National Library Board
Singapore

PREMIS
Implementation at NLB

IPRES 2011 Tutorial

Tuesday, 1 Nov 2011

Agenda

- **Background**
- **PREMIS in NLB**
- **Summary & Usability**

Background to Digital Preservation in NLB

NLB Act

- As a collecting institution and custodian of Singapore's Legal Deposits
- To preserve national documentary heritage materials for current and future use.
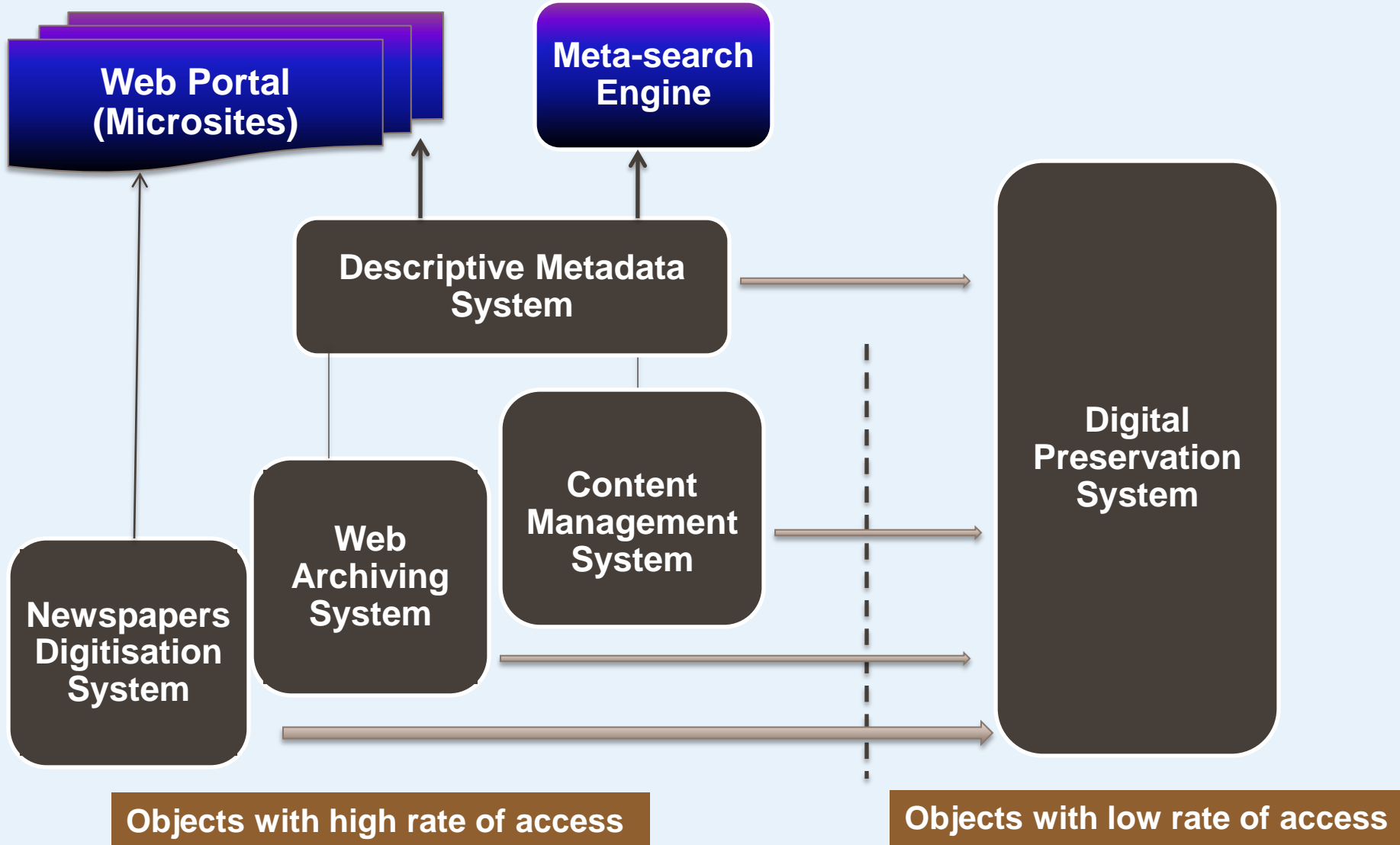
Digital Preservation System

