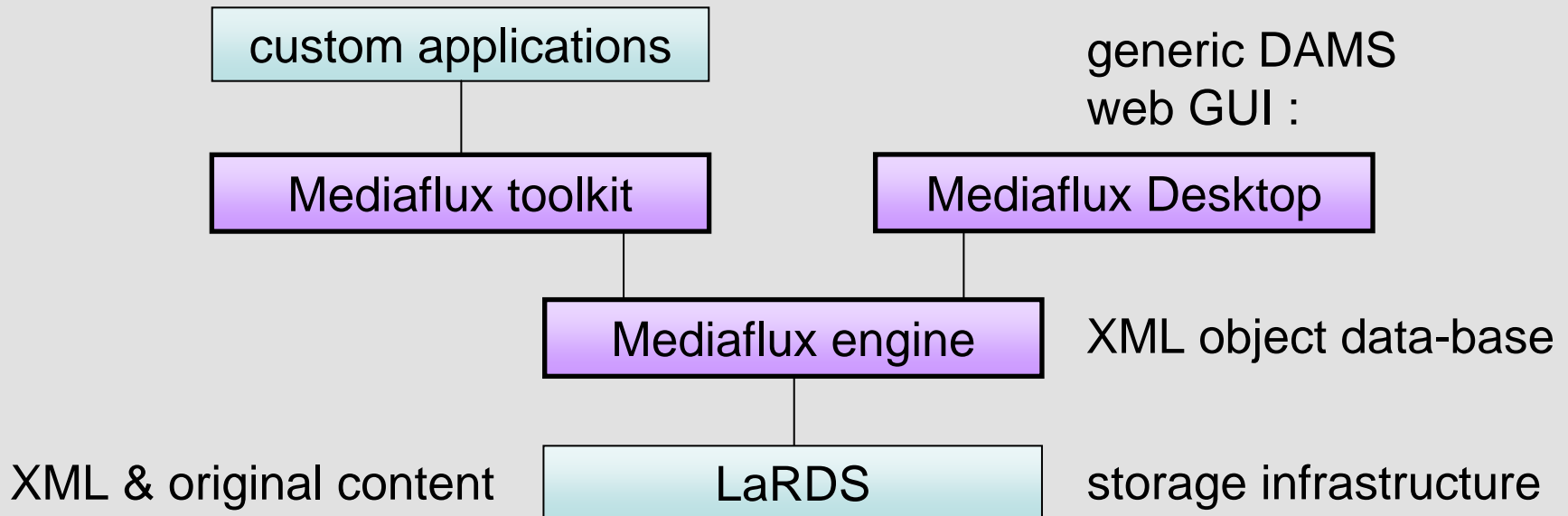
- Store and organize research data objects
- Retrieve and view the contents of data objects
- Search metadata and annotations
- Annotate objects or parts of objects
- Create and modify metadata schemas
- Control who can access what
- Contextualize data objects within a text narrative, or tabular or other document style

How do you view objects using Mediaflux Desktop ?