- Rosetta from Ex Libris
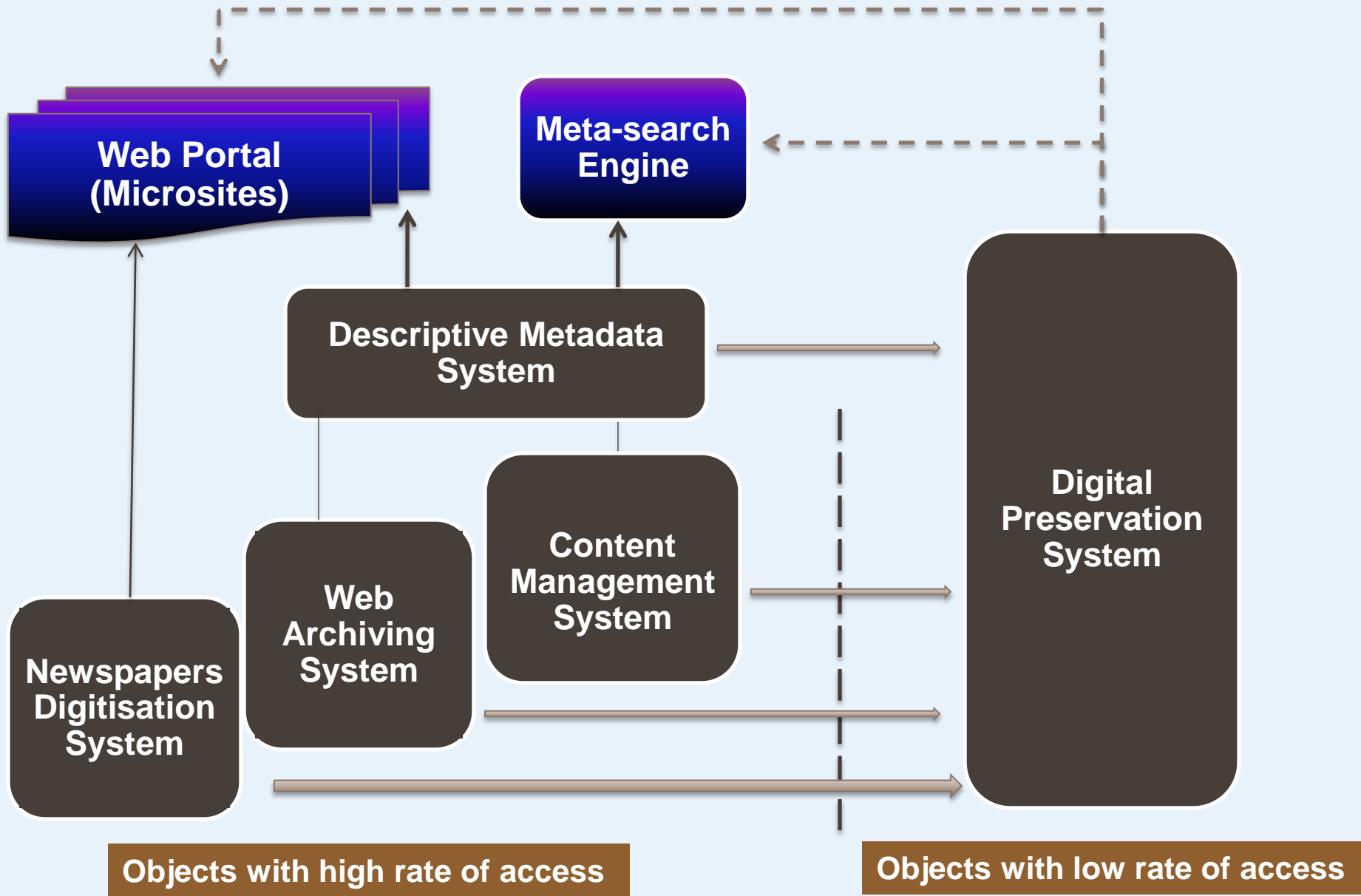 - DNX (preservation metadata based on PREMIS)
  - METS
  - OAIS, etc…..

Current architecture

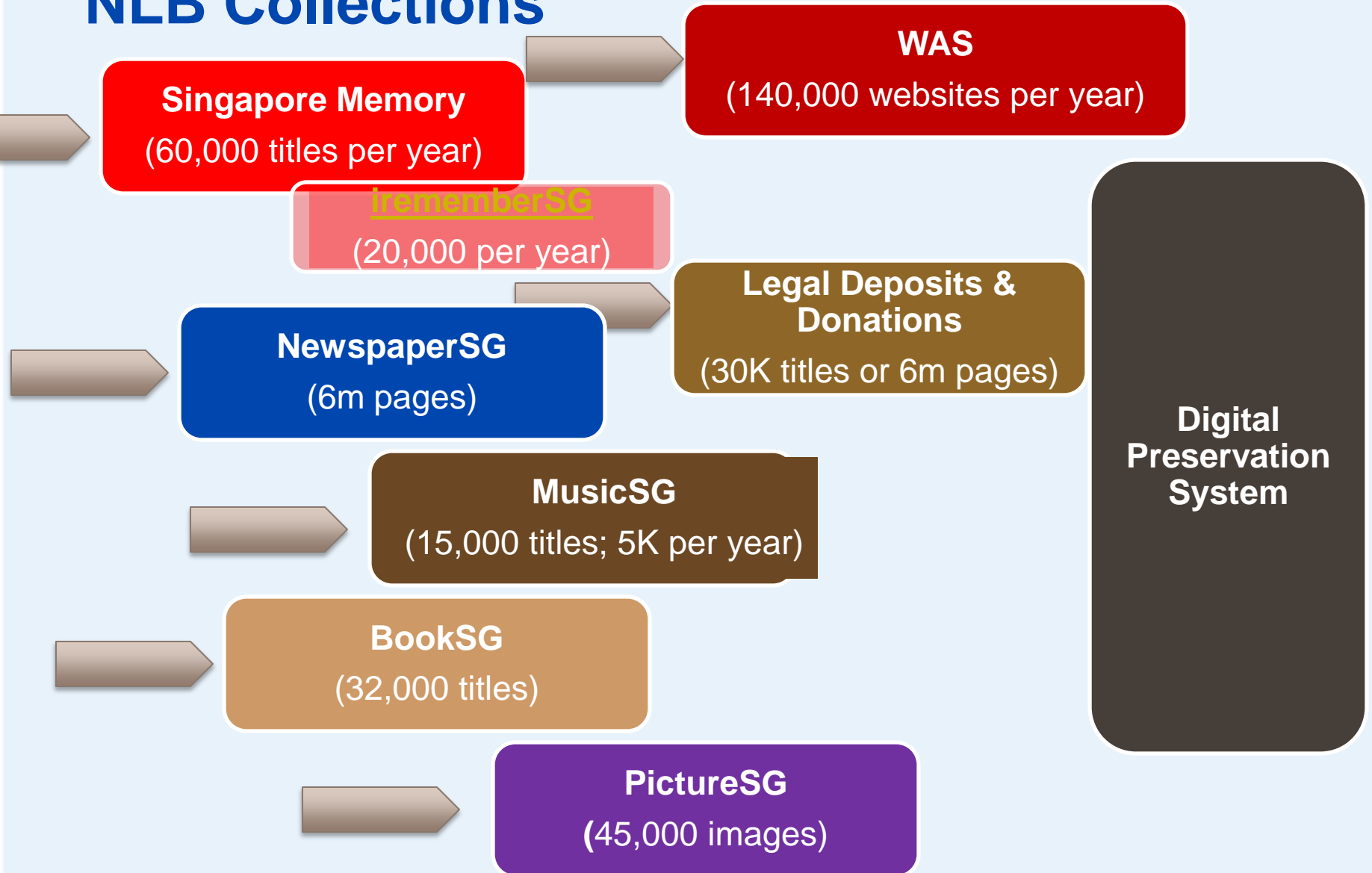- Objects with high rate of access
- Objects with low rate of access

NLB Collections

NLB Current Architecture

Web Portal (Microsites)

Meta-search Engine

Descriptive Metadata System

Content Management System

Newspapers Digitisation System

Web Archiving System

Digital Preservation System

Objects with high rate of access

Objects with low rate of access

Web Portal (Microsites)

Meta-search Engine

Descriptive Metadata System

Content Management System

Web Archiving System

Newspapers Digitisation System

Digital Preservation System

Objects with high rate of access

Objects with low rate of access

NLB Collections

Singapore Memory
(60,000 titles per year)

WAS
(140,000 websites per year)

IrememberSG
(20,000 per year)

Legal Deposits & Donations
(30K titles or 6m pages)

NewspaperSG
(6m pages)

Digital Preservation System

MusicSG
(15,000 titles; 5K per year)

BookSG
(32,000 titles)

PictureSG
(45,000 images)

National Library Board Singapore

Descriptive Metadata

DC terms

identifier	created
title	modified
creator	valid
contributor	available
subject	issued
description	medium
publisher	extent
type	bibliographicCitation
language	spatial
rights	temporal
source	accessRights
audience	license
provenance	rightsHolder
alternative	+ 13 relation fields…
abstract	
tableOfContents	

MODS

location
edition
dateCaptured

NLB terms

originalPublisher
digitalPublisher
secondaryPublisher
nlbshtx
sash
nlbdltx
nlbdltxch
nlbdltxmy
nlbdltxtm
nlbt
apa
collections
sgnt
nlbaud

NLB-AP

Metadata in Rosetta

Descriptive Metadata System

PREMIS	
Object Entity	eventDateTime
objectIdentifier	eventDetail
objectCategory	eventOutcomeInformation
preservationLevel	linkingAgentIdentifier
significantProperties	linkingObjectIdentifier
objectCharacteristics	**Agent Entity**
originalName	agentIdentifier
storage	agentName
environment	agentType
signatureInformation	agentNote
relationship	agentExtension
linkingEventIdentifier	linkingEventIdentifier
linkingIntellectualEntityIdentifier	linkingRightsStatementIdentifier
linkingRightsStatementIdentifier	**Rights Entity**
Event Entity	rightsStatement
eventIdentifier	rightsExtension
eventType	