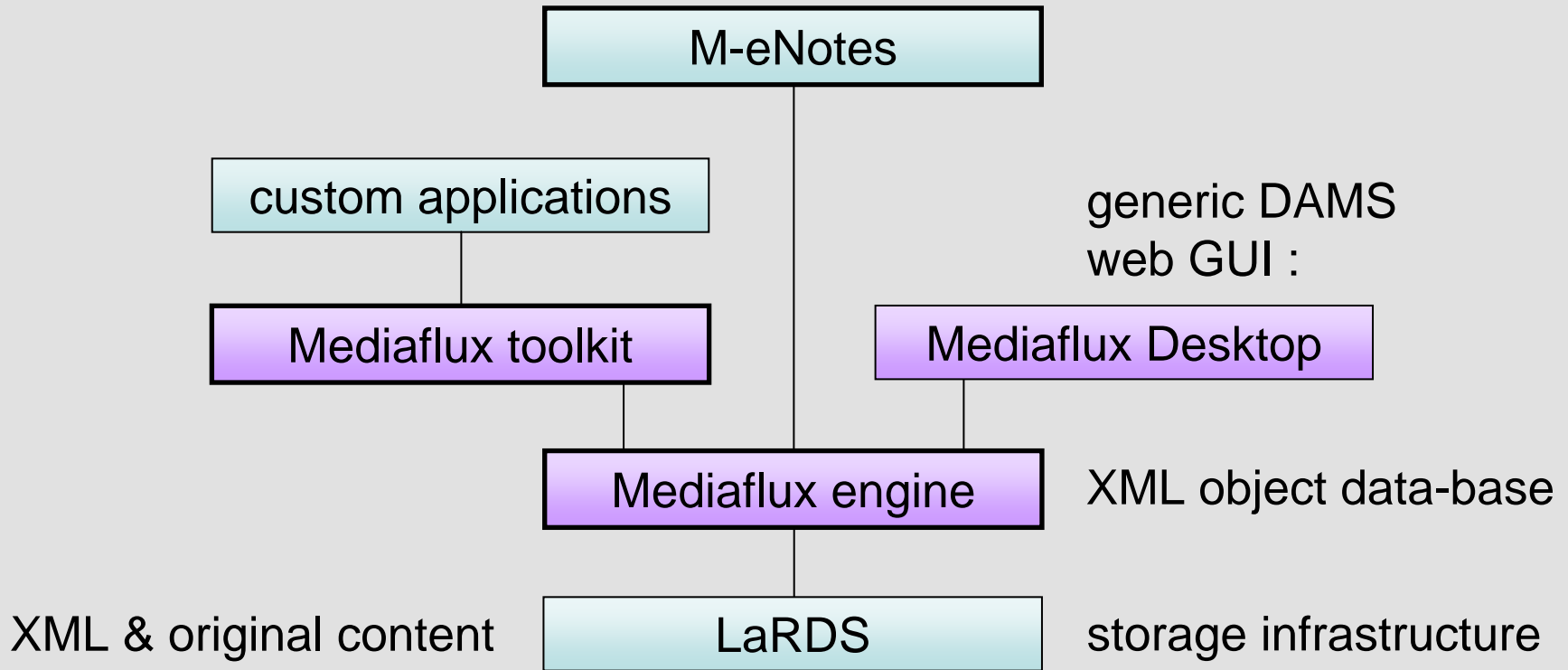
- Assets are organized into a hierarchical tree of 'Collections' (like MS-Windows 'Folders')
- The main browser screen has panes to:
 - navigate through the tree of Collections
 - view a list of Assets within a Collection, including thumbnail and metadata summary
 - view an Asset in detail, including all its metadata and annotations
 - search for an Asset
- Geobrowser: for Assets that have geospatial coordinate metadata, you can browse for the Assets via a clickable map
- Text explorer: you will also be able to contextualize and view Assets within a linked text narrative, or tabular or other HTML document style

Mediaflux – Overview

- Mediaflux Desktop uses only a fraction of the capabilities of the entire Mediaflux system*



Mediaflux – access from other applications



Mediaflux Desktop main panes – overview

viewers

Imaging viewer

with annotation

Video player

with annotation

Audio player

with annotation

browsers

Text explorer

Contextualizes assets within a linked text narrative

Geo browser

Explore assets via a map

Collections tree

Thumbnails, metadata viewer, sort, search, annotations

controls

Access rights

Metadata editor

Metadata library

Mediaflux – important points

- All data is stored in its original format/s
- All metadata is stored in open XML format
 - which can be read by other XML tools
- Nothing is captive to any proprietary format, system or database
- Create your own metadata schema – via GUI
- Can handle very large datasets
- Is also a toolkit – to create custom applications

Mediaflux – Data import

- From user's workstation via the Mediaflux Desktop web GUI
- Directly from main LaRDS A450 (sirius) file-system
 - Mediaflux is DMF aware
- By reference to data held in an external server
- Coming soon:
 - FTP, GridFTP
 - SRB
- Metadata:
 - Direct data entry from Mediaflux Desktop web GUI
 - Import from CSV file