NLB-AP

identifier — source
title — audience
creator — provenance
contributor — alternative
subject — abstract
description — tableOfContents
publisher — created
type — modified
language — valid
rights — available
— issued
NLB terms — medium
originalPublisher — extent
digitalPublisher — bibliographicCitation
secondaryPublisher — spatial
nlbshtx — temporal
sash — accessRights
nlbdltx — license
nlbdltxch — rightsHolder
nlbdltxmy — + 13 relation
nlbdltxtm — fields...
nlbt
apa — **MODS**
collections — location
sgnt — edition
nlbaud — dateCaptured

NLB-AP + Metadata Identifier

Collections AP + Metadata Identifier

Manuscripts AP + Metadata Identifier

Collections Metadata
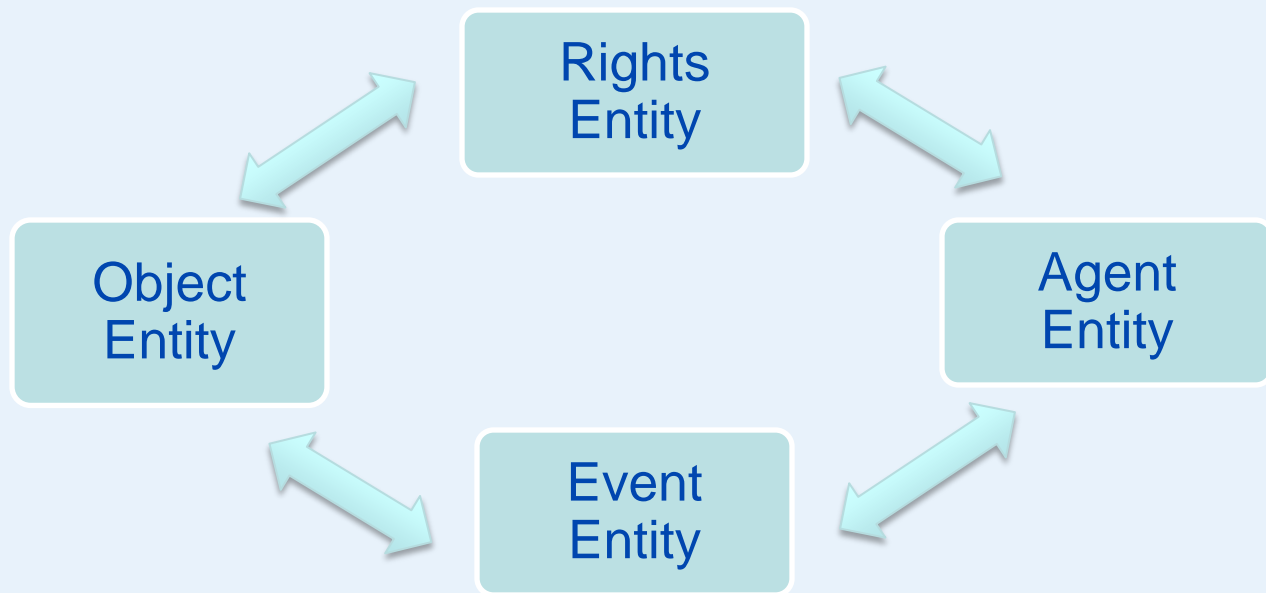
Manuscripts Metadata

PREMIS support in Rosetta

Rights Entity

Agent Entity

Event Entity

Object Entity

Some Qualifiers…

The following slides are not meant as criticism of either PREMIS or Rosetta. Comparison is done to show how an actual preservation system implements PREMIS, where operational consideration affects the design of the preservation metadata.

 Semantic units not found in DNX can be incorporated as:
(i)   New service request
(ii)   Expression of existing relationships built into the system
(iii) Some are already available in METS format

National Library Board
Singapore

Object Entity in Rosetta DNX

Qualifier: Semantic units at Bitstream level are not captured as no data for now

PREMIS	Rosetta		
Object Entity	Rep	File	Remarks
objectIdentifier (Rep, File)			
objectIdentifierType	√	√	All semantic units available
objectIdentifierValue	√	√	All semantic units available
objectCategory	√	√	All semantic units available
preservationLevel (Rep, File)			
preservationLevelValue	√	√	All semantic units available
preservationLevelRole	√	√	All semantic units available
preservationLevelRationale	√	√	All semantic units available
preservationLevelDateAssigned	√	√	All semantic units available
significantProperties (Rep, File)			
significantPropertiesType	X	√	Extracted at File Level only
significantPropertiesValue	X	√	Extracted at File Level only
significantPropertiesExtension	X	X	Not in DNX
ObjectCharacteristics Note: Not applicable at Rep level			…see following slides

Object Entity in Rosetta DNX

PREMIS			Rosetta
Object Entity	**Rep**	**File**	**DNX Fields**
objectCharacteristics (File)			
compositionLevel	**NA**	√	General File Characteristics : compositionLevel
fixity	**NA**	√	file Fixity
size	**NA**	√	General File Characteristics: size
format			file Format General File Characteristics: MIME type & extension
formatDesignation			
formatName	**NA**	√	file Format: format Name
formatVersion	**NA**	√	file Format: formatVersion: 3 (e.g.)
formatRegistry			
formatRegistryName	**NA**	√	file Format: format Registry: PRONOM (e.g.)
formatRegistryKey	**NA**	√	file Format: Registry ID: fmt/7 (e.g.)
formatRegistryRole	**NA**	√	file Format: format Registry Role
formatNote	**NA**	√	file Format: format Note
			…con't on slide 16

National Library Board Singapore

Next

Object Entity	Rep	File	DNX Fields
Object characteristics (File level only)	√	√	General *IE* Characteristics
			General *Representation* Characteristics
			General *File* Characteristics

MID DNX_FL18932   **Metadata Type** dnx:dnx   **Description**
Created by nlb_producer_agent   **Creation Date** 20/09/2011 13:40:04
Updated by   **Update Date** 20/09/2011 13:40:04

Help | Logout

Site Map

preservation Level ▼
Add

internal Identifier
object Characteristics
file Fixity
event
vs Outcome
file Virus Check
file Format
general File Characteristics
file Validation
significant Properties

File Entity Type	None
composition Level	-
file Location Type	File
file Location	-
file Original Name	9000000026_0002.tif
file Original Path	-
file Original ID	/exlibris_apps_stage/deposit_area_1//1001-2000/dep_1293/deposit/content/streams/9000000026_000
file Extension	tif
file MIME Type	image/tiff
file Size Bytes	35007024
storage ID	4
stream Ref Id	-
format ID	fmt/7
risk Identifiers	-

access Rights Policy

User Defined B
User Defined C

Back

Object Entity	Rep	File	DNX Fields
fixity	NA	√	file Fixity
messageDigestAlgorithm	NA	√	file Fixity: fixity Type: MD5 (e.g.)
messageDigest	NA	√	file Fixity: fixity Value: e82aa847ae722a14265251bf8bd0ac92 (e.g.)
messageDigestOriginator	NA	√	file Fixity: agent: REG_SA_JAVA5_FIXITY (e.g.)

Home > Submissions > Approve Sampled SIPs > DNX Record

MID	DNX_FL18931	**Metadata Type**	dnx:dnx	**Description**	guage: English ▼	Help	Logout
Created by	nlb_producer_agent	**Creation Date**	20/09/2011 13:40:04		
Updated by		**Update Date**	20/09/2011 13:40:04		Site Map

preservation Level ▼
Add

internal Identifier
object Characteristics
file Fixity

agent   REG_SA_JAVA5_FIXITY
fixity Type   MD5
fixity Value   ff3003f085cf6c2ef2f974f39a684eb1

Description

preservation Level ▼
Add

internal Identifier
object Characteristics
file Fixity

	Description	Action	
1	fixity Type=MD5, fixity Value=ff3003f085cf6c2ef2f974f39a684eb1	View	
2	fixity Type=SHA1, fixity Value=4cec3bfd496cb3a47a3e5ce4df400a619d9239fd	View	
3	fixity Type=CRC32, fixity Value=2bae3253	View	

Object Entity	Rep	File	DNX Fields
fixity	√	√	vs Outcome
messageDigestAlgorithm	NA	√	vs Outcome: Type: Checksum (e.g.)
messageDigestOriginator	NA	√	vs Outcome: Agent: REG_SA_JAVA5_FIXITY (e.g.)
			vs Outcome: check Date: Tue Sept 20 12:30:15 (e.g.)
			Vs Outcome: result: PASSED (e.g.)

eventOutcomeInformation

Home > Submissions > Approve Sampled SIPs > DNX Full View

Home > Submissions > Approve Sampled SIPs > DNX Record

Description

MID	DNX_FL18932	Metadata Type	dnx:dnx	Description
Created by	nlb_producer_agent	Creation Date	20/09/2011 13:40:04	
Updated by		Update Date	20/09/2011 13:40:04	

preservation Level

Add

internal Identifier
object Characteristics
file Fixity
event
vs Outcome

check Date	Tue Sep 20 13:40:15 SGT 2011
vs Agent	REG_SA_JAVA5_FIXITY
type	CHECKSUM
result	PASSED
result Details	-
vs Evaluation	PASSED
vs Evaluation Details	-

Action

View
View
View
View
View

general File Characteristics
file Validation
significant Properties

vs Outcome	Rep	File	Subfields	
CHECKSUM	NA	√	check Date:	**eventOutcomeInformation**
VIRUSCHECK	NA	√	vs Agent: type:	**eventOutcome**
FILE_ FORMAT	NA	√	result:	**eventOutcomeDetail**
TECHMD	NA	√	result Details:	**eventOutcomeDetailNote**
RISK_ANALYSIS	NA	√	vs Evaluation: vs Evaluation Details:	**eventOutcomeDetailExtension**