Mediaflux – User authentication & access

- **Authentication**

- Stand-alone user database
- Coming soon:
 - Monash LDAP
 - Shibboleth and AAF

- **Access controls**

- Directly by the research team leader via the Mediaflux Desktop web GUI

Mediaflux XML DAMS

– digital image repository

The screenshot displays the Mediaflux XML DAMS interface. On the left, the 'Asset Finder' window shows a search results list with columns for identifier, type, and namespace. The main area features an 'Asset Viewer' window displaying a large image of a sliced watermelon. To the right of the viewer is a 'Metadata' panel with a 'Summary' tab. The 'Metadata' section is circled in red. The 'Summary' tab shows the following information:

Asset Summary	
Type:	image/jpeg
Namespace:	/stuff
Created:	30-Apr-2008 17:36:42
Modified:	30-Apr-2008 17:36:42
Content Type:	image/jpeg
Size:	162.35 KB
Identifier:	608
Version:	1

The 'Metadata' section contains the following information:

```
mf-revision-history
user
  id: 2
  domain: system
  name: manager
  type: create
  source: file:/Users/jason/stuff/D5C000:

mf-image
width: 640
height: 480
colour:
  space: RGB
  components:
    colour
    number: 3
  depth: 24
```

Mediaflux XML DAMS

– audio and video archive

The screenshot displays the Mediaflux XML DAMS interface. The main window is titled "Asset Finder" and shows a search results list on the left and an "Asset Summary" panel on the right. The search results list includes several audio files, such as "MusicSurround.wav" (ID 889) and "Sumatra 0041 mp3" (ID 954). The "Asset Summary" panel for "Sumatra 258 mp3" (ID 961) shows details like Type (audio/mpeg), Namespace (/audio), Name (Sumatra 258 mp3), Created (28-Jul-2008 16:09:13), Modified (28-Jul-2008 16:09:13), Content Type (audio/mpeg), Size (9.33 MB), Identifier (961), and Version (1). Below the summary is a "Metadata" section with "mf-revision-history" showing a user.

Overlaid on the bottom right is the "Asset Viewer" window, which shows the "Annotations" tab selected. The "Annotations" tab contains a table with the following data:

ID	Typ	Author	Created
1		demo:monash	07-Aug-2008

Below the table, there is a "Comment" section with the following details:

- ID:** 1
- Created:** 07-Aug-2008 17:49:13
- Author:** demo:monash
- Location:**
- timecode:** start: 00:00:24:55

At the bottom of the interface, there is a video player control bar with a progress bar showing 00:00:44:46 out of 00:10:13:10. A red circle highlights the play/pause button in the control bar.

Mediaflux XML DAMS

– metadata library

The screenshot shows the 'demo:monash' interface. In the sidebar, the 'Metadata Library' option is circled in red. The main window displays a table titled 'Metadata Library' with columns 'Type' and 'Generated By'. The table lists various metadata types and the user who generated them.

Type	Generated By
mf-image	system
mf-image-exif	system
mf-revision-history	system
mf-note	user
mf-name	user
mf-document	user
mf-user	user
mf-snapshot-account	user
mf-snapshot-file	user
mf-forum	user
mf-forum-topic	user
mf-forum-participant	user
mf-forum-topic-post	user
mf-dicom-project	user
mf-dicom-subject	user
mf-dicom-patient	user
mf-dicom-study	user
mf-dicom-series	user
mf-dicom-prefs	user
mf-www-upload	user
arc-dts-submission	user
dins-session	user

Document management

- Byproduct applications of Mediaflux Desktop include:
 - Document management
 - Contract management
 - Committee management