Home > Submissions > Approve Sampled SIPs > DNX Record

MID	DNX_FL18931	Metadata Type	dnx:dnx	Description
Created by	nlb_producer_agent	Creation Date	20/09/2011 13:40:04	
Updated by		Update Date	20/09/2011 13:40:04	

preservation Level ▼ | Add

internal Identifier
object Characteristics
file Fixity
event
vs Outcome

check Date	Tue Sep 20 13:40:13 SGT 2011
vs Agent	REG_SA_DPS
type	RISK_ANALYSIS
result	PASSED
result Details	-
vs Evaluation	PASSED
vs Evaluation Details	-

event
vs Outcome
file Virus Check
file Format
general File Characteristics
file Validation

vs Outcome

NLB | National Library Bo
Singapore

Back

Object Entity in Rosetta DNX

PREMIS			Rosetta
Object Entity	**Rep**	**File**	**DNX fields (File Format)**
not in PREMIS	**NA**	√	format Description: Tagged Image File Format (e.g.)
	NA	√	exact Format Identification: false (e.g.)
	NA	√	mime Type: image/tiff (e.g.)
	NA	√	agent Version: 5.0 (e.g.)
	NA	√	agent Signature Version: 50 (e.g.)

MID	DNX_FL18931	**Metadata Type**	dnx:dnx	**Description**
Created by	nlb_producer_agent	**Creation Date**	20/09/2011 13:40:04	
Updated by		**Update Date**	20/09/2011 13:40:04	

preservation Level [▼] **Add**

internal Identifier
object Characteristics
file Fixity
event
vs Outcome
file Virus Check
file Format
general File Characteristics
file Validation
significant Properties

agent	REG_SA_DROID
format Registry	PRONOM
format Registry Id	fmt/7
format Registry Role	-
format Name	fmt/7
format Version	3
format Description	Tagged Image File Format
format Note	-
exact Format Identification	false
mime Type	image/tiff
agent Version	5.0
agent Signature Version	50

NLB | Singapore

Object Entity in Rosetta DNX

PREMIS			Rosetta
Object Entity	**Rep**	**File**	**Remarks**
creatingApplication	**NA**	**√**	All semantic units available
inhibitors	**NA**	**√**	All semantic units available
objectCharacteristicsExtension	**X**	**X**	DNX field: object Characteristics (meta-metadata) .. found in IE, Rep and File levels .. as a repeated object characteristics field

MID   DNX_IE18929   Metadata Type   dnx:dnx   Description
Created by   nlb_producer_agent   Creation Date   20/09/2011 13:40:04
Updated by   Update Date   20/09/2011 13:40:04

web Harvesting ▼   Add

producer
producer Agent
internal Identifier
object Characteristics

general IE Characteristics   Delete

access Rights Policy

object Characteristics

Description

object Type   Intellectual Entity
parent ID   -
group ID   -
creation Date   20/09/2011 13:40:04
created By   nlb_producer_agent
modification Date   20/09/2011 13:40:04
modified By   nlb_producer_agent
owner   CRS00.INS00.DEP00

Description

Description

vs Outcome
file Virus Check

owner   CRS00.INS00.DEP00

Object Entity in Rosetta DNX

PREMIS	Rosetta		
Object Entity	**Rep**	**File**	**Remarks**
objectCharacteristicsExtension .. for additional object characteristics not covered by PREMIS e.g. detailed, format-specific technical metadata such as those found in NISO Technical Metadata for Digital Still Images (Z39.87-2006)	**X**	**X**	Detailed NISO image properties are in significant properties

MID: DNX_FL18931   Metadata Type: dnx:dnx   Description
Created by: nlb_producer_agent   Creation Date: 20/09/2011 13:40:04
Updated by   Update Date: 20/09/2011 13:40:04

preservation Level [dropdown]
Add

internal Identifier
object Characteristics
file Fixity
event
vs Outcome
file Virus Check
file Format
general File Characteristics
file Validation
significant Properties

6   Type=nisoImage.bitsPerSample, Value=[8, 8, 8]   View
7   Type=nisoImage.byteOrder, Value=little-endian   View
8   Type=nisoImage.class, Value=class edu.harvard.hul.ois.jhove.NisoImageMetadata   View
9   Type=nisoImage.colourSpace, Value=RGB   View
10   Type=nisoImage.compressionScheme, Value=uncompressed   View
11   Type=nisoImage.dateTimeCreated, Value=2007-05-16T11:50:46   View
12   Type=nisoImage.imageLength, Value=4370   View
13   Type=nisoImage.imageProducer, Value=Zeutschel Omniscan 11   View
14   Type=nisoImage.imageWidth, Value=2646   View
15   Type=nisoImage.mimeType, Value=image/tiff   View
16   Type=nisoImage.orientation, Value=1   View
17   Type=nisoImage.planarConfiguration, Value=1   View

Object Entity in Rosetta DNX

PREMIS			Rosetta
Object Entity	**Rep**	**File**	**Remarks/DNX Fields**
originalName (Rep, File)	√	√	Representation Original Name; File Original Name
storage			
contentLocation (File)	NA	√	All semantic units available
storageMedium (File)	√	X	Representation level: physical Carrier Media
Environment (Rep, File)	√	√	All semantic units available
environmentExtension	X	X	Not in DNX
signatureInformation (File)	NA	√	All semantic units available
Relationship (Rep, File)	√	√	All semantic units in except **relatedEventIdentification**
linkingEventIdentifier (Rep, File)	X	X	Not in DNX
linkingIntellectualEntityIdentifier (Rep, File)	X	X	Not in DNX
linkingRightsStatementIdentifier (Rep, File)	X	X	Not in DNX

Event Entity in Rosetta DNX

PREMIS	Rosetta		
Event Entity	**Rep**	**File**	**Remarks/DNX Fields**
<u>Note</u>: "Event metadata are normally recorded and stored separately from the digital object"			Event metadata recorded at File level
eventIdentifier	**NA**	√	All semantic units available
eventType	**NA**	√	All semantic units available
eventDateTime	**NA**	√	All semantic units available
eventDetail	**NA**	√	All semantic units available
eventOutcomeInformation	**NA**	√	All semantic units available
linkingAgentIdentifier	**NA**	√	All semantic units available
linkingObjectIdentifier	**X**	**X**	Not in DNX

National Library Board Singapore

Agent Entity in Rosetta DNX

PREMIS			Rosetta
Agent Entity	**Rep**	**File**	**Remarks/DNX Fields**
agentIdentifier			Agent metadata are associated with events such as file fixity, virus check, file format, checksum, techMD outcomes e.g. agent: REG_SA_VC_SOPHOSAV (for virus check)
agentName			
agentType			Not in DNX
agentNote			Not in DNX
agentExtension			Not in DNX
linkingEventIdentifier			Not in DNX
linkingRightsStatementIdentifier			Not in DNX
			producer
			producer Agent
			created by, modified by, owner,

Agent Entity in Rosetta DNX

PREMIS			
Agent Entity	**Rep**	**File**	**Rema**
agentIdentifier			Agent
			fixity,
agentName			e.g. ag
agentType			Not in
			producer Agent
			created by, modified by, owner,

MID DNX_IE18929   Metadata Type dnx:dnx
Created by nlb_producer_agent   Creation Date 20/09/2011 13:40:04
Updated by   Update Date 20/09/2011 13:40:04

web Harvesting ▾ Add

Producer Id 40283
User Id 40279
default Language English
Authorative Name NLB Internal
user Name -
Institution NLB Internal
Department NLB Internal
middle Name -
job Title -
address1 NLB
address2 -
address3 -

Delete

MID DNX_IE18929   Metadata Type dnx:dnx
Created by nlb_producer_agent   Creation Date 20/09/2011 13:40:04
Updated by   Update Date 20/09/2011 13:40:04

web Harvesting ▾ Add

producer
producer Agent

first Name Producer Agent
last Name NLB
middle Name -

MID DNX_REP18930   Metadata Type dnx:dnx   Descri
Created by nlb_producer_agent   Creation Date 20/09/2011 13:40:04
Updated by   Update Date 20/09/2011 13:40:04

preservation Level ▾
Add

general Rep Characteristics   Delete
object Identifier
internal Identifier
object Characteristics

object Type Representation
parent ID -
group ID -
creation Date 20/09/2011 13:40:04
created By nlb_producer_agent
modification Date 20/09/2011 13:40:04
modified By nlb_producer_agent
owner CRS00.INS00.DEP00

23

Rights Entity in Rosetta DNX

PREMIS			Rosetta
Rights Entity	Rep	File	Remarks/DNX Fields
rightsStatement			Not in DNX
rightsStatementIdentifier			Not in DNX
rightsBasis			Not in DNX
copyrightInformation			Not in DNX
licenseInformation			Not in DNX
statuteInformation			Not in DNX
rightsGranted			Not in DNX
linkingObjectIdentifier			Not in DNX
linkingAgentIdentifier			Not in DNX
rightsExtension			Not in DNX
			Access Rights Policy

Rights Entity in Rosetta DNX

PREMIS   **Rosetta**

Home > Submissions > Approve Sampled SIPs > Edit Access Rights Exceptions

MID	AR_EVERYONE	Metadata Type	policy:accessrights	Description	No restrictions
Created by	SYSTEM	Creation Date	26/04/2010 05:53:25		
Updated by	SYSTEM	Update Date	26/04/2010 05:53:25		

Copyright Template
Description   No restrictions
When view is restricted show the following message

Group 1   Everyone equal

IE (IE18	...	...
Prese...	...	...
(REP1893	...	...
File	...	...
File	...	...