Contract records management

e.g. software licences and hardware maintenance

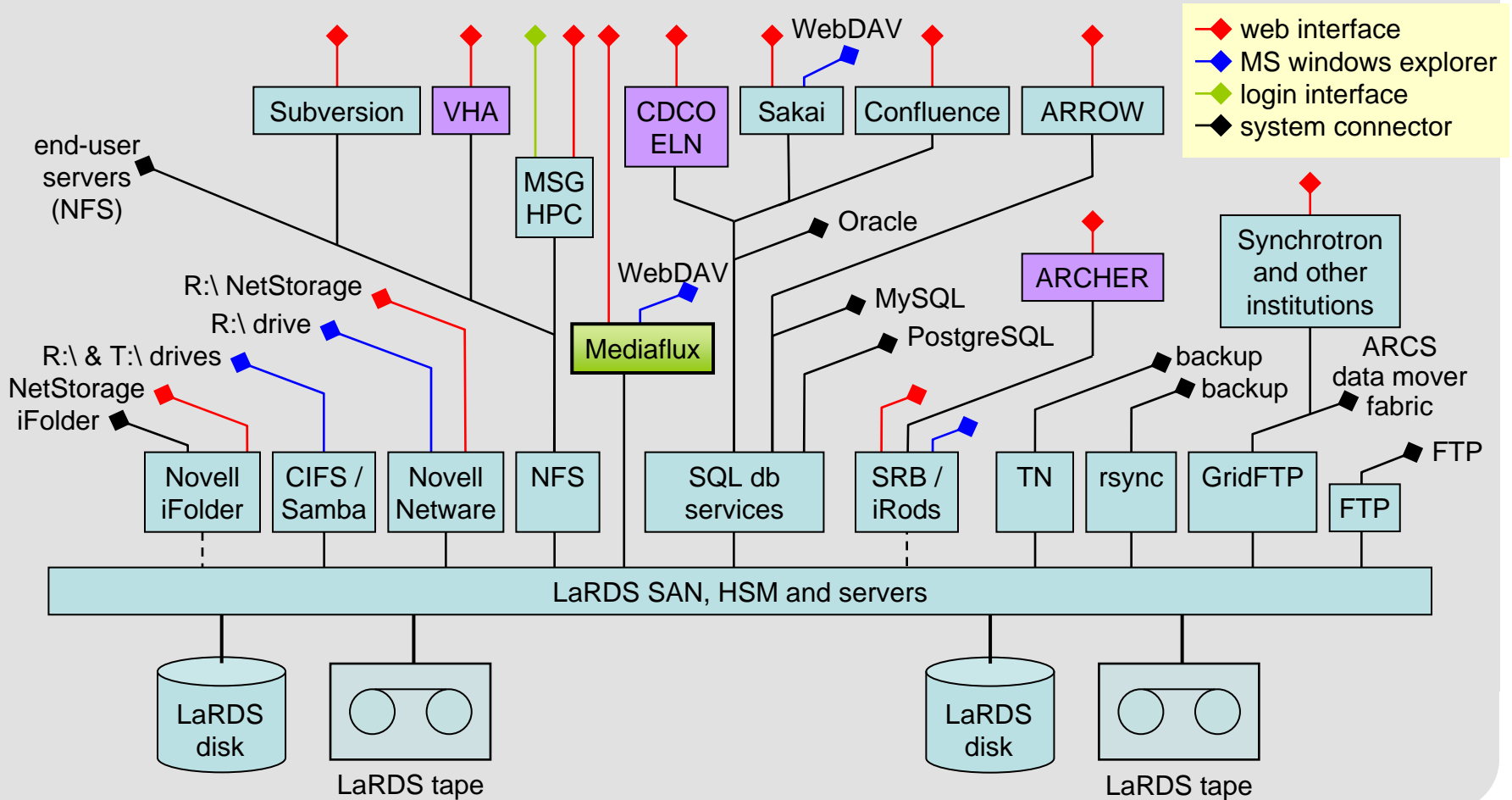
- **Contracts – *sort by metadata***
 - Contract – *collection* of material re one contract
 - *metadata includes*
 - » *key contract attributes*
 - » *in particular: **expiry / renewal date***
 - + PDF of scanned Contract
 - + PDF of scanned P/O
 - + PDF of scanned Invoice
 - + Product specification / brochure
 - + etc.