...producer_agent
15

sign AR
ceptions

Edit   Rem...s
Edit
Edit
Policy   Access Rights   AR_EVERYONE   View

linkingAgentIdentifier			Not in DNX
rightsExtension			Not in DNX
			Access Rights Policy

Web harvesting data in Rosetta DNX

PREMIS			Rosetta
Event Entity	Rep	File	Remarks/DNX Fields
not in PREMIS			Web harvesting (IE level)

MID	DNX_IE18929	**Metadata Type**	dnx:dnx	**Description**
Created by	nlb_producer_agent	**Creation Date**	20/09/2011 13:40:04	
Updated by		**Update Date**	20/09/2011 13:40:04	

producer
producer Agent
internal Identifier
object Characteristics
general IE Characteristics   Delete
access Rights Policy
web Harvesting   Delete

primary Seed URL

WCT Identifier

target Name

group

harvest Date

harvest Time

Summary & Usability

Summary

- **Differing structure**
    - _General File Characteristics_ – includes not just _objectCharacteristics_ but other Object Entity semantic units e.g. _storage_ and _originalName_
    - Sequential structure in PREMIS are applied according to need e.g. Agent metadata are associated with Events such as file fixity, file validation, virus check, etc.

3.1	agentIdentifier (M, R)	
	3.1.1	agentIdentifierType (M, NR)
	3.1.2	agentIdentifierValue (M, NR)
3.2	agentName (O, R)	
3.3	agentType (O, NR)	
3.4	agentNote (O, R)	
3.5	agentExtension (O, R)	
3.6	linkingEventIdentifier (O, R)	
	3.6.1	linkingEventIdentifierType (M, NR)
	3.6.2	linkingEventIdentifierValue (M, NR)
3.7	linkingRightsStatementIdentifier (O, R)	
	3.7.1	linkingRightsStatementIdentifierType (M, NR)
	3.7.2	linkingRightsStatementIdentifierValue (M, NR)

agent	REG_SA_DROID
format Registry	PRONOM

agent	JHOVE , TIFF-hul 1.5 , Plugin Version 2.0
status	-
format	TIFF
version	6.0

check Date	Tue Sep 20 13:40:13 SGT 2011
vs Agent	REG_SA_DPS
type	RISK_ANALYSIS
result	PASSED
result Details	-
vs Evaluation	PASSED
vs Evaluation Details	-

Summary (con't)

- **System-specific requirements**
  - e.g. storage semantic units:

    file Original Path vs file Original ID vs File Location

    temporary storage   e.g. Deposit Area 1   null value (?)
    (recycled – misleading)

file Location	-
file Original Name	9000000026_0001.tif
file Original Path	-
file Original ID	/exlibris_apps_stage/deposit_area_1//1001-2000/dep_1293/deposit/content/streams/9000000026_0001.tif

- **Risk assessment**

  Business rules - not in PREMIS, but noted as "essential for preservation"
  e.g. retention periods, risk assessment, schedules for media refreshment
  Rosetta : Risk Identifiers = Risk Levels – aligned to Preservation Planning

Summary (con't)

- **Linking identifiers not in Rosetta**

  e.g. linkingEventIdentifier, linkingIntellectualEntityIdentifier, linkingRightsStatementIdentifier, linkingObjectIdentifier

- **Properties not from PREMIS**

  e.g. semantic units under General Representation Characteristics, Hardware Registry ID, Software Registry ID

- **Link to external identifiers**

  Use of <objectIdentifier> at Representation level for DMD identifier
  - PREMIS Extensions: METS <mdRef> not used in Rosetta
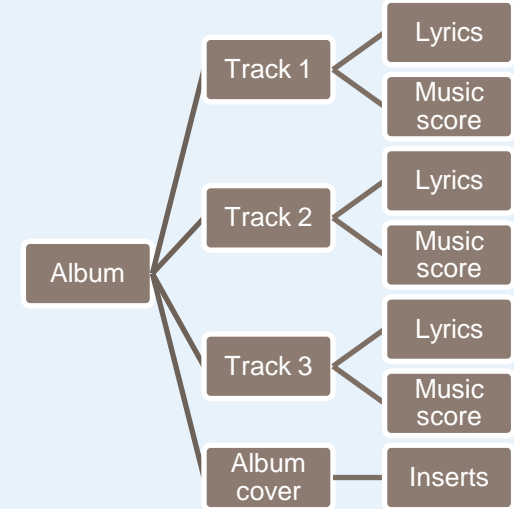
- **Most PREMIS extensions not in DNX**

  - significantPropertiesExtension
  - environmentExtension
  - agentExtension
  - rightsExtension

Usability

Complex objects

- relationshipType e.g. structural (for MusicSG)
- relationshipSubType – isPartOf, hasPart, isIncludedIn, includes, etc.
- relatedObjectIdentification – Type, Value
- environment dependency & sw dependency
  – no immediate need for NLB at the moment, will need in near future
- compositionLevel
  - presently no compression, encryption or derivation is done for NLB objects in Rosetta (will perform derivation in near future)

Thank You...
for your attention

Questions: haliza@nlb.gov.sg

NLB | National Library Board Singapore