Committee records management

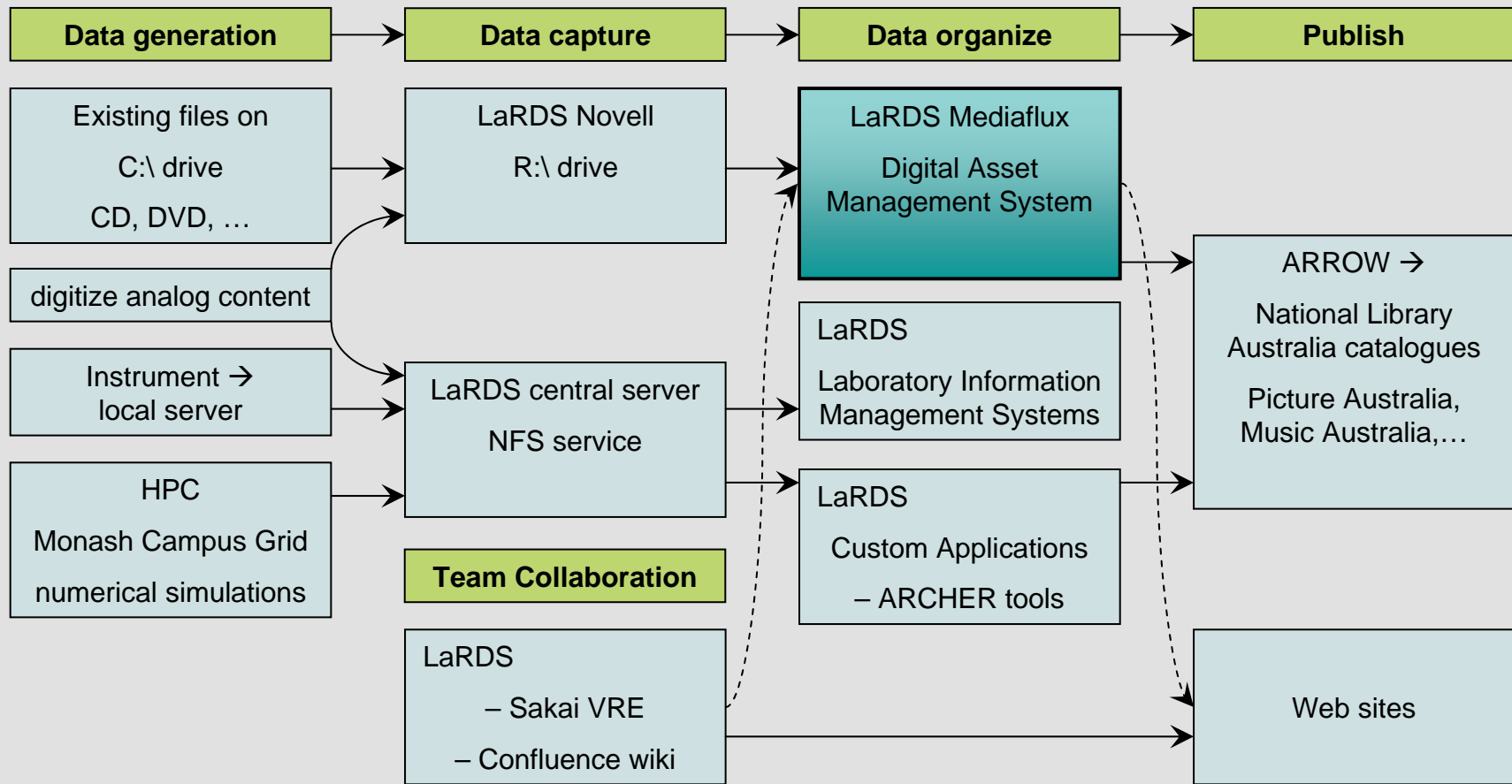
inc. inter-institutional collaborative research team

- e.g. Committee Folder structure template
 - Agendas
 - Minutes
 - + Audio logs, with annotations and event markers
 - Presentations
 - Policies
 - Discussion Papers
 - Treasury
 - General Annual Meetings
 - Annual reports
 - Archive

LaRDS access points & connectors



Data ingestion: the data life-cycle continuum



LaRDS – important points

- LaRDS is different things to different people
 - Mediaflux is just one of the ways of accessing LaRDS
- LaRDS is secure and reliable:
 - all data is backed up to two data centres
 - data is only available to the people you want it be available to :
 - individual
 - workgroup
 - public

LaRDS is available for use

- LaRDS is available for use by all Monash researchers, including HDR students
- Current capacity:
 - + 400 TB on-line disk
 - + 1600 TB near-line tape ... and growing ...
- Aim: Keep capacity ahead of demand
- So: Please contact us for access:
 - + Neil Clarke, Earle Lawrence, MeRC
 - + Library faculty liaison, Local IT support

Questions ...

www.monash.edu/eresearch/services/lards/

Neil.Clarke@its.monash.edu